Problem Set - Instrumental Variables

1. Wooldridge exercise 5.1

2. Consider a simple model to estimate the effect of personal computer (PC) ownership on college grade point average for graduating seniors at a large public university:

$$GPA = \beta_0 + \beta_1 PC + u$$

where PC is a binary variable indicating PC ownership.

   (a) Why might PC ownership be correlated with u?

   (b) Explain why PC is likely to be related to parents' annual income. Does this mean parental income is a good IV for PC? Why or why not?

   (c) Suppose that, four years ago, the university gave grants to buy computers to roughly one-half of the incoming students, and the students who received grants were randomly chosen. Carefully explain how you would use this information to construct an instrumental variable for PC.

3. Suppose that you wish to estimate the effect of class attendance on student performance, as in the Lecture. A basic model is

$$examscore = \beta_0 + \beta_1 attendance + \beta_2 priorGPA + u$$

where *examscore* is students score on the exam (from 1 to 6), *attendance* is the number of seminar meetings attended (from 1 to 12), and $priorGPA$ is the average exam grade last year.

   (a) Let *dist* be the distance from the students' living quarters to the lecture hall. Do you think *dist* is uncorrelated with u?

   (b) Assuming that *dist* and u are uncorrelated, what other assumption must *dist* satisfy in order to be a valid IV for *attendance*?

   (c) Suppose, we add the interaction term $priorGPA \times attendance$. If *attendance* is correlated with u, then, in general, so is $priorGPA \times attendance$. What might be a good IV for $priorGPA \times attendance$?

4. (Difficult) Consider the following structural equation

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

where $cov(x_{i1}, u_i) \neq 0$ and $cov(x_{i2}, u_i) = 0$. You have an instrumental variable z_i for x_{i1}:

$$cov(z_i, x_{i1}) \neq 0 \text{ and } cov(z_{i1}, u_i) = 0.$$

Which you would use to estimate the first-stage

$$x_{i1} = \pi_0 + \pi_1 z_i + \pi_2 x_{i2} + v_i$$

(a) Show that plim $\hat{\beta}_1 = \beta_1$ but plim $\hat{\beta}_2 \neq \beta_2$ when x_2 is included in the first-stage estimation, but not used to construct the instrument $\hat{x}_1$ (i.e., $\hat{x}_1 = \hat{\pi}_0 + \hat{\pi}_1 z_1$)

(b) Show that both plim $\tilde{\beta}_1 \neq \beta_1$ and plim $\tilde{\beta}_2 \neq \beta_2$ when x_2 is not included in the first-stage estimation

Hint: Remember that $x_1 = \hat{x}_1 + (x_1 - \hat{x}_1)$ which means that the second stage becomes

$$y_i = \beta_0 + \beta_1 \hat{x}_{i1} + \beta_2 x_{i2} + \underbrace{u_i + \beta_1(x_{i1} - \hat{x}_{i1})}_{\tilde{u}_i}$$

Hint: The 2SLS estimate of β_1 is $\hat{\beta}_1 = \sum_{i=1}^n \hat{r}_{i1} y_i / \sum_{i=1}^n \hat{r}_{i1}^2$ where $\hat{r}_{i1} = \hat{x}_{i1} - \hat{\alpha}_0 - \hat{\alpha}_1 x_{i2}$, the residual from the following regression: $\hat{x}_{i1} = \alpha_0 + \alpha_1 x_{i2} + r_{i1}$. The 2SLS estimate of β_2 has a similar expression.