

CERGE-EI  
Summer 2014  
Econometrics II  
Instructors: Nikolas Mittag, Dragana Stanišić  
TA: Jelena Plazonja, Gega Todua

## Homework #1

1. (5 points) Advanced Wooldridge 2.1
2. (15 points) Advanced Wooldridge 4.8
3. (10 points) Use the dataset NLS80.dta as in the previous question.
  - (a) Estimate a linear regression model for log wages on education, experience, and experience squared. Report the estimates and standard errors. Also report the full variance/covariance matrix for all coefficients.
  - (b) Predict the exact effect on average log earnings of increasing everybody's education level by 2 years. Assume that work experience is given by the number of years since leaving the school.
4. (20 points) Consider the "true" model of earnings to be given by equation 1:

$$\ln(wage) = \beta_0 + \beta_1 exper + \beta_2 urban + \beta_3 black + \beta_4 educ + \alpha IQ + \epsilon \quad (1)$$

and the relationship between IQ and the rest of explanatory variables is given by equation 2:

$$IQ = \delta_0 + \delta_1 exper + \delta_2 urban + \delta_3 black + \delta_4 educ + \eta \quad (2)$$

- (a) Use the NLS80.dta data to estimate the following "incorrect" equation (3) (omitting the IQ variable):

$$\ln(wage) = \beta_0 + \beta_1 exper + \beta_2 urban + \beta_3 black + \beta_4 educ + u \quad (3)$$

- (b) Express the reduced form  $\delta$  coefficients and error term  $u$  from equation 3 in terms of the structural parameters and the structural error terms from equations 1 and 2
- (c) Express the size of the bias in the  $\delta$  coefficients in terms of the structural parameters. Give interpretation to the two sources of the bias. Under what conditions omitting IQ from the equation yields consistent estimates of the coefficients of the rest of the explanatory variables.

- (d) Add IQ to your explanatory variables and estimate the structural model 1. Compare the two sets of estimates from the 2 models and calculate the bias of each of the coefficients in equation 3( as a difference between the "true" and the biased estimates)
- (e) Estimate equation 2 and use your results together with the results from equation 1 to calculate the bias according to the expression from part (b). Compare with the bias computed in part (c).
5. (15 points) Advanced Wooldridge 4.14.
6. (5 points) Assume you use region-level cross-sectional data to estimate the effect of punishment on the regional crime rate. You use arrest rate as measure of punishment, and also include unemployment rate as another variable in your crime rate regression. However, you know that conviction rate (not present in data) also deters criminality and it turns out to be negatively correlated with the arrest rate.
- (a) Derive the formula for the sign and the size of the bias in your coefficient estimates caused by omitting conviction rate in you model. Is the negative effect of arrest rate on criminality underestimated or overestimated?
- (b) is the effect of unemployment rate on crime rate estimated consistently? Justify your answer
7. (15 points) Introductory Wooldridge (International Edition) 9.4 (i) and (ii) plus:
- (a) Assume  $tvhours^*$  are measured with an error, so that  $tvhours = tvhours^* + \epsilon_y$  and the CEV assumptions hold.
- Are the OLS estimates of the coefficients consistent?
  - What is the error variance in this model?
  - What can we conclude about the standard errors of the OLS coefficient estimated?
  - Are the right hand side variables (on average) more or less likely to be found significant when  $tvhours$  is measured with abovementioned error.
- (b) Now assume that  $tvhours^*$  is observed and measured correctly, but  $motheduc$  is measured with an error.
- Under what assumptions are the OLS estimates of all the coefficients consistent? Is standard OLS-based inference affected by the presence of a measurement error that satisfies these assumptions? How?
  - Assume that the true mother's education and the error in its reported value are uncorrelated. What is the relationship between the OLS estimate of the effect of mother's education and the true effect captured by  $\beta_3$ ? Under which conditions are the OLS estimates of the remaining coefficients consistent? Are these conditions likely to hold?
8. (15 points) Introductory wooldridge (International edition) Chapter 9 C7 plus:

- (a) Provide detailed summary statistics (minimum, maximum, mean, 5th and 95th percentile, standard deviation) for all variables used in this regression.
  - (b) Inspect all the variables used in the regression in terms of their missing values.
9. (extra 5 points) consider a regression  $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$  and suppose that  $x_{1i}$  is measured with error. (Clearly state all necessary assumptions)
- (a) Are  $\beta_1$  and  $\beta_2$  unbiased or biased? under which conditions?
  - (b) Suppose you omit  $x_2$  from the regression. Will the bias in  $\beta_1$  smaller or bigger? Under which conditions? Explain.