Topics in Applied Econometrics

MIT 14.387
Fall 2014

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:

$$Y_i = \alpha + \rho S_i + A_i \gamma + \varepsilon_i \tag{1}$$

As always, the population regression coefficients α, ρ and γ are defined so that ε_i is uncorrelated with S_i and A_i.

   a. Suppose you run a bivariate regression of Y_i on S_i instead. Derive the bivariate population regression coefficient on S_i in terms of the parameters in equation (1). When does the short-regression coefficient equal the long?

   b. Modify (1) to include a vector of maintained controls with coefficient α. Compare the schooling coefficient with and w/o ability in this case. Can the addition of controls change the sign of OVB?

   c. Generalize the OVB formula to multivariate S_i and A_i. (This is a matrix formula for which I give an indulgence).

   d. Why is the long regression more likely to have a causal interpretation? Or is it?

2. Limited dependent variables

   1. Derive the probit and logit likelihood functions and show that maximum likelihood estimation of the index coefficients in these models can be interpreted as a weighted nonlinear least squares procedure. Give a GLS interpretation of these MLEs.
   2. Derive a simple rule of thumb for logit marginal effects.
   3. Consider probit with a single Normally-distributed regressor. Show that OLS estimates the average derivative for this model.
   4. 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. What do Tobit coefficients tell us?
   5. Use the data from Angrist-Evans (1998) to produce your own version of MHE Table 3.4.2.

3. Discuss the relationship between regression and matching, as follows:

   a. Suppose you're trying to estimate a treatment effect conditional on discrete covariates. Prove that if the regression model for covariates is saturated, then the 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. Propose a weighted matching estimator that estimates the same thing as regression.

c. Why might you prefer regression estimates to matching estimates, even if you're primarily interested in the effect of treatment on the treated or the population average treatment effect?

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

4. A stylized version of the Dale and Krueger (2002) regression discussed in class can be written like this

$$Y_i = \alpha'X_i + \rho C_i + \text{DUMMY CONTROLS} + \varepsilon_i \tag{3}$$

where C_i indicates private school attendance. The DUMMY CONTROLS in (3) indicate sets of schools to which applicants have applied and been accepted.

a. Use regression anatomy formulas to show that the DK estimator of private school effects discards students in application/admissions groups that consist entirely of private or public schools.

b. Explain why the result in (a) facilitates interpretation of the DK regression estimates as a kind of matching estimator.

5. IV estimation

a. Construct an extract from the 1980 Census similar to the one used by Angrist and Krueger (1991). Use this extract to replicate MHE Table 4.6.2.
b. "Prove" that LIML is approximately unbiased by replicating Figure 4.6.2 in chapter 4 of MHE (i.e., do the same Monte Carlo experiment).

6. Properties of 2SLS

Use linear algebra (with a further indulgence granted) to show:

a. That 2SLS is an IV estimator.
b. That 2SLS with one instrument is the same as indirect least squares.

7. 2SLS and grouping

a. Define S to be the N×J "summer" matrix such that S'y transforms an N×1 vector y into a J×1 vector of sums.

i. Show that $H = S(S'S)^{-1}S'$ is an N×N matrix that replaces individual observations with averages.

ii. Use the notation above to show that GLS on J grouped means is the same as 2SLS. What is the instrument matrix in this analogy? What are the fitted values?

v. Show that GLS for grouped data minimizes a weighted sum of squares and that this "GLS minimand" is the same as the chi-square statistic that tests instrument-error orthogonality, where the instruments are those in part (iv)

14.387 Applied Econometrics: Mostly Harmless Big Data
Fall 2014