MARCO ANGRISANI
Practical Tutorial, G023 Econometric Theory and Methods

In this handout you can find STATA commands to solve question 3 in exercise 2, the output given by the program and brief comments to the main results.

3.2a

To import the dataset of Table1 in STATA, open the program, click on the *Data Editor* button, write Table1 and save the dataset. The result is a dataset in STATA format (extension .dta).

Remark1: if you write the data in a spreadsheet (e.g. Excel), you can use the command *insheet* to read the data in STATA and *save* to save the dataset in STATA format.

Remark2: you can use STATA help to learn the syntax of the commands.

Opening the data
Click on the *Open (use)* button and select the file you want to open. Once the data are loaded, you can check that the organization of the dataset is the desired one by clicking on the *Data Browser* button, or simply typing the command *browse*.

Estimation of the linear model:

$$C = \alpha + \beta X + \varepsilon \qquad\qquad (1)$$

with $E\left[\varepsilon \middle| X = x\right] = 0$ and $Var\left[\varepsilon \middle| X = x\right] = \sigma^2$


```
. reg C X
```

Source	SS	df	MS
Model	791.853592	1	791.853592
Residual	1057.34998	12	88.1124982
Total	1849.20357	13	142.246428

Number of obs =	14
F(1, 12) =	8.99
Prob > F =	0.0111
R-squared =	0.4282
Adj R-squared =	0.3806
Root MSE =	9.3868

C	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
X	.012894	.0043011	3.00	0.011	.0035226	.0222653
_cons	42.52377	8.325339	5.11	0.000	24.38441	60.66312

Comments

- α can be interpreted as a fixed cost.

- The confidence interval provided by STATA is valid under the assumption that the disturbance term is normally distributed. We did not make any assumption about the distribution of ε, therefore we use approximate inference. In order to construct a 95% confidence interval for α we use:

$$\Pr\left[-1.96 \leq \frac{42.52377 - \alpha}{8.325339} \geq 1.96\right] \approx 0.95$$

  from which we get the 95% confidence interval $[26.206, 58.841]$.

- To test $H_0 : \alpha = 25$, we use the statistic $T = (\hat{\alpha} - \alpha)/Std.Err.$
  We obtain $T = 2.10 > 1.96$, therefore we reject the null.

3.2b

In order to exclude Eastern and Southern companies (efficient companies), we create a binary variable *d* taking value 1 for the efficient companies and 0 otherwise. Then we can run a regression under the condition that d=0. In this case, STATA will run a regression using all the observations except the ones corresponding to the efficient companies.

```
.  gen d=(PES==1 | PES ==9)

.  reg C X if d==0
```

```
      Source |       SS       df       MS                  Number of obs =      12
-------------+------------------------------              F(  1,    10) =   34.85
       Model |  1422.43908      1  1422.43908             Prob > F      =  0.0002
    Residual |  408.120926     10  40.8120926             R-squared     =  0.7771
-------------+------------------------------              Adj R-squared =  0.7548
       Total |  1830.56001     11  166.414546             Root MSE      =  6.3884


------------------------------------------------------------------------------
           C |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |   .0238541   .0040406     5.90   0.000     .0148512    .0328571
       _cons |   26.08391   7.057263     3.70   0.004     10.35935    41.80847
------------------------------------------------------------------------------
```

Comments

- In order to construct a 95% confidence interval for α we use again:

$$\Pr\left[-1.96 \leq \frac{26.084-\alpha}{7.057} \geq 1.96\right] \approx 0.95$$

from which we get the 95% confidence interval $[12.24, 39.92]$.

- To test $H_0 : \alpha = 25$, we use the statistic $T = (\hat{\alpha} - \alpha)/Std.Err.$
  We obtain $T = 0.15 < 1.96$, therefore we do not reject the null.

