Quantitative Techniques – Term 2

Laboratory 3

2 February 2006

Overview

The objectives of this lab are:
- Provide an overview of stationarity analysis and Granger causality test;
- Estimate demand elasticity.

Start Stata 9 from the Windows Start menu (Programs, F. Departmental Software, E. Social Science, Stata). Do not update the version that is installed on your machine.

Before starting your tasks, make sure that the following windows are visible on your screen and choose their size so that they do not overlap: Review, Variables, Results, Command.

Task 1 – Stationarity Analysis

Go to www.staff.city.ac.uk/a.banal-estanol/teaching.htm and download Lab3Prices.dta to your directory. This contains the same data we used in Lab 2.[1]

In order to carry out this task, we need to:
(1) determine whether the time series of prices are stationary or integrated[2];
(2) verify that their ratio is stationary.

This task is at a very introductory level and has the only objective of illustrating the steps required, while it does not take into account possible complications of unit root tests.

    1. Dickey-Fuller tests for unit roots.[3]
Reminder from lecture: The null hypothesis is that the series is <u>non-stationary</u>.

As last week, consider gasoline prices on the East Coast (P1) and on the West Coast (P2).

You can run Dickey-Fuller the test in Stata by clicking on the Menu bar… Statistics… Time Series… Tests… Augmented Dickey Fuller unit-root test.
- Choose the variable you want to run the test for (e.g. eastcoast);
- Tick "Display regression table";

---

[1] The source of the data is http://www.economagic.com/doewkly.htm

[2] This is needed because, if the two time series are stationary, then their ratio is also going to be stationary and the test will not be very informative of whether the products are in the same market.

[3] This is only one of the available tests for unit roots in time-series data. In our example, we are testing whether the series is $x\_t = a + x\_t\text{-}1 + e\_t$.

References: Greene, W. (2003), "Econometric Analysis", Prentice Hall, Section 20.3.4; Charemza, W., and D. Deadman (1992), "New Directions in Econometric Practice", Edward Elgar.
http://www.stata.com/help.cgi?dfuller

- Set the maximum number of lags to zero (setting the lags > 0 means that you are running an "Augmented Dickey Fuller" test – we do not consider this case).[4]

```
. dfuller eastcoast, regress lags(0)

Dickey-Fuller test for unit root                   Number of obs   =        95

                         ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic          Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -1.662           -3.517            -2.894            -2.582
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4506


------------------------------------------------------------------------------
D.eastcoast |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
  eastcoast |
       L1. |  -.0724087    .043557    -1.66   0.100    -.1589042    .0140868
      _cons |   8.058461   4.840097    1.66   0.099    -1.553013   17.66993
------------------------------------------------------------------------------
```

The test statistic (-1.66) is larger (less negative) than the critical values and therefore we cannot reject the null hypothesis of non-stationarity.

We use the difference operator and we repeat the test on the differenced series, i.e. eastcoast_t – eastcoast_t-1.

```
. dfuller d.eastcoast, regress lags(0)

Dickey-Fuller test for unit root                   Number of obs   =        94

                         ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic          Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -7.495           -3.518            -2.895            -2.582
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000


------------------------------------------------------------------------------
D2.eastcoast |    Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   eastcoast |
        LD. |  -.7907026   .1054998    -7.49   0.000    -1.000234   -.5811708
       _cons |   .0167986    .04227     0.40   0.692    -.0671533    .1007505
------------------------------------------------------------------------------
```

We can now reject the null hypothesis that (eastcoast_t – eastcoast_t-1) is non-stationary because the test statistic is more negative than the critical value. Given that the first difference is stationary, we can conclude that eastcoast is an integrated series.

Repeating the steps above for westcoast, we obtain analogous results. Both series are then integrated of the same order and we can proceed to the next step.

2. Stationarity of price ratio

In order to test the null hypothesis of stationarity of the price ratio, we first need to generate a new variable, e.g. we can call it 'ratio'. This can be done in two ways:

---

[4] The same results are obtained if you tick "Random walk with drift". In order to test for random walk, it is necessary to remove the constant from the regression.

