Econometrics I : Sample Questions

Susan Thomas

Fri Dec 5 10:34:29 IST 2008

- There is no use for long or essay-type answers. Keep your answers crisp and to the point. Perfect answers to all the questions can be done using very few words (in many cases, perfect answers are shorter than the questions).
- Use diagrams and example calculations to illustrate your arguments.
- Wisdom is valued over exactitude.
- Approximations are permitted. State your assumptions and proceed.
- Keep your work neat and make your answer paper attractive. Organise all subparts of a question neatly. Start each new question on a new page.

1. A multiple regression of y on a constant, x_1 and x_2 produces the following results:

$$\hat{y} = 4 + 0.4x_1 + 0.9x_2 \qquad R^2 = 8/60, e'e = 520, n = 29$$

$$X'X = \begin{bmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{bmatrix}$$

(a) Test the hypothesis that the two slopes sum to 1.

(b) Test the hypothesis that the slope on x_1 is zero. Calculate the results of running the restricted regression, and compare the two sums of squared deviations.

2. In Malaysia, there are three races: Chinese, Malay and Indian. We are interested in studying "earnings functions", or models of the form $\log e = \beta_0 + \beta_1 \text{exp} + \beta_2 \text{exp}^2 + \epsilon$ where e is earnings per year, and `exp` stands for years of experience.

   (a) Sketch the picture this model aims to convey.

   (b) We suspect that race matters in the earnings function. Show two different methods through which you could estimate models where race-heterogeneity enters into the model. In each method, show how you could test the null hypothesis that there is no heterogeneity.

   (c) Suppose, in addition, we feel that there may be an education effect.

   We create two dummy variables, `SCHOOL` which is true for people who have got past secondary school, and `COLLEGE` which is true for people who have got a college education. How would you modify the specification to include education into the model?

   (d) Does smoking hurt human capital?

      i. What are all the different ways in which smoking could impact on human capital?

      ii. Suppose you are given an "earnings function" dataset (earnings, experience) coupled with one dummy variable `SMOKER` which is 1 for people who smoke. How would you test for the (different) effects of smoking?

3. In each of the model specifications below, everything other than x and y is a parameter to be estimated.

   Show how you would do estimation using OLS, *if* that is possible.

   Say it is impossible if you think it cannot be done. Do not worry about distributional assumptions about the error term.

   (a) $y = a_0 \log(a_1 \, x) + a_2$

   (b) $e^{\frac{y}{\log y}} = a + b \log x$

   (c) $a_0 + y^{a_1} = b_0 + b_1 \, x$

   (d) $y = a_0 + a_1 \sin(x) + a_2 \sin(2x) + a_3 \sin(3x)$

4. All functions have taylor expansions, i.e. all functions can be approximated by polynomials. Hence any nonlinear relationship can be estimated using OLS by estimating a tenth or twentieth degree polynomial. Comment.

5. Suppose $x \sim U(a, b)$. For a sample size of $T = 25$ we wish to estimate the mean of the distribution of $1/x^2$. Student A suggests using the formula $(1/\bar{x})^2$ and student B suggests using $\sum(1/x^2)/25$. Show a Monte Carlo setup to examine the two alternative estimators.

6. Random variables x and y have variances 4 and 16. You need to estimate the difference between their means but can only afford to take 30 observations. How many should you draw of x and how many of y?

7. A MonteCarlo question: suppose you have a program that does the following:

- Draw 20 x values from a uniform distribution between 2 and 8.
- Draw 20 z values from a standard normal distribution.
- Calculate 20 w values using $5 + 2x + 9z$.
- Draw 20 ϵ values from a standard normal distribution.
- Calculate 20 y values using $1 + 4x + 3\epsilon$.
- Regress y on x and save the R^2 value, as $R1$, and the adjusted R^2 values, as $R1_{adj}$.
- Regress y on w and save the R^2 value, as $R2$, and the adjusted R^2 values, as $R2_{adj}$.
- Repeat the above steps 3000 times, to get 3000 values of $R1$, $R1_{adj}$, $R2$ and $R2_{adj}$.
- Calculate the averages of all the four sets of 3000 numbers to get $\hat{R}1$, $\hat{R}1_{adj}$, $\hat{R}2$ and $\hat{R}2_{adj}$.

(a) What should be the relative magnitudes of $\hat{R}1$ and $\hat{R}2$? Explain why you think so.

(b) What should be the relative magnitudes of $\hat{R}1adj$ and $\hat{R}2_{adj}$? Explain why you think so.

(c) Would your answer to part 1) be different if the regressions had been run using 2000 draws instead of just 20 draws each of the variables?

8. Suppose the OLS model holds for $y = \alpha + \beta x + \epsilon$ except that you think that 22 observations has a variance which is different from the next 32 observations. For the first 22 observations, the data (expressed as deviations from their means) have the following numbers: $\sum xy = 100, \sum x^2 = 10, \sum y^2 = 1040$.

   For the remaining data, the numbers are $\sum xy = 216, \sum x^2 = 16, \sum y^2 = 3156$.

   (a) Perform a statistical test to test whether the error variances are the same in both sets of data at the 5% significance level.

   (b) Assuming the error variances differ between the two sets, what is β_{gls}?

   (c) What estimate of the variance of β_{ols} would you use if you believed that the error variances differ between the two sets?

9. Two samples of 50 observations each produce the following moment matrices (in each case, X has a constant and one variable.

	Sample 1	Sample 2
$X'X$	$\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}$	$\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}$
$y'X$	$[300 \quad 2000]$	$[300 \quad 2000]$
$y'y$	2100	2500

(a) Test the hypothesis that the same regression coefficients aply in both data sets assuming the disturbance variances are the same.

(b) Test the hypothesis that the two coefficient vectors are the same *without* assuming that the disturbance variances are the same.