3.2c

Estimation of the model:

$$C = \alpha + \beta\, X + \gamma\, dX = \varepsilon \tag{2}$$

$$where\, d = \begin{cases} 1 & if\ PES = efficient \\ 0 & otherwise \end{cases}$$

```
. gen dX=d*X

. reg C X dX
```

Source	SS	df	MS		Number of obs =	14
					F(2, 11) =	19.39
Model	1440.63924	2	720.319621		Prob > F =	0.0002
Residual	408.564328	11	37.1422116		R-squared =	0.7791
					Adj R-squared =	0.7389
Total	1849.20357	13	142.246428		Root MSE =	6.0944

C	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
X	.0237994	.0038219	6.23	0.000	.0153875	.0322112
dX	-.0094533	.0022619	-4.18	0.002	-.0144316	-.004475
_cons	26.18303	6.671092	3.92	0.002	11.50005	40.866

```
. vce
```

	X	dX	_cons
X	.000015		
dX	-5.9e-06	5.1e-06	
_cons	-.024595	.008843	44.5035

Comments

- In order to construct a 95% confidence interval for α we use again:

$$\Pr\left[-1.96 \leq \frac{26.18303 - \alpha}{6.671092} \geq 1.96\right] \approx 0.95$$

  from which we get the 95% confidence interval $[13.11, 39.25]$.

- The slope of the efficient frontier is $b = \beta + \gamma$, therefore the estimated slope of the efficient frontier is $\hat{b} = \hat{\beta} + \hat{\gamma} = 0.0143$.

- We can obtain the standard error for this estimate from the variance/covariance matrix (*vce*) above. We can use the fact that the variance of $\hat{b}$ is equal to the variance of $\hat{\beta}$ plus the variance of $\hat{\gamma}$ plus 2 times the covariance between $\hat{\beta}$ and $\hat{\gamma}$, to get an estimated standard error of 0.00288 and then the 95% confidence interval $[0.0088, 0.0197]$.

3.2d

Estimation of the linear model:

$$\log C = \theta + \phi \log X + \upsilon \qquad \qquad \textbf{(3)}$$

with $\quad E[\upsilon|X = x] = 0 \quad$ *and* $\quad Var[\upsilon|X = x] = \lambda^2$

Generate the two new variables expressed in logs:

```
. gen logX=log(X)

. gen logC=log(C)


. reg logC logX
```

Source	SS	df	MS		Number of obs =	14
					F(1, 12) =	16.09
Model	.240644792	1	.240644792		Prob > F =	0.0017
Residual	.179455193	12	.014954599		R-squared =	0.5728
					Adj R-squared =	0.5372
Total	.420099985	13	.032315383		Root MSE =	.12229

logC	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
logX	.3815141	.0951064	4.01	0.002	.174295	.5887331
_cons	1.331399	.7107562	1.87	0.086	-.2172061	2.880003

Comments

- In this case is zero as long as $\phi > 0$.

- From equation (3), solve for C and take the derivative of the expression you get with respect to X; you obtain:

$$\frac{\partial C}{\partial X} = \phi X^{\phi-1} e^{\theta+v}$$

  If $\phi < 1$, the marginal cost computed above decreases with the number of customers served: economies of scale.

- In our case the estimated coefficient is $\hat{\phi} = 0.3815 < 1$ which suggests that there are actually economies of scale.

- In order to obtain a more precise conclusion about the presence of economies of scale we test the hypothesis $H_0 : \phi = 1$ *against* $H_1 : \phi < 1$.
  This is a "one-sided" test, therefore we reject the null if

$$T = \frac{\hat{\phi} - \phi}{Std.Err.} \leq z_L(\alpha),$$

  where $z_L(\alpha)$ is the $1 - \alpha$ quantile of a standard normal.

  If the size of the test is 5% ($\alpha = 0.95$), then $z_{1-\alpha} = -1.64$ and our test becomes:

$$T = \frac{\phi - \hat{\phi}}{Std.Err.} \geq 1.64 \Rightarrow \frac{1 - 0.38}{0.95} = 6.50 \geq 1.64$$

- We reject the null, therefore we have evidence that economies of scale exist.