ECON 31720: Problem Set 3

Nadia Lucas

December 1, 2020

Problem 1

Consider the nonparametric Roy model

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

Define MTE as $m(u) \equiv E[Y(1) - Y(0) \mid U = u]$ for any $u \in [0, 1]$.

a

$$ATU \equiv E[Y(1) - Y(0) \mid D = 0]$$

By LIE

$$= E[E[Y(1) - Y(0) \mid D = 0, U = u]]$$
$$= \int_0^1 E[Y(1) - Y(0) \mid D = 0, U = u]du$$

Using Bayes' rule for conditional expectation

$$= \int_0^1 E[Y(1) - Y(0) \mid U = u]\frac{P(D = 0 \mid U = u)}{P(D = 0)}du$$

Using the fact that $P(D = 0 \mid U) = P(p(Z) < U)$

$$= \int_0^1 m(u)\frac{P(p(Z) < U)}{P(D = 0}du$$

b

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:

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

Intuitively, the weights capture the complier group between the two instrument values. Focusing on the numerator we can break the Ys into the potential outcomes using the law of total probability

$$\begin{aligned}E[Y \mid Z = z_1] &= p(z_1)E[Y(1) \mid D = 1, Z = z_1] + (1 - p(z_1))E[Y(0) \mid D = 0, Z = z_1] \\ &= p(z_1)E[Y(1) \mid U \le p(z_1)] + (1 - p(z_1))E[Y(0) \mid U > p(z_1)]\end{aligned}$$

Analogously

$$E[Y \mid Z = z_0] = p(z_0)E[Y(1) \mid U \le p(z_0)] + (1 - p(z_0))E[Y(0) \mid U > p(z_0)]$$

Now differencing $E[Y \mid Z = z_1] - E[Y \mid Z = z_0]$ and using that $p(z_1) > p(z_0)$ we have,

$$\begin{aligned}E[Y \mid Z = z_1] - E[Y \mid Z = z_0] =& p(z_1)E[Y(1) \mid U \le p(z_1)] + (1 - p(z_1))E[Y(0) \mid U > p(z_1)] \\ & - p(z_0)E[Y(1) \mid U \le p(z_0)] - (1 - p(z_0))E[Y(0) \mid U > p(z_0)] \\ =& p(z_1)E[Y(1) \mid p(z_0) < U \le p(z_1)] + (p(z_1) - p(z_0))E[Y(1) \mid U \le p(z_0)] \\ & - (p(z_1) - p(z_0))E[Y(0) \mid U > p(z_1)] + p(z_0)E[Y(0) \mid p(z_0) < U \le p(z_1)] \\ =& (p(z_1) - p(z_0))E[Y(1) - Y(0) \mid p(z_0) < U \le p(z_1)] \\ =& (p(z_1) - p(z_0)) \int_0^1 m(u) \frac{\mathbb{1}(p(z_0) < U \le p(z_1))}{p(z_1) - p(z_0)}\end{aligned}$$

Now the denominator is $E[D \mid Z = z_1] - E[D \mid Z = z_0] = p(z_1) - p(z_0)$ by definition of propensity scores giving us exactly what we wanted.

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

c

Note that $F_P^-(u) \equiv lim_{v \uparrow u} F_P(v)$ implies that $P[u \le P(Z)] = 1 - F_P^-(u)$ which we use to simplify the PRTE.

$$\begin{aligned}PRTE \equiv& \frac{E[Y^*] - E[Y]}{E[D^*] - E[D]} \\ =& \frac{E[D^*Y(1) + (1 - D^*)Y(0)] - E[D(Y(1) + (1 - D)Y(0)]}{E[D^*] - E[D]} \\ =& \frac{E[(D^* - D)Y(1) - (D^* - D)Y(0)]}{E[D^*] - E[D]} \\ =& \frac{E[E[(D^* - D)Y(1) - (D^* - D)Y(0) \mid U = u]]}{E[D^*] - E[D]} \\ =& \frac{1}{E[p^*(Z^*)] - E[p(Z)]} \int_0^1 E[(D^* - D)Y(1) - (D^* - D)Y(0) \mid U = u] du \\ =& \frac{1}{E[p^*(Z^*)] - E[p(Z)]} \int_0^1 E[Y(1) - Y(0) \mid U = u](P[D^* = 1 \mid U = u] - P[D = 1 \mid U = u]) du \\ =& \frac{1}{E[p^*(Z^*)] - E[p(Z)]} \int_0^1 E[Y(1) - Y(0) \mid U = u](E[\mathbb{1}[u \le p^*(Z)] \mid U = u] - E[\mathbb{1}[u \le p(Z)] \mid U = u]) du \\ =& \frac{1}{E[p^*(Z^*)] - E[p(Z)]} \int_0^1 m(u)(1 - F_P^{-*}(u) - (1 - F_P^-(u)) du \\ =& \int_0^1 m(u) \frac{F_P^-(u) - F_P^{-*}(u)}{E[p^*(Z^*)] - E[p(Z)]} du\end{aligned}$$

