Problem Set 3
Due Tuesday, December 1 at 11:59 AM (noon) via Canvas
ECON 31720, University of Chicago, Fall 2020

Assignments must be typeset in nicely formatted LaTeX. Programming assignments must use low-level commands, be commented clearly (not excessively), formatted nicely in 80 characters per column, etc. **You will be graded on the exposition of your written answers, the clarity of your code, and the interpretability and beauty of your tables and graphics.** The problem sets are individual assignments, but you may discuss them with your classmates. Submit the problem sets through Canvas in a single zip/tar/rar file. **Late problem sets will not be accepted under any circumstances.**

1. Consider the nonparametric Roy model

$$\begin{aligned} Y &= DY(1) + (1 - D)Y(0) \\ D &= \mathbb{1}[U \leq p(Z)], \end{aligned}$$

where $(Y(0), Y(1), U)$ are unobserved and assumed to be independent of Z, and $p(z) \equiv \mathbb{P}[D = 1|Z = z]$ is the propensity score. Suppose that U is continuously distributed and that it has been normalized to have a uniform distribution over $[0, 1]$. Suppose that $\mathbb{E}[Y(d)^2]$ exists for both $d = 0, 1$. Define the marginal treatment effect (MTE) as $m(u) \equiv \mathbb{E}[Y(1) - Y(0)|U = u]$ for any $u \in [0, 1]$.

   (a) Show that the average treatment on the untreated (ATU) can be written as the following weighted average of the MTE function:

$$\text{ATU} \equiv \mathbb{E}[Y(1) - Y(0)|D = 0] = \int_0^1 m(u) \frac{\mathbb{P}[p(Z) < u]}{\mathbb{P}[D = 0]} \, du.$$

   (b) Show that the Wald estimand from a contrast between two instrument values z_0 and z_1 with $p(z_1) > p(z_0)$ can be written as the following weighted average of the MTE function:

$$\frac{\mathbb{E}[Y|Z = z_1] - \mathbb{E}[Y|Z = z_0]}{\mathbb{E}[D|Z = z_1] - \mathbb{E}[D|Z = z_0]} = \int_0^1 m(u) \left(\frac{\mathbb{1}[p(z_0) < u \leq p(z_1)]}{p(z_1) - p(z_0)} \right) du.$$

   Explain the connection with what we know about the local average treatment effect (LATE).

   (c) Let $D^\star \equiv \mathbb{1}[U \leq p^\star(Z^\star)]$ and $Y^\star \equiv D^\star Y(1) + (1 - D^\star)Y(0)$ for some function $p^\star$, where $Z^\star \perp\!\!\!\perp (Y(0), Y(1), U)$. Show that the policy-relevant treatment effect (PRTE) can be written as the following weighted average of the MTE function:

$$\text{PRTE} \equiv \frac{\mathbb{E}[Y^\star] - \mathbb{E}[Y]}{\mathbb{E}[D^\star] - \mathbb{E}[D]} = \int_0^1 m(u) \left(\frac{F_P^-(u) - F_{P^\star}^-(u)}{\mathbb{E}[p^\star(Z^\star)] - \mathbb{E}[p(Z)]} \right) du,$$

   where $F_P^-(u) \equiv \lim_{v \uparrow u} F_P(v)$, and similarly for $F_{P^\star}^-(u)$.

(d) Define the marginal treatment response (MTR) function as $m(d|u) \equiv \mathbb{E}[Y(d)|U = u]$. Let s be a function of D and Z. Show that

$$\mathbb{E}[s(D,Z)Y] = \int_0^1 m(1|u) \times \mathbb{E}\left[s(1,Z)|P \geq u\right](1 - F_P^-(u))\,du + \int_0^1 m(0|u) \times \mathbb{E}\left[s(0,Z)|P < u\right] F_P^-(u)\,du,$$

where (as usual) $P \equiv p(Z)$ is the propensity score viewed as a random variable. *Minor mathematical note: If $\mathbb{P}[P \geq u] = 0$ or $\mathbb{P}[P < u] = 0$ then, strictly speaking, the conditional expectations are not well-defined. However, in either case they are multiplied by 0, so we can safely ignore this little detail.*

(e) Construct an example under which the two-stage least squares (TSLS) estimand is a weighted average of $m(u)$ with weights that are negative for some values of u.

