Quantitative Analysis of Finance I

ECON90033

WEEK 11

COINTEGRATION (cont.)

EQUILIBRIUM DYNAMICS AND ERROR CORRECTION

COINTEGRATION TESTING

Reference:

HMPY: § 6.1-6.5

Dr László Kónya
January 2023

COINTEGRATION (cont.)

- As we discussed in the last lecture, variables that are individually nonstationary, e.g. *I*(1), might have a stationary, i.e., *I*(0), linear combination, called cointegrated relationship.

  We illustrated the concept of cointegration with two examples. Let's now consider a third one.

<u>Ex 1</u>:

Consider the present value model of equities.

According to this model, the price (*P*) of an equity is equal to the conditional expected value of the discounted future stream of dividend payments (*D*), i.e.,

$$P_t = E\left(\frac{D_{t+1}}{1+\delta_t} + \frac{D_{t+2}}{(1+\delta_t)^2} + \frac{D_{t+3}}{(1+\delta_t)^3} + \ldots \mid \Omega_t \right)$$

where δ_t is the discount rate and Ω_t is the information set at time *t*.

Assuming that the conditional expectations of future dividends are the same as the present dividend, i.e., $E(D_{t+i} \mid \Omega_t) = D_t$,

$$P_t = \frac{D_t}{1+\delta_t}\left(1 + \frac{1}{1+\delta_t} + \frac{1}{(1+\delta_t)^2} + \dots\right) = \frac{D_t}{1+\delta_t}\sum_{i=0}^{\infty}\frac{1}{(1+\delta_t)^i}$$

$$= \frac{D_t}{1+\delta_t}\frac{1}{1-\frac{1}{1+\delta_t}} = \frac{D_t}{1+\delta_t}\frac{1+\delta_t}{1+\delta_t-1} = \frac{D_t}{\delta_t} \qquad (\delta_t > 0)$$

$$\downarrow$$

$$\ln P_t = -\ln\delta_t + \ln D_t$$

$\longrightarrow$ The corresponding statistical model is

$$\ln P_t = \beta_0 + \beta_1 \ln D_t + \varepsilon_t \longrightarrow p_t = \beta_0 + \beta_1 d_t + \varepsilon_t$$

where β_0 and β_1 are treated as unknown parameters and ε_t is assumed to be a white noise error term.

i.e. setting up for ECM

If the theory is correct, any deviation from the mean, i.e., from the long-run equilibrium, must be temporary in nature, implying that ε_t is stationary. By contrast, if ε_t were a random walk, it would have a stochastic trend and there would be no tendency for p_t to return to $E(p_t)$.

need to verify in practice

$$\varepsilon_t = p_t - \beta_0 - \beta_1 d_t \quad : I(0)$$

Yet, like many macroeconomic and finance variables, p_t and d_t might be $I(1)$. However, an $I(0)$ variable cannot be equal to an $I(1)$ variable (unbalanced equation). Hence, $\varepsilon_t : I(0)$ if and only if the linear combination of p_t and d_t, defined by the right side of this equality, is stationary.

↳ i.e. If there is a genuine equil. r/ship

- In general, equilibrium refers to a state in which there is no tendency to change.

Consequently, equilibrium theories involving variables x_{1t}, x_{2t}, …, x_{nt} that are likely random walks require that these variables have a stationary linear combination,

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_n x_{nt} = 0$$

} deterministic def of stationarity in a system

← In matrix form $\boxed{\boldsymbol{\beta}'\mathbf{x}_t = 0}$ where **β** and $\mathbf{x}_t$ are (n×1) vectors.

However, an economic system never settles down to such a state, there is always some ever-changing deviation from equilibrium.

For this reason, equilibrium is meant to be a long-run (or permanent) relationship to which the system tends to return time to time.

The deviation from this long-run relationship is called equilibrium error.

If equilibrium is meaningful, any deviation from it should be temporary (or transient), so the equilibrium error must be stationary.

defining cointegration

- Variables $x_{1t}, x_{2t}, \ldots, x_{nt}$ are said to be cointegrated of order (d,b), denoted as $(x_{1t}, x_{2t}, \ldots, x_{nt}) \sim CI\,(d,b)$, if they satisfy two conditions:

  i. Each of them is integrated of order d; → $I_a(n) = I_b(n)$

  ii. They have at least one non-trivial linear combination that is integrated of order (d - b), where $d \geq b > 0$.
  → Some $\beta_i \neq 0$

In the special though frequent case of $d = b = 1$, each x_{it} is a random walk, I(1), but the variables have a stationary, I(0), linear combination.