d

Define MTR as $m(d \mid u) \equiv E[Y(d) \mid U = u]$. Show that

$$E[s(D,Z)Y] = \int_0^1 m(1 \mid u) \times E[s(1,Z) \mid P \geq u](1 - F_P^-(u))du$$
$$+ \int_0^1 m(0 \mid u) \times E[s(0,Z) \mid P < u]F_P^-(u)du$$

Starting with the LHS, we split up the Y into its potential outcomes and then take the LIE over U

$$\begin{aligned} E[s(D,Z)Y] &= E[s(D,Z) * (Y(1)\mathbb{1}(U \leq p(Z)) + s(D,Z) * Y(0)\mathbb{1}(U > p(Z))] \\ &= \int_0^1 E[s(D,Z) * (Y(1)\mathbb{1}(U \leq p(Z)) + s(D,Z) * Y(0)\mathbb{1}(U > p(Z)) \mid U = u]du \end{aligned}$$

Then split up the integral into two things that look very close to the MTRs and plug in the CDFs.

$$\begin{aligned} E[s(D,Z)Y] &= \int_0^1 E[s(D,Z) * (Y(1)\mathbb{1}(U \leq p(Z)) \mid U = u]du + \int_0^1 s(D,Z) * Y(0)\mathbb{1}(U > p(Z)) \mid U = u]du \\ &= \int_0^1 E[Y(1) \mid U = u]E[s(D,Z) * \mathbb{1}(U \leq p(Z) \mid U = u]du \\ &+ \int_0^1 E[Y(0) \mid U = u]E[s(D,Z)\mathbb{1}(U > p(Z)) \mid U = u]du \\ &= \int_0^1 m(1 \mid u)E[s(D,Z) * \mathbb{1}(U \leq p(Z) \mid U = u]du + \int_0^1 m(1 \mid u)E[s(D,Z)\mathbb{1}(U > p(Z)) \mid U = u]du \\ &= \int_0^1 m(1 \mid u)E[s(D,Z) \mid u \leq P](1 - F_P^-) + \int_0^1 m(1 \mid u)E[s(D,Z) \mid u > p(Z)]F_P^-(u)du \end{aligned}$$

which is the desired result.

e

An example under which the TSLS estimand is a weighted average of $m(u)$ with weights that are negative for some values of u. From Mogstad, Torgovitsky, and Walters (2019 Working Paper). Recall the TSLS estimand as a weighted average of $m(u)$

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

Recall we are given that $p(z_1) > p(z_0)$. When there are two treatments, the weights on any treatment 1 compliers will be non-negative because it is easier for the population to take up treatment under the treatment 1 regime. However, since there are two instruments, if there is a subset of the population that take up treatment under regime 0 and not regime 1, despite on average having $p(z_1) > p(z_0)$, these people will be given negative weights in the TSLS estimand. If we think of the weights not as an indicator function but as bounds of integration, we can see this is analogous to flipping the bounds for this subset of the population, resulting in a negative weight on those observed MTEs.

Problem 2

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

a

Suppose also that for $d = 0, 1$

$$E[Y(d) \mid U = u] = \alpha_d + \beta_d u$$

If there is no unobserved heterogeneity in the causal effect of D on Y, then the MTE function is a constant function of u. This is the case when $\beta_1 = \beta_0$

$$\begin{aligned} MTE(u) &= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)u \\ &= \alpha_1 - \alpha_0 \end{aligned}$$

Now we show that $\beta_1 = \beta_0$ if and only if

$$E[Y \mid D = 1, Z = 1] - E[Y \mid D = 1, Z = 0] = E[Y \mid D = 0, Z = 1] - E[Y \mid D = 0, Z = 0]$$

