IMF INSTITUTE
FOR CAPACITY DEVELOPMENT

Modeling Non-stationary Variables

Joint Vienna Institute/ IMF ICD
Macro-econometric Forecasting and Analysis,
JV16.12, L02, Vienna, Austria, May 17, 2016

Presenter
Mikhail Pranovich

Lecture Objectives

- Revisit the concept of non-stationary (unit root) process and its implications for analysis and forecasting
- Understand key tests for unit root
- Revisit the concept of cointegration
- … and testing for cointegration

Outline

- Stationary and non-stationary variables
- Testing for unit roots
- Cointegration
- Testing for cointegration

Introduction

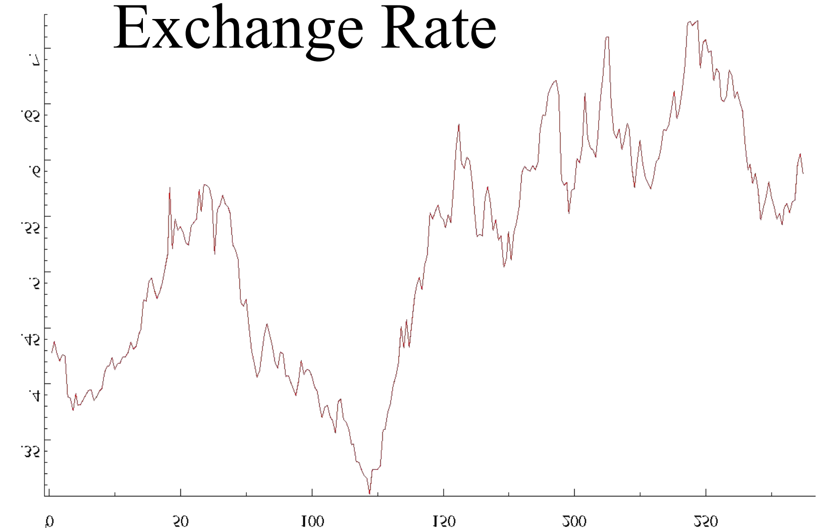
- Many economic (macro/financial) variables exhibit *trending* behavior
  - e.g., real GDP, real consumption, assets prices, dividends…

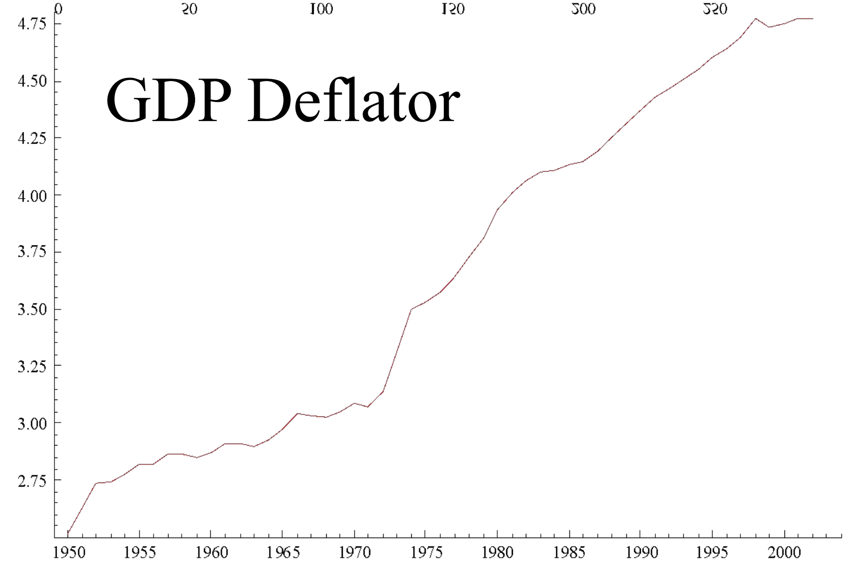
- Key issue for estimation/forecasting:
  - the nature of this trend….
  - … is it *deterministic* (e.g., linear trend) or *stochastic (*e.g., random walk*)*

- The nature of the trend has important implications for the model's parameters and their distributions…

- … and thus for the statistical procedures used to conduct inference and forecasting

Key Macro Series Appear to have trends





Deterministic and Stochastic Trends in Data

- Two types of trends: deterministic or stochastic
- A Deterministic trend is a non-random function of time
  - Example: linear time-trend

$$y_t = \mu_1 + \mu_2 t + \varepsilon_t$$

- A stochastic trend is random, i.e. varies over time
- Examples:
  - (Pure) Random Walk Model: a time series is said to follow a pure random walk if the change is *i.i.d.*