Level of cointegration must be lower than I(n) for each variable

- Vector $\mathbf{\beta} = [\beta_1, \beta_2, \ldots, \beta_n]'$ that defines the cointegrating linear combination of variables $(x_{1t}, x_{2t}, \ldots, x_{nt})$ is called cointegration vector.

  $\mathbf{\beta}$ is not unique, because if it is a cointegration vector, so is $\lambda\mathbf{\beta}$ for any $\lambda \neq 0$.

  $\longrightarrow$ This ambiguity can be avoided by normalizing the cointegration vector, i.e., by selecting one of the variables, say x_{1t}, and setting its coefficient equal to one ($\lambda = 1/\beta_1$ and $\beta_1^* = \lambda\beta_1 = 1$).

(Ex 1)

In our example, $\boxed{p_t = \beta_0 + \beta_1 d_t + \varepsilon_t}$

$\longrightarrow$ $\mathbf{x}_t = [p_t, 1, d_t]'$. Given that each variable in $\mathbf{x}_t$ is $I(1)$ but ε_t is $I(0)$, the components of $\mathbf{x}_t$ are $CI(1,1)$ and the cointegrating vector is $\mathbf{\beta} = [1, -\beta_0, -\beta_1]'$, i.e., it is normalized with respect to p_t.

- There are several important lemmas concerning cointegration.

For example, it can be shown that

i. If x_{1t}, x_{2t} are $CI(1,1)$, then so are x_{1t} and $x_{2,t-i}$ are for any $i = 1, 2, \ldots$.

ii. Up to a scalar, cointegrated variables share the same stochastic trend(s).

iii. Two $I(1)$ variables might have at most one linearly independent cointegration vector.

Or, in general:

$n > 1$ number of $I(1)$ variables might have at most n -1 linearly independent cointegration vectors.

The number of linearly independent cointegration vectors of variables (x_{1t}, x_{2t}, …, x_{nt}) is called the cointegration rank (r).

↳ How many linearly independent vectors do those variables share?

Ex 2: Draw four independent series of 200-200 random numbers from the standard normal distribution, $\{\varepsilon_{1t}\}$, $\{\varepsilon_{2t}\}$ and $\{\xi_{1t}\}$, $\{\xi_{2t}\}$. Using $\{\xi_{1t}\}$ and $\{\xi_{2t}\}$, simulate two independent random walk series, $\{\mu_{1t}\}$ and $\{\mu_{2t}\}$, assuming that $\mu_{1,0} = \mu_{2,0} = 0$,

$\mu_{1t} = \mu_{1,t-1} + \xi_{1t}$ and $\mu_{2t} = \mu_{2,t-1} + \xi_{2t}$, and then three random walks,

$$y_{1t} = \mu_{1t} + \varepsilon_{1t} \qquad y_{2t} = \mu_{2t} + \varepsilon_{2t} \qquad y_{3t} = 2\mu_{1t} + \varepsilon_{2t}$$

→ y_{1t} and y_{2t} have different stochastic trends, so they are not cointegrated, while y_{1t} and y_{3t} share the same stochastic trend, so they are *CI*(1,1).

```
eps1 = ts(rnorm(200))
eps2 = ts(rnorm(200))
xi1 = ts(rnorm(200))
xi2 = ts(rnorm(200))

mu1 = ts(0, start = 1, end = 200)
mu2 = ts(0, start = 1, end = 200)
for (t in 2:200) {
   mu1[t] = ts(mu1[t-1] + xi1[t])
   mu2[t] = ts(mu2[t-1] + xi2[t])
}
```

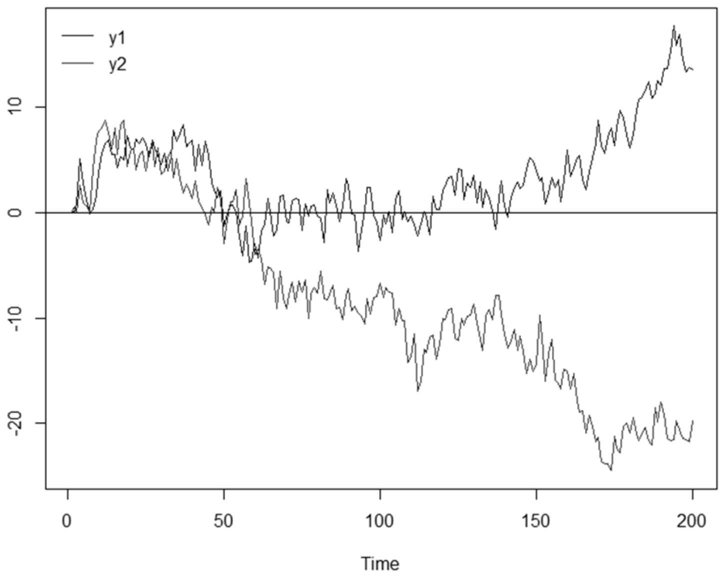
```
y1 = ts(0, start = 1, end = 200)
y2 = ts(0, start = 1, end = 200)
y3 = ts(0, start = 1, end = 200)
for (t in 2:200) {
   y1[t] = mu1[t] + eps1[t]
   y2[t] = mu2[t] + eps2[t]  y3[t] = 2*mu1[t] + eps2[t]
}
```

