# ES1004 Econometrics by Example

## Lecture 10: Cointegration and Error Correction

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition [chapter 14]

1st October 2016

# Time Series Econometrics

13. stationary and nonstationary time series
14. cointegration and error correction models
15. asset price volatility: the ARCH and GARCH models
16. economic forecasting

- previous course on time series econometrics

ES1002 Lectures

ES1002 EViews

The Economic Society

# Stationarity

- regression analysis of time series data assumes that series are stationary
  - its mean and variance are constant over time
  - covariance depends only on the distance between the two periods and not on time
  - a time series with these characteristics is know as weakly or covariance stationary

# Nonstationarity and Superior Regression

- regressions of nonstationary time series may result in
  - a high $R^2$ value
  - statistically significant regression coefficients
- these results are more likely to be misleading or spurious
  - regressions of trending variables often give sginficant t and F statistics and a high $R^2$
  - there is no real relationship between them because each variable is growing over time
  - usually associated with very low durbin-watson $d$ statistic $[R^2 > d]$
  - a false, spurious, misleading regression

# Superior Regression: Example

- egyptian infant mortality rate (Y), 1971-1990, on gross aggregate income of american farmers (I) and total Honduran money supply (M)

$$\hat{Y} = \underset{(16.63)}{179.9} - \underset{(-2.32)}{.2952}\, I - \underset{(-4.26)}{.0439}\, M, \;\; D/W = .4752, F = 95.17$$

$$Corr = .8858, -.9113, -.9445$$

$t$ ratios in parentheses

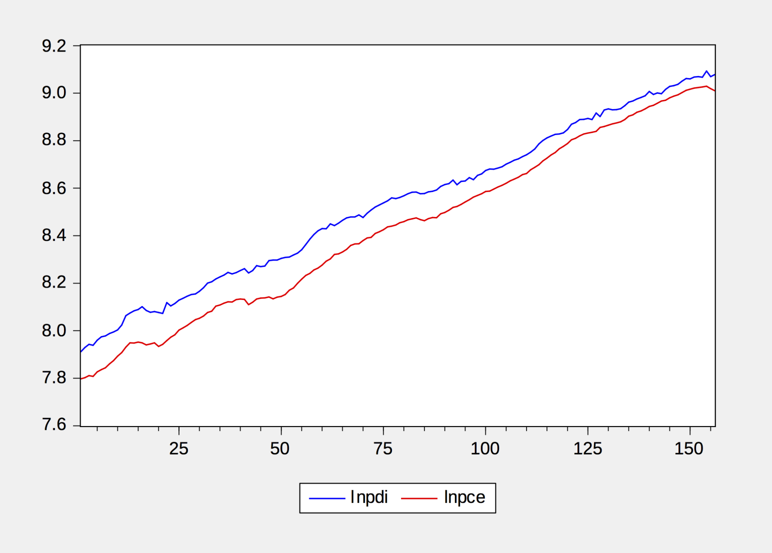
- no logical reason for the observed relationship among the variables
- variables seem to be trending over time
- see more examples of spurious regression [click here]

The Economic Society

# Consumption Expenditure on Disposable Income

- `table14.1` usa quarterly data 1970-2008 [156 observations]
- personal consumption expenditure and personal disposable income
- data are in billions of 2000 dollars

# PCE and PDI Plot I



# PCE and PDI Plot II

- the figure shows that both lnpdi and lnpce are trending series
- suggests that theses series are not stationary
- can be confirmed by unit root analysis (e.g., ADF test)
- both series has have a unit root, or stochastic trend

# Nonstationary PDI

Null Hypothesis: LNPDI has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on AIC, maxlag=1)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.774819 | 0.2089 |
| Test critical values: | 1% level | -4.018748 | |
| | 5% level | -3.439267 | |
| | 10% level | -3.143999 | |

*MacKinnon (1996) one-sided p-values.

# Nonstationary PCE

Null Hypothesis: LNPCE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on AIC, maxlag=1)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.038426 | 0.5754 |
| Test critical values: | 1% level | -4.018748 | |
| | 5% level | -3.439267 | |
| | 10% level | -3.143999 | |

*MacKinnon (1996) one-sided p-values.

