

Lecture 11

Unit Roots and Cointegration

Brooks (4th edition): Chapter 8

© R. Susmel, 2020 (for private use, not to be posted/shared online).

1

Spurious Regression

- Suppose y_t and x_t are non-stationary, $I(1)$. That is, we differentiate them and the changes become stationary, or $I(0)$. We regress y_t against x_t : What happens?
- The usual t -tests on regression coefficients can show statistically significant coefficients, even if in reality it is not so.
- This is the *spurious regression problem* (Granger and Newbold (1974)): We find a statistically significant relation between unrelated variables.
- In a Spurious Regression contexts, the regression errors would be highly correlated and the standard *t-statistic* will be wrongly calculated because the variance of the errors is not consistently estimated.

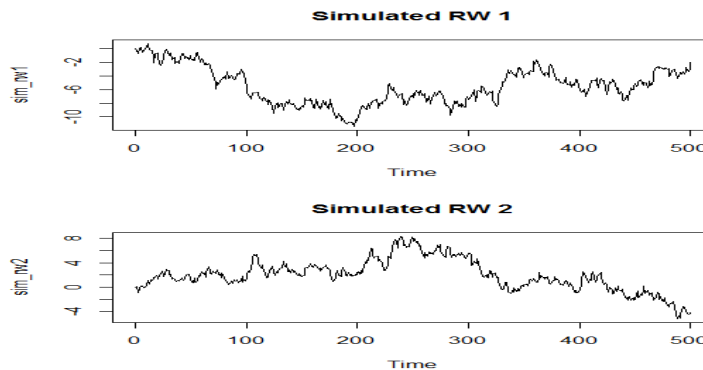
Spurious Regression – Simulated Example

Example: We simulate two independent RW ($RW_{1,t}$, $RW_{2,t}$) and then regress one against the other:

$$RW_{1,t} = \mu + \beta RW_{2,t} + \varepsilon_t$$

With test the true $H_0: \beta = 0$.

```
sim_rw1 <- arima.sim(list(order=c(0,1,0)), sd=.5, n=500) # simulate RW series 1
sim_rw2 <- arima.sim(list(order=c(0,1,0)), sd=.5, n=500) # simulate RW series 2
```



Spurious Regression – Simulated Example

Example:

```
fit_sim_rw <- lm(sim_rw1 ~ sim_rw2) # Regression of two RWs
res_sim_rw <- fit_sim_rw$residuals # Extract residuals
> summary(fit_sim_rw)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-4.61541	0.13188	-35.00	<2e-16 ***	
sim_rw2	-0.47384	0.04076	-11.62	<2e-16 ***	⇒ Reject $H_0: \beta = 0$.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.356 on 499 degrees of freedom
Multiple R-squared: **0.2131**, Adjusted R-squared: 0.2115
F-statistic: **135.1** on 1 and 499 DF, p-value: < 2.2e-16

Note: Very significant t-value (& F-goodness of fit stat), and a good R^2 . But, the model makes no sense.

Spurious Regression – Real Examples

Examples:

(1) Egyptian infant mortality rate (Y), 1971-1990, annual data, on Gross aggregate income of American farmers (I) and Total Honduran money supply (M):

$$\hat{Y}_t = 179.9 - .2952 I_t - .0439 M_t \quad R^2 = .918, DW = .4752, F = 95.17$$

(16.63) (-2.32) (-4.26) $\text{Corr}(Y_t, X_{1,t}) = .8858, -.9113, -.9445$

(2). US Export Index (Y), 1960-1990, annual data, on Australian males' life expectancy (X):

$$\hat{Y}_t = -2943. + 45.7974 X_t \quad R^2 = .916, DW = .3599, F = 315.2$$

(-16.70) (17.76) $\text{Corr}(Y_t, X_t) = .9570$

(3) Total Crime Rates in the US (Y), 1971-1991, annual data, on Life expectancy of South Africa (X):

$$\hat{Y}_t = -24569 + 628.9 X_t \quad R^2 = .811, DW = .5061, F = 81.72$$

(-6.03) (9.04) $\text{Corr}(Y_t, X_t) = .9008$

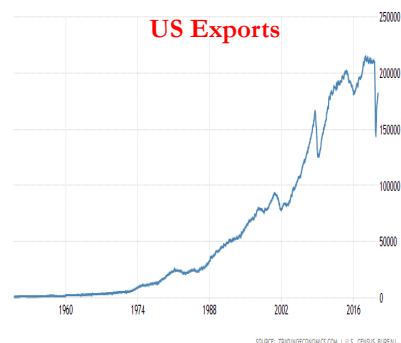
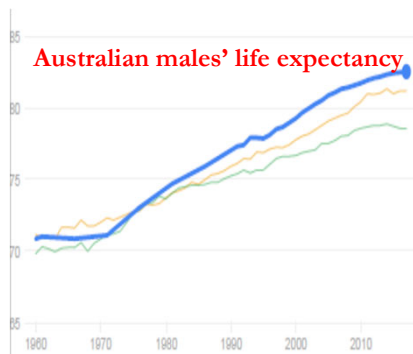
Spurious Regression – Real Examples

Examples (continuation):

(2). US Export Index (Y), 1960-1990, annual data, on Australian males' life expectancy (X):

$$\hat{Y}_t = -2943. + 45.7974 X_t \quad R^2 = .916, DW = .3599, F = 315.2$$

(-16.70) (17.76) $\text{Corr}(Y_t, X_t) = .9570$



Note: It looks like the trend is the common element between Y_t & X_t

Spurious Regression – Statistical Implications

- Suppose y_t & x_t are unrelated I(1) variables. We run the regression:

$$y_t = \beta x_t + \varepsilon_t$$

- True value of $\beta=0$. The above is a spurious regression and $\varepsilon_t \sim I(1)$.

- Technical points: Phillips (1986) derived the following results:

- $\hat{\beta} \xrightarrow{p} \neq 0$ & $\hat{\beta} \xrightarrow{d}$ Non-normal RV not necessarily centered at 0.
 \Rightarrow This is the *spurious regression* phenomenon.

- The OLS t -statistics for testing $H_0: \beta=0$ diverge to $\pm\infty$ as $T \rightarrow \infty$.
 Thus, with a large enough T it will appear that β is significant.

- The usual $R^2 \xrightarrow{p} 1$ as $T \rightarrow \infty$. The model appears to have good fit well, even though it is a bad (nonsense) model.

Spurious Regression – Detection and Solutions

- Given the statistical implications, the typical symptoms are:

- High R^2 , t -values, & F -values.
- Low DW values.

- Q: How do we detect a spurious regression (between I(1) series)?

- Check the correlogram of the residuals.



- Test for a unit root on the residuals. (This lecture.)

Spurious Regression – Detection and Solutions

- Statistical solution:

When series (y_t & x_t) are $I(1)$, work with first differences, instead:

$$\Delta y_t = y_t - y_{t-1} \quad \& \quad \Delta x_t = x_t - x_{t-1}$$

If the relation between the series y_t & x_t exists, β should be the same in levels (y_t, x_t) or first differences ($\Delta y_t, \Delta x_t$).

Levels: $y_t = \beta x_t + \varepsilon_t$ (*)

Lagged Levels: $y_{t-1} = \beta x_{t-1} + \varepsilon_{t-1}$ (**)

Subtract (**) from (*): We have a regression with 1st differences:

First Differences: $\Delta y_t = \beta \Delta x_t + u_t$, where $u_t = \varepsilon_t - \varepsilon_{t-1}$

Now, we have a valid regression, since both regressors are $I(0)$. But, the economic interpretation of the regression changes.

Spurious Regression – Simulated Example

Example: We regress the two RW in first differences:

```
diff_rw1 <- diff(sim_rw1)           # First differences for RW 1
diff_rw2 <- diff(sim_rw2)           # First differences for RW 2
fit_diff_rw <- lm(diff_rw1 ~ diff_rw2)
> summary(fit_diff_rw)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.003199	0.023481	-0.136	0.8917	
diff_rw2	0.106339	0.044773	2.375	0.0179 *	\Rightarrow Reject $H_0: \beta = 0$.

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.525 on 498 degrees of freedom

Multiple R-squared: **0.0112**, Adjusted R-squared: 0.009215

F-statistic: 5.641 on 1 and 498 DF, p-value: 0.01792

Note: Still significant (by chance), but the β coefficient changes sign. Clear indication that something is wrong with regression.

Spurious Regression – Remarks

- The message from spurious regression: Regression of I(1) variables can produce nonsense.

Q: Does it make sense a regression between two I(1) variables?

Yes, only if the regression errors are I(0). That is, when the variables are **cointegrated**.

I(1) Process – Autoregressive Unit Root

- A shock is usually used to describe an unexpected change in a variable or in the value of the error terms at a particular time period.
- When we have a stationary system, the effect of a shock will die out gradually. But, when we have a non-stationary system, the effect of a shock is permanent.
- We have two types of non-stationarity. In an AR(1) model we have:
$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$
 - Unit root: $|\phi_1| = 1$: homogeneous non-stationarity
 - Explosive root: $|\phi_1| > 1$: explosive non-stationarity
- In the last case, a shock to the system become more influential as time goes on. It is not seen in real life. We will not consider them.

I(1) Process – Autoregressive Unit Root

- Consider the AR(p) process:

$$\phi(L)y_t = \mu + \varepsilon_t, \text{ where } \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

As we discussed before, if one of the roots equals 1, $\phi(1) = 0$, or

$$\phi_1 + \phi_2 + \dots + \phi_p = 1$$

- We say y_t has a **unit root**. In this case, y_t is **non-stationary**.

Example: AR(1): $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$. \Rightarrow Unit root: $\phi_1 = 1$.

H_0 (y_t non-stationarity): $\phi_1 = 1$ (or, $\phi_1 - 1 = 0$)

H_1 (y_t stationarity): $\phi_1 < 1$ (or, $\phi_1 - 1 < 0$)

- A *t-test* seems natural to test H_0 . But, the ergodic theorem & MDS CLT do not apply: the *t*-statistic does not have the usual distributions.

Autoregressive Unit Root – Testing

- Now, let's reparameterize the AR(1) process. Subtract y_{t-1} from y_t :

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \mu + (\phi_1 - 1) y_{t-1} + \varepsilon_t$$

$$= \mu + \alpha_0 y_{t-1} + \varepsilon_t$$

- Unit root test: H_0 (unit root process): $\alpha_0 = \phi_1 - 1 = 0$
 H_1 (stationary process): $\alpha_0 < 0$.

- Under H_0 (unit root process): $\alpha_0 = (\phi_1 - 1) = 0$, the model is stationary in Δy_t . Then, if y_t has a unit root:

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

That is, Δy_t is a stationary process with a drift and WN innovations.

- If we reject H_0 : $\alpha_0 = 0$, then y_t is a stationary AR(1) process:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

Autoregressive Unit Root – Testing

- We have a linear regression framework:

$$\Delta y_t = \mu + \alpha_0 y_{t-1} + \varepsilon_t$$

- The natural test for H_0 (unit root process): $\alpha_0 = \phi_1 - 1 = 0$: A *t-test*.

$$t_{\phi=1} = \frac{\hat{\phi}_1 - 1}{SE(\hat{\phi}_1)}$$

- We call this *t-test* the **Dickey-Fuller (DF)** test.
- As mentioned above, under H_0 , the *t-test* does not have the usual *t*-distribution. We use a distribution tabulated (simulated) by DF.

Autoregressive Unit Root – Testing

- We derived the DF test from an AR(1) process for y_t , but y_t may follow a more general AR(p) process:

$$\phi(L)y_t = \mu + \varepsilon_t$$

We rewrite the process using the **Dickey-Fuller reparameterization**:

$$\Delta y_t = \mu + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \cdots + \alpha_{p-1} \Delta y_{t-(p-1)} + \varepsilon_t$$

Note: Both AR(p) formulations are equivalent.

- It can be shown that $\phi(1) = -\alpha_0$. (Roots of $\phi(L) = 0$ equal to 1!)
 \Rightarrow A unit root hypothesis can be stated, again, as $H_0: \alpha_0 = 0$
 $H_1: \alpha_0 < 0$.
- Like in the DF test, we have a linear regression framework. A *t-test* for H_0 (unit root) is the **Augmented Dickey-Fuller (ADF)** test.

Autoregressive Unit Root – Testing: DF

- The **Dickey-Fuller (DF)** test is a special case of the ADF: No lags are included in the regression.

- From our previous AR(1) process, we have:

$$\Delta y_t = \mu + (\phi_1 - 1) y_{t-1} + \varepsilon_t = \mu + \alpha_0 y_{t-1} + \varepsilon_t$$

- If $\alpha_0 = 0$, y_t has a unit root: $H_0: \alpha_0 = 0$
 $H_1: \alpha_0 < 0$.

- We can test H_0 with a *t-test*: $t_{\phi=1} = \frac{\hat{\phi}_1 - 1}{SE(\hat{\phi}_1)}$

Note: There is another associated test with H_0 , the ρ -test: $(T-1)(\hat{\phi}-1)$.