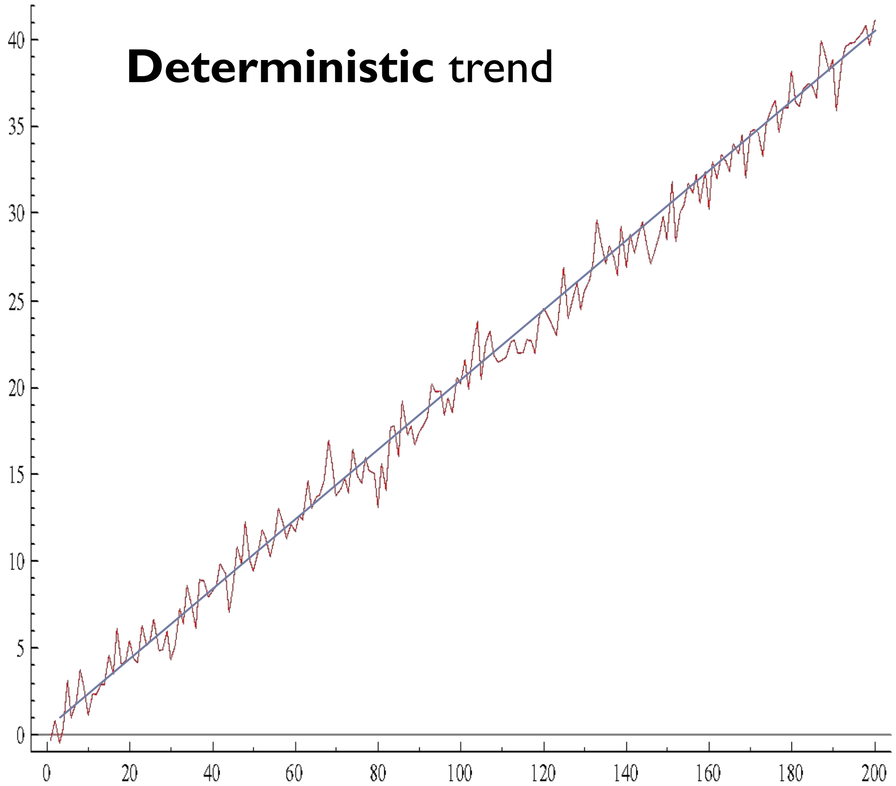
$$y_t = y_{t-1} + \varepsilon_t$$

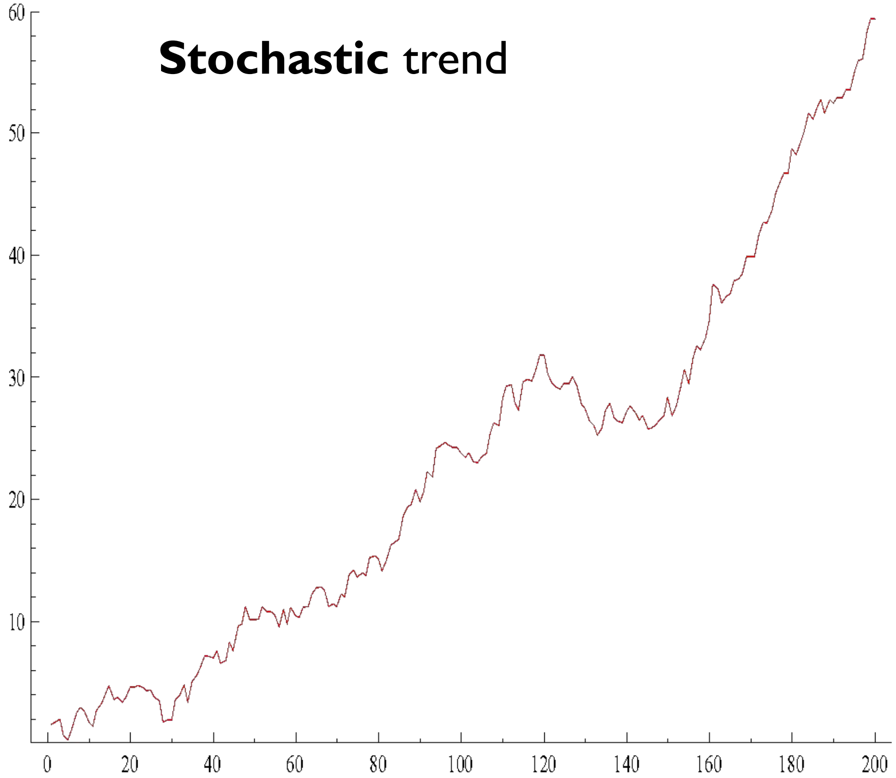
  - Random Walk with a Drift

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

μ is a 'drift'. If $\mu > 0$, then y_t increases on average

Example: Processes with Trends





Stationary and non-stationary processes (1)

- Consider the data generation process (DGP)

$$y_t = \delta + \theta y_{t-1} + v_t$$

- If $|\theta| < 1.0$, the variable is stationary (i.e., y_t, has finite mean and variance)

- Standard econometric procedures may be used to estimate/forecast this model

Stationary and non-stationary processes (2)

- If $|\theta| \geq 1.0$, model is said to be *non-stationary* and its associated (statistical) distribution theory is non-standard.

- In particular:

  - Sample moments do not have finite limits, but converge (weakly) to random quantities;

  - Least squares estimate of θ is super consistent with convergence rates greater than $\sqrt{T}$ (stationary case);

  - Asymptotic distribution of the least squares estimator is non-standard (i.e., non-normal).

- Bottom line: nature of the trend has important implications for hypothesis testing and forecasting, especially in multivariate settings (e.g., VARS).

Reminder: Autoregressive AR(p) Process

- We shall check how shocks affect stationary and non-stationary variables, but first recall what is an AR(p) process

- An AR(p) autoregressive process (AR-process of order p):

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + \varepsilon_t$$

- The error ε_t, is assumed to be independently and identically distributed (*i.i.d.*), with a zero mean and a constant variance

Stochastic trends, autoregressive models and a unit root

- The condition for stationarity in an AR(p) model: roots z of the characteristic equation

$$1 - \theta_1 z - \theta_2 z^2 - \theta_3 z^3 - \ldots - \theta_p z^p = 0$$

must all be greater than one in absolute value: $|z| > 1$

- If an AR(p) process has z=1 => variable has a unit root

- Example: AR(1) process $y_t = \mu + \theta y_{t-1} + v_t$
  - A special case is $\theta = 1 \Rightarrow z = 1 \Rightarrow$ y_t has unit root (stochastic trend)
  - Stationarity requires that $|\theta| < 1$ for $|z|>1$

The Impact of Shocks on Stationary and Non-stationary variables

- Consider a simple AR(1):

$$y_t = \theta y_{t-1} + v_t,$$

where θ takes any value for now

- We can write:

$$y_{t-1} = \theta y_{t-2} + v_{t-1}$$
$$y_{t-2} = \theta y_{t-3} + v_{t-2}$$

- Substituting yields:

$$y_t = \theta(\theta y_{t-2} + v_{t-1}) + \varepsilon_t = \theta^2 y_{t-2} + \theta v_{t-1} + v_t$$

- Successive substituting for y_{t-2}, y_{t-3},... gives an representation in terms of initial value y_{-1} and past errors $v_{t-1}, v_{t-2},\ldots,v_0$

$$y_t = \theta^{t+1} y_{-1} + \theta v_{t-1} + \theta^2 v_{t-2} + \theta^3 v_{t-3} + \ldots + \theta^t v_0 + v_t$$

The Impact of Shocks for Stationary and Non-stationary Series (2)

- Representation at $t=T$: $y_T = \theta^{T+1} y_{-1} + \theta v_{T-1} + \theta^2 v_{T-2} + \theta^3 v_{T-3} + \ldots + \theta^T v_0 + v_T$
- At $t=0$ the variable is hit by a non-zero shock v_0
- We have 3 cases (depending on value of θ):

1. $|\theta| < 1 \Rightarrow \theta^T \rightarrow 0$ and $\theta^T v_0 \rightarrow 0$ as $T \rightarrow \infty$

   Shocks have only a transitory effect (gradually dies away with time)

2. $\theta = 1 \Rightarrow \theta^T = 1$ and $\theta^T v_0 = v_0 \quad \forall \ T$

   Shocks have a permanent effect in the system and never die away:

$$y_T = y_{-1} + \sum_{i=0}^{T} v_i$$

   ... just a sum of past shocks plus some starting value of y_{-1}. The variance grows without bound ($T\sigma^2 \rightarrow \infty$) as $T \rightarrow \infty$

3. $|\theta| > 1$. Now shocks become <u>more</u> influential as time goes on (explosive effect), since if $\theta > 1$, then $|\theta|^T > \ldots > |\theta|^3 > |\theta|^2 > |\theta|$ etc.

Integration

- Another way to write the stochastic trend model is:

$$\Delta y_t = y_t - y_{t-1} = \delta + v_t$$

- Thus the first difference of y_t is stationary provided v_t is stationary ("difference stationary" process). Also referred to as an *I*(1) variable.

- Similarly, in the case of the deterministic trend model, y_t is interpreted as trend stationary

  - because removal of the deterministic trend from y_t renders it a stationary random variable

Order of Integration: I(d)

- In general, if y_t is I(d) then:

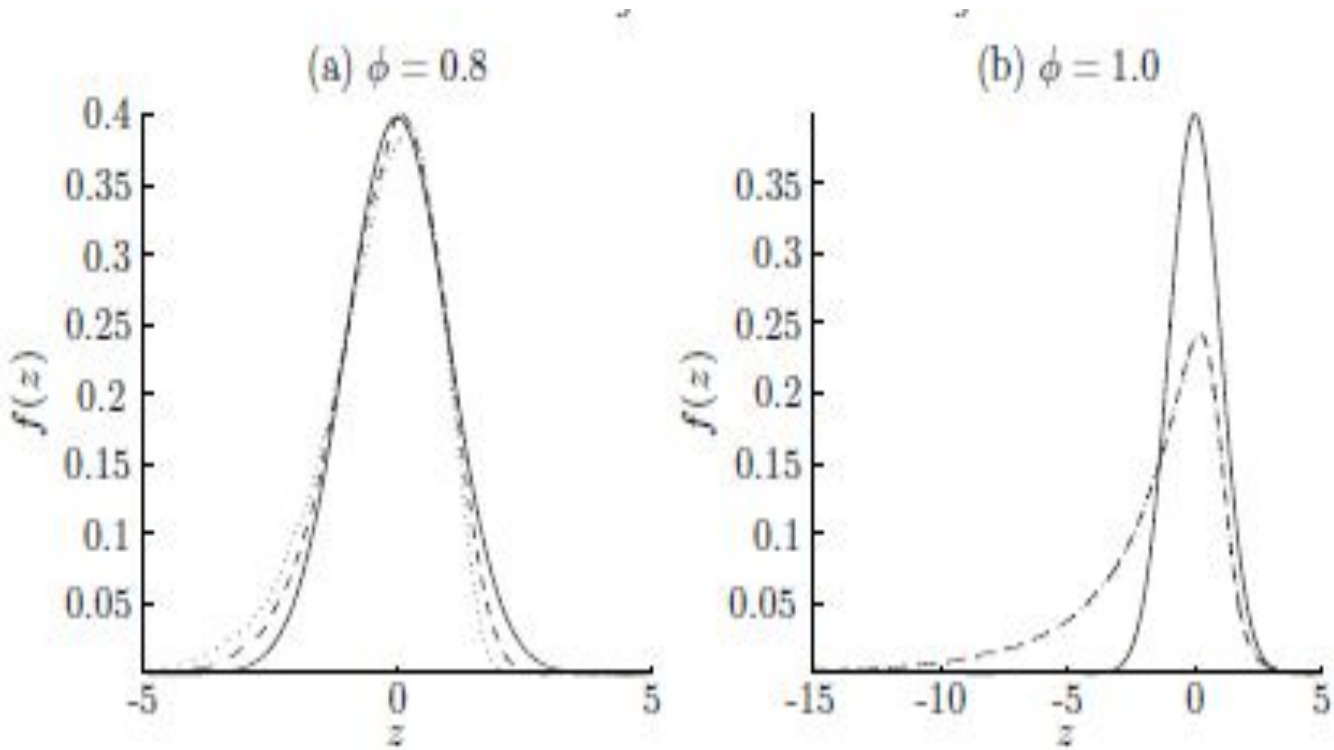
$$\Delta^d y_t = (1 - L)^d y_t = \delta + \nu_t$$

- If d=0, then the series is already stationary

Problems due to Stochastic Trends *(from a statistical perspective)*

- Non-standard distribution of test statistics
- Spurious regression:
  - in a simple linear regression, two (or more) non-stationary time series may appear to be related even though they are not
- Need to use special modeling techniques when dealing with non-stationary data (VARs in differences or VECMs)
- Need to distinguish btw. stochastic and deterministic trends as it may affect estimates of policy-relevant variables
  - e.g. estimate of an output gap or of a structural budget deficit
  - … for that we need unit root tests…

Figure 5: Distribution of OLS estimator for θ



Testing For Unit Roots

- Previous section suggests that I(1) variables need special handling

- So how do we identify I(1) processes, i.e., test for unit roots?

- Natural test is to consider the *t*-statistic for the null-hypothesis of a unit root, i.e., $\hat{\theta}=1$

- Given the previous graph, it is not surprising that the t-distribution for $\hat{\theta}=1$ is non-normal

Testing for Unit Roots: Procedures

- Dickey Fuller
- Augmented Dickey Fuller
- Phillips Perron
- Kwiatkowski, Phillips, Schmidt and Shin (KPSS)

Dickey Fuller Test

- Fuller (1976), Dickey and Fuller (1979)

- Example:
  - consider a particular case of an AR(1) model:

$$y_t = \theta y_{t\text{-}1} + \varepsilon_t$$

  - We test a hypothesis

H_0: θ =1 → the series contains a unit root/stochastic trend (is a random walk)

against

H_1: $|\theta|$ <1 → the series is a zero-mean stationary AR(1)

Dickey-Fuller Test (2)

- For the purpose of testing we reformulate the regression:

$$\Delta y_t = y_t - y_{t-1} = \theta y_{t-1} - y_{t-1} + v_t = (\theta - 1) y_{t-1} + v_t =$$

$$= \psi y_{t-1} + v_t$$

so that the test of H_0: $\theta = 1 \Leftrightarrow H_0$: $\psi = 0$

- The test is based on the *t*-ratio for ψ
  - this *t*-ratio does not have the usual *t*-distribution under the H_0
  - critical values are derived from Monte Carlo experiments, and are tabulated (known): see appendix A
- The test is not invariant to the addition of deterministic components (more general formulation: intercept + time-trend)

Dickey-Fuller Test (3)

- Important issue – shall deterministic components be included in the test model for y_t. Is this

$$\Delta y_t = \psi y_{t-1} + v_t$$

or

$$\Delta y_t = \mu_1 + \psi y_{t-1} + v_t$$

or

$$\Delta y_t = \mu_1 + \mu_2 t + \psi y_{t-1} + v_t \,?$$

- Two ways around:
  - Use prior information/assume whether the deterministic components are included, i.e. use the restrictions (easy to implement in Eviews):
    - $\mu_1 \neq 0$ and $\mu_2 \neq 0$
    - $\mu_1 \neq 0$ and $\mu_2 = 0$
    - $\mu_1 = 0$ and $\mu_2 = 0$
  - Allow for uncertainty about deterministic components (more complicated in Eviews) and implement a testing strategy to find out:
    - restrictions on deterministic components
    - if y_t is non-stationary

DF-Test (3): Deterministic Components are Known

- Say, we assume y_t includes an intercept, but not a time trend

$$y_t = \mu_1 + \theta y_{t-1} + v_t$$

- We test a hypothesis:

  H_0: $\theta = 1 \rightarrow$ the series has a unit root/stochastic trend

  against

  H_1: $|\theta| < 1 \rightarrow$ the series is zero-mean stationary AR(1)

- Reformulate:

$$\Delta y_t = \mu_1 + \psi y_{t-1} + v_t$$

- Test H_0: $\psi = 0 \rightarrow$ the series has a unit root (stochastic trend) against

  H_1: $\psi < 0 \rightarrow$ the series has no unit root (is stationary)

- This way is easy – it is ready for you in Eviews
- But, there are risks involved...

DF-Test (4): Risks Posed by Deterministic Components

- If deterministic components are not included in the test, when they should be, then the test is not correctly sized:
  - The test will reject the H_0: ψ =0, although it is in fact true and should not be rejected (y_t is non-stationary) – type I error
- If deterministic components are included but they should not be, then the test has low power (especially in finite (short) samples):
  - The test will not reject the H_0: ψ =0, although it is false and must be rejected (y_t is stationary) – type II error
- This is why we may prefer (a degree of) uncertainty about deterministic components and use testing strategies (see appendix A for details):
  - Enders Strategy
  - Elder and Kennedy Strategy

The Augmented Dickey Fuller (ADF) Test

- The DF-test above is only valid if ε_t is a white noise: $\varepsilon_t \approx i.i.d(o, \sigma_\varepsilon^2)$

- ε_t will be autocorrelated if there was autocorrelation in the first difference (Δy_t), and we have to control for it

- The solution is to "augment" the test using p lags of the dependent variable. The alternative model (including the constant and the time trend) is now written as:

$$\Delta y_t = \mu_1 + \mu_2 t + \psi y_{t-1} + \sum_{i=1}^{p} a_i \Delta y_{t-i} + \varepsilon_t$$

The ADF-Test (2)

- Again, we have three choices:
  - (1) include neither a constant nor a time trend
  - (2) include a constant
  - (3) include a constant and a time trend

- Again, we either:
  - use prior information and impose a model from the beginning, or
  - remain uncertain about deterministic components and follow one of the Strategies

- *Useful result*: Critical values for the ADF-test are the same as for DF-test

- Note, however, that the test statistics are sensitive to the lag length p

The ADF-Test: Lag Length Selection

- Three approaches are commonly used:
  - Akaike Information Criterion (AIC)
  - Schwarz-Bayesian Criterion (SBC)
  - General-to-Specific successive *t*-tests on lag coefficients

- AIC and BIC are statistics that favour fit (smaller residuals) but penalize for every additional parameter that needs to be estimated:
  - So, we prefer a model with a smaller value of a criterion statistic

- General-to-Specific: begin with a general model where p is fairly large, and successively re-estimate with one less lag each time (keeping the sample fixed)

