Problem Set 7
Econ 312, Spring 2019
James J. Heckman
Due May 23rd, 2019 by class time
This draft, May 17, 2019

Consider a Generalized Roy model

$$Y_1 = \alpha_1' X + \beta + U_1 \tag{A}$$

$$Y_0 = \alpha_0' X + U_0 \tag{B}$$

$$D = \begin{cases} 1 \text{ if } D^* > 0 \\ 0 \text{ if } D^* \leq 0 \end{cases} \tag{C}$$

where $D^* = Y_1 - Y_0 - C(Z)$,

where $C(Z) = \gamma Z + U_Z$; $(U_0, U_1, U_Z) \perp\!\!\!\perp$(X,Z).

$$\begin{pmatrix} U_1 \\ U_0 \\ U_Z \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0, \Sigma_U \\ 0 \end{pmatrix} \tag{1}$$

observed $Y = DY_1 + (1 - D)Y_0$.

1. [10 pts] What is *the* causal effect of D on Y?

   (a) Define at the individual level.

   (b) Define at the aggregate level.

   (c) Using the hypothetical model framework of Heckman and Pinto (see the class slides "Causality in Econometrics and Statistics, Part III:

Structural Models are Causal Models") define $\tilde{D}$ and define the parents of D and of Y; of (Y_0, Y_1).

(d) Write the model (A), (B), (C) in structural equation form.

2. [65 pts] Using the posted data sets I, II, and III, and assuming observations are independent, for the Generalized Roy model,

   (a) Estimate the identified parameters of $Pr(D = 1|X, Z)$. Be explicit about what is identified and what is not.

   (b) What is the sample support of $Pr(D = 1|X, Z)$? Plot the density of $Pr(D = 1|X, Z)$.

   (c) Estimate the parameters of (A), (B) and (C) by maximum likelihood. Write out the likelihood and discuss what parameters are identified and what parameters are not.

   (d) In terms of the model, write out the expressions for

$$E(Y|D = 1, X, Z)$$

and

$$E(Y|D = 0, X, Z)$$

   Estimate the parameters of the model using regression analysis (use your estimates of $Pr(D = 1|X, Z)$ from (a)). What parameters of the full model are identified from these regressions?

   (e) Express $E(U_1|D = 1, X, Z)$ as a function of $Pr(D = 1|X, Z)$. Express $E(U_0|D = 0, X, Z)$ as a function of $Pr(D = 1|X, Z)$.

- What are the estimates of these parameters? How do they compare with the estimates you obtain from Part (c).
- Plot these expressions as functions of $Pr(D = 1|X, Z)$, for X set at the sample mean.

(f) Estimate ATE using the "propensity score" you derived in (a). Plot the ATE as a function of the propensity score for each sample. What do these plots tell you about the appropriateness of the matching assumption for each sample?

(g) Estimate ATE using a regression of Y on P: X, Z. Plot your estimate against P.

3. [20 pts] Using your estimates from MLE, regression and matching just obtained, compute:

(a) The MTE as a function of $V = -[U_1 - U_0 - U_Z]$.

(b) The policy relevant treatment effect for a 10% upward shift in all arguments of Z.

(c) Estimate β from all 3 samples using IV (i) using Z_1 as an instrument, (ii) using Z_2 as an instrument and (iii) using $Pr(D = 1|X, Z)$ as an instrument. Interpret your estimates.

(d) Compare your estimate LATE using $Pr(D = 1|X, Z)$ as an instrument with the derivation of $E(Y_1 - Y_0|X, Z)$ formed from your answer to (d) and (e) above.

4. [5 pts] Suppose that

$$\gamma \sim N\left(\bar{\gamma}, \Sigma_\gamma\right)$$

$$\gamma \perp\!\!\!\perp (U_0, U_1, U_Z, X, Z), Z \perp\!\!\!\perp (U_1 - U_0 - U_Z)$$

Does the Imbens-Angrist monotonicity condition hold? Prove or disprove.