Autoregressive Unit Root – Testing: DF

- $\hat{\phi}$ is not asymptotically normally distributed and $t_{\phi=1}$ is not asymptotically standard normal.
- The limiting distribution of $t_{\phi=1}$ is the DF distribution, which does not have a closed form representation. Then, quantiles of the distribution must be numerically approximated or simulated.
- The distribution of the DF test is *non-standard*. It has been tabulated under different scenarios.

Case 1: no constant: $y_t = \phi_1 y_{t-1} + \varepsilon_t$

Case 2: with a constant: $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$

Case 3: with a constant and a trend: $y_t = \mu + \phi_1 y_{t-1} + \delta t + \varepsilon_t$

Note: In general, the tests with no constant are not used in practice; unless, theoretical reasons impose a constant in the model.

Autoregressive Unit Root – Testing: DF

- Critical values of the $DF(T)$ test under different scenarios.

Table 1: Selected Critical Values of Unit-Root Test Statistics

sample size	Probability			
T	0.01	0.025	0.05	0.10
Model without constant				
100	-2.60	-2.24	-1.95	-1.61
250	-2.58	-2.23	-1.95	-1.62
500	-2.58	-2.23	-1.95	-1.62
∞	-2.58	-2.23	-1.95	-1.62
Model with constant				
100	-3.51	-3.17	-2.89	-2.58
250	-3.46	-3.14	-2.88	-2.57
500	-3.44	-3.13	-2.87	-2.57
∞	-3.43	-3.12	-2.86	-2.57
Model with time trend				
100	-4.04	-3.73	-3.45	-3.15
250	-3.99	-3.69	-3.43	-3.13
500	-3.98	-3.68	-3.42	-3.13
∞	-3.96	-3.66	-3.41	-3.12

Autoregressive Unit Root – Testing: DF

- Which version of the three main variations of the test should be used is not a minor issue. The decision has implications for the size and the power of the unit root test.
- For example, an incorrect exclusion of the time trend term leads to (omitted variables) bias in the coefficient estimate for ϕ , leading to size distortions (false negatives/positives) and reductions in power.
- Technical note: The normalized bias $(T - 1) * (\hat{\phi} - 1)$ 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 $H_0: \phi = 1$. It is usually referred as the ρ -test.

The distribution of this ρ -test is also *non-standard*. It is also simulated.

Autoregressive UR – DF: Case 2

- **Case 2.** DF with a constant term in DGP: $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$

The hypotheses to be tested:

$$\begin{aligned} H_0: \phi_1 = 1, \mu = 0 &\Rightarrow y_t \sim I(1) \text{ without drift} \\ H_1: \phi_1 < 1 &\Rightarrow y_t \sim I(0) \text{ with mean.} \end{aligned}$$

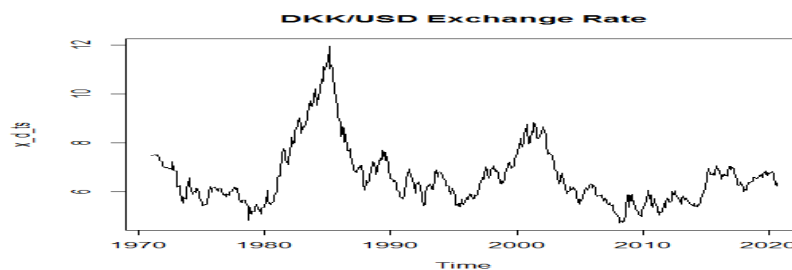
This formulation is appropriate for non-trending economic and financial series like interest rates, exchange rates for developed (& stable) countries and spreads.

- The test statistics $t_{\phi=1}$ is computed from the estimation of the AR(1) model with a constant.

Autoregressive UR – DF: Example Case 2

Example: We do a DF test for the monthly DKK/USD exchange rate, S_t , (1971-2020), $T = 597$. We use **Case 2** (constant in DGP):

```
PPP_da <-
read.csv("http://www.bauer.uh.edu/rsusmel/4397/ppp_2020_m.csv",head=TRUE,sep=",")
x_d <- PPP_da$DKK_USD
x_d_ts <- ts(x_d,start=c(1971,1),frequency=12)
plot(x_d_ts, main = "DKK/USD Exchange Rate")
```



Autoregressive UR – DF: Example Case 2

Example (continuation):

```
T_d <- length(x_d)
x_S <- x_d[-T_d]
diff_x <- x_d[-1] - x_d[-T_d]
fit_df <- lm(diff_x ~ x_S)
> summary(fit_df)> summary(fit_df)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.093170	0.045975	2.027	0.0432 *	
xx	-0.014382	0.006838	-2.103	0.0359 *	⇒ DF test stat

Residual standard error: 0.03326 on 591 degrees of freedom
 Multiple R-squared: 0.009852, Adjusted R-squared: 0.008176
 F-statistic: 5.88 on 1 and 591 DF, p-value: 0.01561

Autoregressive UR – DF: Example Case 2

Example (continuation):

Compare $t_{\phi=1} = \mathbf{-2.103}$ to the values in DF distribution at 5% level

Critical values at 5% level: $\mathbf{-2.87}$ for $T = 500$ & $\mathbf{-2.88}$ for $T = \infty$

$t_{\phi=1} > \mathbf{-2.87} \Rightarrow$ Cannot reject $H_0: \phi_1 = 1$, with a t-test. Then, take 1st differences (changes in S_t) to model the series.

Note: The R library *urca* computes the DF test (and other unit root tests) with the *ur.df* function. There are other R libraries that do unit root tests. The *ur.df* function with **lags=0** is the DF test, with lags>0 is the “*Augmented Dickey-Fuller*” (or ADF) test.

```
library(urca)
lc.df <- ur.df(y=x_d, lags=0, type='drift')
> summary(lc.df)
```

Autoregressive UR – DF: Example Case 2

Example (continuation): `> summary(lc.df)`

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Coefficients:
              Estimate Std. Error t value Pr(> |t|)
(Intercept)  0.093170   0.045975   2.027   0.0432 *
z.lag.1      -0.014382   0.006838  -2.103   0.0359 *      => DF test stat
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2047 on 594 degrees of freedom
Multiple R-squared:  0.007392, Adjusted R-squared:  0.005721
F-statistic: 4.424 on 1 and 594 DF, p-value: 0.03586

Value of test-statistic is: -2.1033 2.2377

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

Autoregressive UR – DF: Example Case 2

Example (continuation):

Check: OLS regression of S_t against a constant and S_{t-1} . The S_{t-1} coefficient should be very close to 1.

