Lecture 9: Nonstationarity. Error Correction Models

Econometric Methods

Andrzej Torój

SGH Warsaw School of Economics – Institute of Econometrics

Outline

Outline

Time series – definition

Y_t – random variable – takes values with some probabilities (described by density function $f(y)$/ distribution function $F(y)$)

values + probabilities: **distribution of a random variable**

$\{Y_t\}$ – stochastic process – sequence of random variables Y_t ordered by time

$\{y_t\}$ – time series – draw from a stochastic process in one sample

Key parameters of random variable's distribution:

Expected value: $E(Y) = \int_{-\infty}^{+\infty} yf(y)\,dy$
Variance: $D^2(Y) = E(Y - E(Y))^2$

Stationarity

Type I (in the strict sense / strong)

Distribution of the stochastic process is time-invariant (in every period, y_t is a value taken by identically distributed random variable Y_t)

Type II (in the large sense / weak)

– mean and variance constant over time$E(Y_t) = \mu < \infty D^2(Y_t) = \sigma^2 < \infty$
– covariance between variables depends on their distance in time (not the moment in time)$Cov(Y_t, Y_{t+h}) = Cov(Y_{t+k}, Y_{t+k+h}) = \gamma(h)$

„White noise"

„White noise" – definition

$E(\varepsilon_t) = 0$ [fluctuations around zero]
$D^2(\varepsilon_t) = \sigma^2 < \infty$ [homoskedasticity]
$Cov(\varepsilon_t, \varepsilon_{t+h}) = 0, \;\; h \neq 0$ [no serial correlation]
...like the disturbances in the classical linear regression model.

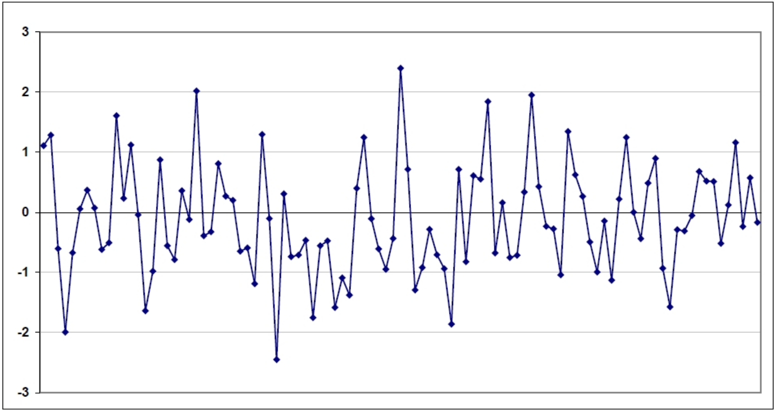
$\varepsilon_t \sim IID(0, \sigma^2)$

I – independent

I – indentically

D – distributed

Stationary series – „white noise”



Nonstationary process – "random walk"

Random walk – definition

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

$y_t = y_{t-1} + \varepsilon_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \ldots =$

$y_0 + \sum_{t=1}^{T} \varepsilon_t \overset{y_0=0}{=} \underbrace{\sum_{t=1}^{T} \varepsilon_t}_{\text{stochastic trend}}$

Properties of random walk:

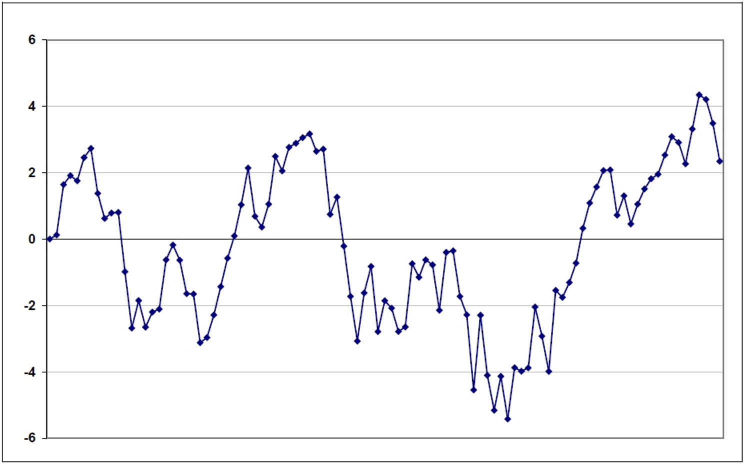
$E(y_t) = E(\sum_{t=1}^{T} \varepsilon_t) = \sum_{t=1}^{T} E(\varepsilon_t) = 0$