- In the Command window, type **gen ratio= eastcoast/ westcoast**
- On the Menu bar, click on Data… Create or change variables… Create new variable. In the window you get, choose the name of the new variable, e.g. **ratio**, and then write **eastcoast/westcoast** in 'contents of the new variable'.

Now perform a Dickey-Fuller unit-root test on the new variable 'ratio'.

```
. dfuller ratio, lags(0)

Dickey-Fuller test for unit root                     Number of obs   =        95

                            ---------- Interpolated Dickey-Fuller ---------
                Test        1% Critical       5% Critical      10% Critical
             Statistic         Value             Value             Value
------------------------------------------------------------------------------
 Z(t)          -1.526          -3.517            -2.894            -2.582
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5208
```

Given that the test statistic (highlighted in green) is less negative that the critical value, we cannot reject the hypothesis that the ratio of prices has a unit root. The evidence indicates that the prices on the East Coast and on the West Coast do not move together over time.

Summary of Task 1:
- When two products (or areas) belong to the same market, the price of one constrains the price of the other, i.e. the prices move together over time;
- The stationarity of the ratio is taken as an indication that the prices move together and therefore the two products or areas belong to the same market;
- In our example, the test results for the individual price series have indicated non stationarity;
- On the basis of the unit-root test on the price ratio, we cannot reject the null hypothesis of stationarity of the ratio and therefore we cannot conclude that the two areas belong to the same market.

On the basis of the test results on the price ratio, what can we conclude about geographic market definition for retail gasoline?
What happens to your results if you perform a stationarity test on the ratio of prices on the West Coast over East Coast?
Given the output above for the unit-root test on 'ratio', what results would you obtain if you were to run the following regression: d.ratio = constant + β*l.ratio + error? In particular, what would the value of the t-statistic on β be?

Task 2 – Granger causality[5]
- Using the same data as in Task 1, we carry out a bivariate Granger causality test (time series of the gasoline prices on the East Coast and on the West Coast);

---

[5] There is an alternative way of testing for Granger causality directly, but it requires installing an additional routine from the web. For this reason, it is not feasible in the University computer room. The description of how to install this package follows, for your information. Type **net install gcause.pkg** in the Command window. Alternatively, from the Menu bar, Help… Search… Search all: the keyword to search for is "gcause". Then follow the instructions to install the package and use the help file for its syntax.

- Note that, for simplicity, we are not removing the effects of any <u>common factors</u> except for the time trend;
- The first step is to estimate the following model by OLS (from LECG, 1999, page 59). The errors have zero mean and constant variance.[6]

$$P_1 = \sum_{s=1}^{T} \beta_s P_{1t-s} + \sum_{s=1}^{T} \gamma_s P_{2t-s} + u_t$$

- In our example, P1 is the price on the East Coast and P2 the price on the West Coast. Assume that we set the lag length = 3, so we have that T = 3.
- In order to simplify the notation we use the lag operator, **L.** or **l.**, followed by the name of the variable. Note the syntax of the lag operator when more than one lag is included in a regression. This syntax is not valid when performing a test.
- Note that we are running a simple OLS regression, using the command 'regress' or 'reg'.[7] This can alternatively be run using the Menu bar instead of typing in the Command window.

```
. reg eastcoast L(1/3).eastcoast L(1/3).westcoast time

      Source |       SS       df       MS              Number of obs =      93
-------------+------------------------------           F(  7,    85) =   82.76
       Model | 79.9069453     7  11.4152779           Prob > F      =  0.0000
    Residual | 11.7248659    85  .137939599           R-squared     =  0.8720
-------------+------------------------------           Adj R-squared =  0.8615
       Total | 91.6318112    92  .995997947           Root MSE      =  .3714

------------------------------------------------------------------------------
    eastcoast |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    eastcoast |
         L1. |   1.044364   .1066324     9.79   0.000     .8323503    1.256378
         L2. |   .1012993   .1501328     0.67   0.502    -.1972051    .3998036
         L3. |  -.3537807   .1092032    -3.24   0.002     -.570906   -.1366555
    westcoast |
         L1. |   .0362912   .0667641     0.54   0.588    -.0964537    .1690361
         L2. |  -.0004602   .1021941    -0.00   0.996    -.2036495    .2027291
         L3. |   .0131382   .0676851     0.19   0.847     -.121438    .1477144
        time |  -.0038358   .0023731    -1.62   0.110    -.0085542    .0008826
       _cons |   17.31988   5.844017     2.96   0.004     5.700412    28.93936
------------------------------------------------------------------------------
```