- It is advised to use AIC
  - Tendency of SBC to select too parsimonious of a model
  - The ADF-test is biased when any autocorrelation remains in the residuals

- Note: the test critical values do not depend on the method used to select the lag length

Dickey-Fuller (and ADF) Test: Criticism

- The power of the tests is low if the process is stationary but with a root “close” to 1 (so called “near unit root” process)
  - e.g. the test is poor at rejecting $\theta = 1$ (ψ=0), when the true data generating process is

$$y_t = 0.95 y_{t-1} + \varepsilon_t$$

- This problem is particularly pronounced in small samples

The Phillips Perron (PP) test

- Rather popular in the analysis of financial time series
- The test regression for the PP-tests is

$$\Delta y_t = \mu_1 + \mu_2 t + \psi y_{t-1} + \varepsilon_t$$

- PP modifies the test statistic to account for any serial correlation and heteroskedasticity of ε_t
- The usual t-statistic in the DF-test $t_{\psi=0}$ …
- … is modified:

$$Z_t = \left(\frac{\hat{\sigma}^2}{\hat{\lambda}^2}\right)^{1/2} \cdot t_{\psi=0} - \frac{1}{2}\left(\frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2}\right) \cdot \left(\frac{T \cdot SE(\hat{\psi})}{\hat{\sigma}^2}\right)$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 - \text{estimate of variance}$$

$$\hat{\lambda}^2 = \hat{\sigma}^2 + 2\sum_{j=1}^{q}[1 - \frac{j}{q+1}]\hat{\gamma}_j,\ \hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\hat{\varepsilon}_t\,\hat{\varepsilon}_{t-j} - \text{estimate of autocovariance of order } j,$$

q – is a number of lags, up to which errors autocorrelation might be present

The PP test (2)

- Under the null hypothesis that $\psi = 0$, Z_t statistic has the same asymptotic distribution as the ADF t-statistic

- Advantages:
  - PP-test is robust to general forms of heteroskedasticity in ε_t
  - No need to specify the lag length for the test regression

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

- The KPSS test is a stationarity test. The H_0 is: $y_t \sim I(0)$
- Start with the model:

$$\Delta y_t = \mu D_t + \phi_t + \varepsilon_t$$

$$\phi_t = \phi_{t-1} + u_t, \quad u_t \approx i.i.d(0, \sigma_u^2),$$

D_t contains deterministic components, ε_t is $I(0)$ and may be heteroskedastic

- The test is then H_0: $\sigma_u^2 = 0$ against the alternative H_1: $\sigma_u^2 > 0$
- The KPSS test statistic is:

$$KPSS = \left(T^{-2} \sum_{t=1}^{T} \hat{S}_t^2 \right) / \hat{\lambda}^2$$

where $\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j$ is a cumulative residual function and $\hat{\lambda}^2$ is a long-run variance of ε_t as defined earlier (see slide 32)

- See Appendix C on some details w.r.t. critical values

Testing for Higher Orders of Integration

- Just when we thought it is over... Consider:

$$\Delta y_t = \psi y_{t-1} + \varepsilon_t$$

  we test H_0: $\psi=0$ vs. H_1: $\psi<0$

- If H_0 is rejected, then y_t is stationary
- What if H_0 is not rejected? The series has a unit root, but is that it? No! What if $y_t \sim I(2)$? So we now need to test

$$H_0: y_t \sim I(2) \text{ vs. } H_1: y_t \sim I(1)$$

- Regress $\Delta^2 y_t$ on Δy_{t-1} (plus lags of $\Delta^2 y_t$, if necessary)
- Test H_0: $\Delta y_t \sim I(1)$, which is equivalent to H_0: $y_t \sim I(2)$
- So, if we do not reject, then we conclude y_t is at least $I(2)$...

Working with Non-Stationary Variables

- Consider a regression model with two variables; there are 4 cases to deal with:

- **Case 1:** Both variables are stationary=> classical regression model is valid

- **Case 2:** The variables are integrated of different orders=> unbalanced (meaningless) regression

- **Case 3:** Both variables are integrated of the same order; regression residuals contain a stochastic trend=> spurious regression

- **Case 4:** Both variables are integrated of the same order; the residual series is stationary=> *y* and *x* are said to be cointegrated and…

- You will have more on this in *L*-5, *L*-8 and *L-9*

Cointegration

- Important implication is that non-stationary time series can be rendered stationary by differencing

- Now we turn to the case of N>1 (i.e., multiple variables)

- An alternative approach to achieving stationarity is to form linear combinations of the $I(1)$ series – this is the essence of "cointegration" [Engle and Granger (1987)]

Cointegration

- Three main implications of cointegration:

  - Existence of cointegration implies a set of dynamic long-run equilibria where the weights used to achieve stationarity are the parameters of the long-run (or equilibrium) relationship.

  - The OLS estimates of the weights converge to their population values at a super-consistent rate of “T” compared to the usual rate of convergence, $\sqrt{T}$

  - Modeling a system of cointegrated variables allows for specification of both the long-run and short-run dynamics. The end result is called a “Vector Error Correction Model (VECM)”.

Cointegration

- We will see that cointegrated systems (VECMs) are special VARS.

- Specifically, cointegration implies a set of non-linear cross-equation restrictions on the VAR.

- Easiest/most flexible way to estimate VECM's is by full-information maximum likelihood.

Long-Run Equilibrium Relationships: Examples

- Permanent Income Hypothesis (PIH)

  - Postulates a long-run relationship between log real consumption and log real income:

$$\log(rc_t) = \beta_c + \beta_y \log(ry_t) + u_t$$

  - Assuming real consumption and income are non-stationary (I(1)) variables, then the PIH is postulating that real consumption and income move together over time and that u_t is a stationary series.

Term Structure Of Interest Rates

- Models the relationship between the yields on bonds of differing maturities.

- Prior is that yields of different (longer) maturities can be explained in terms of a single (typically shorter) maturity yield.

- For example:

$$r_{3,t} = \beta_{c,1} + \beta_{1,1} r_{1,t} + u_{1,t}$$

$$r_{2,t} = \beta_{c,2} + \beta_{2,1} r_{1,t} + u_{2,t}$$

- All the yields are assumed to be I(1), but the residuals are I(0) [stationary]. This is an example of a system of three variables with two (2) long-run relationships

VECM

- Cointegration postulates the existence of long-run equilibrium relationships between non-stationary variables where short-run deviations from equilibrium are stationary.

- What is the underlying economic model?

- How do we estimate such a model?

