

Quantitative Techniques – Term 2

Laboratory 5

16 February 2006

Overview

The objectives of this lab are to:

- Carry out price-concentration analysis;
- Learn how to generate dummy variables;
- Obtain correct standard errors in the presence of serial correlation or heteroscedasticity;
- Interpret coefficients on dummy variables.

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 and download Lab5.dta to your directory, then open it in Stata.

The file contains (weekly) time series data on:

- price1: price in geographic market 1;
- price2: price in geographic market 2;
- hhi1: HHI in geographic market 1;
- hhi2: HHI in geographic market 2;
- time: counts the number of weeks in the sample;
- period: divides the sample into 5 periods taking the following values
 - 1: for $t \leq 20$
 - 2: for $21 \leq t \leq 40$
 - 3: for $41 \leq t \leq 60$
 - 4: for $61 \leq t \leq 80$
 - 5: for $81 \leq t \leq 100$
- inventratio1: ratio of inventory to predicted demand in market 1;
- inventratio2: ratio of inventory to predicted demand in market 2.

Note that, as in previous labs, we analyze the same product in different geographic areas. We carry out a similar task as in the GAO report.

Task 1 – Generating a dummy variable

Depending on the variables currently in your dataset, you have two main cases:

1. Assume that we want to include in our analysis a dummy variable. For instance, this could refer to the liberalization of the market for the product/service we are analyzing. The dummy variable takes value 1 from time t=70 onwards. We call this dummy 'lib'.

- We want 'lib' to be equal to 1 from t=70 onwards:

```
gen lib=1 if time>69
```

You can obtain the same result by: `gen lib=1 if time<=70`

- Stata has created the variable, but has inserted missing values, indicated by a dot (.), for t < 70. We want 'lib' to have value 0 in this interval:

```
replace lib=0 if lib==.
```

Note that the syntax of the 'if' condition: two = instead of one.¹

2. Alternatively, if you already have a variable which only takes a given number of values, Stata can automatically create a dummy variable for each value by using 'tabulate'. Suppose that we wanted to use the categorical variable 'period' to create 5 dummies, one for each of its values (see description above).

- Type 'tabulate' followed by the name of the variable.² Then, after the comma, type generate (or gen) followed by the name you want to give to the dummies.

Note that you can use the name of the categorical variable (e.g. period) and Stata will automatically create 5 dummies, period1 to period5.

```
. tabulate period, gen(period)
```

Period	Freq.	Percent	Cum.
1	20	20.00	20.00
2	20	20.00	40.00
3	20	20.00	60.00
4	20	20.00	80.00
5	20	20.00	100.00
Total	100	100.00	

- (Remember not to use all the dummies in a regression, as you would have problems of perfect multicollinearity – see Ashenfelter et al. Section 12.2).

Task 2 – Price Concentration Analysis

We follow a similar structure as the GOA report:

$$\text{Price1} = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{time} + \beta_4 * \text{Inventratio1} + \beta_5 * \text{Lib} + \text{error}$$

¹ <http://www.stata.com/help.cgi?if>

² <http://www.ats.ucla.edu/stat/stata/faq/dummy.htm>

We estimate the above model by OLS.

```
. reg price1 hhil time inventratio1 lib
```

Source	SS	df	MS	Number of obs = 100		
Model	27.1438436	4	6.7859609	F(4, 95)	=	13.11
Residual	49.1778217	95	.517661281	Prob > F	=	0.0000
				R-squared	=	0.3557
				Adj R-squared	=	0.3285
				Root MSE	=	.71949

price1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hhil	.0006129	.0002162	2.83	0.006	.0001836	.0010422
time	-.0214874	.0042378	-5.07	0.000	-.0299005	-.0130743
inventratio1	.1725362	.2707454	0.64	0.525	-.3649614	.7100337
lib	.4850164	.2629582	1.84	0.068	-.0370216	1.007054
_cons	9.819478	.3905742	25.14	0.000	9.04409	10.59487

Note that the coefficient on the HHI in market 1 is significant at the 5% level and positive. Therefore, on average, a higher HHI would result in higher prices in this market.

