Quantitative Techniques – Term 2

Laboratory 2

26 January 2006

Overview

The objectives of this lab are:

- Apply known statistical concepts (correlation, tests with two samples) to competition policy;
- Guide you through an overview of other statistical tests of prices and price trends.

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 – From last week: import a data file[1]

- Go to www.staff.city.ac.uk/a.banal-estanol/teaching.htm and download Lab2Prices.csv to your directory.
  Note that the names of the variables are in the first row, immediately followed by your data.

  The file contains time series of weekly prices of retail gasoline prices in $cents per gallon in different areas of the United States:[2]
  - East Coast;[3]
  - Midwest;[4]
  - Gulf Coast;[5]
  - Rocky Mountains;[6]
  - West Coast.[7]

  In addition, the dates when prices are measured are also included, as is a variable "Time", which simply counts the number of weeks.

- Type in the Command window:
  insheet using "*your directory*\Lab2Prices.csv"

---

[1] Reference: http://www.stata.com/help.cgi?infiling

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

[3] CT, MA, VT, NH, ME, RI, NY, NJ, PA, DC, MD, DE, WV, VA, NC, SC, GA, FL.

[4] MI, OH, KY, TN, IN, WI, IL, MN, IO, MO, ND, SD, NE, KS, OK.

[5] Al, MS, AR, LA, TX, NM.

[6] MT, WY, CO, ID, UT.

[7] AK, WA, OR, NV, CA, AZ, HI.

This command loads the data in Stata. Note that Stata recognizes commands in lower case letters. Use " " when you specify a filename.
Alternatively, you can use the Menu bar: File, Import, ASCII data created by a spreadsheet.

- Save the file as Lab2Prices.dta. This is the extension used by Stata for datasets.

Task 2 – Cross-sectional price tests: independent samples[8]

This task builds on the statistical concepts you have learned in the first term. We want to test the null hypothesis that the average price in Midwest (μ_M) is equal to the average price on the East Coast (μ_E) (e.g. H0: $\mu_M = \mu_E$). Stata allows you to test this hypothesis against three alternative hypotheses at the same time.

You can:

- Go to the Command window and type **ttest eastcoast== midwest, unpaired**; OR
- From the Menu bar, go to Statistics… Summaries, tables and tests… Classical tests of hypotheses… Two-sample mean comparison test. Then choose the variables and tick "unpaired".

```
. ttest eastcoast== midwest, unpaired

Two-sample t test with equal variances

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
eastco~t |      96    111.1375    .1012666    .9922064    110.9365    111.3385
 midwest |      96     109.001    .1331427    1.304526    108.7367    109.2654
---------+--------------------------------------------------------------------
combined |     192    110.0693    .1137245    1.575813     109.845    110.2936
---------+--------------------------------------------------------------------
    diff |            2.136458    .1672779                1.806498    2.466419
------------------------------------------------------------------------------
Degrees of freedom: 190

               Ho: mean(eastcoast) - mean(midwest) = diff = 0

     Ha: diff < 0               Ha: diff != 0              Ha: diff > 0
       t =  12.7719               t =  12.7719              t =  12.7719
   P < t =   1.0000          P > |t| =   0.0000          P > t =   0.0000
```

Task 3 – Hedonic price analysis[9]

This exercise is based on Exercise 3, Chapter 4 of Berndt, E.R. (1991), "The Practice of Econometrics: Classic and Contemporary", Addison-Wesley.
The purpose of the exercise is to construct a price index for computers using hedonic regression techniques.

- Download Lab2Hedonic.dta from www.staff.city.ac.uk/a.banal-estanol/teaching.htm

[8] http://www.stata.com/help.cgi?ttest
[9] Reference on the use of hedonic price indexes in practice: http://www.bls.gov/ppi/ppicomqa.htm

Open Stata in a new window and then you can open the file by either:

- Open – using the Menu bar or the icon on the toolbar; or
- Type in the command window:
  use "*your directory*\Lab2Hedonic.dta"
  Again, note the use of " " when you specify a filename.

The file contains the following variables:

- Year: year in which the model was introduced (from 1954 to 1965);
- Lnrent: natural log of monthly rental of computers;
- Lnmult: natural log of average time to obtain and complete multiplication instructions;
- Lnaccess: natural log of average time to access information from memory;
- Lnmem: natural log of memory size;
- D61: a variable which takes value 1 when year=61 and zero otherwise;
- D62, d63, d64, d65: constructed similarly as for d61.