Bivariate VECMs

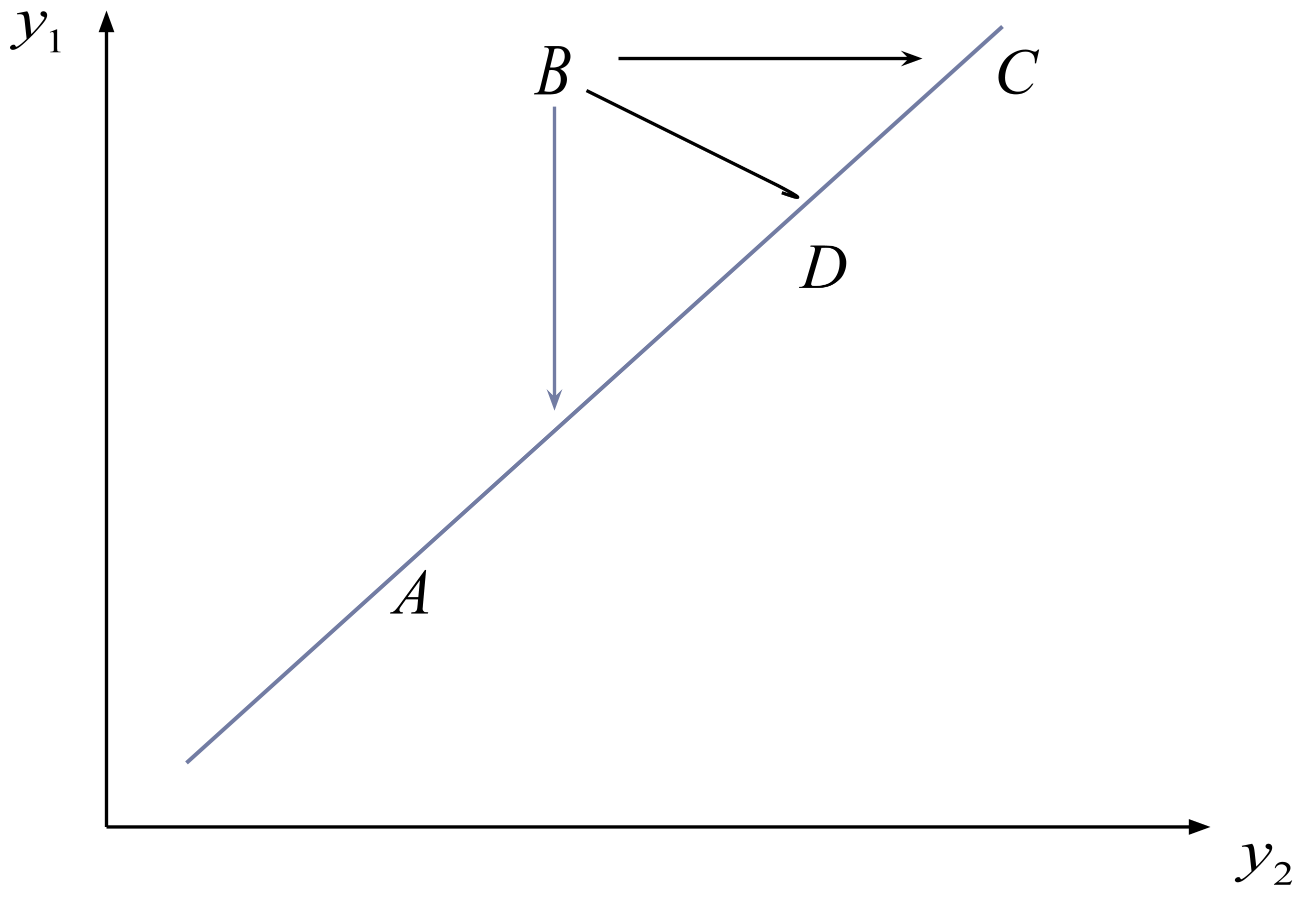
- Consider a bivariate model containing two I(1) variables, say $y_{1,t}$ and $y_{2,t}$.

- Assume the long-run relationship is given by

$$y_{1,t} = \beta_c + \beta_y y_{2,t} + u_t$$

- Here $\beta_c + \beta_y y_{2,t}$ represents the long-run equilibrium, and u_t represents the short-run deviations from the long-run equilibrium (see next slide).

Phase Diagram: VECM



Adjusting Back To Equilibrium

- Suppose there is a positive shock in the previous period, raising $y_{1,t}$ to point B while leaving $y_{2,t-1}$ unchanged.

- How can the system converge back to its long-run equilibrium?

- There are three possible trajectories…

Adjustments Are Made by $Y_{1,t}$

- Long-run equilibrium is restored by $y_{1,t}$ decreasing toward point A *while $y_{2,t}$ remains unchanged at its initial position*.

- Assuming that the short-run change in $y_{1,t}$ are a linear function of the size of the deviation from the LR equilibrium, u_{t-1}, the adjustment in $y_{1,t}$ is given by:

$$y_{1,t} - y_{1,t-1} = \alpha_1 u_{t-1} + v_{1,t} = \alpha_1 \left(y_{1,t-1} - \beta_c - \beta_y y_{2,t-1} \right) + v_{1,t}$$

- where $\alpha_1 < 0$ is a parameter to be estimated.

Adjustments Are Made by $Y_{2,t}$

- Long-run equilibrium is restored by $y_{2,t}$ increasing toward point C *while $y_{1,t}$ remains unchanged after the initial shock*.

- Assuming that the short-run movements in $y_{2,t}$ are a linear function of the size of shock, u_t, the adjustment in $y_{2,t}$ is given by:

$$y_{2,t} - y_{2,t-1} = \alpha_2 u_{t-1} + v_{2,t} = \alpha_2 \left(y_{1,t-1} - \beta_c - \beta_y y_{2,t-1} \right) + v_{2,t}$$

$$\alpha_2 > 0$$

- where is a parameter to be estimated.

Adjustments are made by both $Y_{1,t}$ and $Y_{2,t}$

- The previous two equations may operate simultaneously with both $y_{1,t}$ and $y_{2,t}$ converging to a point on the long-run equilibrium path such as D.

- The relative strengths of the two adjustment paths depend on the relative magnitudes of the adjustment parameters, α_1 and α_2.

- The parameters α_1 and α_2 are known as the "error-correction parameters" or short-run adjustment coefficients.

VECM = Special VAR

- A VECM is actually a special case of a VAR where the parameters are subject to a set of cross-equation restrictions because all the variables are governed by the same long-run equations. Consider what we have when we put the two equations together:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} -\alpha_1 \beta_c \\ -\alpha_2 \beta_c \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\beta_y \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

- or in terms of a VAR…

VECM = Special VAR

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} -\alpha_1\beta_c \\ -\alpha_2\beta_c \end{bmatrix} + \begin{bmatrix} 1+\alpha_1 & -\alpha_1\beta_y \\ \alpha_2 & 1-\alpha_2\beta_y \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}$$

which is clearly a first-order VAR

$$y_t = \mu + \Phi y_{t-1} + v_t$$

VECM = Special VAR

- Obviously, we have a first order VAR with two restrictions on the parameters.

- In an unconstrained VAR of order one, no cross-equation restrictions are imposed, implying 6 unknown parameters.

- However, a VECM – owing to the cross-equation restrictions – has only ***four*** unknown parameters. Less restrictions are needed to identify the model.

Multivariate Methods: N > 2

- Can easily generalize the relationship between a VAR and a VECM to N variables and p lags.

- Assume first that p = 1: $y_t = \mu + \Phi_1 y_{t-1} + v_t$

- Subtracting y_{t-1} from both sides:

$$y_t - y_{t-1} = \mu - \left(I_N - \Phi_1\right) y_{t-1} + v_t$$

- or $\Delta y_t = \mu - \Phi(1) y_{t-1} + v_t, \text{ where } \Phi(1) = \left(I_N - \Phi_1\right)$

- This is a VECM, but with p = 0 lags.

VAR with p lags > 1

- Allowing for p lags gives:

$$\Phi(L)y_t = \mu + v_t$$

- where v_t is an N dimensional vector of *iid* disturbances and $\Phi(L) = I_N - \Phi_1 L - \ldots - \Phi_p L^p$ is a p-th order polynomial in the lag operator.

- The resulting VECM has p-1 lags given by:

$$\Delta y_t = \mu - \Phi(1)y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t, \text{ where } \Gamma_j = -\sum_{i=j+1}^{p} \Phi_i$$

Cointegration

- If the vector time series y_t is assumed to be I(1), then y_t is cointegrated if there exists an N x r full column rank matrix, β, such that the r linear combinations:

$$\beta' y_t = u_t$$

are I(0).

- The dimension "r" is called the cointegrating rank and the columns of β are called the co-integrating vectors.

- This implies that (N – r) common trends exist that are I(1).

