

# Quantitative Techniques – Term 2

## Laboratory 7

2 March 2006

### Overview

The objective of this lab is to:

- Estimate a cost function for a panel of firms;
- Calculate returns to scale;
- Introduce the command “cluster” to calculate robust standard errors in a panel.

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.

Go to [www.staff.city.ac.uk/a.banal-estanol/teaching.htm](http://www.staff.city.ac.uk/a.banal-estanol/teaching.htm), download Lab7.dta to your directory.

The source of the dataset is Greene (2003), Example 13.1.<sup>1</sup>

The file contains data on 6 airlines over 15 years on the following variables:

- c: total cost of production (in \$ 000s);
- q: output, measured as ‘revenue passenger miles’ – it is an index;
- pf: price of input (i.e. the price of fuel);<sup>2</sup>
- lf: the load factor is average capacity utilization of the fleet.

In addition, you have an indicator for the number of airlines (idairline) and the number of time periods (time). These are needed to identify the structure of the dataset.

Our sample is a balanced panel, which means that we have the same number of observations for each firm. The techniques for balanced panels can easily accommodate panels in which the group sizes differ (unbalanced panels).

### Task 1 – Estimation of a cost function for a panel of airlines

In this lab, we estimate a Cobb-Douglas cost function which we linearize by taking logarithms:

$$\ln C_{it} = \beta * \ln Q_{it} + \gamma * \ln PF_{it} + \delta * LF_{it} + u_i + \varepsilon_{it}$$

---

<sup>1</sup> Greene, W. (2003), *Econometric Analysis*, 5<sup>th</sup> edition, Prentice Hall.

<sup>2</sup> Note that the price of fuel differs across airlines due to different mixes of planes and regional variation in supply characteristics. See Greene (2003).

where  $i$  indicates the firm and  $t$  refers to the time period. Note that the intercept  $u_i$  is firm-specific in the above equation. Finally, the error terms  $\varepsilon_{it}$  are assumed to be homoscedastic and uncorrelated.

Before proceeding to the analysis you need to:

- Generate the logarithms of the variables, e.g.  
`gen lnc=ln(c)`  
and similarly for output and the price of fuel;
- Declare the dataset to be a panel. From the Menu bar, you can choose Statistics... Longitudinal / Panel Data Analysis... Declare dataset to be cross-sectional time-series. Alternatively, you can type in the Command window:

```
tsset idairline time
```

Note that this is the same command we used for time-series, but in this case we also have the cross-section dimension.

## 1. Pooled OLS

The first estimator we use on the cost function is OLS, which is equivalent to assume that all the firms have the same intercept, i.e.  $u_i = u$  for all firms. Note that this is a way of exploring the characteristics of the data, but does not exploit the panel structure of our sample.

```
. reg lnc lnq lnpf lf
```

Source	SS	df	MS	Number of obs = 90		
Model	112.705452	3	37.5684839	F( 3, 86)	=	2419.34
Residual	1.33544153	86	.01552839	Prob > F	=	0.0000
				R-squared	=	0.9883
				Adj R-squared	=	0.9879
Total	114.040893	89	1.28135835	Root MSE	=	.12461

  

	lnc	lnq	lnpf	lf	_cons	[95% Conf. Interval]	
Coef.	.8827385	.453977	-1.62751	9.516923			
Std. Err.	.0132545	.0203042	.345302	.2292445			
t	66.60	22.36	-4.71	41.51			
P> t	0.000	0.000	0.000	0.000			
	.8563895	.4136136	-2.313948	9.0612			
	.9090876	.4943404	-.9410727	9.972645			

As expected, the coefficients on output and on the price of input are positive. Moreover, capacity utilization is negatively related to costs.

## 2. Fixed effects (or least squares dummy variable)<sup>3</sup>

For this estimator, we go back to the initial formulation, where  $u_i$  varies for each

<sup>3</sup> Explanation of intercept in the fixed effect model + comparison with estimating OLS with dummies: <http://www.stata.com/support/faqs/stat/xtreg2.html>

individual in the sample. Intuitively, this formulation captures the fact that two observations for the same individual are likely to be “more similar” than observations from different individuals.

In addition, the fixed effects estimator takes account of the fact that  $u_i$  may be correlated with the regressors, i.e.  $\ln Q_{it}$ ,  $\ln PF_{it}$  and  $LF_{it}$ .

