Mastering Mostly Harmless Econometrics

MIT 14.381 **J. Angrist**
Fall 2019 **D. Hughes**
Due: Tues, Nov. 12

PROBLEM SET I

1. Regression basics

You're interested in the regression of log wages, Y_i, on years of schooling, S_i, controlling for another variable related to schooling and earnings that we'll call A_i for "ability". Write this regression as:

$$Y_i = \alpha + \rho S_i + A_i\gamma + \varepsilon_i \qquad (1)$$

Recall that regression coefficients are defined so that ε_i is uncorrelated with S_i and A_i.

a. Modify (1) to include a vector of *k* controls, X_i, with coefficient β. Compare the schooling coefficient with and w/o ability in this case (call the short regression coefficient ρ_0 and the long regression coefficient ρ_1). Can the addition of controls change the sign of OVB?

b. Generalize the OVB formula with controls to allow for multivariate S_i and A_i. (This produces a matrix formula for which we give an indulgence).

c. Go back to the model without controls other than ability. Suppose schooling is mismeasured, with measurement error independent of ability. Show how this sort of measurement error affects the schooling coefficient in models like (1), with and without ability controls.

d. Suppose instead that schooling is well-measured, but the only ability variable available is a GRE math score. GRE and IQ are highly correlated, yet IQ is probably a better control for purposes of estimating an equation like (1). Why? Use a model like that at the end of Section 2.2.1 in AK-99 to show how use of GRE instead of IQ affects the schooling coefficient identified by a model like (1).

2. Limited dependent variables

a. Derive the probit and logit likelihood functions and give a GLS interpretation of the MLEs.

b. Derive a simple rule of thumb for logit marginal effects.

c. Consider probit with a single Normally distributed regressor, X, and dummy dependent variable, Y. Show that the OLS slope coefficient from a regression of Y on X equals the average derivative for a probit CEF.

d. Derive the Tobit CEF and a simple rule of thumb for Tobit marginal effects. Use this to show that Tobit marginal effects must be smaller in magnitude than the corresponding Tobit coefficients, with the same sign. What do Tobit coefficients tell us? (for hints, see MHE 3.4.2)

e. Use the data from Angrist-Evans (1998) to produce your own version of MHE Table 3.4.2.

3. Regression and matching

a. Consider an additive treatment effect model conditional on discrete covariates. Prove that if the regression model for covariates is saturated, then matching and regression estimands are the same in either of the following two cases: (i) treatment effects are independent of covariates; (ii) treatment assignment is independent of covariates.

b. Why might you prefer regression estimates to matching estimates, even if you're primarily interested in the population average treatment effect?

c. Calculate matching and regression estimates in the empirical application of your choice. Explain the difference between the two estimates with the aid of a figure like the one used for this purpose in Angrist (1998).

4. A stylized version of the MM Chapter 2.2 regression discussed in class can be written:

$$\ln Y_i = \alpha + \beta P_i + \text{DUMMY CONTROLS}_i + e_i, \tag{3}$$

where P_i indicates private school attendance. The notation DUMMY CONTROLS_i in (3) is short for dummy variables indicating sets of schools to which applicants have applied and been admitted.

a. Show that OLS estimates of this model discard information on students in application/admission groups that consist entirely of private- or public-school graduates.

b. Why is this fact relevant for the analogy between regression and matching?

5. 2SLS and grouping

a. Define S to be an $N \times J$ "summer matrix" such that $S'y$ transforms an $N \times 1$ vector y into a $J \times 1$ vector of sums with n_j terms in each. Show that $H = S(S'S)^{-1}S'$ is an idempotent $N \times N$ matrix that replaces N individual observations with group averages.

b. Assuming the underlying data are homoscedastic, use matrix S to show that GLS on J grouped means is the same as 2SLS. What are the instruments in this 2SLS procedure? What are the first-stage fitted values?

c. Show that GLS for grouped data minimizes a weighted sum of squares and that this GLS minimand is the same as the over-identification test statistic that tests instrument-error orthogonality, where the instruments are those in (b). How is this test statistic distributed?