Unit Roots and Cointegration

In the following you will find a reprint of the e-Tutorial No. 9 written by Roberto Perrelli as used in his lectures "Introduction to Applied Econometrics", Course Econ 371 - Fall 2003.

Download the data "eggs.dta" from the Econ 371 web site, *http://www.econ.uiuc.edu/~perrelli/econ371.html*, and go along the STATA instructions. Perform the same investigation within the R-environment. Use the R-function **`read.dta`** from the **`foreign`** package to load binary STATA files. [Diethelm Würtz, May 2004]

e-Tutorial 9:

This issue focuses on time series models, with special emphasis on the tests of unit roots and cointegration. The examples are done with the help of the statistical software package STATA. Next you need to declare your data as time series in STATA:

```
tsset year
```

I. Unit Root: Augmented Dickey-Fuller Test

In STATA, you have two ways to perform the test:

(1) using the command **dfuller**, or

(2) using OLS (but checking for significance in the Dickey-Fuller tables).

I suggest you to consider 3 variations of the test:

(a) models with intercept and trend;

(b) models with intercept, but without trend;

(c) models without intercept and trend.

a) <u>Models including intercept and trend:</u> For example, using 1 lag in the chicken series, you will have the following result

```
dfuller  chic, regress trend lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        52

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.998            -4.146            -3.498            -3.179
------------------------------------------------------------------------------
* MacKinnon approximate p-value for Z(t) = 0.6030

------------------------------------------------------------------------------
D.chic   |      Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
chic     |
      L1 |  -.1820551   .0911164     -1.998   0.051       -.365257    .0011467
      LD |  -.0861985   .1435294     -0.601   0.551      -.3747837    .2023867
_trend   |  -315.6405   266.9686     -1.182   0.243      -852.4168    221.1358
_cons    |   83287.07   42600.86      1.955   0.056      -2367.711    168941.8
------------------------------------------------------------------------------
```

Here the null hypothesis is the presence of unit root. Thus, the augmented Dickey-Fuller statistic is -1.998, and lies inside the acceptance region at 1%, 5%, and 10%. Therefore, we cannot reject the presence of unit root.

b) <u>Models including intercept but not trend</u>: Same rationale, but adjusting the command to:

```
dfuller  chic, regress lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        52

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.618            -3.577            -2.928            -2.599
------------------------------------------------------------------------------
* MacKinnon approximate p-value for Z(t) = 0.4737

------------------------------------------------------------------------------
D.chic   |      Coef.   Std. Err.      t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
chic     |
      L1 |  -.1282545   .0792599     -1.618   0.112      -.2875333    .0310243
      LD |  -.1141494   .1421427     -0.803   0.426      -.3997958    .1714969
_cons    |   51982.91   33508.86      1.551   0.127      -15355.67    119321.5
------------------------------------------------------------------------------
```

What can you conclude from the null hypothesis here?

c) Models excluding both intercept and trend: Idem, but adjusting the command to:

dfuller chic, noconstant regress lags(1)

```
Augmented Dickey-Fuller test for unit root         Number of obs   =        52

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.712            -2.619            -1.950            -1.610

------------------------------------------------------------------------------
D.chic   |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
chic     |
      L1 |  -.0059671   .0083782     -0.712   0.480      -.0227951     .010861
      LD |  -.1757909   .1383822     -1.270   0.210      -.4537398     .102158
------------------------------------------------------------------------------
```

And here, what can you conclude?

Comments on Unit Root Tests:

Note 1: Unit root tests are very sensitive to the number of included lags and/or constant and trends. Very likely, some of the results will indicate the presence of unit root while others will not.

Note 2: *How to make a general conclusion on the test results with so many models available?* Johnston & DiNardo (1997, p.226), for example, mention that one of the objectives of including lags is to achieve white noise residuals. Other authors recommend the use AIC or SIC in the model selection.

Note 3: It is quite simple to calculate information criteria in ADF tests. Each output of "dfuller" corresponds to a linear regression on the lags, constant, and/or trend of the series (for a time trend, you can "approximate" the regression coefficient by using a vector from 1 to 54, instead of years).

Example: The ADF test for unit root on the egg series, using 4 lags, but no constant nor trend is as follows:

dfuller egg, noconstant regress lags(4)

```
Augmented Dickey-Fuller test for unit root         Number of obs   =        49
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)              1.033            -2.622            -1.950            -1.610
------------------------------------------------------------------------------
D.egg    |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
egg      |
      L1 |    .005339    .005167      1.033   0.307      -.0050744    .0157524
      LD |   .3691248   .1547069      2.386   0.021       .0573335    .6809162
     L2D |  -.0210851   .1709519     -0.123   0.902       -.365616    .3234457
     L3D |  -.0248243   .1758323     -0.141   0.888      -.3791909    .3295423
     L4D |  -.0593437   .1599065     -0.371   0.712      -.3816141    .2629267
------------------------------------------------------------------------------
```