On the other hand, prices seem to decline over time, while liberalization has the effect of increasing prices (significant at 10%).

Note that the inventories ratio does not have any significant effect on price in market 1, so we can try to estimate the same model removing the variable 'inventratio1'.

```
. reg price1 hhil time lib
```

Source	SS	df	MS	Number of obs = 100		
Model	26.9336188	3	8.97787295	F(3, 96)	=	17.45
Residual	49.3880465	96	.514458817	Prob > F	=	0.0000
				R-squared	=	0.3529
				Adj R-squared	=	0.3327
				Root MSE	=	.71726

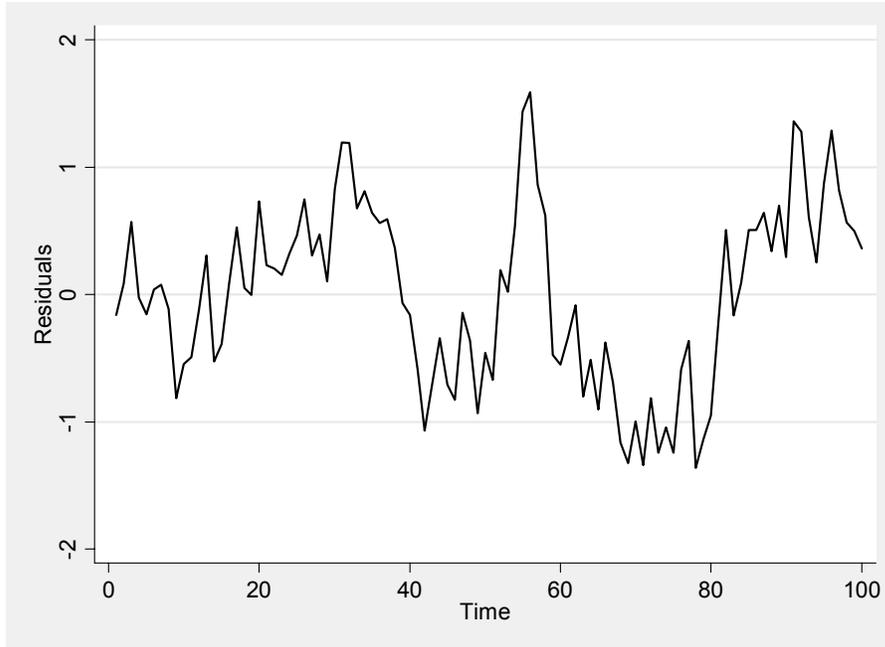
price1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hhil	.0005843	.0002109	2.77	0.007	.0001657	.0010028
time	-.021957	.0041603	-5.28	0.000	-.0302152	-.0136989
lib	.5094629	.259339	1.96	0.052	-.0053209	1.024247
_cons	9.956824	.3247163	30.66	0.000	9.312268	10.60138

Task 3 – Correction of standard errors

As we saw last week, serially correlated errors often occur with time series data. Plot the errors from the regression above and notice that they do not seem random.

Reminder from previous labs:

- How to obtain fitted residuals – predict *name of variable*, residuals
- How to plot the residuals - line *name of your variable* time
You can save the graph by typing - graph save "*YourDirectory\NameOfFile.gph*"



We now use the option ‘robust’³ in conjunction with the usual OLS regression. This calculates standard errors adjusted for potential heteroscedasticity and serial correlation. Note that the coefficients are the same as what we got in Task 2. Reminder from first term: the OLS estimates are unbiased also if errors are heteroscedastic or serially correlated (Ashenfelter et al., Chapter 14).

```
. reg price1 hhi1 time lib, robust
```

```
Regression with robust standard errors
```