```
fit_S_1 <- lm(x_d[-1] ~ x_d[-T_chf])
> summary(fit_S_1)
```

Coefficients:

```
              Estimate Std. Error t value Pr(> |t|)
(Intercept)  0.093170   0.045975   2.027   0.0432 *
x_d[-T_chf]  0.985618   0.006838 144.140 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.2047 on 594 degrees of freedom
Multiple R-squared: 0.9722, Adjusted R-squared: 0.9722
F-statistic: 2.078e+04 on 1 and 594 DF, p-value: < 2.2e-16

Autoregressive UR – DF: Case 3

- **Case 3.** With constant and trend term in the DGP.

The test regression is $y_t = \mu + \phi_1 y_{t-1} + \delta t + \varepsilon_t$

and includes a constant and deterministic time trend to capture the deterministic trend under the alternative. The hypotheses to be tested:

$$H_0: \phi_1 = 1, \delta = 0 \quad \Rightarrow y_t \sim I(1) \text{ with drift}$$

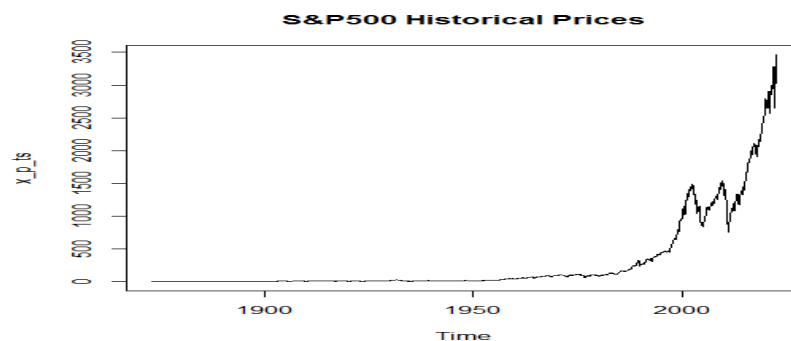
$$H_1: |\phi_1| < 1 \quad \Rightarrow y_t \sim I(0) \text{ with deterministic time trend.}$$

- This formulation is appropriate for trending time series like asset prices or the levels of macroeconomic aggregates like real GDP. The test statistics $t_{\phi=1}$ is computed from the above regression.

Autoregressive UR – DF: Example Case 3

- **Example:** We do a DF test for the monthly stock index price, P_t , (1873 - 2020), $T=1796$. We use **Case 3** (constant and trend in DGP):

```
Sh_da <-
read.csv("http://www.bauer.uh.edu/rsusmel/4397/Shiller_2020data.csv",head=TRUE,sep=",")
x_p <- Sh_da$P
T_p <- length(x_p)
library(urca)
```



Autoregressive UR – DF: Example Case 3

- **Example (continuation):** We use function *ur.df* in the library *urca*.

```
lc.df_p <- ur.df(y=x_p, lags=0, type='trend')
```

```
> summary(lc.df_p)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.252237	1.367601	-0.184	0.854	
z.lag.1	0.006243	0.001380	4.526	6.42e-06 ***	⇒ DF test stat
tt	0.000329	0.001627	0.202	0.840	

Value of test-statistic is: **4.5257** 16.6595 20.1626

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-3.96	-3.41	-3.12

Compare $t_{\phi=1} = \mathbf{4.5257}$ to the values in DF distribution (5%): **-3.41**.

$t_{\phi=1} > \mathbf{-3.41} \Rightarrow$ Cannot reject $H_0: \phi_1 = 1$, with a t-test at 5% level.

Then, take 1st differences –i.e, returns!– to model the series.

Autoregressive UR – DF: Example Case 3

- **Example (continuation):**

OLS regression of P_t against a constant and P_{t-1} . The P_{t-1} coefficient should be very close to 1.

```
fit_p_1t <- lm (x_p[-1] ~ x_p[-T_p] + trend_p)
```

```
> summary(fit_p_1t)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.020395	1.346116	-0.758	0.449
x_S1	1.006243	0.001380	729.401	<2e-16 ***
trend	0.000329	0.001627	0.202	0.840

Residual standard error: 26.28 on 1794 degrees of freedom

Multiple R-squared: **0.9982**, Adjusted R-squared: 0.9982

F-statistic: **4.93e+05** on 2 and 1794 DF, p-value: < 2.2e-16

Autoregressive UR – Testing: ADF

- The DF tests is valid if ε_t is white noise. In particular, ε_t will show autocorrelation if there was also autocorrelation in the dependent variable, Δy_t . The solution is to “*augment*” the DF test using p lags of the dependent variable.
- Back to the general, $AR(p)$ process. We can rewrite the equation as the *Dickey-Fuller reparameterization*:

$$\Delta y_t = \mu + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \dots + \alpha_p \Delta y_{t-p} + \varepsilon_t$$
- Under H_0 (unit root): $\alpha_0 = 0$, Δy_t is a stationary $AR(p-1)$ process. Under H_1 : $\alpha_0 < 0$, y_t is a standard $AR(p)$ stationary process.
- The usual t -test for H_0 is the *Augmented Dickey-Fuller* (ADF) test. Similar to the DF test case, we have a non-normal distribution.

Autoregressive UR – ADF: Example Case 2

Example: We do an ADF test, with **12 lags** for the monthly DKK/USD exchange rate. We use **Case 2** (constant in DGP):

```
lc.df_lag12 <- ur.df(y=x_d, lags=12, type='drift')
```

```
> summary(lc.df_lag12)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.129831	0.049055	2.647	0.00835 **	
z.lag.1	-0.019787	0.007317	-2.704	0.00705 **	⇒ ADF test stat
z.diff.lag1	-0.004393	0.041620	-0.106	0.91597	
z.diff.lag2	0.103244	0.041562	2.484	0.01327 *	
z.diff.lag3	0.049343	0.041872	1.178	0.23911	
z.diff.lag4	0.013162	0.041917	0.314	0.75363	
z.diff.lag5	0.016241	0.041920	0.387	0.69859	
z.diff.lag6	-0.021091	0.041876	-0.504	0.61470	
z.diff.lag7	0.060137	0.041854	1.437	0.15131	
[...]					⇒ None of the omitted coefficients is significant

Autoregressive UR – ADF: Example Case 2

Example (continuation):

Value of test-statistic is: **-2.7041** 3.6606

Critical values for test statistics:

1pct	5pct	10pct	
tau2	-3.43	-2.86	-2.57
			⇒ Cannot reject H_0 : unit root at 5% level
phi1	6.43	4.59	3.78

• ADF test with 4 lags.