Granger Representation Theorem

- Suppose y_t, which can be I(1) or I(0), is generated by

$$\Delta y_t = \mu - \Phi(1) y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

- Three important cases:
  - (a) If $\Phi(1)$ has full rank, i.e., r = N, then y_t is I(0)
  - (b) If $\Phi(1)$ has reduced rank 0 < r < N,

    $\Phi(1) = -\alpha\beta'$ where α and β are each $(N \ x \ r)$ matrices with full column rank.

  - then y_t is I(1) and $\beta' y_t$ is I(0) with cointegrating vectors given by the columns of β
  - (c) if $\Phi(1)$ has zero rank, r = 0, $\Phi(1) = 0$ and y_t is I(1) and not cointegrated.

Examples: Rank of Long-Run Models

- The form of $\Phi(1)$ for the two long-run models we considered above:
- Permanent Income: (N=2, r=1)

$$\Phi(1) = -\alpha\beta' = -\begin{bmatrix} \alpha_{1,1} \\ \alpha_{2,1} \end{bmatrix} \begin{bmatrix} 1 \\ -\beta_y \end{bmatrix}'$$

- Term structure: (N = 3, r = 2)

$$\Phi(1) = -\alpha\beta' = -\begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \\ \alpha_{3,1} & \alpha_{3,2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \beta_{3,1} & \beta_{3,2} \end{bmatrix}'$$

Key Implications of the GE Representation Theorem

- The Granger-Engle theorem suggests the form of the model that should be estimated given the nature of the data.

- If $\Phi(1)$ has full rank, N, then all the time series must be stationary, and the original VAR should be specified in levels. This is the "unrestricted model".

- If $\Phi(1)$ has reduced rank, with 0 < r < N, then a VECM should be estimated subject to the restrictions

$$\Phi(1) = -\alpha\beta' \text{ , viz:}$$

$$\Delta y_t = \mu + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

Key Implications of the GE Representation Theorem

- If $\Phi(1)=0$, then the appropriate model is:

$$\Delta y_t = \mu + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

- In other words, if all the variables in y_t are I(1) and not cointegrated, we should estimate a VAR(p-1) in first differences.

- Note that this is the most restricted model compared to the previous two, which is important when calculating likelihood ratio tests for cointegration.

Dealing With Determinstic Components

- We can easily extend the base VECM to include a deterministic time trend, viz:

$$\Delta y_t = \mu_0 + \mu_1 t + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

- where now μ_0 and μ_1 are (N x 1) vectors of parameters associated with the intercept and time trend.

- The deterministic components can contribute both to the short-run and the long-run components of y_t

Deterministic Components

- Suppose we can decompose these parameters into their short-run and long-run components by defining:

$$\mu_j = \delta_j + \alpha\beta_j', \quad j = 0,\ 1$$

- where δ_j (N x 1) is the short-run component and $\alpha\beta_j'$ is the long-run component.

- We can rewrite the model as:

$$\Delta y_t = \delta_0 + \delta_1 t + \alpha\left(\beta_0' + \beta_1' t + \beta' y_{t-1}\right) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

Deterministic Components

- The term $\left(\beta_0' + \beta_1' t + \beta' y_{t-1}\right)$ represents the long-run relationship among the variables.

- The parameter δ_0 provides a drift component in the equation of Δy_t , so it contributes a trend to y_t

- Similarly $\delta_1 t$ allows for linear time trend in Δy_t and a quadratic trend to y_t

- By contrast, β_0 contributes a constant to the EC-Eq and $\beta_1' t$ contributes a linear time trend to EC-Eq

Deterministic Components

- The equation

$$\Delta y_t = \delta_0 + \delta_1 t + \alpha\left(\beta_0' + \beta_1' t + \beta' y_{t-1}\right) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

- contains five important special cases summarized on the next slide.
- Model 1 is the simplest (and most restricted) as there are no deterministic components.
- Model 2 allows for r intercepts in the long-run equations.
- Model 3 (most common) allows for constants in both the short-run and the long-run equations – total of N+r intercepts.

Alternative Deterministic Structures

Summary of alternative VECM specifications with $\mu_j = \delta_j + \alpha\beta_j', j = 0, 1$.

Model	Specification
1	$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	Restrictions: $\{\delta_0 = 0, \delta_1 = 0, \beta_0 = 0, \beta_1 = 0\}$
2	$\Delta y_t = \alpha(\beta_0' + \beta' y_{t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	Restrictions: $\{\delta_0 = 0, \delta_1 = 0, \beta_1 = 0\}$
3	$\Delta y_t = \delta_0 + \alpha(\beta_0' + \beta' y_{t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	$= \mu_0 + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	Restrictions: $\{\delta_1 = 0, \beta_1 = 0\}$
4	$\Delta y_t = \delta_0 + \alpha(\beta_0' + \beta_1' t + \beta' y_{t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	$= \mu_0 + \alpha(\beta_1' t + \beta' y_{t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	Restrictions: $\{\delta_1 = 0\}$
5	$\Delta y_t = \delta_0 + \delta_1 t + \alpha(\beta_0' + \beta_1' t + \beta' y_{t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	$= \mu_0 + \mu_1 t + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$
	Restrictions: None

Estimating VECM Models

- If you are willing to assume that the error term v_t is white noise and $N(0,\sigma^2)$, the parameters of the VECM can be estimated directly by full-information maximum likelihood techniques.

- Basically, one estimates a traditional VAR subject to the cross-equation restrictions implied by cointegration.

- Using FIML is the most flexible approach, but it requires one to ensure that the parameters of the overall model are identified (via exclusion restrictions). More on this later.

Three Cases: $\Phi(1)$ can be inverted.

- VECM is equivalent to the unconstrained VAR. No restrictions are imposed on the VAR.
- Maximum likelihood estimator is obtained by applying OLS to each equation separately.
- The estimator is applied to the levels of the data, since they are (must be) stationary.

Reduced Rank (Cointegration) Case: FIML

- If $\Phi(1)$ cannot be inverted (i.e., reduced rank case, or we are dealing with a cointegrated system), we impose the cross-equation restrictions coming from the lagged ECM term(s), and then estimate the system using full-information maximum likelihood methods.

- The VECM is a restricted model compared to the unconstrained VAR.

Reduced Rank Case: Johansen Estimator

- We can also use the Johansen (1988) estimator.

- This differs from FIML in that the cross-equation identifying restrictions are NOT imposed on the model before estimation.

- The Johansen approach estimates a basis for the vector space spanned by the cointegrating vectors, and THEN imposes identification on the coefficients.

Zero-Rank Case for $\Phi(1)$

- When $\Phi(1)=0$, the VECM reduces to a VAR in first differences.

- As with the full-rank model, the maximum likelihood estimator is the ordinary least squares estimator applied to each equation separately.

- This is the most constrained model compared to a VECM/unconstrained VAR in levels.

Identification

- The Johansen procedure requires one to normalize the cointegrating vectors so that one of the variables in the equation is regarded as the dependent variable of the long-run relationship.

- In the bi-variate term structure and the permanent income example, the normalization takes the form of designating one of variables in the system as the dependent variable.

Identification: Triangular Restrictions

- Suppose there are r long-run relationships.

- Identification can be achieved by transforming the top (r x r) block of $\hat{\beta}$ (the long-run parameters) to the identity matrix.

- If r = 1, this corresponds to normalizing one the coefficients to unity.

Triangular Restrictions

- If there are N = 3 variables and r = 2 cointegrating equations, one sets $\hat{\beta}$ to:

$$\hat{\beta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \hat{\beta}_{3,1} & \hat{\beta}_{3,2} \end{bmatrix}$$

- This form of the normalized estimated co-integrated vector is appropriate for the tri-variate term structure model introduced earlier.

Structural Restrictions

- Traditional identification methods can also be used with VECM's, including exclusion restrictions, cross-equation restrictions, and restrictions on the disturbance covariance matrix.