$D^2(y_t) = D^2(\sum_{t=1}^{T} \varepsilon_t) \overset{cov(\varepsilon_t, \varepsilon_{t-h})=0}{=} \sum_{t=1}^{T} D^2(\varepsilon_t) = T\sigma^2$

$cov(y_t, y_{t-h}) = E(y_t, y_{t-h}) - E(y_t)E(y_{t-h}) =$

$E(\sum_{t=1}^{T-h} \varepsilon_t \sum_{t=1}^{T-h} \varepsilon_t) - \underbrace{E(\sum_{t=1}^{T-h})}\underbrace{E(\sum_{t=1}^{T-h})} = D^2(\sum_{t=1}^{T-h} \varepsilon_t) = \sum_{t=1}^{T-h} D^2(\varepsilon_t) = (T-h)\sigma^2$

Nonstationary series – random walk



Integration order

Stationary variable is called **integrated of order 0**, notation: I(0).

Integration order – definition

Variable y_t is integrated of order d ($y_t \sim I(d)$) if it can be transformed into a stationary variable after differencing d times.

E.g. variable y_t generated by the process $y_t = y_{t-1} + \varepsilon_t$ is integrated of order 1 ($y_t \sim I(1)$), as $y_t - y_{t-1} = \varepsilon_t$, and $\varepsilon_t \sim I(0)$ by definition.

First differences: $\Delta y_t = y_t - y_{t-1}$

Second differences:
$\Delta\Delta y_t = \Delta y_t - \Delta y_{t-1} = y_t - y_{t-1} - y_{t-1} + y_{t-2} = y_t - 2y_{t-1} + y_{t-2}$
("filter (1,-2,1)")

Nonstationary process – random walk with drift

Random walk with drift – definition

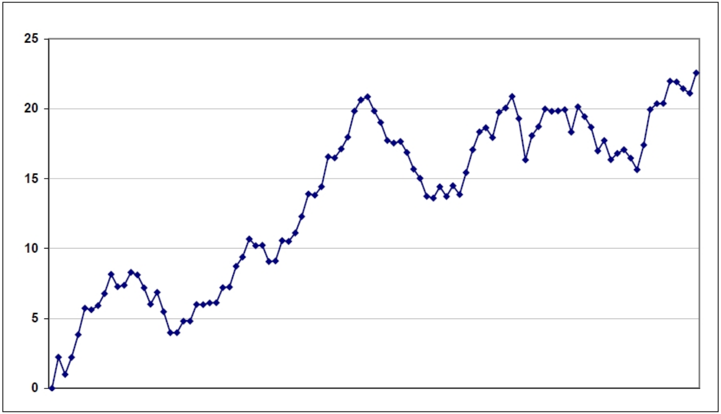
$$y_t = \alpha_0 + y_{t-1} + \varepsilon_t$$
$$y_t = \alpha_0 + y_{t-1} + \varepsilon_t = \alpha_0 + \alpha_0 + y_{t-2} + \varepsilon_{t-1} + \varepsilon_t =$$
$$\alpha_0 + \alpha_0 + \alpha_0 + y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \ldots \overset{y_0=0}{=} T\alpha_0 + \underbrace{\sum_{t=1}^{T} \varepsilon_t}_{\text{stochastic trend}}$$

Random walk with drift – properties:

$E(y_t) = E(T\alpha_0 + \sum_{t=1}^{T} \varepsilon_t) = T\alpha_0 + \sum_{t=1}^{T} E(\varepsilon_t) = T\alpha_0$

Variance and covariance – the same as in the case of random walk (adding a constant does not affect the dispersion).