Deriving what each component is, we define each component with the potential outcomes framework and then plug in the MTR function given in the problem

$$\begin{aligned} E[Y \mid D = 1, Z = 1] &= E[Y(1) \mid D = 1, Z = 1] \\ &= E[Y(1) \mid u \leq p(1)] \\ &= \beta_1(E[u \mid u \leq p(1)] \\ &= \beta_1 * \frac{p(1)}{2} \end{aligned}$$

where the last equality follows from the uniform distribution of U. Similarly, extending this same logic to the rest of the components,

$$\begin{aligned} E[Y \mid D = 1, Z = 0] &= \beta_1 * \frac{p(0)}{2} \\ E[Y \mid D = 0, Z = 1] &= \beta_0 * \frac{p(1)}{2} \\ E[Y \mid D = 0, Z = 0] &= \beta_0 * \frac{p(0)}{2} \end{aligned}$$

Giving that

$$\begin{aligned} E[Y \mid D = 1, Z = 1] - E[Y \mid D = 1, Z = 0] &= \beta_1 \left(\frac{p(1) - p(0)}{2} \right) \\ E[Y \mid D = 0, Z = 1] - E[Y \mid D = 0, Z = 0] &= \beta_0 \left(\frac{p(1) - p(0)}{2} \right) \end{aligned}$$

Therefore

$\beta_1 - \beta_0 \implies E[Y \mid D = 1, Z = 1] - E[Y \mid D = 1, Z = 0] = E[Y \mid D = 0, Z = 1] - E[Y \mid D = 0, Z = 0]$

And given that $p(1) > p(0)$, this means that $\frac{p(1)-p(0)}{2} \neq 0$ so

$E[Y \mid D = 1, Z = 1] - E[Y \mid D = 1, Z = 0] = E[Y \mid D = 0, Z = 1] - E[Y \mid D = 0, Z = 0] \implies \beta_1 = \beta_0$

And the desired result that $MTE(u)$ is a constant function of u iff this conditions holds.

b

If we continue to assume that the MTR functions are linear as in (a). Regress Y on $p(Z)$ for each of the subpopulations $d = 0, 1$ to point identify α_d and β_d and use these to construct an implied LATE. The LATE rewritten as a weighted average of MTEs

$$\begin{aligned}
LATE &= \int_0^1 m(u) \frac{\mathbb{1}[p(0) < u \leq p(1)]}{p(1) - p(0)} du \\
&= \frac{1}{p(1) - p(0)} \int_{p(0)}^{p(1)} m(u) du \\
&= \frac{1}{p(1) - p(0)} \int_{p(0)}^{p(1)} E[Y(1) - Y(0) \mid U = u] du \\
&= \frac{1}{p(1) - p(0)} \int_{p(0)}^{p(1)} \alpha_1 - \alpha_0 + (\beta_1 - \beta_0) u du \\
&= \frac{1}{p(1) - p(0)} \left[(\alpha_1 - \alpha_0)(p(1) - p(0)) + (\beta_1 - \beta_0) \frac{p(1)^2 - p(0)^2}{2} \right] \\
&= (p(1) - p(0)) + (\beta_1 - \beta_0) \frac{p(1) + p(0)}{2}
\end{aligned}$$

Note the Wald given by problem 1

$$\frac{E[Y \mid Z = 1] - E[Y \mid Z = 0]}{E[D \mid Z = 1] - E[D \mid Z = 0]}$$

Plugging in that $p(1) = E[D \mid Z = 1]$ and $p(0) = E[D \mid Z = 0]$ toe the LATE as weighted MTE interpretation we have

$$LATE = \frac{1}{E[D \mid Z = 1] - E[D \mid Z = 0]} \left[(\alpha_1 - \alpha_0)(p(1) - p(0)) + (\beta_1 - \beta_0) \frac{p(1)^2 - p(0)^2}{2} \right]$$

So it will suffice to show that

$$E[Y \mid Z = 1] - E[Y \mid Z = 0] = (\alpha_1 - \alpha_0)(p(1) - p(0)) + (\beta_1 - \beta_0) \frac{p(1)^2 - p(0)^2}{2}$$

