

Answer each of the following questions, clearly and thoroughly. Your answers to each of the first two questions will determine 30% of your grade. Your answer to the third question will determine 40% of your grade on the final.

- 1) We have spent substantial time studying estimation problems that arise if we are unable to fully specify a model, and omit an important variable. We know that under certain conditions, omitting an important variable can bias ordinary estimates of relationships between an independent and dependent variable.

We have discussed several estimation strategies useful in the context :

- Instrumental variables
- Selection correction techniques
- Difference-in-differences estimators
- Fixed effects
- Natural experiments

- A) Choose any two of these estimation strategies. For each, clearly describe why it may improve on ordinary parameter estimates.
- B) Because we cannot measure an omitted variable, it's mere presence (and influence) cannot be directly tested. The ability of any of the estimation strategies listed above to improve on ordinary estimates rests on assumptions about the omitted variable and/or additional information employed in the estimation. Describe the key assumptions or information that your chosen estimators draw on? What are the limitations of these estimators?
- 2) In their paper, "Sex, Drugs and Catholic Schools: Private Schools and Non-Market Adolescent Behavior," David Figlio and Jens Ludwig examine the effects of enrollment in a Catholic school on certain types of behavior: drinking alcohol, and criminal activity. To study this question, Figlio and Ludwig used a sample of more than 10,000 students enrolled in either Catholic schools or public high schools, and who were seniors in 1992. [Note, all those enrolled in non-Catholic private high schools were dropped from the sample.]

Figlio and Ludwig were interested in identifying the effects of attendance at Catholic schools on limiting undesirable behaviors among teens. A principal hypothesis motivating this question is that exposing students to more rigid standards of conduct and dress, as well as providing moral or religious education might limit the extent to which students engaged in drinking, sexual activity, and illegal behavior.

The authors estimated a model (among others) in which the independent variables include measures of race, gender, parents' education, religion, urbanicity, behavior in 8th grade (like skipping school and fighting). Most importantly, the independent variables

included a dummy variable equal to one if the student attended Catholic school, and zero if s/he attended public school. The dependent variable was equal to one if the student reported drinking alcohol in the last year, zero otherwise.

The authors estimated this model via OLS. [Models with dummy dependent variables estimated via OLS are called linear probability models. Such models are estimated instead of logit or probit because they are easy to interpret. The OLS coefficients can be interpreted in the usual way. Note that if the mean value of the dependent variable is 0.25, that means that on average, there is a 0.25 probability of Y occurring (or 25% of the sample have a value of 1 on the dependent variable)]

The authors recognized that they faced potentially serious problems because enrollment in Catholic schools was not likely to be randomly determined. For example, attendance in a Catholic school might be more common among children whose parents themselves provide substantial moral or behavioral guidance. Or, attendance at Catholic schools, may be more common if a student lives in a neighborhood with many social problems, and poor public schools, because the Catholic school serves as a refuge from neighborhood problems.

To develop better estimates of the effect of Catholic school attendance on drinking, the authors employed an instrumental variables estimator. To instrument for Catholic school attendance, the authors used measures of the quality of the public transit system in the students' cities. Good public transit decreases barriers to Catholic school enrollment, because it enables kids to get to school without incurring additional costs (in time or money) on the family.

The results of the OLS and IV estimation are summarized below. Only the coefficients on the Catholic school dummy is reported.

The Effect of Private School Attendance on Drinking

Dependent variable = Drank Last Year

Independent variable	<u>OLS</u>	<u>IV</u>
Catholic School Attendance	-0.037 * <i>(0.022)</i>	0.028 <i>(0.062)</i>

Remember, model controls for a variety of other measurable student, parental, and school characteristics. Coefficients are reported in bold, and standard errors on those coefficients are italicized in parentheses below.

** Indicates significance at the 0.10 level.*

*** Indicates significance at the 0.05 level.*

A) Focusing on the OLS estimates, how do you interpret the effect of Catholic school

attendance on drinking behavior?

- B) Focusing on the IV estimates, how do you interpret the effect of Catholic school attendance on drinking behavior?
 - C) Lay out a plausible explanation for the differences between the OLS and IV coefficients of Catholic school attendance on students' drinking behavior.
 - D) Describe how you would assess the quality of the instrumental variables estimate.
- 3) You have been hired as an analyst with expertise in understanding the impact of training on job performance by a very large employer, the U.S. Army. The Army is interested in identifying how the training it provides to enlisted men and women improves performance on the job.

They would like you to evaluate whether soldiers who engage in training beyond basic training perform better than soldiers who do not. Assume that the training soldiers engage in beyond basic training is homogeneous. Soldiers have either engaged in it or not. The training is not compulsory. In order to engage in training, soldiers must volunteer, and their commanding officer must approve the time spent away from normal duties. So, training takes a commitment from both individual soldier, and from the commanding officer. The Army informs you that 37 percent of the soldiers currently enlisted have engaged in training.

Performance of each soldier in his or her job is measured annually by an objective outside reviewer. A soldier's performance score is recorded on a percentile scale of 0 to 100. You can treat this as a continuous variable. [For the purposes of answering the questions below, assume the performance assessments are valid and reliable.]

The Army has informed you that performance scores are likely to be a function of the following things:

- whether a soldier has participated in training other than basic training
- a soldier's rank
- the number of months a soldier has spent in the Army

The Army have provided you with records for all enlisted personnel for each of the past 12 years. Their records contain information on performance, and on the variables listed above. These records constitute an unbalanced panel. They contain information on all soldiers in each of these years. In any one year, these data contain observations on soldiers who will be enlisted only for two or three years, as well as observations on soldiers who are making a career out of military service.

The records also contain one additional key piece of information: The military base the

soldier was assigned to. This information is included because the Army reports to you that commanding officers at some bases are particularly reluctant to assign soldiers to training, while at other bases, they are much more liberal in assigning soldiers to training. Soldiers are assigned to bases at random. So, the likelihood that soldiers get training is at least in part explained by forces beyond their own control; the process used to assign them to bases.

As a final bit of background, officials at the Army tell you that while they are interested in the effect of training on productivity, they are also worried about its effect on retention. Military service has often been a stepping stone to civilian jobs. But, they worry, training improves soldiers skills and improves their prospects in the civilian labor market, so at least some of the trained soldiers leave the army soon after.

- A) With this in mind, carefully describe the estimation problems you confront. In doing so, lay out the basic model. Discuss how the problems you have identified affect your ability to draw inference from OLS about the effect of participating in training on performance.
- B) Describe the approach(es) you would employ to estimate the effect of training on performance.