Econometrics A: Problemset 3

LAURA MAYORAL
Institute for Economic Analysis and Barcelona GSE

Winter 2021

Deadline: February 4th, before 15:00. Please submit your answers in paper or electronically to your TA, sanghyun.park@insead.edu.

1. Let y be a random scalar, and let $\mathbf{x}$ be a $1\times$k random vector, such that $\mathbf{x}_1 = 1$. Consider the population model:

$$E(y|\mathbf{x}) = \mathbf{x}\beta$$
$$Var(y|\mathbf{x}) = \sigma^2$$

a) Explain in your own words the conditional independence assumption and why it permits to provide a causal interpretation to your regression results. If the model above verifies the CIA, explain what this entails with respect to the matrix of regressors $\mathbf{x}$ above.

b) Assume that you have some data and you have obtained the OLS estimator of β, $\hat{\beta_N}$. Explain the meaning of consistency, clearly state the conditions that you need for the OLS estimator to be consistent and sketch the proof of consistency.

c) Without introducing additional restrictions, could you compute the exact distribution of $\hat{\beta_N}$ in finite samples? Explain how you could construct an approximation to that distribution if N, the sample size, is large enough. Derive the asymptotic distribution of $\hat{\beta_N}$ under the assumptions above (Hint: notice the assumption about the conditional variance of y and what this implies for the conditional variance of u.) Using the analogy principle, provide an estimator for the variance-covariance matrix of $\hat{\beta_N}$.

d) Explain the difference between the standard error that you've derived in c) and the heteroscedasticity-robust standard errors. If the condition $Var(y|\mathbf{x}) = \sigma^2$ doesn't hold, explain what's the correct standard error associated to $\hat{\beta_N}$ and why.

e) Explain the conditions under which the OLS estimator is BLUE and also explain the meaning of this term.

2. You've estimated the following model by OLS: $y = 1.5 + 2x_1 + 3.5x_2 + \hat{u}$

a. What's the mean of the estimated residuals $\hat{u}$?

b. By construction, the OLS estimator is computed in a such a way that $(x_1\ x_2)'\hat{u} = 0$. Does this imply that $x'u = 0$? (where u is the population error term)

c. Assume that you've obtained an estimate for heroscedasticity-robust variance-covariance matrix of $\hat{\beta}$ given by the matrix below. Explain the meaning of the elements of the main diagonal and the elements off the main the diagonal.

(1)
$$\begin{pmatrix} 0.2 & -0.1 & 0.3 \\ -0.1 & 0.3 & 0.2 \\ 0.3 & 0.2 & 0.1 \end{pmatrix}$$

d) Explain the difference between point estimation and interval estimation. Using the data above, provide a confidence interval at the level of confidence of 95% for β_1 and β_2. Provide an interpretation. Are these estimates statistically significant?

e) Compute the p-value associated to β_2 and interpret its meaning.

Computer Practise

3. You would like to estimate the effect of an additional year of education on wages. Use the dataset NLS80.dta, which gathers data collected from a population of men.

Note: there is plenty of information on the web that can be of help. For instance, this file can be of help: https://www.princeton.edu/~otorres/Regression101.pdf

A) Using the stata command *estout* produce a regression table (in word or latex or your favourite word processor) whose columns introduce regressors sequentially. The table should contain the estimated coefficients, their associated standard errors (or p-values, as you prefer), "stars" indicating the level of significance (*, ** or *** if significant at 10%, 5% or 1%, respectively). At the bottom of the table, include the value of the R^2 statistic as well as the number of observations. Use robust standard errors. The regressors you should include are as follows:

column 1: lwage on education

column 2: lwage on education, IQ, and Kww

column 3: lwage on education, IQ, and Kww, age

column 4: lwage on education, IQ, and Kww, age, tenure, exper

column 5: lwage on education, IQ, and Kww, age, tenure, exper, hours

column 6: lwage on education, IQ, and Kww, age, tenure, exper, hours, married

column 6: lwage on education, IQ, and Kww, age, tenure, exper, hours, married

column 7: lwage on education, IQ, and Kww, age, tenure, exper, hours, married, black, south, urban

Provide a text that describes the table.

B) Using the results in column 7:

i) provide an interpretation of the direction of the relationship of educ and wages, as well as of the magnitude of the effect.

ii) Provide an interpretation for the R^2 statistic and for the F-test of joint statistical significance. Do you reject the null hypothesis of that test?

iii) Test the hypothesis that black, south, urban are all equal to zero using an F-test.

iv) test the hypothesis that the coefficient of education is 0.5 using a two-tailed t-test, $\alpha = 0.05$. Repeat the same exercise using where the alternative hypothesis is $H_1 : \beta > 0.5$

v) Use the command predict to obtain the predicted values for lwage. Also obtain the residuals.

vi) Use the command estat hettest to test for heteroskedasticity in the residuals. What do you conclude from this test?

C) Introduce conditions in your regression: compute again the regression in column 7 in two different cases: for married men and for unmarried men (to do this, use the command if at the end of your regression, before the ",r" option). What do you observe?

D) Use the binscatter command to produce a plot of lwage versus education, controlling by all the remaining variables.

i) Plot the scatter plot of lwage versus education. Also compute the binscatter of the same variables. Finally, plot a new binscatter, controlling this time for all the regressors in column 7.

ii) Use the commands pwcorr and graphic matrix to compute the correlation matrix of the regressors in column 7.