By using OLS, estimate the parameters of the equation for years '60 to '65:[10]
Lnrent = constant + a*d61 + b*d62 + c*d63 + d*d64 + e*d65 + f*lnmult + g*lnaccess + h*lnmem + error

(You can also click on the variables in the left-hand side Variables window.)

```
. reg  lnrent d61 d62 d63 d64 d65 lnmult lnaccess lnmem if year>59

      Source |       SS       df       MS              Number of obs =      82
-------------+------------------------------           F(  8,    73) =   88.77
       Model |  106.679895     8  13.3349869           Prob > F      =  0.0000
    Residual |   10.965562    73  .150213178           R-squared     =  0.9068
-------------+------------------------------           Adj R-squared =  0.8966
       Total |  117.645457    81  1.45241305           Root MSE      =  .38757

------------------------------------------------------------------------------
      lnrent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d61 |  -.1386585   .1679443    -0.83   0.412     -.473371    .1960541
         d62 |  -.4916432    .175347    -2.80   0.006    -.8411092   -.1421771
         d63 |  -.5948134   .1676572    -3.55   0.001    -.9289537   -.2606731
         d64 |  -.9265809   .1678509    -5.52   0.000    -1.261107   -.5920546
         d65 |   -1.14627   .1671586    -6.86   0.000    -1.479416   -.8131232
      lnmult |  -.0711882   .0282091    -2.52   0.014    -.1274089   -.0149675
    lnaccess |  -.1349376   .0292385    -4.62   0.000    -.1932099   -.0766654
       lnmem |   .5770442   .0357484    16.14   0.000     .5057978    .6482907
       _cons |  -.0733124   .3176761    -0.23   0.818    -.7064399    .5598152
------------------------------------------------------------------------------
```

- Note that the coefficients highlighted above can be interpreted as elasticities.

- The coefficients on d61, d62, d63, d64 and d65 can be used to construct quality-adjusted price indices with respect to 1960, as follows:

[10] http://www.stata.com/help.cgi?regress

Year	Coefficient (e.g. α)	Price index - antilogarithm (e.g. e^{α})	% Change with respect to 1960
1961	-.1386	0.871	-12.9%
1962	-.49164	0.612	-38.8%
1963	-.5948	0.552	-44.8%
1964	-.9265	0.396	-60.4%
1965	-1.1462	0.318	-68.2%

Task 4 – Price correlation[11]

- Using the same data as in Task 2, calculate the correlation between the prices in the different geographic areas. (Remember you can also use the Menu bar: Statistics… Summaries, tables and tests…Summary statistics… Correlations and covariances).

```
. correlate eastcoast midwest westcoast gulfcoast rocky
(obs=96)

             | eastco~t  midwest westco~t gulfco~t    rocky
-------------+---------------------------------------------
   eastcoast |   1.0000
     midwest |   0.6658   1.0000
   westcoast |   0.2490   0.5745   1.0000
   gulfcoast |   0.7134   0.9225   0.6322   1.0000
       rocky |   0.1291   0.6182   0.8434   0.6280   1.0000
```

- Repeat the correlation analysis for the log of prices (LECG report, page 54). Reminder: to generate a new variable, you need the command "generate". E.g. **gen lneastcost=ln(eastcoast)**

```
. correlate  lneastcoast lnmidwest lnwestcoast lngulfcoast lnrocky
(obs=96)

             | lneast~t lnmidw~t lnwest~t lngulf~t  lnrocky
-------------+---------------------------------------------
 lneastcoast |   1.0000
   lnmidwest |   0.6647   1.0000
 lnwestcoast |   0.2469   0.5761   1.0000
 lngulfcoast |   0.7122   0.9224   0.6335   1.0000
     lnrocky |   0.1340   0.6231   0.8451   0.6340   1.0000
```

- Repeat the analysis for the differences in the logarithms of prices (approximation of growth rates).
- In order to do this, we will use the difference operator. First, you need to declare your dataset to be time-series by typing **tsset time**

```
tsset time
    time variable:  time, 1 to 96
```

- The difference operator is written **D.** or **d.** followed by the name of the variable. You don't need to generate a variable separately, you can simply used the command you need followed by the difference operator.

[11] http://www.stata.com/help.cgi?corr

```
. correlate D.lneastcoast D.lnmidwest D.lnwestcoast D.lngulfcoast D.lnrocky
(obs=95)

             |        D.       D.       D.       D.       D.
             | lneast~t lnmidw~t lnwest~t lngulf~t  lnrocky
-------------+---------------------------------------------
lneastcoast  |
          D1 |   1.0000
lnmidwest    |
          D1 |   0.6454   1.0000
lnwestcoast  |
          D1 |   0.2513   0.2297   1.0000
lngulfcoast  |
          D1 |   0.5669   0.6636   0.4231   1.0000
lnrocky      |
          D1 |   0.1064   0.2458   0.3826   0.2795   1.0000
```

Task 5 – Granger causality[12]

- Using the same data as in Task 2, we carry out a bivariate Granger causality test, assuming a lag length 3;
- Note that, for simplicity, we are not removing the effects of any common factors except for the time trend, to demonstrate the syntax of the command;
- The first step is to estimate the following model by OLS (from LECG, 1999, page 59).[13]

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

- In our example, P1 is the price on the East Coast and P2 the price on the West Coast. As we have set the lag length = 3, 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.

```
. 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
```

[12] 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.