# PCE and PDI Regression

Dependent Variable: LNPCE
Method: Least Squares
Date: 09/30/16 Time: 17:47
Sample: 1 156
Included observations: 156

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -0.842510 | 0.033717 | -24.98751 | 0.0000 |
| LNPDI | 1.086822 | 0.003949 | 275.2418 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.997971 | Mean dependent var | 8.430699 |
| Adjusted R-squared | 0.997958 | S.D. dependent var | 0.366642 |
| S.E. of regression | 0.016567 | Akaike info criterion | -5.350030 |
| Sum squared resid | 0.042269 | Schwarz criterion | -5.310929 |
| Log likelihood | 419.3024 | Hannan-Quinn criter. | -5.334149 |
| F-statistic | 75758.02 | Durbin-Watson stat | 0.367188 |
| Prob(F-statistic) | 0.000000 | | |

# PCE and PDI Regression

- notice that $R^2 > d = 0.3672$
  - this raises the possibility that this regression might be spurious
  - due to regressing stochastic trend series
- durbin-watson $d$ statistic suggests that the error term suffers from first-order autocorrelation
- elasticity of personal consumption expenditure of 1.08 with respect to PDI [greater than 1 which seems high]

# PCE and PDI Regression

- both series are trending
- let us add a trend variable
  - a catch-all for all other variables might affect both regressand and regressors
  - e.g., as population increases both PCE and PDI also increase
  - note that we could have either added population as an additional regressor or preferably to consider PCE and PDI on a per capita basis

# PCE and PDI Regression with A Trend

Dependent Variable: LNPCE
Method: Least Squares
Date: 10/01/16 Time: 15:32
Sample: 1 156
Included observations: 156

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.672968 | 0.487339 | 3.432860 | 0.0008 |
| LNPDI | 0.770241 | 0.061316 | 12.56179 | 0.0000 |
| TIME | 0.002366 | 0.000457 | 5.172271 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.998273 | Mean dependent var | 8.430699 |
| Adjusted R-squared | 0.998251 | S.D. dependent var | 0.366642 |
| S.E. of regression | 0.015335 | Akaike info criterion | -5.498352 |
| Sum squared resid | 0.035978 | Schwarz criterion | -5.439701 |
| Log likelihood | 431.8715 | Hannan-Quinn criter. | -5.474531 |
| F-statistic | 44226.64 | Durbin-Watson stat | 0.261692 |
| Prob(F-statistic) | 0.000000 | | |

# PCE and PDI Regression with A Trend

- the elasticity of PCE with respect to PDI is now less than unity
- the trend variable is statistically significant
- allowing for linear trend, the relationship between the two variables is strongly positive
- but notice again the low $d$ value, which suggests that results are plagued by autocorrelation
- or maybe this regression too is spurious

# Stationary Error

$$LPCE_t = \beta_1 + \beta_2 LPDI_t + \beta_3 t_t + u_t$$

rewrite this model as

$$u_t = LPCE_t - \beta_1 - \beta_2 LPDI_t - \beta_3 t_t$$

- suppose after estimating the first eq. we found that $u_t (= e_t)$ to be stationary i.e., $I(0)$
  - although LPCE and LPDI are individually $I(1)$ , they have stochastic trends, their combination shown in the second equation is $I(0)$
  - their linear combination cancels out the stochastic trends in the two series

# Stationary Error

- in that case, the regression of LPCE and LPDI is not spurious - they are cointegrated
- this can be seen clearly in the graph
  - although both series are stochastically trending, they do not drift apart substantially
- two variables will be cointegrated if they have a long-run or equilibrium, relationship between them
  - according to economic theory, there is a strong relationship between consumption and disposable income

# Cointegrating Regression

$$LPCE_t = \beta_1 + \beta_2 LPDI_t + \beta_3 t_t + u_t$$

- cointegrating regression
- slope parameters $\beta_2 1$ and $\beta_3$ are cointegrating parameters

# Engle-Granger Test

- there are several tests of cointegration but we will consider only EG test
- implement ADF unit root tests on the residuals estimated for the cointegration regression

# Engle-Granger: Example

Null Hypothesis: E has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=0)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -3.392600 | 0.0008 |
| Test critical values: | 1% level | -2.579967 | |
| | 5% level | -1.942896 | |
| | 10% level | -1.615342 | |

*MacKinnon (1996) one-sided p-values.