```
lc.df_lag4 <- ur.df(y=x_d, lags=4, type='drift')
```

```
> summary(lc.df_lag4)
```

Value of test-statistic is: **-2.4313** 2.9757

Critical values for test statistics:

1pct	5pct	10pct	
tau2	-3.43	-2.86	-2.57
			⇒ Cannot reject H_0 : unit root at 5% level
phi1	6.43	4.59	3.78

Autoregressive UR – ADF: Example Case 3

Example: We do a ADF test with **12 lags** for the monthly stock index price (1873 - 2020). We use **Case 3** (constant & trend in DGP):

```
lc.adf_p12 <- ur.df(y=x_p, lags=12, type='trend')
```

```
> summary(lc.adf_p12)
```

Value of test-statistic is: **4.1425** 12.1762 15.4108

Critical values for test statistics:

1pct	5pct	10pct	
tau3	-3.96	-3.41	-3.12
			⇒ Cannot reject H_0 : unit root at 5% level

```
lc.adf_p4 <- ur.df(y=x_p, lags=4, type='trend')
```

```
> summary(lc.adf_p4)
```

Value of test-statistic is: **5.0976** 19.1402 23.7456

Critical values for test statistics:

1pct	5pct	10pct	
tau3	-3.96	-3.41	-3.12
			⇒ Cannot reject H_0 : unit root at 5% level

Autoregressive Unit Root – Other Tests

- Phillips and Perron (*PP*) developed a more comprehensive theory of unit root nonstationarity.

The PP tests are similar to ADF tests, but they incorporate an automatic correction to the DF test procedure that allows for autocorrelation in the residuals.

- The tests usually produce the same conclusions as the ADF tests, but the calculation of the test statistics is more complicated.

Note: There are other unit root tests: ADF-GLS (also called ERS, which corrects ADF test for heteroscedasticity), KPSS (Kwaitowski, Phillips, Schmidt and Shin), Zivot and Andrews (ZA, which allows for structural breaks), etc. We will not cover them in this class.

Autoregressive UR – Testing: Issues

- Q: How many lags in the ADF test?
Most popular strategy: Select lags using AIC (or BIC).

Some answers include an automatic selection of lags, similar to the selection of lags for the Newey and West SE.

- Q: Do these tests have much power?
It is difficult to distinguish $\phi_1 = 0.98$ or $\phi_1 = 0.99$ from $\phi_1 = 1$. The implications are completely different. In general, we find that these tests have low power against “*near unit root*” alternatives.

If the tests have low power, we may keep as stationary a time series that is a non-stationary process. A nonsense regression can happen!

Autoregressive UR – Testing: Remarks

- It is not always easy to affirm that a process has a unit roots. These tests have *low power* against near-unit-root alternatives.
- There are also *size problems* (false positives) because we cannot include an infinite number of augmentation lags as might be called for with MA processes.
- However, the ADF test is a critical tool we use to identify the underlying time series model: ARMA, or trend + ARMA, or ARIMA?
- If we select an ARIMA, what is the order of the integration, d ?
- In addition, we if we use an $AR(p)$ to approximate an $ARMA(p, q)$, the ADF can help us determine the right order of approximation, p .

Cointegration

- Integration: In a univariate context, y_t is $I(d)$ if its $(d - 1)$ -th difference is $I(0)$. That is, $\Delta^d y_t$ is stationary.

$$\Rightarrow y_t \text{ is } I(1) \text{ if } \Delta y_t \text{ is } I(0)$$

- In many time series, integrated processes are considered together and they form equilibrium relationships:
 - Short-term and long-term interest rates
 - Inflation rates and interest rates.
 - Income and consumption.
 - Spot and Forward rates.
 - Dividends and Earnings.

Idea: Although a time series vector is integrated, certain linear transformations of the time series may be stationary.

Cointegration – Definition

- An $m \times 1$ vector time series \mathbf{Y}_t is said to be **cointegrated of order** (d, b) , or $CI(d, b)$, where $0 < b \leq d$, if each of its component series Y_{it} is $I(d)$ but some linear combination $\alpha' \mathbf{Y}_t$ is $I(d - b)$ for some constant vector $\alpha \neq 0$.
- Terminology: α is called **cointegrating vector** (*long-run parameter*).

Example: We have 2 $I(1)$ variables: Spot, S_t , and 3-mo forward rates, $F_{t,T=3-mo}$, for the USD/NZD exchange rate. That is, $m = 2$.

If $\varepsilon_t = S_t - \beta F_{t,T=3-mo} \sim I(0) \Rightarrow S_t \text{ \& } F_{t,T=3-mo}$ are cointegrated of order(1, 1) or $CI(1, 1)$.

$$\alpha' = [1 \quad -\beta] \quad \Rightarrow \text{cointegrating vector}$$

$$\mathbf{Y}_t = \begin{bmatrix} S_t \\ F_{t,T=3-mo} \end{bmatrix} \quad \Rightarrow \alpha' \mathbf{Y}_t = [1 \quad -\beta] \begin{bmatrix} S_t \\ F_{t,T=3-mo} \end{bmatrix} = \varepsilon_t \sim I(0)$$

Cointegration – Definition

- The cointegrating vector is not unique. For any scalar c

$$c \alpha' \mathbf{Y}_t = \alpha^{*'} \mathbf{Y}_t \sim I(d - b) \alpha^{*'}$$
