Economics 8379
Prof. Williams
Homework 5
due Friday, April 24

For this assignment you will use the dataset `SIMCE_25.dta` Please read the file `simple_codebook.pdf` to familiarize yourself with the data. Also look at my sample do file.

1. Estimate the effect of attending a voucher school (`voucher_pub`) on the math test score (`mate`) using OLS. What assumption is required in order to consider this a causal estimate? Is this assumption plausible? What problem, if any, does observed heterogeneity in the causal effect pose for interpretation of the OLS coefficient? What problem, if any, does unobserved heterogeneity in the causal effect pose for interpretation of the OLS coefficient?

2. Using `voucher_2002` as an excluded instrument, estimate the effect of attending a voucher school (`voucher_pub`) on the math test score (`mate`) using 2SLS. Is there any concern about weak instrument bias? Under what conditions can this be interpreted as an estimate of the ATE?

3. Let X denote the "control variables" in the outcome equation (i.e., the regressors excluding `voucher_pub`) and let Z denote the regressors in the first stage equation – X *and* the excluded instrument, $\tilde{Z}$. Under the assumptions

   A0 $Y_d = \beta_d' X + U_d$ and $D = \mathbf{1}(\gamma' Z \geq V)$

   A1 $(U_0, U_1, V) \perp\!\!\!\perp Z \mid X$

   A2 the distribution of $\gamma' Z \mid X$ is nondegenerate (relevance and exclusion restriction)

   the IV estimate in 2 is equal to $\int_0^1 w(u) MTE(u) du$ where

   $$w(u) = \frac{E(\tilde{Z} - E(\tilde{Z}) \mid P(Z) > u_D) Pr(P(Z) > u_D)}{Cov(\tilde{Z}, D)}$$

   See Heckman, Urzua, and Vytlacil (2006, Review of Economics and Statistics). In this light, how should we interpret the IV estimate in question 2? (Hint: First think about what these weights look like when $\tilde{Z}$ is binary.)

4. Define a binary instrument via $\tilde{Z}_i = \mathbf{1}(\texttt{voucher\_2002} > 0.3)$ and estimate the model using this instrument. What is the LATE interpretation? How does this change if you instead define the binary instrument via $\tilde{Z}_i = \mathbf{1}(\texttt{voucher\_2002} > 0.5)$? For each case, estimate the proportion of compliers in the population and compare the average family size among compliers to the average family size in the population.

5. Use the `heckman` command in Stata to estimate the selection model. What is the sign of the coefficient on the inverse mills ratio term in each regression? How can you interpret the sign of the coefficients in the Roy model? In the generalized Roy model?

6. Estimate the MTE two ways: under the normality assumption and semiparametrically. Does your answer change if you add `avg_l_diff`, `avg_m_diff`, and `avg_c_diff` as additional instruments?

7. Can you redo this graph with confidence bands around the MTE curve? (You can do this one of two ways. One way would be to recognize that the MTE as a particular value for u is linear in the estimated parameters. So if you have an estimated variance covariance matrix for the parameters ($V = Var(\psi)$) then $Var(c'\psi) = c'Vc$. The second way would be to use a bootstrap.)