

**Economics 371**  
**Problem Set #10**

(1) Consider a simple model to estimate the effect of personal computer ownership (denoted “PC”) on college grade point average (GPA):

$$GPA_i = \beta_0 + \beta_1 PC_i + u_i.$$

In this model, PC is a binary variable taking the value 1 if the student owns a computer, and is zero otherwise.

(1a) Why might you be concerned about using OLS as a method to estimate  $\beta_0$  and the “causal” effect of computer use on GPA,  $\beta_1$ ? (*Hint:* Think about how mean-independence may potentially be violated.)

(1b) Suppose that you also had data on the parental income of each student. Someone suggests that you should use the parental income variable as an instrument for computer ownership. What conditions would need to be satisfied in order to obtain consistent parameter estimates when parental education is used as an instrumental variable (IV)? Which of these conditions are likely to be violated or satisfied if this particular instrument is used?

(1c) Now, suppose that incoming students are randomly assigned to residence halls. Some of the residence halls are “older” and do not contain their own computer labs, while others contain large computer labs. Explain how you might use this data to construct an instrumental variables estimator. How important is the *random* assignment to residence halls in this context?

(2) That problem will make use of the data, WAGE2.dta, which is provided on the course website.

Consider the regression model

$$\log(wage)_i = \beta_0 + \beta_1 educ_i + u_i.$$

(2a) Run this regression using OLS, and report your estimate of the return to education (*educ*). Discuss why one might be concerned that *educ* is endogenous. Also note that  $\log(wage)$  will need to be generated in STATA.

(2b) The variable *brthord* in this data set is the birth order of the respondent. (If you are the first-born child in the family, *brthord* = 1, if you are the second child, *brthord* = 2, etc). Explain why *brthord* and *educ* may be negatively correlated. Regress *educ* on *brthord* to see if, in fact, the correlation is negative and statistically significant.

(2c) Use *brthord* as an instrument for *educ* and obtain the IV estimator of  $\beta_1$ . Here is how you do this in STATA:

1. Under the “Statistics” menu, choose “endogenous covariates” and then “instrumental variables and two stage least squares.”
2. In the menu that pops up, choose “logwage” as the dependent variable [ you will have to generate this], leave the independent variables blank (since there are no exogenous variables in our regression equation), choose *educ* as the endogenous variable and *brthord* as the instrument. Obtain your IV estimates, and compare them to (3a). Under what conditions will *brthord* be a valid instrument? Do you think those conditions are likely to apply for this application?

(2d) Now, estimate the model

$$\log(wage)_i = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + \beta_3 sibs_i + \beta_4 exper_i + u_i$$

again using instrumental variables. In this regression, *IQ* is the IQ of the respondent, *sibs* is the respondent's number of siblings, and *exper* is years of labor market experience.

To perform this IV regression in STATA, follow the steps of (3c), but this time, add *IQ*, *sibs*, and *exper* as independent variables. (Keep the dependent and endogenous variables the same, and also keep the same instrument). Note that STATA automatically performs two stage least squares here, using the newly included exogenous variables as candidate instruments.

Interpret your IV results. What remains significant?