We can back out and show using our functional forms and the uniform distribution of the propensity scores and some nasty algebra

$$
\begin{aligned}
E[Y \mid Z=1] - E[Y \mid Z=0] &= p(1)E[Y \mid D=1, Z=1] + (1-p(1))E[Y \mid D=0, Z=1] \\
&\quad - p(0)E[Y \mid D=1, Z=1] - (1-p(0))E[Y \mid D=0, Z=1] \\
&= p(1)(\alpha_1 + \beta_1 \frac{p(1)}{2}) + (1-p(1))(\alpha_0 + \beta_0 \frac{1+p(1)}{2}) \\
&\quad - p(0)(\alpha_1 + \beta_1 \frac{p(0)}{2}) - (1-p(0)(\alpha_0 + \beta_0 \frac{1+p(0)}{2}) \\
&= (p(1)-p(0))(\alpha_1 - \alpha_0) + \beta_1 \frac{p(1)^2}{2} + \beta_0 \frac{1+p(1)}{2} - \beta_0 \frac{p(1)+p(1)^2}{2} \\
&\quad - \beta_1 \frac{p(0)^2}{2} - \beta_0 \frac{1+p(0)}{2} + \beta_0 \frac{p(0)+p(0)^2}{2} \\
&= (p(1)-p(0))(\alpha_1 - \alpha_0) + \beta_1 \left(\frac{p(1)^2 - p(0)^2}{2} \right) \\
&\quad + \beta_0 \left(\frac{p(0)^2 - p(1)^2}{2} \right) \\
&= (p(1)-p(0))(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \frac{p(1)^2 - p(0)^2}{2}
\end{aligned}
$$

Which is what we wanted. Going back, this gives us that

$$
\frac{E[Y \mid Z=1] - E[Y \mid Z=0]}{p(1)-p(0)} = \alpha_1 - \alpha_0 + (\beta_1 - \beta_0) \frac{p(1)+p(0)}{2}
$$

Our Wald is on the LHS and what we analytically derived for the LATE is on the right hand side.

c

This proof would generalize to any MTR function (not necessarily just linear ones) because we can always pull out $\frac{1}{p(1)-p(0)}$ from the LATE as weighted MTEs and from there we can express the numerator of the Wald as MTEs weighted by terms that depend on $p(1), p(0)$, not using the linearity assumption

Problem 3

Suppose $D \in \{0,1\}$ is a binary treatment and Y is a continuously distributed outcome with potential outcomes $Y(0)$ and $Y(1)$. R is another continuously distributed observed random variable and suppose $D = \mathbb{1}[R \geq c]$. There is an unobserved binary variable $M \in \{0,1\}$ with the property that $M = 1$ only above the cutoff. $f_{R|M}(r \mid 0)$ is continuous at $r = c$ and $f_{R|M}(r \mid 1)$ is right continuous at $r = c$.

a

Given the assumption that $f_{R|M}(r \mid 0)$ is continuous at $r = c$. We can observe the limit from the left of

$$
\lim_{r \uparrow c} f_{R|M}(r \mid 0)
$$

since only $M = 0$ is observed in the data to the left of $r = c$. And we know that if we all of a sudden have a jump at $f_R(c)$ then it must be due to $\pi = P[M = 1 \mid R = c]$. So

$$\pi = 1 - \frac{lim_{r \uparrow c} f_{R|M}(r \mid 0)}{f_R(c)}$$

b

Going back to problem set 1, question 2c, we have a set of bounds that tracks closely with this problem. We want to compute bounds on

$$\delta \equiv E[Y(1) - Y(0) \mid R = c, M = 0]$$

We know $E[Y(0) \mid R = c, M = 0]$ because we are given in the assumptions that $E[Y(0) \mid R = c, M = 0]$ is continuous so

$$\lim_{r \uparrow c} E[Y \mid R = r] = E[Y(0) \mid R = c, M = 0]$$

So we can observe that term from data. All that is left is to find the bounds on $E[Y(1) \mid R = c, M = 0]$. We have already found π which we can think of as always takers/(compliers + always takers) which is the same way π is defined in 2c. Thus tracking closely from 2c where $G^{-1}(q)$ is the q-th quantile distribution of $Y \mid M = 1, R = c$. Intuitively, we bound $E[Y(1) \mid R = c, M = 0]$ between the two extremes where $E[Y(1) \mid R = c, M = 1]$ are all below $E[Y(1) \mid R = c, M = 0]$ or all above $E[Y(1) \mid R = c, M = 0]$.