Number of obs =	100
F(3, 96) =	24.75
Prob > F =	0.0000
R-squared =	0.3529
Root MSE =	.71726

price1	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
hhi1	.0005843	.0001944	3.01	0.003	.0001985 .0009701
time	-.021957	.0040023	-5.49	0.000	-.0299016 -.0140125
lib	.5094629	.2981717	1.71	0.091	-.0824033 1.101329
_cons	9.956824	.2883269	34.53	0.000	9.3845 10.52915

Note that ‘robust’ standard errors can be greater or lower than the incorrect ones obtained in the simple regression. This may result in lower t-statistics for the coefficients and even make those coefficients insignificant.

³ <http://www.stata.com/help.cgi?robust>

Task 4 – An Alternative Specification

In this task, we use the dummies created in Task 1 to subdivide the period under consideration into 5 sub-periods. Instead of including ‘time’ as an explanatory variable, we can include these dummies. Unlike the time trend, the dummies do not constrain the effect of time on price1 to be the same in each period.

In the equation below, we include only 4 out of 5 dummies to avoid perfect multicollinearity.

$$\text{Price1} = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{lib} + \beta_4 * \text{period2} + \beta_5 * \text{period3} + \beta_6 * \text{period4} + \beta_7 * \text{period5} + \text{error}$$

```
. reg price1 hhi1 lib period2 period3 period4 period5
```

Source	SS	df	MS	Number of obs = 100		
Model	55.3541793	6	9.22569655	F(6, 93)	=	40.92
Residual	20.967486	93	.225456839	Prob > F	=	0.0000
-----				R-squared	=	0.7253
-----				Adj R-squared	=	0.7076
Total	76.3216653	99	.770925912	Root MSE	=	.47482

price1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hhi1	.0005941	.0001417	4.19	0.000	.0003127	.0008755
lib	-.0078524	.2135862	-0.04	0.971	-.4319922	.4162874
period2	.0849578	.1502245	0.57	0.573	-.2133583	.383274
period3	-.9480087	.1511996	-6.27	0.000	-1.248261	-.6477564
period4	-1.854567	.1906136	-9.73	0.000	-2.233088	-1.476047
period5	-.6423179	.2617683	-2.45	0.016	-1.162138	-.122498
_cons	9.667907	.2125559	45.48	0.000	9.245813	10.09

For instance, for period 3, the additional impact of ‘period’ on price1 is given on average by -0.948, keeping the other variables constant. The reference point is period one (whose dummy we omitted). (In this example, the coefficient in period2 is not significant)

$$\text{Period 1: } E[\text{price1} | \text{HHI1}, \text{Lib}, \text{period1}] = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{lib}$$

$$\text{Period 2: } E[\text{price1} | \text{HHI1}, \text{Lib}, \text{period2}] = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{lib} + \beta_4$$

$$\text{Period 3: } E[\text{price1} | \text{HHI1}, \text{Lib}, \text{period3}] = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{lib} + \beta_5$$

$$\text{Period 4: } E[\text{price1} | \text{HHI1}, \text{Lib}, \text{period4}] = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{lib} + \beta_6$$

$$\text{Period 5: } E[\text{price1} | \text{HHI1}, \text{Lib}, \text{period5}] = \beta_1 + \beta_2 * \text{HHI1} + \beta_3 * \text{lib} + \beta_7$$

As the betas above are different from one another, we can capture particular events in a sub-period. Note that the betas may also have different signs.

On the basis of the results above, would a merger in market 1 have a likely impact on P1?

Repeat the above analysis (Tasks 2-3) for market 2. In a dominance case relating to market 2, what would you conclude about the link between market power and prices?

As an exercise, you can create a 'do-file' with the commands we have used in this lab. Include also a log file. (You can check your do-file with Lab5.do – which you can download from www.staff.city.ac.uk/a.banal-estanol/teaching.htm).