Stata has a command to perform panel-data estimation: `xtreg`.<sup>4</sup> The syntax is similar to the “`reg`” command we have been using for OLS. In order to choose the fixed effects option, we need to specify “, fe” after the list of variables:<sup>5</sup>

```
. xtreg lnc lnq lnpf lf, fe

Fixed-effects (within) regression              Number of obs   =          90
Group variable (i): idairline                 Number of groups =           6

R-sq:  within = 0.9926                          Obs per group:  min =          15
        between = 0.9856                          avg             =         15.0
        overall = 0.9873                          max             =          15

corr(u_i, Xb) = -0.3475                          F(3, 81)        =       3604.80
                                                Prob > F         =         0.0000

-----+-----
      lnc |          Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      lnq |   .9192846     .0298901    30.76  0.000     .8598126   .9787565
      lnpf |   .4174918     .0151991    27.47  0.000     .3872503   .4477333
       lf |  -1.070396     .20169      -5.31  0.000    -1.471696  -.6690963
   _cons |   9.713528     .229641    42.30  0.000     9.256614  10.17044
-----+-----
sigma_u |   .1320775
sigma_e |   .06010514
   rho  |   .82843653   (fraction of variance due to u_i)
-----+-----
F test that all u_i=0:   F(5, 81) =    57.73          Prob > F = 0.0000
```

The overall significance of the regression is good, as shown by the F-test on the right-hand side of the output above. This F test does not include the firm-specific effects, but only `lnq`, `lnpf` and `lf`. Individual coefficients are also significant and of the expected sign, broadly in line with the OLS estimates.

You can see the estimated fixed effects by typing:

```
predict fixed, u
```

This is the same command we have used for fitted values and residuals, the only difference is the “u” option instead of “residuals”. You can choose any name other than “fixed”.

Note that Stata automatically tests the null hypothesis that all the firm-specific intercepts are jointly zero (see last line in output above). In this case, we can reject the null hypothesis that the fixed effects are zero.

<sup>4</sup> <http://www.stata.com/help.cgi?xtreg>

<sup>5</sup> The estimation can also be carried out from the Menu bar, Statistics...Longitudinal/Panel data...Linear regression.

### 3. Random effects

The random effects estimator assumes that  $\alpha_i$  is uncorrelated with the regressors, i.e.  $\ln Q_{it}$ ,  $\ln PF_{it}$  and  $LF_{it}$ .

We use the same Stata command as before. If no option is specified, then the default estimator for xtreg is random effects.<sup>6</sup>

```
. xtreg lnc lnq lnpf lf

Random-effects GLS regression                Number of obs   =          90
Group variable (i): idairline              Number of groups =           6

R-sq:  within = 0.9925                      Obs per group:  min =          15
        between = 0.9856                      avg           =         15.0
        overall = 0.9876                      max           =          15

Random effects u_i ~ Gaussian              Wald chi2(3)    =    11091.33
corr(u_i, X) = 0 (assumed)                 Prob > chi2     =         0.0000
```

	lnc	lnq	lnpf	lf	_cons
Coef.	.9066805	.4227784	-1.064499	9.627909	
Std. Err.	.025625	.0140248	.2000703	.210164	
z	35.38	30.15	-5.32	45.81	
P> z	0.000	0.000	0.000	0.000	
[95% Conf. Interval]	.8564565 .9569045	.3952904 .4502665	-1.456629 -.672368	9.215995 10.03982	

```
sigma_u | .12488859
sigma_e | .06010514
rho     | .81193816 (fraction of variance due to u_i)
```

The results from the random effects model are similar to the estimates from the fixed effects model. The main difference is in the underlying assumptions. Note that zero correlation between the firm-specific intercepts and the regressors is assumed to be zero.

The option “predict *name\_of\_variable*, u” can also be used after the random effects estimator.

#### Summary of Task 1:

- Estimates from pooled OLS, fixed effects and random effects models have produced similar results for the Cobb-Douglas cost function;
- The assumption of uncorrelation between the firm-specific effect and the other independent variables is often too strong and fixed effects estimators are commonly used in studies on cost functions;
- The fixed effects estimator is more appropriate when the model applies only to the firms in the sample, but not for out-of-sample prediction. However, it is costly in terms of degrees of freedom;
- There are tests to check which assumptions are more appropriate, e.g. Hausman test – see Greene (2003), Section 13.4.4.