- Then we perform an F-test for causality, i.e. the null hypothesis is that the coefficients on P2 are jointly zero.[8]

```
. test l.westcoast l2.westcoast l3.westcoast

 ( 1)  L.westcoast = 0
 ( 2)  L2.westcoast = 0
 ( 3)  L3.westcoast = 0

       F(  3,    85) =    1.55
            Prob > F =   0.2070
```

[6] References for Granger causality test: Greene, W. (2003), "Econometric Analysis", Prentice Hall; Hamilton, J.D. (1994), "Time Series Analysis", Princeton University Press.
[7] http://www.stata.com/help.cgi?regress
[8] http://www.stata.com/help.cgi?test

The test statistic (+1.55) is too low to reject the null hypothesis that the coefficients on l.westcoast, l2.westcoast and l3.westcoast are jointly zero. Therefore we cannot conclude that prices on the West Coast do not Granger cause prices on the East Coast.

This conclusion therefore leaves open the possibility that the West Coast and the East Coast form a single geographic market.

Summary of Task 2:

- Causality tests may be used to investigate correlations between the prices of two products (or in two areas);
- If the hypothesis of no causality is rejected, then the two products (or areas) are likely to be in the same market;
- In our example, we were unable to reject the hypothesis that the prices on the West Coast could not be used to forecast prices on the East Coast.

In our example, is the outcome of the F-test consistent with the t-tests on each coefficient? Do you think that the result on Granger causality is symmetric if we swap prices on the East Coast and on the West Coast? Is it sufficient to perform the Granger test in Task 2 to conclude that the East Coast and the West Coast are not part of the same market?

Task 3 – Price Elasticity Analysis
Go to www.staff.city.ac.uk/a.banal-estanol/teaching.htm and download Lab3Demand.dta to your directory.[9] Open it in Stata.

The file contains annual data on the U.S. gasoline market for the years 1960-1995. The variables are (in natural logarithms):
- Lngpop: U.S. gasoline consumption per capita (deflated);
- Lnpg: price of gasoline;
- Lny: per capita income;
- Lnpnc: price of new cars;
- Lnpus: price of used cars;

We will use this dataset to estimate own-price and cross-price elasticities.

"The own-price elasticity of a product (or group of products) is defined as the percentage change in sales for each percentage change in the price. [...]
Cross-price elasticity between two products measures the extent to which the volume of sales of one responds to changes in the price of the other." (Lexecon report, pages 14-15)

1. OLS regression[10]
Under the OLS assumptions, we estimate the following model:
Lngpop = $\beta_1 + \beta_2$*lnpg + β_3*lny + β_4*lnpnc + β_5*lnpuc + + β_6*year + error

---

[9] The source is: Greene (2003), op. cit, Example 2.3., page 12.
[10] Note that the constant reported in Greene (2003), Example 4.4, should be corrected: see http://pages.stern.nyu.edu/~wgreene/Text/Errata/ERRATA5.htm

```
. reg lngpop lnpg lny lnpnc lnpuc year

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  5,    30) =  185.84
       Model | .780166138        5 .156033228           Prob > F      =  0.0000
    Residual | .025187728       30 .000839591           R-squared     =  0.9687
-------------+------------------------------           Adj R-squared =  0.9635
       Total | .805353866       35  .02301011           Root MSE      =  .02898


------------------------------------------------------------------------------
      lngpop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnpg | -.1155303   .0334795    -3.45   0.002    -.1839046   -.0471561
         lny |  1.954627   .1928542    10.14   0.000     1.560766    2.348488
       lnpnc |  .2052824   .1520193     1.35   0.187    -.1051825    .5157473
       lnpuc | -.1292738   .0714118    -1.81   0.080    -.2751161    .0165684
        year | -.0191182   .0059565    -3.21   0.003     -.031283   -.0069533
       _cons |  20.06665   10.11464     1.98   0.056    -.5901884     40.7235
------------------------------------------------------------------------------
```