Yellow highlight are CI

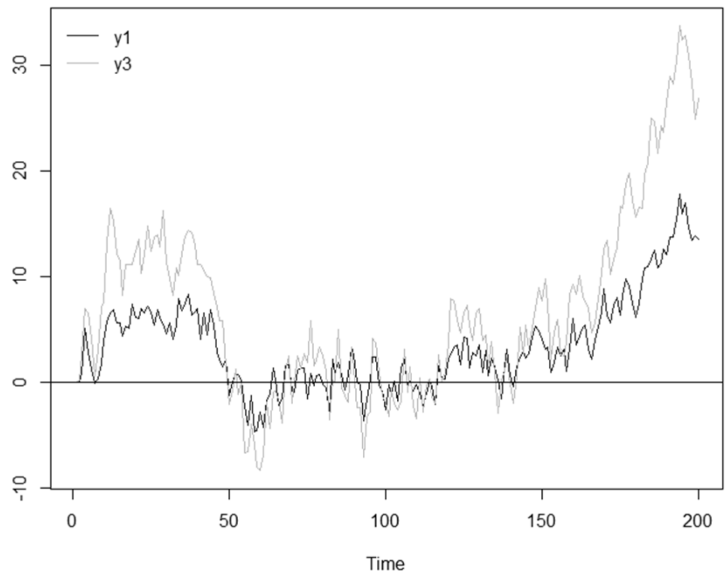
Plot first y_{1t} and y_{2t}, and y_{1t} and y_{3t}.

```
ts.plot(y1, y2, col = c("blue", "red"))
legend("topleft", bty="n", lty=c(1,1),
        col=c("blue", "red"),
        legend = c("y1", "y2"))
abline(h = 0)
```

```
ts.plot(y1, y3, col = c("blue", "green"))
legend("topleft", bty="n", lty=c(1,1),
        col=c("blue", "green"),
        legend = c("y1", "y3"))
abline(h = 0)
```



y_{1t} and y_{2t} are two random walks that are not cointegrated, and they indeed appear to wander independently of each other.