2. Suppose that $D \in \{0,1\}$ is a binary treatment and let $(Y(0), Y(1))$ be the potential outcomes associated with D. Assume that D is determined by

$$D = \mathbb{1}[U \leq p(Z)],$$

where $p(z) \equiv \mathbb{P}[D = 1|Z = z]$ is the propensity score, U is an unobservable random variable that is distributed uniformly over $[0,1]$, and $Z \in \{0,1\}$ is a binary instrument. Assume that $(Y(0), Y(1), U) \perp\!\!\!\perp Z$, and that $p(1) > p(0)$.

(a) Suppose that we also assume that the marginal treatment response functions are linear, i.e.

$$\mathbb{E}[Y(d)|U = u] = \alpha_d + \beta_d u \quad \text{for } d = 0,1.$$

Show that there is no unobserved heterogeneity in the causal effect of D on Y (i.e. the MTE function is a constant function of u) if and only if

$$\begin{aligned}&\mathbb{E}[Y|D = 1, Z = 1] - \mathbb{E}[Y|D = 1, Z = 0] \\ &\quad = \mathbb{E}[Y|D = 0, Z = 1] - \mathbb{E}[Y|D = 0, Z = 0].\end{aligned}$$

(b) Suppose that we continue to assume that the MTR functions are linear as in (a). We regress Y on $p(Z)$ for each of the subpopulations $d = 0,1$, to point identify α_d and β_d for $d = 0,1$, and we use these quantities to construct an implied LATE using the interpretation of the LATE as a weighted average of the MTE functions. Show that this implied LATE is always equal to the Wald estimand, even if our assumption that the MTR functions are linear is not actually correct.

(c) Explain the significance of the finding in part (b) for using a linear MTR model.

3. Suppose that $D \in \{0,1\}$ is a binary treatment, and Y is a continuously distributed outcome with potential outcomes $Y(0), Y(1)$. Let R be another continuously distributed

observed random variable and suppose that $D = \mathbb{1}[R \geq c]$. Suppose there is an unobserved binary variable $M \in \{0, 1\}$ with the property that $\mathbb{P}[R \geq c | M = 1] = 1$. Let $f_{R|M}(r|m)$ denote the density of R conditional on M. Maintain the following assumptions:

- $f_{R|M}(r|0)$ is continuous at $r = c$.
- $f_{R|M}(r|1)$ is continuous from the right at $r = c$, that is, $\lim_{r \downarrow c} f_{R|M}(r|1) = f_{R|M}(c|1)$.
- $f_R(c) \neq 0$.
- $\mathbb{E}[Y(0)|R = r, M = 0]$ is continuous at $r = c$.
- The distribution of Y given r is continuous from the right at $r = c$, that is, $\lim_{r \downarrow c} \mathbb{P}[Y \leq y | R = r] = \mathbb{P}[Y \leq y | R = c]$ for all y.

(a) Show that $\pi \equiv \mathbb{P}[M = 1 | R = c]$ is point identified.

(b) Derive sharp bounds on $\delta \equiv \mathbb{E}[Y(1) - Y(0) | R = c, M = 0]$.
Hint: Apply a result from a previous problem set.

(c) Explain how the assumptions and the result in part b) differ from the usual sharp regression discontinuity design framework.