Similar output can be obtained by linear regression as follows:

```
regress  D.egg L.egg LD.egg L2D.egg L3D.egg L4D.egg, noconstant

  Source |       SS       df       MS              Number of obs =      49
---------+------------------------------           F(  5,    44) =    1.90
   Model |   275576.07     5  55115.2141           Prob > F      =  0.1144
Residual |  1278907.93    44  29066.0893           R-squared     =  0.1773
---------+------------------------------           Adj R-squared =  0.0838
   Total |  1554484.00    49  31724.1633           Root MSE      =  170.49
------------------------------------------------------------------------------
D.egg    |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
egg      |
      L1 |    .005339    .005167      1.033   0.307      -.0050744    .0157524
      LD |   .3691248   .1547069      2.386   0.021       .0573335    .6809162
     L2D |  -.0210851   .1709519     -0.123   0.902       -.365616    .3234457
     L3D |  -.0248243   .1758323     -0.141   0.888      -.3791909    .3295423
     L4D |  -.0593437   .1599065     -0.371   0.712      -.3816141    .2629267
------------------------------------------------------------------------------
```

Did you understand why?

Note that the t-statistic for the lag of egg (L1) is the same as the ADF statistic, but the distribution used in the ADF hypothesis testing procedure is no longer the trivial t-student. Because of the unit root consequences, specific critical values are provided by Dickey and Fuller to test such statistic (See Table B.6 on Hamilton (1994)).

II. Cointegration: Engle-Granger Test

To perform a simplified version of the Engle-Granger cointegration test, just follow the steps below:

(1) Pre-test the variables for the presence of unit roots. Check if they are integrated of the same order;

(2) Regress chickens against eggs (long run equilibrium relationship)

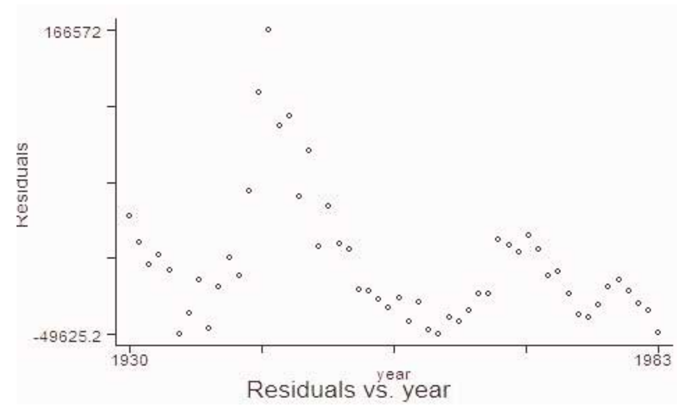
```
regress chic egg

  Source |       SS       df       MS              Number of obs =      54
---------+------------------------------           F(  1,    52) =    2.05
   Model |  4.3347e+09     1  4.3347e+09           Prob > F      =  0.1579
Residual |  1.0981e+11    52  2.1117e+09           R-squared     =  0.0380
---------+------------------------------           Adj R-squared =  0.0195
   Total |  1.1414e+11    53  2.1536e+09           Root MSE      =   45953

------------------------------------------------------------------------------
    chic |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     egg |  -10.21917   7.132592     -1.433   0.158      -24.53176    4.093421
   _cons |   470461.5   36111.96     13.028   0.000       397997.5    542925.4
------------------------------------------------------------------------------
```

(3) Obtain the residuals from this equation

```
predict residual, res
```

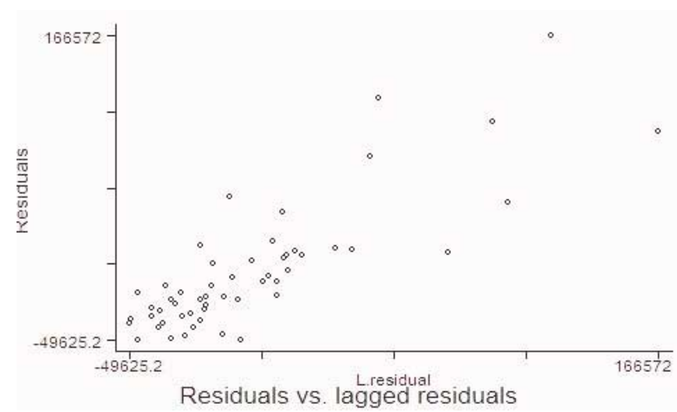
(4) Graph the residuals against time

```
graph residual year, title(Residuals vs. year)
```



And graph the residuals against lagged residuals.

```
graph residual L.residual, title(Residuals vs. lagged residuals)
```



Can you see what's going on in the graphs above?

(5) Proceed with a unit root test on the residuals, considering lags 0 to 4. This is a *residual-based version of the ADF test*. The only difference from the traditional ADF to (this version of) the Engle-Granger test is the critical values. The critical values to be used here are no longer the same provided by Dickey-Fuller, but instead provided by Engle and Yoo (1987) and others (see approximated critical values in Table B.9, Hamilton (1994)). This happens because the residuals above are not the actual error terms, but estimated values from the long run equilibrium equation of chickens against eggs.

Some authors (e.g., Enders (1995)) consider a fourth step, consisting in the estimation of error-correction models and checking of models adequacy. However, you are not required to do it here.

RP

Great thank's to Roberto Perrelli for this e-Tutorial. [DW]