Quantitative Techniques – Term 2

Laboratory 9

23 March 2006

Overview
The purpose of this lab is to estimate the parameters of a production function and introduce the concept of endogenous variable.

Go to www.staff.city.ac.uk/a.banal-estanol/teaching.htm, download Lab9.dta to your directory.
The dataset is from Greene, W., (2003) "Econometric Analysis", Prentice Hall, Example 6.2 and contains cross-section data on output, labour and capital.

Introduction – Production Functions and Endogeneity

In Lecture 7, we studied the Cobb-Douglas production function:

$$Q_i = e^{\alpha} K_i^{\beta_k} L_i^{\beta_l} e^{u_i} \qquad (1)$$

where Q_i is output for firm i, K_i is the capital input (or more generally a fixed input), L_i is the labour input (or more generally a variable input) and u_i is an error term.

Taking logs we linearize the function as:

$$y_i = \alpha + \beta_k k_i + \beta_l l_i + u_i \qquad (2)$$

where $y_i = lnQ_i$, $k_i = lnK_i$ and $l_i = lnL_i$.

As in Lecture 7, assume that the "true model" also included another regressor, X_i, not observable by the analyst. This unobserved variable could be, for instance, different managerial capability across firms or different productivity shocks.

$$y_i = \alpha + \beta_k k_i + \beta_l l_i + \beta_x X_i + \varepsilon_i \qquad (3)$$

If we estimated equation (2), then we would in fact bundle the unobserved variable with the error. If observed inputs were correlated with the unobserved variable, the OLS assumptions would fail. In particular, the right-hand variables would be correlated with the error and the hypothesis of exogeneity would fail. Using the notation above, $E(k_i * u_i) \neq 0$ and/or $E(l_i * u_i) \neq 0$.

From an economic point of view, let's consider the example of a productivity shock which the analyst cannot observe (but the firm observes). For instance, when choosing its level of labour, the firm has observed the shock and therefore the input is correlated with the shock. This means that we cannot assume that the regressors in the model are exogenous. As a result, the OLS estimates are inconsistent.

We saw in Lecture 7 that a way around this problem is to estimate the production function using panel data. Another alternative is to use instrumental variables.

What is an instrument?
In the example of productivity shocks and labour input, consider the price of labour. Assume that the labour market is competitive, i.e. none of our firms can determine the price of labour.

The price of labour is (a) uncorrelated with the error in the firms' production function; (b) correlated with the level of labour.
Therefore the price of labour can be used as an instrument for the level of labour.

Intuitively, using a variable as an instrument means that we are (a) regressing the endogenous variable on the instrument (and the other endogenous variables); (b) use the predicted endogenous variable as a regressor in our initial model.

Another example: demand equations
In the first part of the second term, we studied residual demand analysis and merger simulation. The estimation of demand functions also suffers from an endogeneity problem, since prices and quantities are jointly determined.
When estimating a demand function, we have somehow assumed that it was already an equilibrium relationship and we have not taken the supply side into account. For instance, in the lab of merger simulation we have estimated a model in which price was a regressor but we have not treated it as endogenous.
e.g. LECG report for the OFT (1999), Chapter 9 on residual demand (pages 71 - 72) and Chapter 15 on estimation of demand systems.

Task 1 – OLS Estimation

1. *Obtain estimates of parameters*

- Generate the logarithms of the variables using the 'gen' command;
- Estimate the production function by OLS and correct the standard errors for heteroscedasticity using 'robust';

```
. reg lnoutput lnlabour lncapital, robust

Regression with robust standard errors                 Number of obs =      27
                                                       F(  2,    24) =  382.77
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9435
                                                       Root MSE      =  .18837

------------------------------------------------------------------------------
             |               Robust
    lnoutput |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .6029994   .1751383     3.44   0.002      .2415317    .9644671
   lncapital |   .3757102   .1151868     3.26   0.003      .1379763     .613444
       _cons |   1.170644   .2805119     4.17   0.000      .5916957    1.749592
------------------------------------------------------------------------------
```

- What can you conclude about the overall quality of the regression and the individual coefficients?
- Can you infer anything about returns to scale?