# Unit Root and Cointegraions Tests

- tests for unit roots are performed on single time series
- cointegration deals with the relationship among a group of variables, each having a unit root
- it is better to test each series for unit roots
    - some of the series may have more than one unit root
      $\rightarrow$ need to be differenced more than once to become stationary

# Unit Root and Cointegraions Tests

- if two time series $Y$ and $X$ are integrated of different orders then the error term in the regression of $Y$ and $X$ is not stationary
  - this regression equation is said to be unbalanced
- if the two variables are integrated of the same order, the regression equation is said to be balanced

# Equilibrating Error

- LPCE and LPDI are cointegrated i.e., have a long-term, or equilibrium, relationship
- how is this equilibrium achieved?

$$u_t = LPCE_t - \beta_1 - \beta_2 LPDI_t - \beta_3 t_t$$

- the 'equilibrating' error term that corrects deviations of LPCE from its equilibrium value given by the cointegration regression

# Granger Representation Theorem

- if two variables $Y$ and $X$ are cointegrated, the relationship between the two can be expressed as an error correction mechanism (ECM)

$$\Delta LPCE_t = A_1 + A_2 \Delta LPDI_t + A_3 u_{t-1} + v_t$$

$\Delta$ the first difference operator
$u_{t-1}$ the lagged value of the error correction term
$v_t$ a white noise error term

# Granger Representation Theorem

$$LPCE_t = \beta_1 + \beta_2 LPDI_t + \beta_3 t_t + u_t$$

- the long-run relationship: $\beta_2$ gives the long-run impact of LPDI on LPCE

$$\Delta LPCE_t = A_1 + A_2 \Delta LPDI_t + A_3 u_{t-1} + v_t$$

- the short-run relationship: $A_2$ gives the immediate, or short-run, impact of $\Delta LPDI$ on $\Delta LPCE$

# Granger Representation Theorem

$$\Delta LPCE_t = A_1 + A_2 \Delta LPDI_t + A_3 u_{t-1} + v_t$$

- ECM: changes in LPCE depend on changes on LPDI and the lagged equilibrium error term $u_{t-1}$
- if this error term $u_{t-1}$ is zero
  - there will not be any disequilibrium between the two variables
  - long-run relationship will be given by the cointegrating relationship
- but it the error term $u_{t-1}$ is nonzero
  - the relationship between LPCE and LPDI will be out of equilibrium

# Granger Representation Theorem

$$\Delta LPCE_t = A_1 + A_2 \Delta LPDI_t + A_3 u_{t-1} + v_t$$

- the absolute value $A_3$ will decide how quickly the equilibrium is reached
- the incorporates both the short-run and long-run dynamics
- all variables are $I(0)$ stationary so we can estimate the equation using OLS

# ECM: Example

Dependent Variable: D(LNPCE)
Method: Least Squares
Date: 10/01/16 Time: 17:33
Sample (adjusted): 2 156
Included observations: 155 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.005530 | 0.000626 | 8.831326 | 0.0000 |
| D(LNPDI) | 0.306401 | 0.051589 | 5.939318 | 0.0000 |
| E(-1) | -0.065247 | 0.033504 | -1.947407 | 0.0533 |

| | | | |
|---|---|---|---|
| R-squared | 0.189780 | Mean dependent var | 0.007825 |
| Adjusted R-squared | 0.179120 | S.D. dependent var | 0.006764 |
| S.E. of regression | 0.006129 | Akaike info criterion | -7.332568 |
| Sum squared resid | 0.005709 | Schwarz criterion | -7.273663 |
| Log likelihood | 571.2740 | Hannan-Quinn criter. | -7.308642 |
| F-statistic | 17.80173 | Durbin-Watson stat | 1.707361 |
| Prob(F-statistic) | 0.000000 | | |

The Economic Society

# ECM: Example

- all coefficients are individually statistically significant at the 6% or lower level
- short-run consumption-income elasticity 0.31%
- the long-run value is given by the cointegrating regression, which is about 0.77
- the coefficient of the error-correction term $-0.06$
    - only about 6% of the discrepancy between long-term and short-term PCE is corrected within a quarter
    - i.e., a slow rate of adjustment to equilibrium

The Economic Society

# Engle-Granger Shortcomings

- if you have more than three variables, there might be more than one cointegrating relationship
- once we go beyond two time series, we will have to use Johansen methodology to test for cointegrating relationships among multiple variables.