$$\begin{aligned} E[Y \mid R = c, Y \leq G^{-1}(1 - \pi)] - E[Y(0) \mid R = c, M = 0] \\ \leq \delta \leq E[Y \mid R = c, Y \geq G^{-1}(\pi)] - E[Y(0) \mid R = c, M = 0] \end{aligned}$$

These bounds are sharp because when $\pi = 0$, in other words, there are no observations with $M = 1$ messing with the measurements at the cutoff, then these bounds both exactly equal $E[Y(1) - Y(0) \mid R = c, M = 0]$ as measured at the cutoff as a sharp RD.

c

The assumptions and the result in part b differ from the usual sharp regression discontinuity framework because now there is unobserved heterogeneity in the composition of individuals on either side of the cutoff. Essentially, we no longer have continuity in the running variable so any estimate using the assumptions of continuity would be biased to include a composition effect of $M = 1$ individuals entering the data. Thus the result in b is adjusting for this composition effect and doesn't look the same as the usual sharp regression discontinuity design framework. The bounds reflect the uncertainty as to where in the distribution of $Y(1)$ potential outcomes, the $M = 1$ individuals lie.

Problem 4

Now we have a setting with two different cutoffs $C \in \{c_l, c_h\}$. We assume that

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

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.

$E[Y(1) - Y(0) \mid R = r]$ is point identified for all $r \in [c_l, c_h]$.

For $r < c_l$ below the lowest boundary, we have that

$$E[Y(0) \mid R = r, C = c_l] = g_0(r)$$

is point-identified as this is observed. And we have for $c_l \leq r \leq c_h$, given that we've solved for $g_0(c_l)$ already since we have a continuous function, we can get what α_0 is by looking at outcomes from $D = 0$ right at cutoff c_l and taking

$$E[Y(0) \mid R = c_l] - g_0(c_l) = \alpha_0$$

And then go on to construct

$$E[Y(0) \mid R = r, C = c_h] = g_0(r) + \alpha_0$$

Similarly, we can observe and construct from data between the two cutoffs

$$E[Y(1) \mid R = r, C = c_l] = g_1(r)$$

and then back out

$$E[Y(1) \mid R = c_h] - g_1(c_h) = \alpha_1$$

And use that to construct

$$E[Y(1) \mid R = r, C = c_h] = g_1(r) + \alpha_1$$

which is now all point-identified.

Intuitively, this is different from our usual sharp RD designs because there are two discontinuities. Thus we have some overlapping support for $g_0(r) + \alpha_0$ and $g_1(r)$. In a sense we observe the same segment of the data as treated in one setting and untreated in another so we have to use both cutoffs to back out what the relevant treatment effects are in each setting. But the method for backing out α_0 and α_1 is relatively similar to our usual sharp RD design which employs the continuity in the running variable.

Problem 5

Reported are the list of tables and then subsequent estimates of the MTE with the procedure detailed in the code. We see that the treatment effects all become substantially larger than the TSLS estimators across the board because by using the MTEs, we use the full support of individuals, not just those induced into treatment. For part b, where only the twins instrument is used, conceptually there is no such thing as a never taker, so those who do get treatment always comply so we see steeper MTE plots. The MTE plots are all evaluated using the mean of X.

Table 1: Children and Their Parents' Labor Supply: Various Parameters of Interest

Same sex as instrument: TSLS = -0.11825				
	ATE	ATT	ATU	LATE
Spec 1	-0.18905	-0.1963	-0.18417	-0.19634
Spec 2	-0.18986	-0.19673	-0.18524	-0.19672
Spec 3	-0.18671	-0.18923	-0.18501	-0.18764
Spec 4	-0.18733	-0.18878	-0.18635	-0.18851
Spec 5	-0.18656	-0.18841	-0.18532	-0.1881
Twins as instrument: TSLS = -0.08285				
	ATE	ATT	ATU	LATE
Spec 1	-0.18752	-0.18957	-0.18614	-0.18961
Spec 2	-0.19186	-0.19812	-0.18765	-0.19811
Spec 3	-0.18795	-0.18953	-0.18688	-0.18781
Spec 4	-0.18722	-0.18557	-0.18832	-0.1852
Spec 5	-0.18536	-0.18166	-0.18785	-0.18129
Same sex and twins as instrument: TSLS = -0.09029				
	ATE	ATT	ATU	LATE
Spec 1	-0.18967	-0.19448	-0.18644	-0.1945
Spec 2	-0.19214	-0.19846	-0.1879	-0.19845
Spec 3	-0.18856	-0.19044	-0.18731	-0.18881
Spec 4	-0.18727	-0.18565	-0.18835	-0.18531
Spec 5	-0.18644	-0.18398	-0.18809	-0.1836