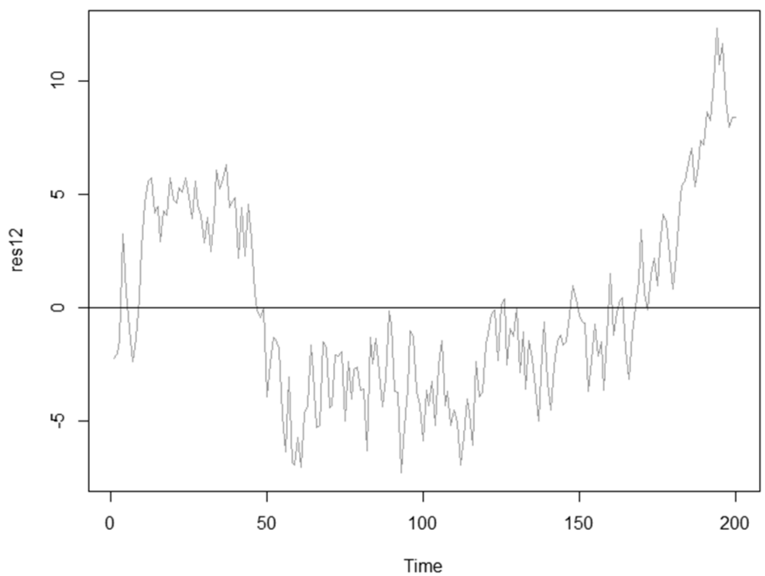


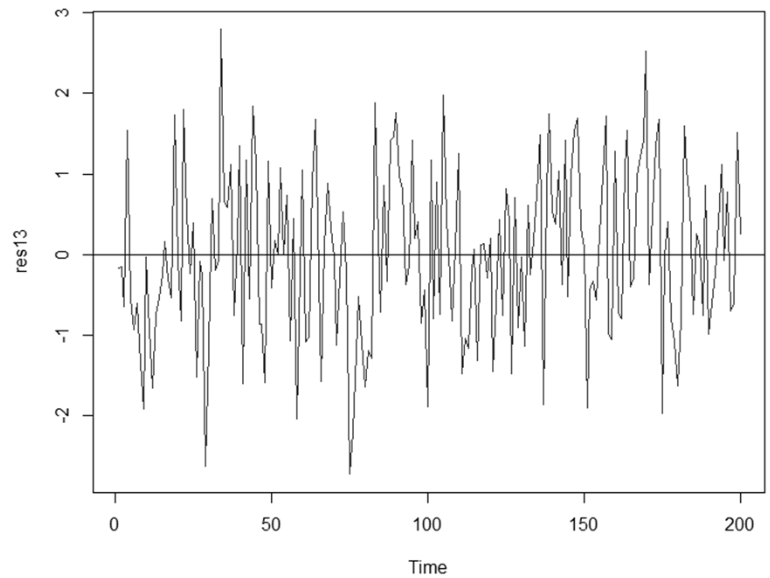
y_{1t} and y_{3t} are two cointegrated random walks and they appear to follow similar time paths.

Regress y_{1t} first on y_{2t} and then on y_{3t}. Plot the residuals.

m12 = lm(y1 ~ y2)
res12 = ts(residuals(m12))
ts.plot(res12, col = "darkorange")
abline(h = 0)

m13 = lm(y1 ~ y3)
res13 = ts(residuals(m13))
ts.plot(res13, col = "darkviolet")
abline(h = 0)





The first regression is spurious, and indeed, *res1* exhibits the characteristics of random walks

The second regression is not spurious, and indeed, *res2* appears to be stationary.

- In case of cointegrated variables the regression on levels is not spurious and the traditional regression methodology (like e.g., the t and F -tests) is valid, granted that the sample size is reasonably large. note !!

← The OLS estimator of the **β** cointegration vector is biased for finite samples, but it is super-consistent in the sense that it converges to the true parameter vector at a much faster rate than in standard models based on stationary variables (these rates are T and $T^{1/2}$, respectively).

On the contrary, if some variables are integrated, say $I(1)$, but not cointegrated, the regression model should be specified in terms of the first differenced, hence $I(0)$, variables.

For this reason, before estimating a time series regression, it is imperative to subject the variables to unit root testing and, if the variables turn out to be nonstationary but integrated, to cointegration testing.

i.e. under normal circumstances

EQUILIBRIUM DYNAMICS AND ERROR CORRECTION

- The main feature of cointegrated variables is that their time paths are influenced by how far they have deviated from their joint equilibrium.

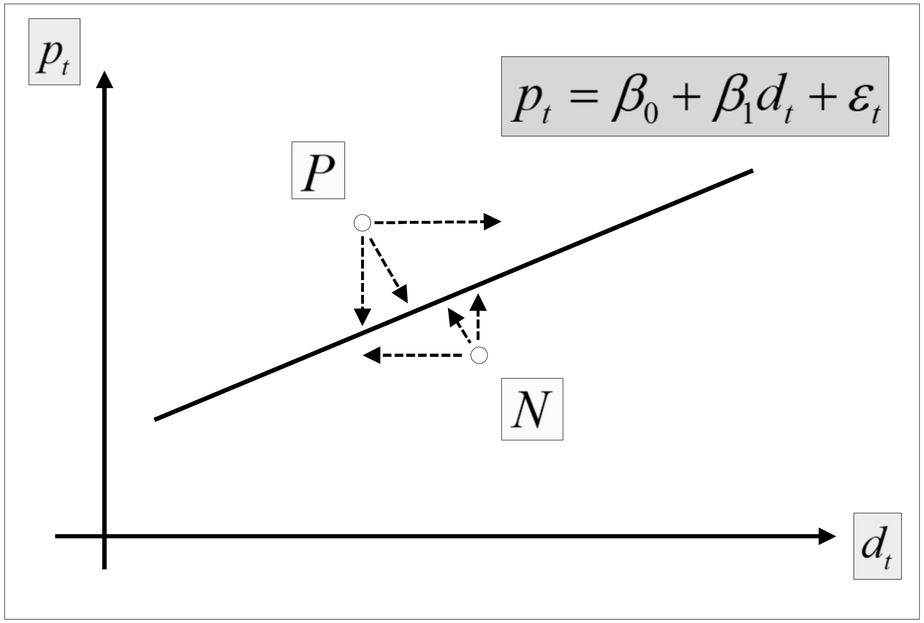
  ← If a system is in disequilibrium but has a tendency to return to long-run equilibrium, at least some of the variables must respond to the magnitude of disequilibrium.