<sup>6</sup> Alternative syntax: `xtreg lnc lnq lnpf lnlf, re`

*Further exercise:*

- Try and create a do file with the commands we have used for this task. Create also a log file in text format (reminder: the file extension is “.log”, otherwise it can only be read in Stata);
- For those interested in cost functions, the translog cost function is often employed in academic papers focusing on regulated sectors. See Berndt (1996),<sup>7</sup> Chapter 9, or Greene (2003), section 14.3;
- Similarly to the fixed effects model, you can test whether the classical regression model is appropriate or the random effects estimator is preferable. This is the so-called Breusch and Pagan test for the random effects model. In this example, we can reject the null hypothesis of the classical regression model in favour of the random effects model. See Greene (2003), Section 13.4.3.<sup>8</sup>

## **Task 2 – Returns to scale**

A direct estimate of returns to scale can be obtained from the cost function as:

Returns to scale =  $(1/\beta)$

Note that, in a Cobb-Douglas cost function,  $\beta$  is constant and returns to scale do not vary with the level of output. If  $(1/\beta) = 1$ , there are constant returns to scale.

Use the estimates obtained above to establish whether the airlines in our sample exhibit constant returns to scale.

### **1. Pooled OLS**

From the pooled OLS estimates, the coefficient on output is 0.8827, which indicates increasing returns to scale  $\rightarrow (1/0.8827) > 1$

### **2. Fixed effects**

The coefficient on output is 0.9192846, and this indicates increasing returns to scale.

### **3. Random effects**

The coefficient on output is 0.9066805, which confirms the finding of increasing returns to scale.

*Further exercise:*

- If you want to test the null hypothesis of constant returns to scale, this is equivalent to test that  $\beta = 1$ . Type in the Command window: `test lnq=1`
- You can practice with two outputs. In this case, you need to replace  $\beta$  with the sum of the coefficients on both outputs;
- In a multi-product cost function, you can test the null hypothesis of constant returns to scale by an F-test. In Stata, type in the Command window: `test output1 +`

---

<sup>7</sup> Berndt, E. (1996), *The Practice of Econometrics*, 2<sup>nd</sup> edition, Addison-Wesley.

<sup>8</sup> After estimating a random effects model, type the command “`xttest0`”.

$output2 = 1$ , where *output1* and *output2* are the names of the two output variables in your dataset.

### Task 3 – “Cluster” option

We now revisit the panel data estimators with a further complication added. In a sample with a panel structure, residuals often are not only heteroscedastic, but also correlated within each individual/firm.<sup>9</sup>

In Stata, standard errors can be adjusted to take this into account using the “cluster” option. These are the standard errors usually reported in academic papers.

The syntax requires the specification of the “group” variable, e.g. the airline in our example.

```
. xtreg lnc lnq lnpf lf, fe cluster(idairline)
```

Fixed-effects (within) regression  
Group variable (i): idairline

Number of obs = 90  
Number of groups = 6

R-sq: within = 0.9926  
between = 0.9856  
overall = 0.9873

Obs per group: min = 15  
avg = 15.0  
max = 15

F(3,87) = 647.73  
Prob > F = 0.0000

corr(u\_i, Xb) = -0.3475

(Std. Err. adjusted for 6 clusters in idairline)

	lnc	lnq	lnpf	lf	_cons	sigma_u	sigma_e	rho
Coef.	.9192846	.4174918	-1.070396	.4286714	9.713528	.1320775	.06010514	
Robust Std. Err.	.0328726	.0193486	.4286714	.3188995				(fraction of variance due to u_i)
t	27.97	21.58	-2.50	30.46				
P> t	0.000	0.000	0.055	0.000				
[95% Conf. Interval]	.8347828 1.003786	.3677546 .4672289	-2.172331 .0315387	8.893771 10.53329				

Note that the coefficients are the same as the estimates we obtained before, while standard errors are larger. Remember from Term 1 that we do not know in advance whether standard errors in the presence of heteroscedasticity and/or serial correlation are larger or smaller.

In the output above, the significance of the coefficient on load factor is reduced once we use robust standard errors.

This command is useful not only with panel data, but also if you have groups and sub-groups, e.g. the observations for firms within the same area may not be independent.

<sup>9</sup> Note that the option “robust” (i.e. without cluster) takes into account heteroscedasticity but not correlation of residuals across observations.