Hint: use the command 'test'[1] or go to Statistics… Postestimation… Tests… Test linear hypotheses… (Select 'linear expressions are equal').

Further exercise:

- A more flexible functional form is given by the translog production function:[2]

$$y_i = \beta_1 + \beta_2 l_i + \beta_3 k_i + \beta_4(\tfrac{1}{2}l_i^2) + \beta_5(\tfrac{1}{2}k_i^2) + \beta_6 l_i k_i + u_i$$

- As an exercise, generate the new variables you need for the estimation of a translog function and estimate the function by OLS;
- Then use an F-test for the null hypothesis that $\beta_4 = \beta_5 = \beta_6 = 0$ (i.e. the null hypothesis that the Cobb-Douglas model is appropriate).

Task 2 – IV (Instrumental Variables) Estimation[3]

In this task, we will use the price of labour (pl in your dataset) as an instrument for the level of labour.

Pl = price of labour
Pk = price of capital

1. Obtain estimates of parameters

- From the introduction, remember the conditions for a variable to be used as an instrument. Look at the correlation between lnlabour and pl reported below. What can you conclude?

```
. corr lnlabour pl
(obs=27)

             | lnlabour       pl
-------------+------------------
    lnlabour |   1.0000
          pl |  -0.9433   1.0000
```

- From the Menu, go to Statistics… Endogenous Covariates… Instrumental variables and 2SLS. Alternatively, you can directly type the following command. Note the syntax for the endogenous variable and the instrument – highlighted below.

```
. ivreg lnoutput lncapital (lnlabour =  pl), robust

Instrumental variables (2SLS) regression               Number of obs =      27
                                                       F(  2,     24) =  481.65
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9430
                                                       Root MSE      =  .18919

------------------------------------------------------------------------------
             |               Robust
    lnoutput |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |   .6603947   .1740447     3.79   0.001     .3011842    1.019605
   lncapital |     .34092   .1184132     2.88   0.008     .0965272    .5853128
       _cons |   1.098882   .2464031     4.46   0.000     .5903309    1.607433
------------------------------------------------------------------------------
Instrumented:  lnlabour
Instruments:   lncapital pl
------------------------------------------------------------------------------
```

[1] http://www.stata.com/help.cgi?test
[2] See Greene (2003).
[3] http://www.stata.com/help.cgi?ivreg

- What has happened to the coefficients?
- Look at the instruments Stata has used in the above model, lncapital and pl. Is it in line with what you expected?

VERY IMPORTANT: http://www.stata.com/support/faqs/stat/ivreg.html

- Now add another instrument, pk, and re-estimate the model.

```
. ivreg lnoutput lncapital (lnlabour =  pl pk), robust

Instrumental variables (2SLS) regression               Number of obs =      27
                                                       F(  2,    24) =  461.64
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9433
                                                       Root MSE      =  .18866

------------------------------------------------------------------------------
             |               Robust
    lnoutput |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnlabour |    .636999   .1778914     3.58   0.002     .2698492    1.004149
   lncapital |   .3551013   .1206419     2.94   0.007     .1061086    .6040941
       _cons |   1.128134   .2542946     4.44   0.000     .6032955    1.652972
------------------------------------------------------------------------------
Instrumented:  lnlabour
Instruments:   lncapital pl pk
------------------------------------------------------------------------------
```

- Test again the null of constant returns to scale and see if your conclusions have changed.

Further exercise:

- Econometricians will tell you that you can test if your regressors are endogenous or not. The Hausman test compares the least squares estimator and the IV estimator (see Greene, 2003, Section 5.5). This is also implemented in Stata: search for ‘ivendog’ and install the package – note that this is not possible in the computer rooms, you need your computer for this.[4]

- Be aware that the estimation of production functions is at the centre of an ongoing debate. IV estimation and panel data techniques have not produced satisfactory results in many applications. Depending on data availability and the context of the analysis, more recent techniques are dynamic panel data (so-called system GMM estimator) or structurally-derived estimates (see Olley and Pakes, 1996; Levinsohn and Petrin, 2003).

- You can perform IV estimation also on panel data using ‘xtivreg’.

[4] http://www.stata.com/search.cgi?query=ivendog