Nonstationary variable – random walk with drift



Order of integration – why it matters

- possible problems:
  - regression I(1) vs I(1) – **spurious regression** (trending variables)
  - regression I(0) vs I(1) – **nonstationary residuals** (economically unlikely), **mistakes in statistical inference** (test statistic for variable significance is not t-distributed)
- solutions:
  - transformation of the series (differencing, logarithm)
  - respecification of the model (for consistency with the theory)
  - error correction and cointegration

Order of integration – why it matters

- possible problems:
  - regression I(1) vs I(1) – **spurious regression** (trending variables)
  - regression I(0) vs I(1) – **nonstationary residuals** (economically unlikely), **mistakes in statistical inference** (test statistic for variable significance is not t-distributed)
- solutions:
  - transformation of the series (differencing, logarithm)
  - respecification of the model (for consistency with the theory)
  - error correction and cointegration

Outline

Dickey-Fuller (DF)

Dickey and Fuller, 1979, 1981

$y_t = \alpha_1 y_{t-1} + \varepsilon_t$

- process is stationary if $\alpha_1 < 1$
- process is nonstationary if $\alpha_1 = 1$ (then random walk)

$H_0 : \alpha_1 = 1$

$H_1 : \alpha_1 < 1$

- true $H_0 \rightarrow$ nonstationary variables $\rightarrow$ bias in OLS $\rightarrow$ hypothesis not verifiable immediately

DF – test statistic

Subtract y_{t-1} from both sides – potentially stationary dependent variable:

$$\Delta y_t = \underbrace{(\alpha_1 - 1)}_{\delta} y_{t-1} + \varepsilon_t$$

$$H_0 : \delta = 0 \iff \alpha_1 = 1 \iff y_t \sim I(1)$$
$$H_1 : \delta < 0 \iff \alpha_1 < 1 \iff y_t \sim I(0)$$

$$DF^{emp} = \frac{\hat{\delta}}{\hat{S}_{\delta}} \sim DF$$

(computed as t-statistic, but with different distribution – see ***MacKinnon (1996)***)

If $DF^{emp} < DF^*$, **reject** H_0 against H_1 and consider the process **stationary**.

DF – algorithm

H_0 not rejected:

- nonstationarity, but unknown order of integration (1 or higher)...

Test again...

- variable is integrated of order 2 ($y_t \sim I(2)$), if stationary after double differentiation

$$\Delta(\Delta y_t) = \delta \Delta y_{t-1} + \varepsilon_t$$

$$H_0 : \delta = 0 \iff y_t \sim I(2)$$

$$H_1 : \delta < 0 \iff y_t \sim I(1)$$

$$DF^{emp} = \frac{\hat{\delta}}{\hat{S}_{\delta}} \sim DF$$

DF – I(2) series

If again H_0 not rejected:

- variable integrated of order 2 or not rejecting H_0 due to low power of the test...

Test again...

$\Delta^3 y_t = \delta \Delta^2 y_{t-1} + \varepsilon_t$

$H_0: \delta = 0 \iff y_t \sim I(3)$

$H_1: \delta < 0 \iff y_t \sim I(2)$

$$DF^{emp} = \frac{\hat{\delta}}{\hat{S}_\delta} \sim DF$$

- failure to reject the null again – too weak power of the test (too rarely rejects false null hypothesis)
- in economics, series integrated of order higher than 2 generally absent

Augmented DF

Said and Dickey, 1985

- The power of DF test substantially lower under serial correlation of residuals in the test regression.
- This serial correlation should be handled.
- The simplest solution: dynamise the model by supplementing lags of the dependent variable.

Augmented Dickey-Fuller test – ADF

$$\Delta y_t = \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \ldots + \gamma_k \Delta y_{t-k} + \varepsilon_t$$

ADF – how many lags?

- in general: the purpose is to eliminate the serial correlation of the error term
- CAUTION! we do not use DW statistic to evaluate it (remember why...?)
- auxiliary algorithms: set the maximum lag length to consider and...
  - ...pick the best regression by means of information criteria (AIC, SIC, HQC)
  - ...see if the last significant; if not, remove it and check again
- rule-of-thumb formula for maximum lag length: $\left(4 \cdot \frac{T}{100}\right)^{\frac{1}{4}}$, where T – sample size (*Schwert, 2002*)

ADF – how many lags?

