Problem Set 2

Chris Conlon

Fall 2025

Econometrics I   Professor Chris Conlon
NYU Stern   Email: cconlon@stern.nyu.edu

1 Prediction errors

The Cornwell and Rupert data for this problem can be downloaded from `http://www.github.com/chrisconlon/applied_metrics`.

Source: Cornwell and Rupert, (1988) "Efficient estimation with panel data: an empirical comparison of instrumental variables estimators"

(a) For this part of the assignment, you are to replicate the regression "Mincerian Regression, Cornwell and Rupert Data" from the Linear Regression slides by obtaining the same coefficients and standard errors. Now that you have replicated the regression, we'll consider a couple of minor extensions.

(b) Functional Form. The example thus far computes a single, generic effect of education on LWAGE. We're interested in determining if there is a different effect for men (FEM=0) and women (FEM=1). One compact way to do this is to add an interaction term, FEM*ED to the model. The different effects are the coefficient on ED which is for men and the sum of the two effects, ED and FEM*Ed, for women. Reestimate your model with this additional effect, and report your result. (Do this both for the continuous measure "years of schooling" and the "levels of schooling completed" dummies).

(c) Standard Errors. Can you compute the bias-corrected bootstrap confidenfce interval for the difference in log wages between men and women college graduates (controlling for all other variables). Thinking about the right regression to run and the correct $g(\cdot)$ function should be what is tricky here.

5 OLS Residuals

(a) Consider the following table of data nad potential residuals:

y	x_1	x_2	e_1	e_2	e_3
?	1	0	1	2	3
?	1	-1	-3	-1	-2
?	1	1	2	-1	1

Which of the potential vectors of residuals e_1, e_2, and e_3 (if any) could be from a regression of y on $\begin{bmatrix} x_1 & x_2 \end{bmatrix}$? Explain. (b) Show that the estimated OLS parameters are unchanged if the dependent variable and all dependent variables are transformed by subtracting their means.

That is, let

$$\begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} = \left(\boldsymbol{X'X} \right)^{-1} \boldsymbol{X'y},$$

where β_0 is the intercept, $\boldsymbol{\beta}$ is a column of the other parameter estimates, and the first column of $\boldsymbol{X}$ is a vector of 1 's as usual. Show that $\boldsymbol{\beta} = \left(\widetilde{\boldsymbol{X}}' \widetilde{\boldsymbol{X}} \right)^{-1} \widetilde{\boldsymbol{X}}' \widetilde{\boldsymbol{y}}$, where $\widetilde{\boldsymbol{X}}$ drops the first column of $\boldsymbol{X}$ and for the other columns, $\widetilde{\boldsymbol{x}}_k = \boldsymbol{x}_k - n^{-1} \sum_{i=1}^{n} x_{ki}$. Similarly, $\widetilde{\boldsymbol{y}} = \boldsymbol{y} - n^{-1} \sum_{i=1}^{n} y_i$.