4. Let Y be an outcome, $D \in \{0, 1\}$ a binary treatment, $Y(0), Y(1)$ potential outcomes, and R a running variable with interval support. Let $C \in \{c_\ell, c_h\}$ be a binary random variable, where c_ℓ, c_h are two known values in the support of R with $c_\ell < c_h$. Suppose that treatment is determined by

$$D = \mathbb{1}[R \geq C].$$

Assume that

$$\mathbb{E}[Y(d)|R = r, C = c] = g_d(r) + \alpha_d \mathbb{1}[c = c_h]$$

for all d, r, and c, where $g_d(r)$ is a continuous, unknown function of r, and α_d is an unknown scalar parameter. The observed data is (Y, D, R, C).

Show that $\mathbb{E}[Y(1) - Y(0) | R = r]$ is point identified for all $r \in [c_\ell, c_h]$. Provide an intuitive explanation of the role of the maintained assumptions in establishing this result and how both the assumptions and result differ from the usual identification argument for sharp regression discontinuity designs.

5. This problem involves the data used in "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" by Angrist and Evans (1998). The data is available on Canvas. (Note that the data has five more observations than reported in the paper. The data comes from Angrist's website, so I'm not sure what the source of this discrepancy is, but it shouldn't matter.)

For each case, the outcome is Y is *worked for pay*, the treatment D is *more than 2 children*, and the covariates X are *age*, *age at first birth*, *ages of the first two children in quarters*, *boy 1st*, *boy 2nd*, *black*, *hispanic*, *other race*. Let $Y(0), Y(1)$ denote potential outcomes, and let Z be the instrument, which varies in the parts below. Assume, as usual, that $D = \mathbb{1}[U \leq p(X, Z)]$, with p the propensity score and the usual exogeneity conditions and normalization. Let $m(d|u, x) \equiv \mathbb{E}[Y(d)|U = u, X = x]$ denote the MTR.

(a) Let Z be *same sex*. Estimate m for each of the following specifications. Use those estimates to estimate the ATE, ATT, ATU, and the (unconditional) LATE for *same sex*. Compare your results to the TSLS estimator with D as the endogenous variable, Z as excluded instrument, and X as controls (included instruments). Present your results in an easy-to-read-table. Also, construct an estimator of $m(u) \equiv \mathbb{E}[Y(1) - Y(0)|U = u]$, and plot it as a function of u for each specification.

   i. $m(d|u, x) = \alpha_d + \beta_d u + \gamma_d' x$.

   ii. $m(d|u, x) = \alpha_d + \beta_d u + \gamma' x$, where γ does not depend on d.

   iii. $m(d|u, x) = \alpha_d + \beta_d u + \gamma_d' x + \delta_d' xu$.

   iv. $m(d|u, x) = \alpha_d + \beta_{d1} u + \beta_{d2} u^2 + \gamma_d' x$.

   v. $m(d|u, x) = \alpha_d + \beta_{d1} u + \beta_{d2} u^2 + \beta_{d3} u^3 + \gamma_d' x$.

(b) Repeat part a) with Z as *twins*. Note any additional challenges or problems that arise with this instrument.

(c) Repeat part a) with Z as both *same sex* and *twins*.

6. Read "Identification and Extrapolation of Causal Effects with Instrumental Variables" (2018, *Annual Review of Economics*) by Mogstad and Torgovitsky. Reproduce Figure 6.

Note #1: These are numerical simulations in which nothing is being estimated. So, the propensity score, as well as all moments of (Y, D, Z), can be taken to be known, but should be computed to be consistent with the data generating process indicated in Section 2.4 of the paper.

Note #2: It is not clear from the discussion in the paper, but these are Bernstein polynomials. You may find it helpful to read Section S.2 of the supplemental appendix of Mogstad, Santos and Torgovitsky (2018), which discusses how Bernstein polynomials can be constrained to be bounded and/or monotone.

Note #3: Completing this problem will require you to solve linear programs. There are many open-source solvers, and any one of them will be sufficient for this problem. However, I encourage you to use a commercial solver like Gurobi or CPLEX, which are ***much*** *faster and provided at no cost to academic researchers. Gurobi in particular is an extremely impressive piece of software that you may find interesting and/or useful in your future research.*