- in general: the purpose is to eliminate the serial correlation of the error term
- CAUTION! we do not use DW statistic to evaluate it (remember why...?)
- auxiliary algorithms: set the maximum lag length to consider and...
  - ...pick the best regression by means of information criteria (AIC, SIC, HQC)
  - ...see if the last significant; if not, remove it and check again
- rule-of-thumb formula for maximum lag length: $\left(4 \cdot \frac{T}{100}\right)^{\frac{1}{4}}$, where T – sample size (*Schwert, 2002*)

ADF – specification of test regression

$$\Delta y_t = \delta y_{t-1} \left(+ \sum_{i=1}^{k} \gamma_i \Delta y_{t-i} \right) + \varepsilon_t$$

possible extensions:

$$\Delta y_t = \beta + \delta y_{t-1} + \left(\sum_{i=1}^{k} \gamma_i \Delta y_{t-i} \right) + \varepsilon_t$$

$\rightarrow H_0$: nonstationary process with drift

additionally: linear trend, quadratic trend, seasonal dummies...

(A)DF – trendstationarity

$$\Delta y_t = \beta + \delta y_{t-1} + \left(\sum_{i=1}^{k} \gamma_i \Delta y_{t-i} \right) + \varepsilon_t$$

extension:

$$\Delta y_t = \beta + \delta y_{t-1} + \left(\sum_{i=1}^{k} \gamma_i \Delta y_{t-i} \right) + \gamma t + \varepsilon_t$$

$\rightarrow$ if H_0 rejected and trend t significant – **trendstationary** series (then include t in the regression)

$\rightarrow$ if H_0 not rejected, series is nonstationary

Beware the difference!

stationary process $\neq$ deterministic trend $\neq$ stochastic trend

Exercise (1)

Test the order of integration of the following variables with ADF test:
– log real wages
– log labour productivity
– unemployment

Outline

Nonstationarity – what then?

- dealing with models based on nonstationary variables:
  - OLS: spurious regression
  - explosive behaviour of dynamic models
  - I(0)+I(1): nonstationary residuals + false statistical inference
- solution: differencing I(1) data
  - OLS, dynamic models: adequate tools
  - **LONG-TERM RELATIONSHIPS** lost from the data

Nonstationarity – what then?

- dealing with models based on nonstationary variables:
  - OLS: spurious regression
  - explosive behaviour of dynamic models
  - I(0)+I(1): nonstationary residuals + false statistical inference
- solution: differencing I(1) data
  - OLS, dynamic models: adequate tools
  - **LONG-TERM RELATIONSHIPS** lost from the data

Cointegration of time series

- dependencies between nonstationary variables – sometimes stable in time
- ...then called **COINTEGRATING RELATIONSHIPS**
- there is a mechanism that brings the system back to equilibrium everytime it is shocked away from it (**Granger theorem**)

Cointegration

Time series y_1 and y_2 are cointegrated (of order d, b), if they are integrated of order d and there exists their linear combination integrated of order $d - b$:

$$y_1, y_2 \sim CI(d, b) \iff y_1 \sim I(d) \wedge y_2 \sim I(d) \wedge \exists_{\beta \neq \mathbf{0}} \underbrace{\mathbf{y}^T \boldsymbol{\beta}}_{y_1\beta_1 + y_2\beta_2} \sim I(d-b)$$

Usually: variables integrated of order 1, their combination – stationary.

Testing for cointegration: Engle-Granger procedure

1. Test the order of integration.
2. If all of them, say, I(1), specify the cointegrating relationship: $y_{1t} = \beta_0 + \beta_1 y_{2t} + \epsilon_t$
    1. estimate parameters β_0, β_1 via OLS
    2. compute the residual series $(\hat{\epsilon}_t)$
3. Test these residuals for stationarity – if stationary, variables are cointegrated.

Testing for cointegration: Engle-Granger procedure

1. Test the order of integration.
2. If all of them, say, I(1), specify the cointegrating relationship: $y_{1t} = \beta_0 + \beta_1 y_{2t} + \epsilon_t$
   1. estimate parameters β_0, β_1 via OLS
   2. compute the residual series ($\hat{\epsilon}_t$)
3. Test these residuals for stationarity – if stationary, variables are cointegrated.

Testing for cointegration: Engle-Granger procedure

1. Test the order of integration.
2. If all of them, say, I(1), specify the cointegrating relationship: $y_{1t} = \beta_0 + \beta_1 y_{2t} + \epsilon_t$
   1. estimate parameters β_0, β_1 via OLS
   2. compute the residual series ($\hat{\epsilon}_t$)
3. Test these residuals for stationarity – if stationary, variables are cointegrated.