  This is due to a mechanism in the system that relates the short-term movements of the variables in any period to the deviation from the long-run equilibrium in the previous period.

  The dynamic model that embodies this idea is known as the vector error correction model (*VECM*).

- Consider again the present value model of equities, in particular the statistical model that describes the cointegrating relationship (long-run equilibrium equation), between the logarithms of the price (p_t) and the dividend payment (d_t) of an equity.

The expected sign of this long-run relationship is positive. Therefore, it can be illustrated by the following diagram:



If the system is in equilibrium at time t, (p_t, d_t) is right on the population regression line.

Suppose, however, that there is a shock at time t (i.e., $\varepsilon_t \neq 0$), which results in some deviation from the long-run relationship.

This shock can be positive (P: $\varepsilon_t > 0$) or negative (N: $\varepsilon_t < 0$).

$$\varepsilon_t = p_t - \beta_0 - \beta_1 d_t$$

→ $\varepsilon_t > 0$ means that p_t is too high compared to d_t. Therefore, to restore equilibrium in time $t + 1$, either p_{t+1} should decrease, or d_{t+1} should increase, or both.

Similarly, $\varepsilon_t < 0$ means that p_t is too low compared to d_t, so to restore equilibrium in time $t + 1$, either p_{t+1} should increase, or d_{t+1} should decrease, or both.

These possible adjustments can be described with the following dynamic bivariate model:

RLEQs

$$\Delta p_{t+1} = a_{10} + \alpha_1 \left(p_t - \beta_0 - \beta_1 d_t \right) + \varepsilon_{1,t+1}$$
$$\Delta d_{t+1} = a_{20} + \alpha_2 \left(p_t - \beta_0 - \beta_1 d_t \right) + \varepsilon_{2,t+1}$$

ε terms are identical

For ECMs

where α_1, α_2 are the so-called speed of adjustment coefficients; $\varepsilon_{1,t+1}$, $\varepsilon_{2,t+1}$ are white-noise error terms that may be contemporaneously correlated with each other,

the expression within the brackets, called error correction (*EC*) term, is the cointegrating equation in time *t*,

⟶ The cointegrating vector is $[1, -\beta_0, -\beta_1]'$.

The speed of adjustment coefficients are expected to be $\alpha_1 \leq 0$, $\alpha_2 \geq 0$, but at least one of them must be different from zero, …

⟵ Otherwise, *EC* drops out from both equations, and either the system is incorrectly specified, or p_t and d_t are not cointegrated.

… and the larger they are in absolute value, the faster the adjustment to equilibrium.

Moreover, given that $\beta_1 > 0$, stability requires $-2 < \alpha_1$ and $\alpha_2 < 2$.

- In general, the basic bivariate *VECM* for *CI*(1,1) variables *Y* and *Z* is

$$\Delta y_t = \alpha_1 \left(y_{t-1} - \beta_1 z_{t-1} \right) + \varepsilon_{1t}$$
$$\Delta z_t = \alpha_2 \left(y_{t-1} - \beta_1 z_{t-1} \right) + \varepsilon_{2t}$$

It is denoted as *VECM*(0) because it does not have lagged differences.

This system has several interesting features:

i. This *VECM*(0) is equivalent to the following *VAR*(1) model:

$$y_t = (1+\alpha_1) y_{t-1} - \alpha_1 \beta_1 z_{t-1} + \varepsilon_{1t}$$
$$z_t = \alpha_2 y_{t-1} + (1-\alpha_2 \beta_1) z_{t-1} + \varepsilon_{2t}$$

This is a level *VAR*(1), but a *restricted* one because it has only 3 independent parameters, instead of 4.

Similarly, it can be shown that every *VECM*(*p*) model has an equivalent restricted level *VAR*(*p* +1) representation.

ii. Given that y_t and z_t are $I(1)$ and that no lag is necessary, ε_{1t}, ε_{2t}, $\Delta y_{t\text{-}1}$ and $\Delta z_{t\text{-}1}$ are all stationary.

$\longrightarrow$ $(y_{t\text{-}1} - \beta z_{t\text{-}1})$ must be also stationary, i.e., y_t and z_t must be $CI(1,1)$. Otherwise, both equations are unbalanced.

iii. This parameterization allows for two different types of dynamics:

adjustment to the long-run equilibrium via the lagged *EC* term,

and additional short-run dynamics captured by the lagged first differences (autoregressive distributed lags).

iv. One of the speed of adjustment coefficients can be zero.