Given that the model is in loglinear form (or constant elasticity form), we can read the elasticities of demand with respect to the regressors directly from the coefficients.
As expected, the own-price elasticity is negative and pretty small in absolute value (-0.11), which indicates that gasoline demand is not very elastic. The income elasticity also has the expected sign.

2. Omitted variables

Note what would happen if income was omitted from the above equation, i.e. if we were to estimate the following model.

```
. reg lngpop lnpg lnpnc lnpuc

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  3,    32) =   11.15
       Model | .411551052        3 .137183684           Prob > F      =  0.0000
    Residual | .393802877       32  .01230634           R-squared     =  0.5110
-------------+------------------------------           Adj R-squared =  0.4652
       Total |  .80535393       35 .023010112           Root MSE      =  .11093


------------------------------------------------------------------------------
      lngpop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnpg |  .0756982   .1061912     0.71   0.481    -.1406062    .2920026
       lnpnc | -.6680671   .4145029    -1.61   0.117    -1.512382    .1762478
       lnpuc |  .4948794   .2484313     1.99   0.055    -.0111585    1.000917
       _cons | -.0870583    .033138    -2.63   0.013    -.1545582   -.0195584
------------------------------------------------------------------------------
```

We obtain the counterintuitive result that the own-price elasticity is positive (+0.07), even if the coefficient is not significant. The coefficient on Lnpg is in fact capturing also the relationship between price and per capita income (the omitted variable). E.g. see Greene (2003), Section 8.2., for a technical explanation.

If per capita income was not included in the model, we would have obtained a non significant and positive estimate for the price elasticity of demand.

3. Violations of OLS assumptions/Misspecification: serially correlated residuals[11]

So far we have estimated the demand model under the OLS assumptions, but we have not verified whether this was correct. Tests on the residuals from the OLS estimation should have been carried out to detect potential problem. For instance, a misspecification of the model may also result in problems of serial correlation in the residuals.

Assume we regressed gasoline consumption only on its price by OLS, i.e.

```
. reg lngpop lnpg

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  1,    34) =   27.38
       Model |  .359241235     1  .359241235           Prob > F      =  0.0000
    Residual |  .446112631    34   .01312096           R-squared     =  0.4461
-------------+------------------------------           Adj R-squared =  0.4298
       Total |  .805353866    35   .02301011           Root MSE      =  .11455

------------------------------------------------------------------------------
      lngpop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnpg |   .1676711   .0320441     5.23   0.000     .1025497    .2327925
       _cons |  -.1167347   .0288282    -4.05   0.000    -.1753206   -.0581489
------------------------------------------------------------------------------
```

In the model above, the price elasticity is again positive, as in point 2 above.

Note that the residuals from this misspecified equation would look serially correlated:[12]
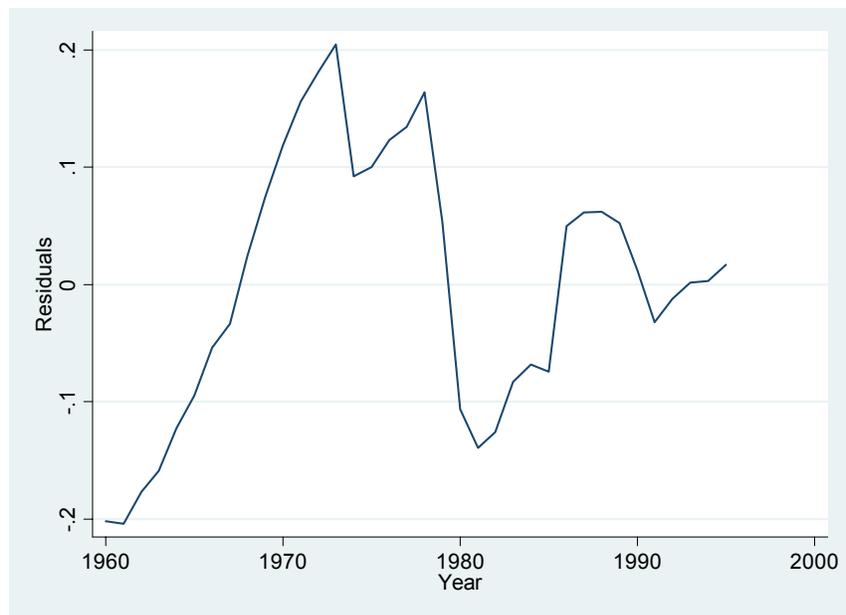
You can obtain the plot in two ways:

---

[11] This point is based on example 12.2. of Greene (2003).
[12] Reminder from Lab1 – After you estimate a model, you can obtain the residuals by typing: predict resid, residuals. Note that 'resid' is the name I give to the new variable that contains the residuals, but you could choose another one.

- In the Command window, type: line resid year
- From the Menu, choose Graphics… Easy Graph… Line graph.

If we compare this with the residuals from the full model, as in point 1, we can see that improving the specification of the model reduces the autocorrelation of the residuals.
The residuals from the first model have already been created in the dataset and are called 'resfull'. Plot them using the commands above.
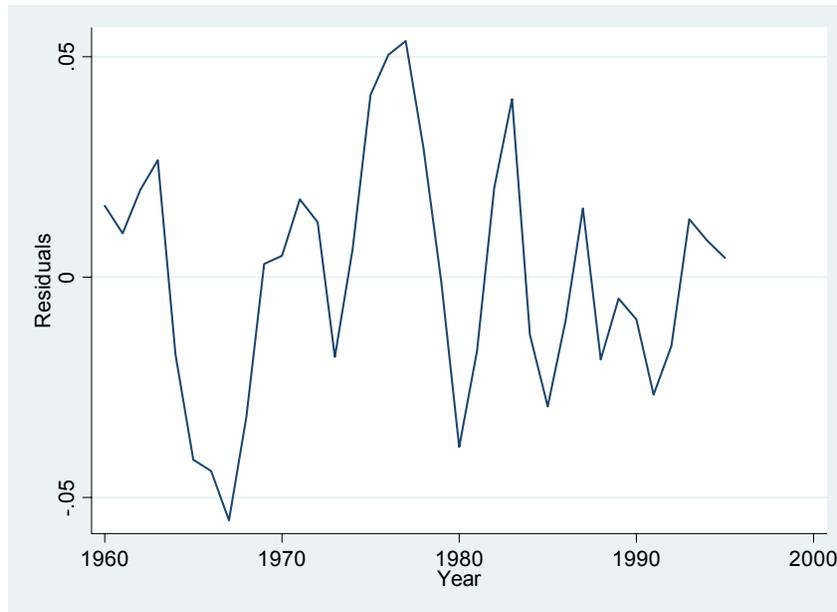
By improving the specification of the model, the degree of serial correlation in the residuals has been reduced and the sign of the own-price elasticity has become negative.

Summary of Task 3:

- The estimation of demand elasticities is carried out in the context of a model of consumption as a function of income, own price and other factors;
- In our example, we found a low elasticity of demand with respect to price. A hypothetical monopolist would therefore find it profitable to increase prices, as the reduction in sales would be outweighed by the higher profits;
- A misspecified model may bias coefficients (example of omitted variables);
- With time series data, serial correlation of the residuals may arise. We have seen an example in which it can be caused by a misspecification of the model.

What sign do you expect for the coefficient β_4?
On the basis of the analysis in point 1, what do you conclude on the cross elasticity of gasoline consumption with respect to the price of new cars?
Given the results in point 1, what can you conclude about the definition of the market for gasoline?
What other example of violation of the OLS assumptions can you think of?