- Example: Johansen and Juselius(1992) propose an open economy model in which $y_t = \{s_t, p_t, p_t^*, i_t, i_t^*\}$ represents, respectively, the spot exchange rate, the domestic price level, the foreign price, the domestic interest rate and the foreign interest rate.
- Thus, N = 5.

Open Economy Model

- Assuming r = 2 long-run equations, the following restrictions consisting of normalization, exclusion and cross-equation restrictions on yield the normalized long-run parameter matrix

$$\beta' = \begin{bmatrix} 1 & -\beta_{2,1} & \beta_{2,1} & 0 & 0 \\ 0 & 0 & 0 & 1 & -\beta_{5,1} \end{bmatrix}$$

- The long-run equations represent PPP and UIP.

$$s_t = \beta_{2,1}\left(p_t - p_t^*\right) + u_{1,t} \quad \text{[PPP]}$$

$$i_t = \beta_{5,1} i_t^* + u_{2,t} \quad \text{[Uncovered IP]}$$

Cointegration Rank

- So far we have taken the rank of the system as given. But how do we decide how many co-integrating vectors are in the vector of N variables?

- Simple approach is to estimate models of different rank and then do a formal likelihood ratio test to decide whether restricted model (i.e., the model with rank r less than N) is appropriate.

- Specifically, one would estimate the most restricted model (r = 0), a model that assumes (r=1), then a model that assumes r = 2, etc. The process ends when we cannot reject the null ($r = r_0$).

Cointegration Rank: Likelihood Ratio Test

- Suppose we estimate the model assuming no cointegration. Let the parameters involved in that model be denoted by $\hat{\theta}_{r=N}$.

- Let the value of the likelihood of this model be denoted by $L_T\left(\hat{\theta}_{r=N}\right)$

- Now estimate the model assuming r ≥ 1. Obviously, this is an restricted model compared to the r = N case. Let the value of the likelihood in this case be denoted by $L_T\left(\hat{\theta}_{r=r_0}\right)$

Cointegration Rank: Likelihood Ratio Test

- Using the standard result for the likelihood ratio test, we get the following LR test statistic:

$$LR = -2\left((T-p)\ln L_T\left(\hat{\theta}_{r=r_0}\right) - (T-p)\ln L_T\left(\hat{\theta}_{r=N}\right)\right)$$

- We reject the restricted model if the likelihood ratio test is greater than the corresponding critical value.

- In this case, imposing the restrictions does not yield a superior model.

Cointegration Rank: Johansen Approach

- A numerically equivalent approach was proposed by Johansen (1988).

- He expressed the problem in terms of the eigen values of the likelihood function – an approach that is *numerically equivalent* to the likelihood ratio test. He termed it the "trace statistic".

- The critical values of the LR test are non-standard, and depend on the structure of the deterministic part of the model. Critical values are shown on the next slide.

Critical Values of the Likelihood Ratio Test

Quantiles of tr$M$ based on a sample of size $T = 1000$ and 100000 replications.

		Number of common trends ($N-r$)					
		1	2	3	4	5	6
Model 1	0.90	2.983	10.460	21.677	36.877	56.041	78.991
	0.95	4.173	12.285	24.102	39.921	59.829	83.428
	0.99	6.967	16.380	29.406	46.267	67.174	92.221
Model 2	0.90	7.540	17.869	32.058	50.206	72.448	98.338
	0.95	9.142	20.205	34.938	53.734	76.559	103.022
	0.99	12.733	25.256	41.023	60.943	84.780	112.655
Model 3	0.90	2.691	13.347	26.948	44.181	65.419	90.412
	0.95	3.822	15.430	29.616	47.502	69.293	95.105
	0.99	6.695	19.810	35.130	54.307	77.291	103.980
Model 4	0.90	10.624	23.224	39.482	59.532	83.681	111.651
	0.95	12.501	25.726	42.585	63.336	88.089	116.781
	0.99	16.500	30.855	49.047	70.842	96.726	126.510
Model 5	0.90	2.706	16.090	31.874	51.136	74.462	101.484
	0.95	3.839	18.293	34.788	54.680	78.588	106.265
	0.99	6.648	22.978	40.776	61.744	86.952	115.570

Tests on the Cointegrating Vector (Long-Run Parameters)

- Hypothesis tests on the cointegrating vector, β, constitute tests of long-run economic theories.

- In contrast to the cointegration rank tests, the asymptotic distribution of the Wald, Likelihood Ratio and Lagrange Multiplier tests χ^2 is under the null hypothesis that the restrictions are valid.

Exogeneity

- An important feature of a VECM is that all of the variables in the system are endogenous.

- When the system is out of equilibrium, all the variables interact with each other to move the system back into equilibrium,

- In a VECM, this process occurs (as we saw) through the impact of lagged variables so that $y_{i,t}$ is affected by the lags of the other variables either through the error correction term, u_{t-1}, or through the lags of $\Delta y_{j,t}, j \neq i$

Weak versus Strong Exogeneity

- If the first channel does not exist, i.e., the lagged error correction term does not influence the adjustment process, the variable concerned is said to be *weakly* exogenous.

- If the first and second channels do not exist, then only the lagged values of a variable can be used to explain its changes. In this case, we say that that variable is *strongly* exogenous.

- Strong exogeneity testing is equivalent to Granger causality testing.

Example: Exogeneity

- Consider the bi-variate term structure model with one cointegrating vector.

$$\Delta y_t^{10} = \alpha_1 (y_{t-1}^{10} - \beta_0 - \beta_1 y_{t-1}^{1}) + \sum_{i=1}^{p-1} \gamma_{10,i} \Delta y_{t-i}^{10} + \sum_{i=1}^{p-1} \phi_{10,i} \Delta y_{t-i}^{1} + \varepsilon_t$$

$$\Delta y_t^{1} = \alpha_2 (y_{t-1}^{10} - \beta_0 - \beta_1 y_{t-1}^{1}) + \sum_{i=1}^{p-1} \gamma_{1,i} \Delta y_{t-i}^{10} + \sum_{i=1}^{p-1} \phi_{1,i} \Delta y_{t-i}^{1} + \varepsilon_t$$

- The ten-year interest rate, y_t^{10}, is said to be *weakly exogenous* if $\alpha_1 = 0$
- Strong exogeneity amounts to the requirement that

$$\alpha_1 = 0,\ \phi_{10,i} = 0\ \forall\ i$$

Impulse Response Functions

- The dynamics of a VECM can be investigated using impulse response functions.

- The approach is to re-express the VECM as a VAR, but preserving the implied restrictions on the parameters.

- For example, consider the VECM

$$\Delta y_t = \mu_0 + \mu_1 t + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + v_t$$

Impulse Response Functions: VECM

- This VECM can be expressed as a VAR in levels:

$$y_t = \mu + \sum_{j=1}^{p} \Phi_j y_{t-j} + v_t$$

- subject to the restrictions:

$$\Phi_1 = \alpha\beta' + \Gamma_1 - I_N$$

$$\Phi_j = \Gamma_j - \Gamma_{j-1}, j = 2, 3, \ldots, p$$

IMF INSTITUTE FOR CAPACITY DEVELOPMENT

Appendices

Appendix A: Process moments, key results: AR(1) model with θ < 1

- Mean (first moment):

$$(8) \qquad E[y_t] = \delta \sum_{j=0}^{t-1} \theta^j + \sum_{j=0}^{t-1} \theta^j v_{t-j} + \theta^t y_0 \to \frac{\delta}{1-\theta} \text{ as t} \to \infty$$

- Variance (second moment):

$$(9) \qquad var[y_t] = E[(y_t - E[y_t])^2] = E\left[\sum_{j=0}^{t-1} \left(\theta^j v_{t-j}\right)^2\right] \to \frac{\sigma^2}{1-\theta^2} \text{ as t} \to \infty$$

- Key point to note is that the first and second moments are converging to finite constants.