- Some normalization assumption is required to uniquely identify α . Usually, α_1 (=the coefficient of the first variable) is normalized to 1. Look at the previous example, where the cointegrating vector is $[1 \quad -\beta]$.
- The most common case is $d = b = 1$.
- If the $m \times 1$ vector time series \mathbf{Y}_t contains more than 2 components, each being $I(1)$, then there may exist $k (< m)$ linearly independent $1 \times m$ vectors $\alpha_1', \alpha_2', \dots, \alpha_k'$, such that $\alpha' \mathbf{Y}_t \sim I(0)$ $k \times 1$ vector process, where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ is a $k \times m$ **cointegrating matrix**.

Cointegration – Definition

- The number of linearly independent cointegrating vectors is called the **cointegrating rank**: \mathbf{Y}_t is cointegrated of rank k .

If the $m \times 1$ vector time series \mathbf{Y}_t is $CI(k, 1)$ with $0 < k < m$ CI vectors, then, we say that there are $m - k$ common $I(1)$ *stochastic trends*.

Cointegration – Example

Example: Consider the following system of three processes ($m = 3$):

$$x_{1t} = \beta_1 x_{2t} + \beta_2 x_{3t} + \varepsilon_{1t}$$

$$x_{2t} = \beta_3 x_{3t} + \varepsilon_{2t}$$

$$x_{3t} = x_{3,t-1} + \varepsilon_{3t}$$

where the error terms are uncorrelated WN processes. Since x_{3t} is a RW –i.e., $I(1)$ –, clearly, all the 3 processes are individually $I(1)$.

- One CI relationship: x_{1t} , x_{2t} , & x_{3t} .

Let $\mathbf{Y}_t = (x_{1t}, x_{2t}, x_{3t})'$ & $\boldsymbol{\alpha} = (1, -\beta_1, -\beta_2)'$ $\Rightarrow \boldsymbol{\alpha}' \mathbf{Y}_t = \varepsilon_{1t} \sim I(0)$.

Note: The coefficient for x_{1t} ($=\alpha_1$) is normalized to 1.

- A second CI relationship: x_{2t} , & x_{3t}

Let $\boldsymbol{\alpha}^* = (0, 1, -\beta_3)'$ $\Rightarrow \boldsymbol{\alpha}^{*'} \mathbf{Y}_t = \varepsilon_{2t} \sim I(0)$.

Cointegration – Example

Example (continuation):

We have 2 C.I. vectors: α & α^* . They are independent, that is,

$$\text{rank} \begin{bmatrix} 1 & 0 \\ -\beta_1 & 1 \\ -\beta_2 & -\beta_3 \end{bmatrix} = 2 \quad (\text{This is the } \textit{cointegrating rank}.)$$

- Summary for the system (with three time series –i.e., $m = 3$):

We have: $k = 2$ independent C.I. vectors: α & α^*
 $m - k = 1$ common stochastic trend (ST): $\sum_{t=1}^T \varepsilon_{3t}$.

Cointegration – Long-Run Relation

- Intuition for $I(1)$ case

$\alpha'Y_t$ forms a *long-run equilibrium*. The system cannot deviate too far from the equilibrium, otherwise economic forces, say arbitrage or competition, will operate to restore the equilibrium. We think of cointegrated variables as variables that “*move together*.”

Example: In the previous example, we have two long-run relationships –i.e., two CI relationships:

$$1) \varepsilon_{1t} = x_{1t} - \beta_1 x_{2t} - \beta_2 x_{3t}.$$

$$2) \varepsilon_{2t} = x_{2t} - \beta_3 x_{3t}$$

Interpretation: Let's look at CI relation 2). In the *long-run*, when x_{3t} changes by 1 unit, x_{2t} changes by β_3 units. (This is why the CI vector is referred as **long-run parameters**.)

Cointegration – Error Correction Model

- When series are cointegrated, the deviations from the long-run equilibrium affect the behavior of the series in the short-run. The model that includes the deviations from equilibrium is called “*error correction model*,” or ECM.

In the ECM, the *lagged deviations* from the long-run equilibrium enter linearly in a model of first differences for the cointegrated variables.

Suppose we have two cointegrated variables, S_t & F_t , with a cointegrating vector $\alpha' = [1 \ -\beta]$. Then,

$$\Delta S_t = \gamma (1 S_{t-1} - \beta F_{t-1}) + \beta_1 \Delta F_t + u_t$$

The deviation from long-run equilibrium –i.e., $(S_{t-1} - \beta F_{t-1})$ above– is called the *error correction term*.

Cointegration – Error Correction Model

Note: Suppose we have two cointegrated variables, say, S_t & F_t , but we model them ignoring the *error correction term*. Then, we have a misspecified model (biased estimation!).

Example: In the previous example, we have x_{1t} , x_{2t} , & x_{3t} are cointegrated, with CI vector: $\alpha = (1, -\beta_1, -\beta_2)'$

Then, the ECM becomes:

$$\Delta x_{1t} = \gamma \epsilon_{1t-1} + \beta_1 \Delta x_{2t} + \beta_2 \Delta x_{3t} + u_t$$

where $\epsilon_{1t} = x_{1t} - \beta_1 x_{2t} - \beta_2 x_{3t}$ is the error correction term.

Technical note: The *Granger representation theorem* states that if there exists a dynamic linear model with stationary disturbances –i.e., $I(0)$ – & the data are $I(1)$, then the variables must be *cointegrated of order (1,1)*.

Residual Based Tests of the Null of No CI

- These are test procedures designed to distinguish a system without cointegration from a system with at least one cointegrating relationship. They do not estimate the number of cointegrating vectors (the k).
- Tests are conditional on pretesting (for unit roots in each variable).