For example, if $\alpha_1 < 0$ and $\alpha_2 = 0$, the system can adjust to deviations from the long-run equilibrium through changes in *Y*, but the development of *Z* is independent of the equilibrium error.

Alternatively, if $\alpha_1 = 0$ and $\alpha_2 > 0$, the system can adjust to deviations from the long-run equilibrium through changes in *Z*, but the development of *Y* is independent of the equilibrium error.

COINTEGRATION TESTING

- The objective of cointegration testing is to find out whether variables that have stochastic trends share a common stochastic trend.

  To do so with reasonable certainty, we need to conduct some test for cointegration. If there are only two variables y_t and z_t, an obvious option is to regress y_t on z_t, or vice versa, and test the OLS residuals, e_t, for a unit root with an *ADF* τ-type test, without constant and trend.

  → i.e use model 1 !!

  This test, called Engle-Granger (*EG*) cointegration test, is based on the following test regression and hypotheses:

  - regress y on z
  - subject $\hat{\varepsilon}$ to ADF

$$\Delta e_t = \gamma e_{t-1} + \sum_{i=2}^{p} \beta_i \Delta e_{t-i+1} + \xi_t$$

$$H_0 : \gamma = 0 \qquad vs \qquad H_A : (-2 <) \gamma < 0$$

e_t has a unit root, so ε_t is likely $I(1)$ and y_t, z_t are not $CI(1,1)$.

e_t does not have a unit root, so ε_t is likely $I(0)$ and y_t, z_t are $CI(1,1)$.

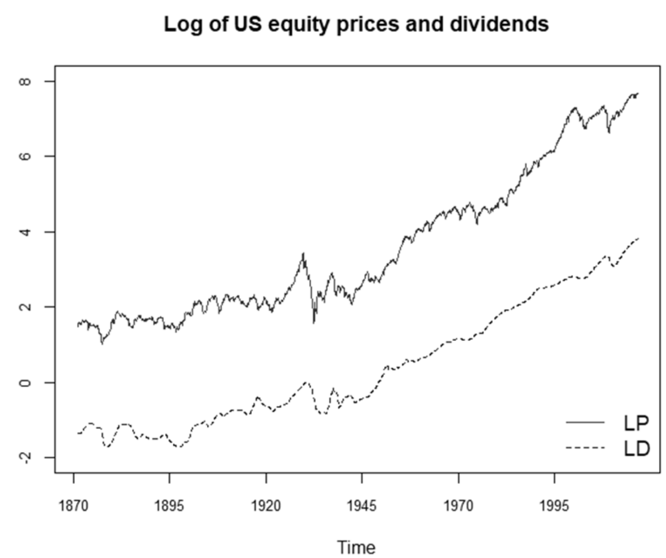
Note: The *EG* test is basically an (*A*)*DF* unit root test on the OLS residuals, but the *DF* critical values are not appropriate in the *EG* test.

Ex 3: (HMPY, pp. 118-126, 139-143)

To illustrate the concept of cointegration we consider monthly S&P500 equity prices and dividends from January 1871 to September 2016. The logarithms of these series are denoted as *LP* and *LD*, respectively.

a) Plot the two series.

```
plot.ts(LP, col = "red", ylim = c(-2, 8), ylab = "",
        xaxt = "n", yaxt = "n“,
        main = "Log of US equity prices and dividends"),
par(cex.axis = 2, cex.lab = 0.5, cex = 0.4)
axis(side = 1, at = seq(1870, 2016, by = 25))
axis(side = 2, at = seq(-2, 8, by = 2))
lines(LD, col = "darkblue", lty = 2 )
legend("bottomright", legend = c(“LP", “LD"),
        col = c("red", "darkblue“),
        lty = 1:2, cex = 2, bty = “n”)
```



Similarly to many other financial time series, *LP* and *LD* have trending characteristics.

b) Test for unit roots by performing the *ADF* and *KPSS* on the levels and first differences.

LP	*Detected unit root (α = 0.10)*	
	ADF	*KPSS*
Level	1	1
First diff.	0	0
	I(1)	*I*(1)

LD	*Detected unit root (α = 0.10)*	
	ADF	*KPSS*
Level	1	1
First diff.	0	1
	I(1)	*I*(2)

Based on the results, *LP* and *LD* behave as *I*(1) variables. Consequently, they might be cointegrated, i.e., *CI*(1,1).

c) Perform the *EG* test with the *coint-test()* function of the *aTSA* package.

If *LP* and *LD* are cointegrated, then in principle it should not make any difference whether *LP* or *LD* is the dependent variable in the long-run equilibrium regression. In practice, however, normalization might matter and the two possible cointegrating regressions might lead to different conclusions.

