CERGE-EI
Summer 2013/2014
Instructors: Nikolas Mittag, Dragana Stanišić
TA: Jelena Plazonja, Gega Todua
Date: 12/05/2014

ECONOMETRICS II

Exercise session #3
(Omitted Variable Bias and Measurement Error)

Problem 1. In a study of gender discrimination a linear regression model is estimated on cross-sectional data, with the logarithm of monthly earnings as the dependent variable and a dummy variable that is 1 for men and 0 for women as independent variable. The relation also has an intercept. The OLS estimate of the intercept is 5.91(0.012) and the OLS estimate of the slope is 0.10(0.025).

(a) What is the interpretation of the slope coefficient and what is the interpretation of the intercept?

(b) What do you conclude about the relative difference in monthly earnings between men and women? Is it significantly different from 0? (Assume number of observations is large.)

(c) It is argued that the estimate of the slope is biased, because work experience (in years) has been omitted from the relation. Write down the true model and the artificial regression that relates the gender dummy to the omitted work experience. Express the omitted variable bias in terms of the coefficients of the true model and the artificial regression. What is the likely sign of the bias?

(d) Assume that an additional year of experience leads to a 2% increase in monthly earnings. Also, assume that men have on average 3 more years of experience than women. What is the true difference between the monthly earnings of a man and a woman with the same work experience.

Problem 2. Assume you have an individual -level cross-sectional data on the amount of safety equipment and whether an individual has an insurance. You are interested in the effect of insurance on the amount of safety equipment, but you worry about omitting an important factor, risk aversion, which you do not observe in the data. Assume that insurance and risk aversion are positively correlated.

(a) What is the likely direction of the bias of your OLS estimate of the effect of insurance on investment in safety? Justify any assumptions you make about the sign of the underlying parameters.

(b) Propose two different estimation strategies that allow you to estimate the true effect of insurance on safety equipment even when risk aversion is not observed. Consider using additional variables (other information that you may have in the data) and clearly state what should be their properties.

Problem 3. Explain the concepts of classical measurement error (CME) and optimal prediction error (OPE) and state the main differences between them.

Problem 4. Consider a regression of consumption on income and a constant. Suppose the standard deviation of income is 3, the covariance between income and consumption is 5.4.

(a) Assume that income is uncorrelated with the error term of the regression. What is the effect of an increase in income by 100 units on the level of consumption, when estimated by OLS?

(b) Now assume that income is measured with an error, which satisfies the classical measurement error assumptions and which has a variance equal to 1. Determine the size and the direction of the bias in the effect of income on consumption estimated by OLS. Is the estimate of the constant term consistent?

(c) Now assume that income is observed accurately but consumption is measured with a classical measurement error with variance equal to 1. The t-statistic of the OLS estimate of the coefficient of income is equal to 2. What is the size of the bias in the effect of income on consumption when estimated by OLS? Can we conclude anything about the significance of the estimated effect? Justify your answer.

Problem 5. Consider a cross-sectional regression of monthly charity contributions (in CZK) on monthly income (in CZK) and a constant. Suppose income is measured with an error that satisfies the classical measurement error (CME) assumptions. The standard deviation of the true income is 2000 CZK and the standard deviation of the measurement error is 1000 CZK.

(a) Express the probability limit of the OLS estimate of the propensity to contribute to a charity as a function of the true coefficient. Determine the size and the direction of the bias. Assume that LS estimation of this regression produces a coefficient on income equal to 0.02. what is the true effect on charity contributions of increasing one's income by 1000 CZK?

(b) Now assume that income is measured accurately but the amount of charity contributions is contaminated with an error. Under what assumptions will the OLS estimate of the effect of income on contributions to charity be consistent? Are they likely to hold?

Problem 6. Consider the following model:

$$y^* = \beta x^* + \epsilon, \text{ where } \mathbb{E}(\epsilon|x^*) = 0$$

(a) Assume x^* is measured with an error. the observed x is determined as follows: $x = x^* + v$, where v is independent of x^* and $v \sim N[0, \sigma_v^2]$. prove that the squared correlation between x and y^* is smaller than the squared correlation between x^* and y^*.

(b) Assume y^* is also measured with an error. The observed y is determined by: $y = y^* + u$, where u is independent of x^* and y^* and $u \sim NN[0, \sigma_u^2]$. How does the squared correlation between x and y compare to that between x^* and y^*?

(c) Compare the squared correlations (between x^* and y^*, x and y^*, and x and y) and comment.

(d) How does your answer change when the two errors are distributed iid instead of Normally with the means and variances as before?

Problem 7. Consider multivariate regression model $y = X^*\beta + \epsilon$, where X^* is measured with an error, that is $X = X^* + u$. Derive plim of the OLS estimator and discuss the consistency problem caused by CME. Additionally, consider the case of two regressors where one is measured with error and one not and discuss how the attenuation bias of error-ridden variable worsens when other variables are included. Discuss how the presence of error-ridden variable causes inconsistency in the OLS estimates of the coefficients on other variables if the variables are correlated.