Solutions: Problem Set #10

(1a) Mean-independence could be violated for a variety of reasons. First, highly-motivated students may be more likely to purchase computers. (They decide to purchase a computer, simply because they think it will give them an "edge," and this motivation also is correlated with GPA). In this case, we have an omitted variables problem leading to a mean-independence violation. Second, students from wealthier families may be more likely to purchase a computer. If wealthier families also tend to invest more in the human capital of their children, and this human capital also helps GPA, then again, we have an omitted variables bias problem and mean-independence violation.

(1b) The second story described in (1) suggests that this may not be a valid instrument. Sure, parental income is probably strongly correlated with PC ownership (instrument relevance), but at the same time, parental income may have a direct effect on GPA independently of its effect on PC ownership. In this case, we would violate instrument exogeneity.

(1c) Construct a dummy variable called z which equals one if the student lives in a hall with a computer lab, and equals zero otherwise. We would then apply the IV estimator:

$$\hat{\beta}_{1,IV} = \frac{\sum_{i=1}^{n}(GPA_i - \overline{GPA})(z_i - \overline{z})}{\sum_{i=1}^{n}(PC_i - \overline{PC})(z_i - \overline{z})}.$$

The instrument is seemingly relevant - students who have access to a computer lab in their residence hall are probably less likely to purchase a PC. At the same time, if there is no lab in the residence hall, students are probably likely to purchase a computer since, if you did not own a PC, you would need to leave the residence hall to do computer-based assignments, check e-mail, etc. (If it is cold, there are non-trivial costs for doing this!)

The random assignment also makes the instrument plausibly exogenous - there is nothing in the model that would generate a correlation between z and GPA. If students get to choose their halls, there is an obvious selection bias problem - it may be the case that high achieving students, or students of lower incomes will select into the halls with computer labs.

(2) OLS estimates of the model produce the following estimated regression equation:

$$log(\widehat{wage}) = 5.97 + .060Educ.$$

This suggests that an added year of education increases ones' wages by about 6 percent. This coefficient is often called the *return to education*.

(2b) Generally speaking, we expect that first-born children will have the undivided attention of the mother and father during their infancy. As a result, these parternal inputs may contribute to the human capital and abilities of the child, perhaps leading that child to acquire more education. Young children with other siblings will not receive as much partental attention, on average. Thus, we might expect to see a negative correlation between birth order and educational attainment.

The estimated regresison equation which relates education to birth order is:

$$\widehat{Educ} = 14.15 - .28birthord.$$

The coefficient on *birthord* is found to be statistically significant at all standard levels, with a $t-$statistic equal to -6.1.

(2c) The IV estimates for this regression are as follows:

$$log(\widehat{wage}) = 5.03 + .13Educ,$$

which suggests a return to education equal to 13 percent.

For *brthord* to be a valid instrument, we first need it to be correlated with education. The results in (3b) suggest that this is not a problem - it is strongly correlated with education. The second condition we need is that it is not correlated with the error in the log wage equation, or, said differently, that once we control for education, *brthord* does not enter the log wage equation. This second condition is a little more difficult to believe, and can not be tested. Our previous argument suggested that *brthord* was correlated to education since parents would devote more time, on average, to first-born children. This suggests that such children might be more "able" and this ability may itself be an important determinant of log wages above and beyond education itself. So, potentially, *brthord* may be correlated with the error term in the log wage equation.

(2d) The 2SLS estimate is as follows:

$$log(\widehat{wage}) = 3.72 + .222Educ - .004IQ + .007Sibs + .047Exper.$$

First, we note that, aside from the intercept, and experience at the 10 percent level, no coefficients in this model are significantly different from zero. A reason for this is high collinearity. In 2SLS, *Educ* is replaced by its fitted value from a regression of *Educ* on *IQ*, *Sibs*, *Exper* and *brthord*. This fitted value is highly correlated with the other variables appearing in our regression equation (*IQ*, *Sibs*, and *Exper*), leading to large standard errors.

This regression, however, is something of an improvement over the one in (3c), since we have added other controls like siblings and IQ which control for family background. This tends to increase the credibility of using *brthord* as an instrument. However, the cost of doing this is that we loose precision in our estimation. This is a common result in IV studies.