Outline

Error correction model (1)

- some "error correction" mechanism directly implied by the Granger theorem
- recall the ADL(1,1) model:
  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$
  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t + y_{t-1} - y_{t-1} + x_{t-1} - x_{t-1} + \beta_0 x_{t-1} - \beta_0 x_{t-1} + \alpha_1 x_{t-1} - \alpha_1 x_{t-1}$
- rearranging terms, we obtain the error correction model:

$$\Delta y_t = (\alpha_1 - 1)\underbrace{(y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} x_{t-1})}_{ECT:\ \hat{\epsilon_{t-1}}} + \beta_0 \Delta x_t + \varepsilon_t$$

Error correction model (1)

- some "error correction" mechanism directly implied by the Granger theorem
- recall the ADL(1,1) model:
  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$
  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} +$
  $\varepsilon_t + y_{t-1} - y_{t-1} + x_{t-1} - x_{t-1} + \beta_0 x_{t-1} - \beta_0 x_{t-1} + \alpha_1 x_{t-1} - \alpha_1 x_{t-1}$
- rearranging terms, we obtain the error correction model:

$$\Delta y_t = (\alpha_1 - 1)\underbrace{(y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} x_{t-1})}_{ECT:\ \hat{\epsilon_{t-1}}} + \beta_0 \Delta x_t + \varepsilon_t$$

Error correction model (1)

- some "error correction" mechanism directly implied by the Granger theorem
- recall the ADL(1,1) model:
  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$
  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} +$
  $\varepsilon_t + y_{t-1} - y_{t-1} + x_{t-1} - x_{t-1} + \beta_0 x_{t-1} - \beta_0 x_{t-1} + \alpha_1 x_{t-1} - \alpha_1 x_{t-1}$
- rearranging terms, we obtain the error correction model:

$$\Delta y_t = (\alpha_1 - 1)\underbrace{(y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1} x_{t-1})}_{ECT:\ \hat{\epsilon_{t-1}}} + \beta_0 \Delta x_t + \varepsilon_t$$

Error correction model (2)

Procedure:

1. estimate the cointegrating relationship
2. estimate the ECM, using differenced variables and lagged residuals from the cointegrating relationship

- variables are cointegrated when $(\alpha_1 - 1) < 0$
  - > 0: disequilibrium expands
  - $= 0$: no error correction
  - $(-1; 0)$: **error correction** (close to 0: slow, close to -1: quick; $half - life = \frac{\ln(0.5)}{ln(\alpha_1)}$)
  - $= -1$: full error correction in 1 period
  - < -1: overshooting, oscillatory adjustment

Error correction model (2)

Procedure:

1. estimate the cointegrating relationship
2. estimate the ECM, using differenced variables and lagged residuals from the cointegrating relationship

- variables are cointegrated when $(\alpha_1 - 1) < 0$
  - > 0: disequilibrium expands
  - $= 0$: no error correction
  - $(-1; 0)$: **error correction** (close to 0: slow, close to -1: quick; $half - life = \frac{\ln(0.5)}{ln(\alpha_1)}$)
  - $= -1$: full error correction in 1 period
  - < -1: overshooting, oscillatory adjustment

Error correction model (2)

Procedure:

1. estimate the cointegrating relationship
2. estimate the ECM, using differenced variables and lagged residuals from the cointegrating relationship

- variables are cointegrated when $(\alpha_1 - 1) < 0$
  - > 0: disequilibrium expands
  - $= 0$: no error correction
  - $(-1; 0)$: **error correction** (close to 0: slow, close to -1: quick; $half - life = \frac{\ln(0.5)}{ln(\alpha_1)}$)
  - $= -1$: full error correction in 1 period
  - < -1: overshooting, oscillatory adjustment

Exercise (2)

Consider the following variables:

– log real wages (dependent)

– log labour productivity

– unemployment

What is their order of integration?

Are they cointegrated?

Are the signs in the cointegrating relationship economically reasonable?

Is there error correction mechanism at work?

What is the half-life of the adjustment?

Example inspired by *Zasova, Melihovs (2005)*

Readings

- Greene:
  - Chapter 20 – Models With Lagged Variables
  - Chapter 21 - Time-Series Models
  - Chapter 22 - Nonstationary Data