- There are two cases to consider.
- **CASE 1** - Cointegration vector is pre-specified/known (say, from economic theory):
Construct the hypothesized linear combination that is $I(0)$ by theory; treat it as data. Apply a DF unit root test to that linear combination.
- The null hypothesis is that there is a unit root, or no cointegration.

47

Residual Based Tests of the Null of No CI

- **CASE 2** - Cointegration vector is *unknown*. It should be estimated. Thus, if there exists a cointegrating relation, the coefficient on Y_{2t} is nonzero, allowing us to express the “static” regression equation as

$$Y_{1t} = \beta Y_{2t} + u_t$$
- Then, apply a unit root test to the estimated OLS residual from estimation of the above equation, but
 - Include a constant in the static regression if the alternative allows for a nonzero mean in u_t .
 - Include a trend in the static regression if the alternative is stochastic cointegration –i.e., a nonzero trend for $\alpha'Y_t$.

Note: Tests for cointegration using a prespecified cointegrating vector are generally more powerful, but they are rarely used.

48

Engle and Granger Cointegration

- Engle and Granger propose a 2-steps cointegration test:

Step 1. Test H_0 (unit root) in each component series Y_{it} individually, using the univariate unit root tests, say ADF, PP tests.

Step 2. If the H_0 (unit root) cannot be rejected, then the next step is to test cointegration among the components, i.e., to test whether $\alpha'Y_t$ is $I(0)$.

- In practice, the cointegration vector is *unknown*. One way to test the existence of cointegration is the regression method –see, Engle and Granger (1986) (EG).

- If $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{mt})$ is cointegrated, $\alpha'Y_t$ is $I(0)$ where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$. Then, $(1/\alpha_1)\alpha$ is also a cointegrated vector where $\alpha_1 \neq 0$.

49

Engle and Granger Cointegration – Step 1

- EG consider the regression model for Y_{1t} :

$$Y_{1t} = \delta D_t + \phi_1 Y_{2t} + \dots + \phi_{m-1} Y_{mt} + \varepsilon_t$$

where D_t : deterministic terms.

- Check whether ε_t is $I(1)$ or $I(0)$:

- If $\varepsilon_t \sim I(1)$, then Y_t is not cointegrated.

- If $\varepsilon_t \sim I(0)$, then Y_t is cointegrated with a **normalized cointegrating vector** $\alpha' = (1, \phi_1, \dots, \phi_{m-1})$.

- Steps:

1. Run OLS. Get estimate $\hat{\alpha} = (1, \hat{\phi}_1, \dots, \hat{\phi}_{m-1})$

2. Use residuals e_t for unit root testing using the ADF or PP tests without deterministic terms (constant or constant and trend).

50

Engle and Granger Cointegration – Step 2

- **Step 2:** Use residuals e_t for unit root test. For example, ADF test:

$$\Delta e_t = \lambda e_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta e_{t-1} + a_t$$

H_0 (unit root in residuals = no cointegration): $\lambda = 0$ H_1
(cointegration): $\lambda < 1$

- t-statistic: $t_\lambda = \frac{\hat{\lambda}}{s_{\hat{\lambda}}}$

- Critical values tabulated by simulation in EG.

Remark: If we reject H_0 , then \mathbf{Y}_t is cointegrated, there is further modeling step: Construct the ECM. That is,

$$\Delta Y_{1t} = \gamma e_{t-1} + \phi_1 \Delta Y_{2t} + \phi_2 \Delta Y_{3t} + \dots + \phi_{m-1} \Delta Y_{mt} + u_t$$

51

Engle and Granger Cointegration – Table

- Critical values tabulated by simulation in EG.

N	Model 1 - constant, no trend				Model 2 - constant, trend			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-3.724	-3.318	-2.986	-2.633	-4.375	-3.943	-3.589	-3.238
50	-3.568	-3.213	-2.921	-2.599	-4.152	-3.791	-3.495	-3.181
100	-3.498	-3.164	-2.891	-2.582	-4.052	-3.722	-3.452	-3.153
250	-3.457	-3.136	-2.873	-2.573	-3.995	-3.683	-3.427	-3.137
500	-3.443	-3.127	-2.867	-2.570	-3.977	-3.670	-3.419	-3.132
>500	-3.434	-3.120	-2.863	-2.568	-3.963	-3.660	-3.413	-3.128

- We expect the usual ADF distribution would apply here. But, Phillips and Ouliaris (PO) (1990) show that is not the case under H_0 (no-cointegration). They tabulated the corrected distributions.

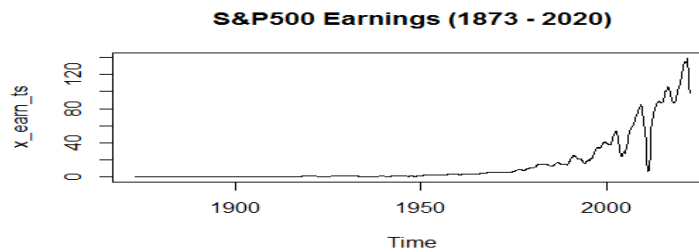
Hansen (1992) improved on these distributions, getting adjustments for different DGPs with trend and/or drift/no drift.

52

Engle and Granger Cointegration – E & D

Example: We test if there is cointegration between dividends and earnings. We use Shiller's monthly dataset (1873:1 - 2020:9).

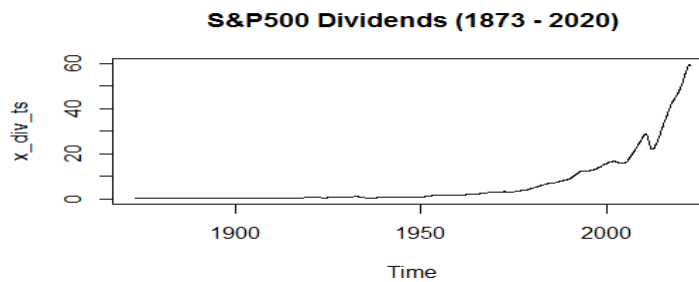
```
Sh_da <-  
read.csv("http://www.bauer.uh.edu/rsusmel/4397/Shiller_2020data.csv", head=TRUE, sep=  
",")  
x_earn <- Sh_da$E          # Extract Earnings  
x_div <- Sh_da$D           # Extract Dividends  
x_earn_ts <- ts(x_earn, start=c(1873,1), frequency=12)  
plot(x_earn_ts, main="S&P500 Earnings (1873 - 2020)")
```



Engle and Granger Cointegration – E & D

Example (continuation):

