Problem Set #10: Specification and Data Problems

Economics 435: Quantitative Methods

Fall 2011

1 Proxy variables

An important policy question in labor economics is the extent to which young people from lower-income families are "credit constrained" in their educational choices.

The regression we would want to estimate is:

$$school = \beta_0 + \beta_1 faminc + \beta_2 ability + u$$

where *school* is the number of years of school received by the respondent, $faminc$ is the household income of the respondent's family, *ability* is the student's ability[1], and

$$E(u|faminc, ability) = 0$$

The parameter of interest is β_1. Suppose we have a random sample of data on $(school, faminc)$, but do not observe ability.

a) Let $\hat{\beta}_1^A$ be the coefficient on $faminc$ from an OLS regression of *school* on $faminc$. Find plim $\hat{\beta}_1^A$

b) Based on your answer, is $\hat{\beta}_1^A$ more likely to overestimate β_1 or underestimate it? Explain your reasoning.

c) One way to address an omitted variable problem is to find a proxy for the omitted variable. Suppose that we have data from a standardized test taken by each respondent. Let *test* be the respondent's test score. Assume that:

$$ability = \gamma_0 + \gamma_1 faminc + \gamma_2 test + v$$

where

$$E(u|faminc, test) = E(v|faminc, test) = 0$$

Find $E(school|faminc, test)$

d) Suppose that we estimate an OLS regression of *school* on $faminc$ and *test*. Let $\hat{\beta}_1^B$ be the coefficient on $faminc$ from that regression. Find plim $\hat{\beta}_1^B$.

e) Based on your answer, what do you think is the direction of bias in using $\hat{\beta}_1^B$ to estimate β_1? Do you think the bias in $\hat{\beta}_1^B$ is likely to be higher or lower than the bias in $\hat{\beta}_1^A$? Explain your reasoning.

[1] "Ability" here is something of a loaded word. In this context, it means

2 Regression with standardized variables

Suppose that we have a random sample of size n on the random variables y and x such that:

$$E(y|x) = \beta_0 + \beta_1 x$$

Now suppose that we standardize both the independent and dependent variable. Let $(\tilde{y}_i, \tilde{x}_i)$ be defined by:

$$\begin{aligned} \tilde{y}_i &= \frac{y_i - \bar{y}}{\sqrt{v\hat{a}r(y)}} \\ \tilde{x}_i &= \frac{x_i - \bar{x}}{\sqrt{v\hat{a}r(x)}} \end{aligned}$$

Let $\hat{\beta}_1$ be the coefficient on $\tilde{x}_i$ from an OLS regression of $\tilde{y}$ on $\tilde{x}$.

a) Find $\hat{\beta}_0$ as a function of $\hat{\beta}_1$.

b) Find the R^2 from the regression as a function of $\hat{\beta}_1$.

c) Find the sample correlation $c\hat{o}rr(\tilde{y}, \tilde{x})$ as a function of $\hat{\beta}_1$.

d) Can $\hat{\beta}_1$ possibly be less than -1 or greater than $+1$? Prove it.