Principles of Econometrics, Part 1
The Finnish Doctoral Programme in Economics (FDPE)

Elias Einiö
University of Helsinki
elias.einio(at)helsinki.fi

Problem Set 1

Sep 24, 2009

Solutions must be handed in before the demonstrations
(or sent by e-mail to elias.einio(at)helsinki.fi).

1. Suppose we estimate the model $y_i = x_i\beta + u_i$ (the simple regression model without a constant term, i.e. only one explanatory variable).

a) Explain on what criterion the ordinary least squares estimator of β is based on.

b) Construct the objective function based on the criterion.

c) Derive the ordinary least squares estimator satisfying the criterion.

2. Suppose we estimate the model $y_i = X_i'\beta + u_i$, where X_i is an $k \times 1$ vector of explanatory variables and β is a $k \times 1$ vector of parameters.

a) For n observations the model may be written in matrix notation as $y = X\beta + u$ where y is a $n \times 1$, vector X is a $n \times k$ matrix, and u is a $n \times 1$ vector. Explain how each term in this equation is constructed.

b) Derive the OLS estimator of β.

c) Show that the projection matrix $P = X(X'X)^{-1}X'$ and the annihilator matrix $M = I - P$ are symmetric and idempotent.

d) Show that (i) $\hat{y} = Py$ where $\hat{y}$ is the vector of fitted values (i.e. $\hat{y} = Xb$, where b is the OLS estimate of β).

e) Suppose that $E[\epsilon\epsilon' \mid X] = \sigma^2 I$, where σ^2 is a scalar. Show that $E[(b - \beta)e' \mid X] = 0$, where e is the vector of residuals (i.e. $e = y - Xb$).

(Tip: Showing that $e = M\epsilon$ might be useful.)

3. (Partitioned regression) Suppose we estimate the model

$$y = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u = X_1\beta_1 + X_2\beta_2 + u$$

where X_1 is a $n \times k_1$ and X_2 a $n \times k_2$ matrix of explanatory variables, and β_1 and β_2 are the corresponding parameter vectors.

a) Show that if the columns of X_1 are orthogonal to the columns of X_2, i.e. $X_1'X_2 = 0$, the OLS coefficients of the regression of y on $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ can be obtained from the separate regressions of y on X_1 and y on X_2.

b) Suppose that $X_1'X_2 \neq 0$ but we erroneously exclude X_2 from the regression (i.e. we regress y only on X_1). Derive the OLS estimator of β_1 from this regression. Show that it is unbiased only if $\beta_2 = 0$.

4. (You will need software to do this exercise. You may use any computer program that you wish.) Download the data file from

http://www.valt.helsinki.fi/staff/eeinio/Etrics09/farms.csv
(A csv file with ";" as a separator and "," as a decimal pointer.)

or

http://www.valt.helsinki.fi/staff/eeinio/Etrics09/farms.R
(An R data file.)

The data file contains observations on 247 dairy farms each observed in each of 6 years. The file contains variables (columns) YEAR, MILK, LABOR, COWS, LAND, FEED signifying the obvious (there are also some other variables in the data file but you may ignore them for the moment).

a) Produce a simple scatter (X-Y) plot with LABOR on the horizontal axis and MILK on the vertical axis. What conclusion do you draw about the relationship between LABOR (input) and MILK (output)?

b) Note that LABOR only takes a few values, 1, 1.2, 2, 2.5, and so on. Compute the mean value of MILK for the different values of LABOR. What do you conclude about the conditional mean?
(R tip: Use the function `tapply()`. Just remember to transform LABOR into a factor by `factor()` in order to use it as a value for the parameter `INDEX`. In order to get rid of marginal differences in the LABOR variable use `round()`. An alternative way is to use the function `which()` to select elements in MILK corresponding to a specific value of LABOR. E.g. `milk_1=data$MILK[which(data$LABOR==1)]`.)

c) Estimate a regression model where MILK is explained by LABOR. What happens when you include COWS in the model? Explain why? Also estimate a model including

all variables and interpret the results.
(R tip: You may want to use year dummies which are implemented by `factor()`)