```
x_div_ts <- ts(x_div, start=c(1873,1), frequency=12)  
plot(x_div_ts, main="S&P500 Dividends (1873 - 2020)")
```



- Both series, Earnings and Dividends, show a clear trend.

Engle and Granger Cointegration – E & D

Example (continuation):

Step 1: ADF Tests for Unit Roots with 4 and 12 lags for Earnings.

```
• lc.df_earn_t <- ur.df(y=x_earn, lags=4, type='trend')
> summary(lc.df_earn_t)
```

Value of test-statistic is: **-2.8983** 3.2865 4.6402 \Rightarrow Cannot reject H_0 : unit root at 5% level

Critical values for test statistics:

```
1pct 5pct 10pct
tau3 -3.96 -3.41 -3.12
```

```
• lc.df_earn_t_12 <- ur.df(y=x_earn, lags=12, type='trend')
> summary(lc.df_earn_t_12)
```

Value of test-statistic is: **-0.7685** 1.9611 1.9807 \Rightarrow Cannot reject H_0 : unit root at 5% level

- The ADF cannot reject H_0 (unit root) for Earnings.

Engle and Granger Cointegration – E & D

Example (continuation):

Step 1: ADF Tests for Unit Roots with 4 and 12 lags for Dividends.

```
• lc.df_div_t <- ur.df(y=x_div, lags=4, type='trend')
> summary(lc.df_div_t)
```

Value of test-statistic is: 1.627 4.3152 4.96 \Rightarrow Cannot reject H_0 : unit root at 5% level

Critical values for test statistics:

```
1pct 5pct 10pct
tau3 -3.96 -3.41 -3.12
```

```
• lc.df_div_t_12 <- ur.df(y=x_div, lags=12, type='trend')
> summary(lc.df_div_t_12)
```

Value of test-statistic is: **3.7214** 10.732 13.512 \Rightarrow Cannot reject H_0 : unit root at 5% level

- The ADF cannot reject H_0 (unit root) for Dividends.

Engle and Granger Cointegration – E & D

Example (continuation):

Step 2:

A. Run cointegrating regression. Get coefficients. Keep residuals.

B. Test residuals for a Unit Root with 4 and 12 lags.

Step 2-A. Regress Dividends against Earnings

```
• coint_div_earn <- lm(x_div ~ x_earn)
```

```
> summary(coint_div_earn)
```

Coefficients:

```
      Estimate Std. Error t value Pr(> |t|)
(Intercept)  0.554824   0.080139   6.923 6.12e-12 ***
x_earn       0.399444   0.002524 158.286 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 3.001 on 1796 degrees of freedom

Multiple R-squared: 0.9331, Adjusted R-squared: 0.9331

Engle and Granger Cointegration – E & D

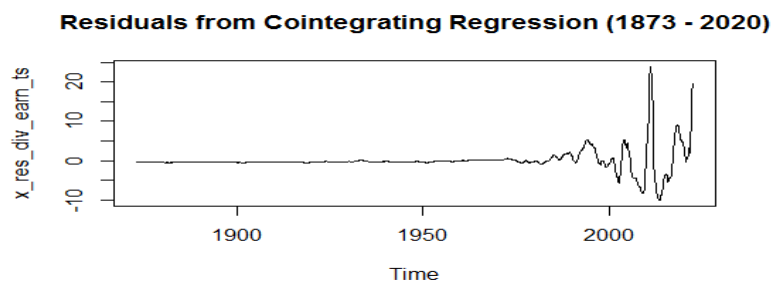
Example (continuation):

Step 2-A. Extract residuals and plot them

```
• res_div_earn <- coint_div_earn$residuals # Extract Residuals
```

```
x_res_div_earn_ts <- ts(res_div_earn, start=c(1873,1), frequency=12)
```

```
> plot(x_res_div_earn_ts, main="Residuals from Cointegrating Regression (1873 - 2020)")
```



- The clear trend in both series is gone from residuals.

Engle and Granger Cointegration – E & D

Example (continuation):

Step 2-B. Test residuals for unit roots

```
• lc.adf_res_div_earn <- ur.df(y=res_div_earn, lags=4, type='drift')
> summary(lc.adf_res_div_earn)
```

Value of test-statistic is: **-9.2459** 42.8014

Critical values for test statistics:

1pct	5pct	10pct	
tau2	-3.43	-2.86	-2.57

⇒ Reject H_0 : unit root at 5% level.

```
• lc.adf_res_div_earn_12 <- ur.df(y=res_div_earn, lags=12, type='drift')
> summary(lc.adf_res_div_earn_12)
```

Value of test-statistic is: **-6.8064** 23.2198 ⇒ Reject H_0 : unit root at 5% level

• We reject H_0 : The linear combination (Dividends_t – **0.554824** – **0.399444** Earnings_t) is stationary. That is, there is a *long-run relation*:
When Earnings increase by \$1, Dividends increase by **\$0.40**.

Engle and Granger Cointegration – E & D

Example (continuation): Now, we build the ECM:

```
d_x_div <- diff(x_div)
d_x_earn <- diff(x_earn)
fit_ecm <- lm(d_x_div ~ res_div_earn[-T] + d_x_earn)
> summary(fit_ecm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.032851	0.002197	14.953	< 2e-16 ***	
res_div_earn[-T]	-0.014309	0.000745	-19.207	< 2e-16 ***	⇒ Significant at 5% level.
d_x_earn	-0.007241	0.002296	-3.153	0.00164 **	⇒ Negative sign?

Residual standard error: 0.09298 on 1794 degrees of freedom

Multiple R-squared: **0.1708**, Adjusted R-squared: 0.1699

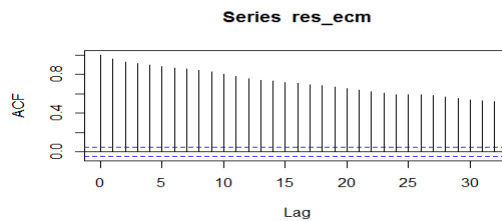
F-statistic: 184.7 on 2 and 1794 DF, p-value: < 2.2e-16

• If error is positive (dividends too high relative to long-run relation), next period there is a negative adjustment in dividends. Good, intuitive. But, sign for changes in Earnings points to misspecification!

Engle and Granger Cointegration – E & D

Example (continuation): Check the residuals for autocorrelation:

```
res_ecm <- fit_ecm$residuals
acf_res <- acf(res_ecm)
```



Highly autocorrelated residuals! Very likely because of misspecified lag structure for dependent variable:

⇒ Add lags of $\Delta \text{Dividends}_t$ in regression.

Engle and Granger Cointegration – E & D

Example (continuation): We add 4 d_x_div lags to regression.

```
p_lag <- 4                                # Select # of lags for test (set p)
div_lag <- matrix(0,T-p_lag,p_lag)        # Matrix to collect lagged residuals
a <- 1
while (a <= p_lag) {                       # Do loop creates matrix (e_lag) with lagged e
  za <- d_x_div[a:(T-p_lag+a-1)]
  div_lag[a,a] <- za
  a <- a+1
}
d_x_earn_p <- d_x_earn[(p_lag+1):T]        # Adjust for new sample size: T - p_lag
ec_term_p <- ec_term[(p_lag+1):T]
dyn_ecm4 <- lm(d_x_div[(p_lag+1):T] ~ div_lag + ec_term_p + d_x_earn_p)
summary(dyn_ecm4)
```

Engle and Granger Cointegration – E & D

Example (continuation):

```
> summary(dyn_ecm4)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.0014896	0.0006921	2.152	0.03150 *	
div_lag1	0.0733874	0.0241628	3.037	0.00242 **	
div_lag2	-0.0117834	0.0321913	-0.366	0.71438	
div_lag3	0.0323275	0.0321303	1.006	0.31449	
div_lag4	0.8526806	0.0245515	34.730	< 2e-16 ***	⇒ Significant lag structure
ec_term_p	-0.0022602	0.0002484	-9.098	< 2e-16 ***	⇒ Negative sign, as expected.
d_x_earn_p	0.0036109	0.0007041	5.128	3.24e-07 ***	⇒ Positive sign!

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02753 on 1786 degrees of freedom

Multiple R-squared: 0.9276, Adjusted R-squared: 0.9274

F-statistic: 3814 on 6 and 1786 DF, p-value: < 2.2e-16

Engle and Granger Cointegration – E & D

Example (continuation):

- The sign for Change in Earnings is reversed. We have a positive relation: A positive (negative) change in Earnings increases (decreases) dividends.

Note: The model was misspecified, the dynamic lag structure for Change in Dividends was missing.

- In the ECM we still have the negative relation (the “*correction*”) between errors and dependent variable, Change in Dividends.