# ECON 21110 - Applied Microeconometrics - Assignment 4

Jack Ogle in collaboration with Eva Haque, Matt Lohrs, and Jack Knickrehm

Problem 1

(a) Population Model (1)

$$JQ_i = \beta_0 + \beta_1 econ_i + \beta_2 drinking_i + u_i$$

$\beta_0$ is the estimator measuring the average baseline of JQ indexes. Holding all other variables constant including holding heavy drinking and being enrolled in applied econometric constant at 0 $\beta_0$ is the average job quality index JQ.

$\beta_1$ is the estimator measuring the average effect of taking applied econometrics on Job quality index holding heavy drinking constant. On average taking econometrics increases Job Quality index by $\beta_1$ standard deviations.

$\beta_2$ is the estimator measuring the average effect of heavy drinking on Job quality index holding taking applied econometrics constant. On average heavy alcohol consumption decreases or increases Job Quality index by $\beta_2$ standard deviations.

(b)

No I do not expect drinking to be truthfully reported by college students because even if the consumption is anonymously reported students still might feel guilty or ashamed of how much they drink. They might think that the researcher would like to hear a certain answer and in order to be helpful report that specific answer. For these reasons I believe that they might under report the amount of alcohol they consume. I also believe that the amount people report drinking is correlated with measurement error. Heavy drinkers are more prone to under report compared to people who seldom drink or don't drink at all. The conditions of an survey that would greatly reduce if not eliminate measurement error would be if student honestly reported drinking habits. Additionally if they surveyed honest students who are part of a randomized sample, such that there was not measurement error or omitted variable bias.

(c)

A violation of MLR.4 (the zero conditional mean) would make the the OLS estimator $\hat{\beta}_1$ of $\beta_1$ (1) biased. I would expect MLR.4 to be Omitted Variable biased here. I expect this because there are many other factors that are in the error term that are correlated to the other independent variables. For example, the Economics department requires all Economic specialization in data science majors to take Applications of Microeconometrics or a similar alternative in a pool of classes. Or maybe one of the students is in a fraternity or sorority and they are required to participate in heavy drinking. So the binary indicator variables of being a Economics specialization Data Science Majors and Being a member of a fraternity or sorority are variables in the error term that would violate MLR.4 because they are correlated to independent variables.

(d)

In order for econ_cost to be a valid instrument for econ if it is correlated with econ and uncorrelated with the error term.

$$Cov(JQ_i, econcost_i) = Cov(\beta_1 econ_i + U_i, econcost_i)$$

$$Cov(U_i, econcost_i) = 0$$

$$\beta_1 = \frac{Cov(JQ_i, econcost_i)}{Cov(econ_i, econcost_i)}$$

A consistent estimate $\hat{\beta_{IV_1}}$ of $\beta_1$ is

$$\hat{\beta_{IV_1}} = \frac{\Sigma(JQ_i - \bar{JQ})(econcost_i - \bar{econcost})}{\Sigma(econ_i - \bar{econ})(econcost_i - \bar{econcost})}$$

I believe that the variable econ_cost or above econcost is not a valid IV. We are not given any data here so I am going off of the follwoing assumptions. The econ variable is an indicator variable indicating a 1 if a student takes econometrics. The econ_cost variable randomly assigns students to colleges where the econometrics is a cheeper class compared to other colleges offering the same class. Econ_cost is not a good IV because it is weakly correlated or not correlated with taking applied econometrics. Students do not care whether the price of the course is cheaper than the same course offered at a different college. Transferring colleges is an too expensive an implicit cost to pay for taking a less explicitly expensive class. The fact that they are in a college with a less expensive econometrics class compared to other colleges does not increase nor decrease the probability that students take the econometric class. Because this IV is not correlated with econ it is not valid.

(e)

First-Stage regression: Isolate the part of econ that is correlated with the error term regress econ on econ_cost using OLS

$$econ_i = \pi econcost_i + v$$
$$\hat{\pi} = \frac{\hat{Cov}(econcost_i, econ_i)}{\hat{Var}(econcost_i)}$$

then use $\hat{\pi}$ to predict the value of econ: $\hat{econ}_i = \hat{\pi} econcost_i$

From here we can calculate the heteroskedasticity-robust F-statistic to determine if it is relevant.

(f)

The main problem with the endogenous variable is that the variable is correlated with the error term resulting in Omitted Variable Bias and MLR.4 is violated. Our estimator $\tilde{\beta}_1$ solves the endogeneity bias problem if econ_cost is a valid estimator because MLR.4 is more credibly satisfied. In using a valid instrumental variable we have successfully taken a variable out of the error term and included it our regression. We must include the variance of the estimator that is uncorrelated with the error term. if econ_cost is a valid IV for econ we can use econ cost to estimate econ. Then we can get the variance of econ uncorrelated with the error term. This allows us to run an unbiased regression because there is overall less correlation between the error term and the independent variables.

(g)

We cannot test exogeneity. If econ_cost is not truly exogenous in the structural model (1) then it would be endogenous and correlated with the error term. This would violate MLR.4 and the estimator would be inconsistent and biased.

(h)

This Cov(Z, U) = 0 must hold for all IV. However, it is impossible to observe the error term therefore we must employ economic theory to intuit an answer. From the response to part d I believe that econ_cost is not a valid instrument for econ. This is because the cost of an econometrics course relative to other econometrics courses at different colleges is weakly correlated if correlated at all with a students choice to take that class. If econ is not truly exogenous to the model then there is omitted variable bias and the variable is correlated to the error term. Meaning that there is(are) a variable(s) that is(are) correlated with econ and JQ, but is not included in the regression (eg. Ability, IQ, KWW). Thus the econ_cost variable is unlikely to be exogenous. The IV is not valid as it is not corrolated with econ nor is it exogenous.

$$\frac{Cov(econ, U)}{Var(econ)} < \frac{Cov(econcost, U)}{Cov(econcost, econ)} => \lim_{n\to\infty} plim\hat{\beta}_1 < \lim_{n\to\infty} plim\tilde{\beta}_1$$

This shows that the estimator is more consistent without the IV and we should stick with the original estimator $\hat{\beta}_1$ in (1).