→ It is recommended to estimate both equilibrium regressions and to test both residual series for a unit root. If H_0 is rejected in at least one case, we can conclude that the variables are cointegrated.

only at least one!!

```
library(aTSA)
eg.1 = coint.test(LP, LD, nlag = 12)
```

```
Response: LP
Input: LD
Number of inputs: 1
Model: y ~ X + 1
-------------------------------
Engle-Granger Cointegration Test
alternative: cointegrated

Type 1: no trend
    lag      EG p.value
  12.00   -4.86    0.01
-----
 Type 2: linear trend
    lag      EG p.value
12.0000 -0.0577  0.1000
-----
 Type 3: quadratic trend
    lag      EG p.value
 12.000  -0.941   0.100
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```

```
eg.2 = coint.test(LD, LP, nlag = 12)
```

```
Response: LD
Input: LP
Number of inputs: 1
Model: y ~ X + 1
-------------------------------
Engle-Granger Cointegration Test
alternative: cointegrated

Type 1: no trend
    lag      EG p.value
  12.00   -4.84    0.01
-----
 Type 2: linear trend
    lag      EG p.value
 12.000   0.489   0.100
-----
 Type 3: quadratic trend
    lag      EG p.value
 12.000   0.428   0.100
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```

There are three sets of results on the printouts, corresponding to equilibrium relationships without trend, with linear trend, and with quadratic trend. In this case we need to consider the first.

The reported *p*-values are not exact but approximate values.

This time, irrespectively of normalization, the *p*-value is not larger than 0.01, so the null hypothesis of no cointegration between *LP* and *LD* is rejected at the 1% significance level.

- If two variables, *Y* and *Z*, prove to be *CI*(1,1), the relationship between them is best captured with a vector error-correction model.

A *VECM*, however, cannot be estimated with OLS straightforwardly because the equilibrium errors, and thus the error-correction term, are unknown.

⟶ A possible solution is the Engle-Granger methodology, which suggests replacing the equilibrium errors with the residuals from the cointegrating regression.

⟶
$$\Delta y_t = a_{10} + \alpha_1 e_{t-1} + \sum_{i=1}^{p} a_{11,i} \Delta y_{t-i} + \sum_{i=1}^{p} a_{12,i} \Delta z_{t-i} + \varepsilon_{yt}$$

$$\Delta z_t = a_{20} + \alpha_2 e_{t-1} + \sum_{i=1}^{p} a_{21,i} \Delta y_{t-i} + \sum_{i=1}^{p} a_{22,i} \Delta z_{t-i} + \varepsilon_{zt}$$

Apart from the estimated error correction term $e_{t\text{-}1}$, these equations constitute a *VAR*(*p*) in the first differences, so they can be estimated one-by-one with OLS.

The optimal lag length (p) can be determined based on some model selection criterion and testing for autocorrelation,

and restrictions concerning the α_1, α_2 speed of adjustment coefficients can be conducted using t-tests.

(Ex 3)

d) Estimate a *VECM* of *LP* and *LD* with the two-step *EG* method.

First, estimate the long-run equilibrium relationship between *LP* and *LD*.

eq.1 = lm(LP ~ LD)
summary(eq.1)

```
Call:
lm(formula = LP ~ LD)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.0783 -0.1998  0.0125  0.2104  0.8164 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.137520   0.007268   431.7   <2e-16 ***
LD          1.195686   0.004424   270.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2958 on 1747 degrees of freedom
Multiple R-squared:  0.9766,	Adjusted R-squared:  0.9766 
F-statistic: 7.304e+04 on 1 and 1747 DF,  p-value: < 2.2e-16
```

This regression looks good, and the estimated equilibrium relationship is

$$\widehat{LP}_t = 3.138 + 1.196 LD_t$$

The slope estimate suggests that a 1% increase of dividends is accompanied on average by an about 1.196% increase of prices.

Second, use the residuals from this regression as equilibrium errors and estimate the two equations of *VECM* one-by-one.

```
e.1 = ts(eq.1$residuals, start = c(1871,1),
         end = c(2016,9), frequency = 12)
le.1 = window(lag(e.1, -1), start = c(1871,2),
         end = c(2016,9), frequency = 12)
DLP = window(diff(LP), start = c(1871,2),
         end = c(2016,9), frequency = 12)
ec.11 = lm(DLP ~ le.1)
summary(ec.11)
```