[13] References for Granger causality test: Greene, W. (2003), "Econometric Analysis", Prentice Hall; Hamilton, J.D. (1994), "Time Series Analysis", Princeton University Press.

```
        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.

```
. 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
```

Task 6 – Dynamic price regressions

Using the same data as in Task 2, we need to test whether the time series of prices are stationary or to identify their order of integration, then verify if they are co-integrated. This task is at a very introductory level and has the only objective of illustrating the steps required in the estimation of dynamic regressions.

1. Dickey-Fuller tests for unit roots.[14]

Consider once again prices on the East Coast (P1) and on the West Coast (P2). In this example, we are going to use lneastcoast and lnwestcoast.
The null hypothesis is that the series is non-stationary.

You can run 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. lneastcoast);
- Tick “Display regression table”;
- 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 for the moment).

[14] 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-1 + e_t
References: Greene, op. cit., Section 20.3.4; Charemza, W., and D. Deadman (1992), “New Directions in Econometric Practice”, Edward Elgar.
http://www.stata.com/help.cgi?dfuller

```
. dfuller lneastcoast, 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)             -1.666            -3.517            -2.894            -2.582
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4487

------------------------------------------------------------------------------
D.           |
 lneastcoast |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lneastcoast  |
          L1 |  -.0724415   .0434823    -1.67   0.099    -.1587886    .0139056
_cons        |   .3413512   .2048254     1.67   0.099    -.0653915    .7480938
------------------------------------------------------------------------------
```

The test statistic (-1.666) is larger (less negative) than the critical values and therefore we cannot reject the null hypothesis of non-stationarity. We use again the difference operator and we repeat the test on the differenced series, i.e. lneastcoast_t – lneastcoast_t-1.

```
. dfuller d.lneastcoast, 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)             -7.519            -3.518            -2.895            -2.582
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D2.          |
 lneastcoast |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lneastcoast  |
          LD |  -.7929278   .1054501    -7.52   0.000    -1.002361   -.5834947
_cons        |   .0001496   .0003803     0.39   0.695    -.0006056    .0009049
------------------------------------------------------------------------------
```

We can now reject the null hypothesis that (lneastcoast_t – lneastcoast_t-1) is non-stationary. This leads us to the conclusion that lneastcoast is a series integrated of order one.
Repeating the steps above for lnwestcoast, we obtain analogous results for lnwestcoast. Both series are then integrated of the same order and we can proceed to the next step.

2. Co-integration test.

This test involves verifying whether there exists a linear combination of the two series that is stationary (see LECG, page 64). We therefore need to test whether the fitted residuals from the linear model lneastcoast_t = a + b*lnwestcoast_t + e_t are stationary.

- Estimate the following model

```
. reg lneastcoast lnwestcoast

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F(  1,    94) =    6.10
       Model |  .000462415     1  .000462415           Prob > F      =  0.0153
    Residual |  .007120168    94  .000075746           R-squared     =  0.0610
-------------+------------------------------           Adj R-squared =  0.0510
       Total |  .007582583    95  .000079817           Root MSE      =   .0087

------------------------------------------------------------------------------
 lneastcoast |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lnwestcoast |   .1073602   .0434519     2.47   0.015     .0210855     .193635
       _cons |   4.194378   .2089846    20.07   0.000     3.779434    4.609322
------------------------------------------------------------------------------
```

- Calculate the residuals

```
. predict resid, residuals
```

- Test for stationarity of the residuals

In order to reject the null hypothesis of non-stationarity, the test statistic has to be more negative that the critical value for this test. At a 5% level, for around 100 observations, the critical value is – 4.14.

Note that there are separate tables for stationarity tests when applied to the residuals. For this reason, the critical value does not correspond to that given by Stata. (For an explanation see Charemza and Deadman, page 129).

```
. dfuller resid, 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)             -1.850            -3.517            -2.894            -2.582
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3559

------------------------------------------------------------------------------
D.resid      |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
resid        |
          L1 |  -.0800248   .0432527    -1.85   0.067    -.1659161    .0058664
_cons        |   .0000516   .0003689     0.14   0.889    -.0006809    .0007841
------------------------------------------------------------------------------
```

As the test statistic is less negative than the critical value, we conclude that the residuals are not stationary and therefore the two variables are not cointegrated.

Answer to question in Lab 1, related to Task 5:

How to test the following hypothesis?

H0: coefficient on educ = 0.1, coefficient on kidsge6 = -0.03

```
. test (educ=0.1) (kidsge6=-0.03)

 ( 1)  educ = .1
 ( 2)  kidsge6 = -.03

       F(  2,   421) =    0.27
            Prob > F =    0.7639
```