- So WLLN applies: $\frac{1}{T}\sum_{t=2}^{T} y_{t-1} \xrightarrow{p} \lim_{t\to\infty} E\left[y_t\right]$ and $\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2 \xrightarrow{p} \lim_{t\to\infty} E\left[y_t^2\right]$

- So any estimator based on these quantities should converge in a similar fashion.

Appendix A: Process moments, Simulation of an AR(1) model

- Assume $\delta = 0.0, \phi = 0.8, \sigma^2 = 1.0$

- It follows that $\lim_{t\to\infty} E[y_t] = \dfrac{\delta}{1-\theta} = \dfrac{0.0}{1-0.8} = 0.0$

- Also $\lim_{t\to\infty} \text{var}(y_t) = \dfrac{\sigma^2}{1-\theta^2} = \dfrac{1.0}{1-0.8^2} = 2.778$

- Note that the sample moments converge to these values as the sample size increases. Also, the variance of the estimator is approaching zero as T increases.

	Sample Mean		Sample Variance	
T	Mean	Variance	Mean	Variance
50	-0.001	0.428	2.62	1.24
100	-0.003	0.231	2.701	0.661
200	-0.002	0.12	2.738	0.34
400	-0.002	0.061	2.756	0.173
800	0	0.031	2.767	0.087
1600	0	0.016	2.772	0.044
	50000 replications			

Appendix A: Process moments, key results: AR(1) model with θ = 1

- First moment: $E[y_t] = \theta^t y_0 + \sum_{j=0}^{t-1} \theta^j \delta = y_0 + \delta t$
- Second moment: $\text{var}(y_t) = \sigma^2 \sum_{j=0}^{t-1} \theta^{2j} = \sigma^2 (1 + \theta^2 + \theta^4 + \dots) = \sigma^2 t$
- Appropriate scaling factors for these moments are $T^{-3/2}$ and T^{-2} respectively.
- Define $m_1 = \frac{1}{T^{3/2}} \sum_{t=2}^{T} y_{t-1}, \; m_2 = \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2$ (sample moments)

Appendix A: Process moments, simulation of an I(1) Process

- Notice that the variances of the first two sample moments do not fall as the sample size is increased (Columns 2 and 4).

- The variances converge to 1/3, so m_1 and m_2 converge to *random* variables in the limit.

	Sample Mean (m_1)				Sample Variance (m_2)		
T	Mean		Variance		Mean		Variance
50	-0.001		0.323		0.49		0.317
100	-0.002		0.329		0.496		0.33
200	-0.003		0.331		0.497		0.334
400	-0.002		0.328		0.494		0.324
800	-0.001		0.335		0.501		0.339
1600	0.003		0.336		0.503		0.339
	50000 replications						

Appendix B: Enders Strategy

Estimate $\Delta y_t = \mu_1 + \mu_2 t + \psi\, y_{t-1} + \varepsilon_t$

Test H_0: $\psi=0$
t-ratio test, 5% Crit. value is -3.45

No unit root (y_t is stationary). Additional testing is needed for deterministic components

Test H_0: $\mu_2=\psi=0$
F-test, 5% Crit. value is 6.49

Test H_0: $\psi=0$ using *N*-distribution
t-test, 5% Crit. value is -1.64

Unit root (y_t has both stochastic and deterministic trends).
$y_t = \mu_1 + \mu_2 t + y_{t-1} + \varepsilon_t$

No unit root (y_t is stationary around deterministic trend).
$y_t = \mu_1 + \mu_2 t + \theta y_{t-1} + \varepsilon_t, |\theta|<1$

Estimate $\Delta y_t = \mu_1 + \psi\, y_{t-1} + \varepsilon_t$

Test H_0: $\psi=0$
t-ratio test, 5% Crit. value is -2.89

No unit root (y_t is stationary).
Additional testing of μ_1 is needed

Test H_0: $\mu_1=\psi=0$
F-test, 5% Crit. value is 4.71

Test H_0: $\psi=0$ using *N*-distribution
t-test, 5% Crit. value is -1.64

Unit root (y_t is non-stationary) $y_t = \mu_1 + y_{t-1} + \varepsilon_t$

No unit root (y_t is stationary).
$y_t = \mu_1 + \theta y_{t-1} + \varepsilon_t, |\theta|<1$

Estimate $\Delta y_t = \psi\, y_{t-1} + \varepsilon_t$

Test H_0: $\psi=0$
t-ratio test, 5% Crit. value is -1.64

No unit root (y_t is stationary).
$y_t = \theta y_{t-1} + \varepsilon_t, |\theta|<1$

Unit root (y_t is non-stationary). $y_t = y_{t-1} + \varepsilon_t$

Appendix B: Enders Strategy (2)

- Enders Strategy was criticized for:
  - triple- and double-testing for unit roots
  - unrealistic outcomes: economic variables unlikely contain both stochastic and deterministic trend as in

$$\Delta y_t = \mu_1 + \mu_2 t + \psi\, y_{t-1} + \varepsilon_t\,,\ \mu_2 \neq 0,\ \psi = 0,$$

    this possibility should be excluded from the test

  - not taking advantage of prior knowledge

- Alternative: Elder and Kennedy Strategy

Appendix B: Elder and Kennedy Strategy

Estimate $\Delta y_t = \mu_1 + \mu_2 t + \psi\, y_{t-1} + \varepsilon_t$

Test H_0: $\psi=0$
t-ratio test, 5% Crit. value is -3.45

No unit root (y_t is stationary).

Test H_0: $\mu_2=0$
double sided t-test,
5% Crit. values are $-1.95<t<1.95$

No unit root (y_t is stationary without deterministic trend):
$y_t = \mu_1 + \theta y_{t-1} + \varepsilon_t, |\theta|<1$

No unit root (y_t is stationary around deterministic trend).
$y_t = \mu_1 + \mu_2 t + \theta y_{t-1} + \varepsilon_t, |\theta|<1$

Unit root (y_t is non-stationary).

Estimate $\Delta y_t = \mu_1 + \varepsilon_t$

Test H_0: $\mu_1=0$
double sided t-test,
5% Crit. values are $-1.95<t<1.95$

Unit root (y_t is non-stationary without intercept):
$y_t = y_{t-1} + \varepsilon_t$

Unit root (y_t is non-stationary with intercept).
$y_t = \mu_1 + y_{t-1} + \varepsilon_t$

Nonstationary Asymptotics

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr$$

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \sigma^2 \int_0^1 W(r)dW(r)$$

where $W(r)$ denotes a standard Brownian motion (Wiener process) defined on the unit interval. Using the above results Phillips showed that under the unit root null $H_0 : \phi = 1$

$$T(\hat{\phi} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} \tag{4.1}$$

$$t_{\phi=1} \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}} \tag{4.2}$$

Nonstationary Asymptotics

The above yield some surprising results:

- $\hat{\phi}$ is *super-consistent*; that is, $\hat{\phi} \xrightarrow{p} \phi$ at rate T instead of the usual rate $T^{1/2}$.

- $\hat{\phi}$ is not asymptotically normally distributed and $t_{\phi=1}$ is not asymptotically standard normal.

- The limiting distribution of $t_{\phi=1}$ is called the Dickey-Fuller (DF) distribution and does not have a closed form representation. Consequently, quantiles of the distribution must be computed by numerical approximation or by simulation[3].

- Since the *normalized bias* $T(\hat{\phi} - 1)$ has a well defined limiting distribution that does not depend on nuisance parameters it can also be used as a test statistic for the null hypothesis $H_0 : \phi = 1$.

Source: *faculty.washington.edu/ezivot/econ584/notes/**unitroot**.pdf*