$$\widehat{DLP}_t = 0.0035 - 0.0011 e_{t-1}$$

```
Call:
lm(formula = DLP ~ le.1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.31078 -0.01900  0.00229  0.02365  0.40282 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.0035390  0.0009734   3.636 0.000285 ***
le.1        -0.0011184  0.0032921  -0.340 0.734097    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0407 on 1746 degrees of freedom
Multiple R-squared:  6.61e-05,  Adjusted R-squared:  -0.0005066 
F-statistic: 0.1154 on 1 and 1746 DF,  p-value: 0.7341
```

```
DLD = window(diff(LD), start = c(1871,2),
         end = c(2016,9), frequency = 12)
ec.12 = lm(DLD ~ le.1)
summary(ec.12)
```

$$\widehat{DLD}_t = 0.0029 + 0.0078 e_{t-1}$$

```
Call:
lm(formula = DLD ~ le.1)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.097458 -0.004073  0.000639  0.005320  0.053637 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.0029487  0.0002638  11.177   <2e-16 ***
le.1        0.0077963  0.0008922   8.738   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01103 on 1746 degrees of freedom
Multiple R-squared:  0.0419,    Adjusted R-squared:  0.04135 
F-statistic: 76.36 on 1 and 1746 DF,  p-value: < 2.2e-16
```

The first *EC* equation is insignificant, but the second is significant (though its explanatory power is also poor).

⟶ α_1 is only insignificantly different from zero, but α_2 is significantly positive.

⟶ The system adjusts to deviations from the long-run equilibrium through changes in *LD*, but *LP* develops independently from the equilibrium error.

The *EC* equations were based on the long-run equilibrium relationship normalized by *LP*. Alternatively, it can be normalized by *LD*.

eq.2 = lm(LD ~ LP)
summary(eq.2)

```
Call:
lm(formula = LD ~ LP)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.61028 -0.17175 -0.01075  0.17007  0.86225 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.553898   0.012325  -207.2   <2e-16 ***
LP           0.816805   0.003022   270.3   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2445 on 1747 degrees of freedom
Multiple R-squared:  0.9766,	Adjusted R-squared:  0.9766 
F-statistic: 7.304e+04 on 1 and 1747 DF,  p-value: < 2.2e-16
```

$$\widehat{LD}_t = -2.554 + 0.817 LP_t$$

e.2 = ts(eq.2$residuals, start = c(1871,1), end = c(2016,9), frequency = 12)
le.2 = window(lag(e.2, -1), start = c(1871,2), end = c(2016,9), frequency = 12)
ec.21 = lm(DLP ~ le.2)
summary(ec.21)

$$\widehat{DLP}_t = 0.0035 + 0.0022 e_{t-1}$$

```
Call:
lm(formula = DLP ~ le.2)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.31057 -0.01906  0.00222  0.02362  0.40216 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.0035391  0.0009734   3.636 0.000285 ***
le.2        0.0022496  0.0039830   0.565 0.572290    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0407 on 1746 degrees of freedom
Multiple R-squared:  0.0001827, Adjusted R-squared:  -0.00039 
F-statistic: 0.319 on 1 and 1746 DF,  p-value: 0.5723
```

ec.22 = lm(DLD ~ le.2)
summary(ec.22)

$$\widehat{DLD}_t = 0.0029 - 0.0084 e_{t-1}$$

```
Call:
lm(formula = DLD ~ le.2)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.097523 -0.004042  0.000915  0.005394  0.053688 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.002948   0.000265  11.124  < 2e-16 ***
le.2        -0.008365   0.001084  -7.713 2.05e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01108 on 1746 degrees of freedom
Multiple R-squared:  0.03295,	Adjusted R-squared:  0.03239 
F-statistic: 59.49 on 1 and 1746 DF,  p-value: 2.055e-14
```

Note:

a) To save time, in this illustrative example we do not worry about potential autocorrelation. In real projects, however, you should test the *EC* equations for autocorrelation, and if necessary, augment them with the lagged value(s) of the left-hand-side variable.

b) In this example the two pairs of *EC* equations are qualitatively very similar. In other cases, however, the normalization of the long-run equilibrium equation might make differences.

d) The *EG* test can be extended to test for cointegration between more than two variables.

Nevertheless, even if more than two variables are involved in the equilibrium regression, the *EG* test does not provide information about the number of independent cointegrating vectors, i.e., about the cointegration rank (r).

In fact, on a system of more than two $I(1)$ variables the *EG* test is valid only if there is at most one cointegration relation of all variables.

An alternative and more general tests for cointegration is provided by the Johansen methodology.

Lo must use J-test for systems > 2 eqns.

WHAT SHOULD YOU KNOW?

- Present value model
- Cointegration
- Error correction
- Vector error correction model
- Cointegration vector, cointegration rank
- Engle-Granger (*EG*) cointegration test