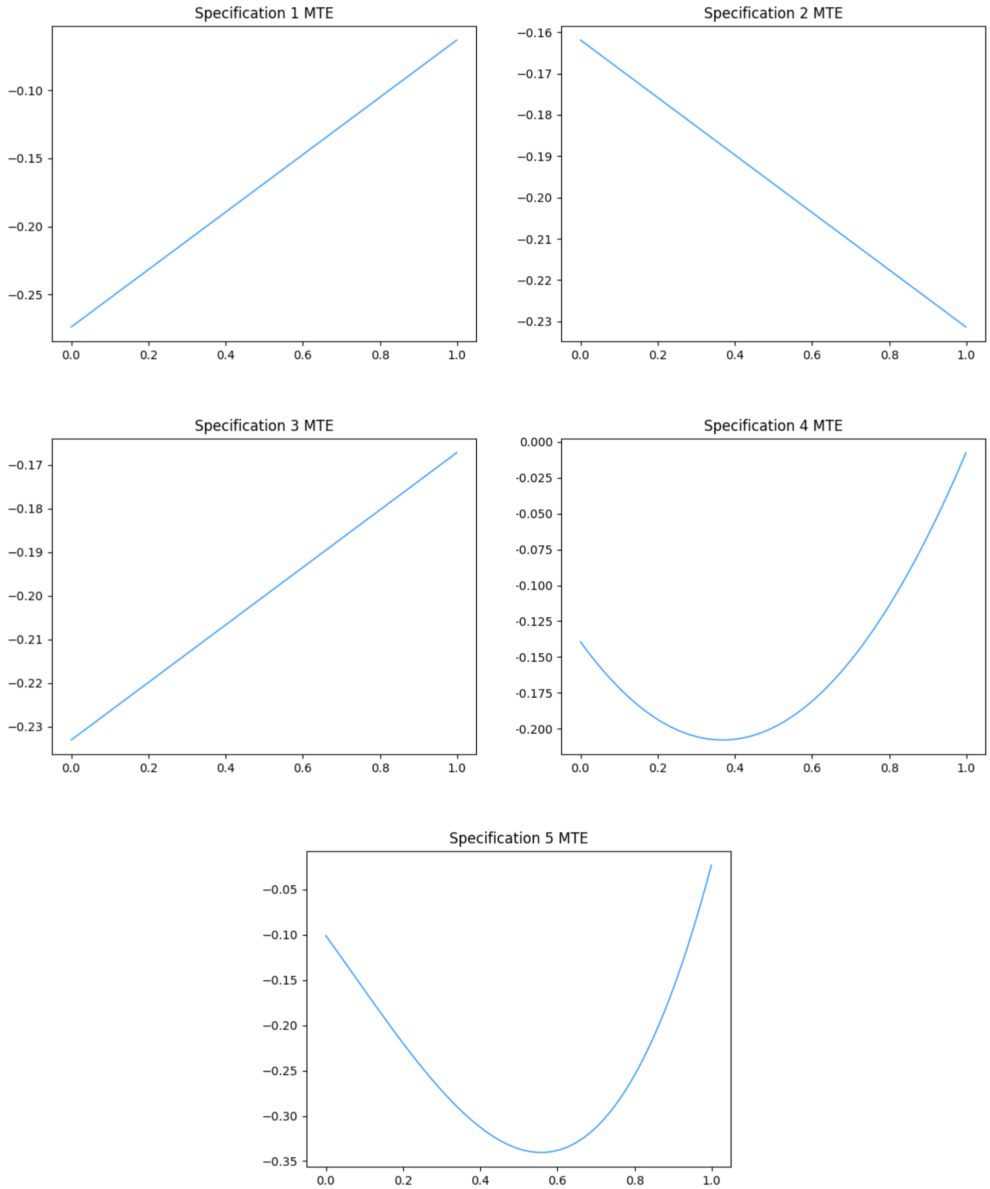

Figure 1: Z = *same sex*

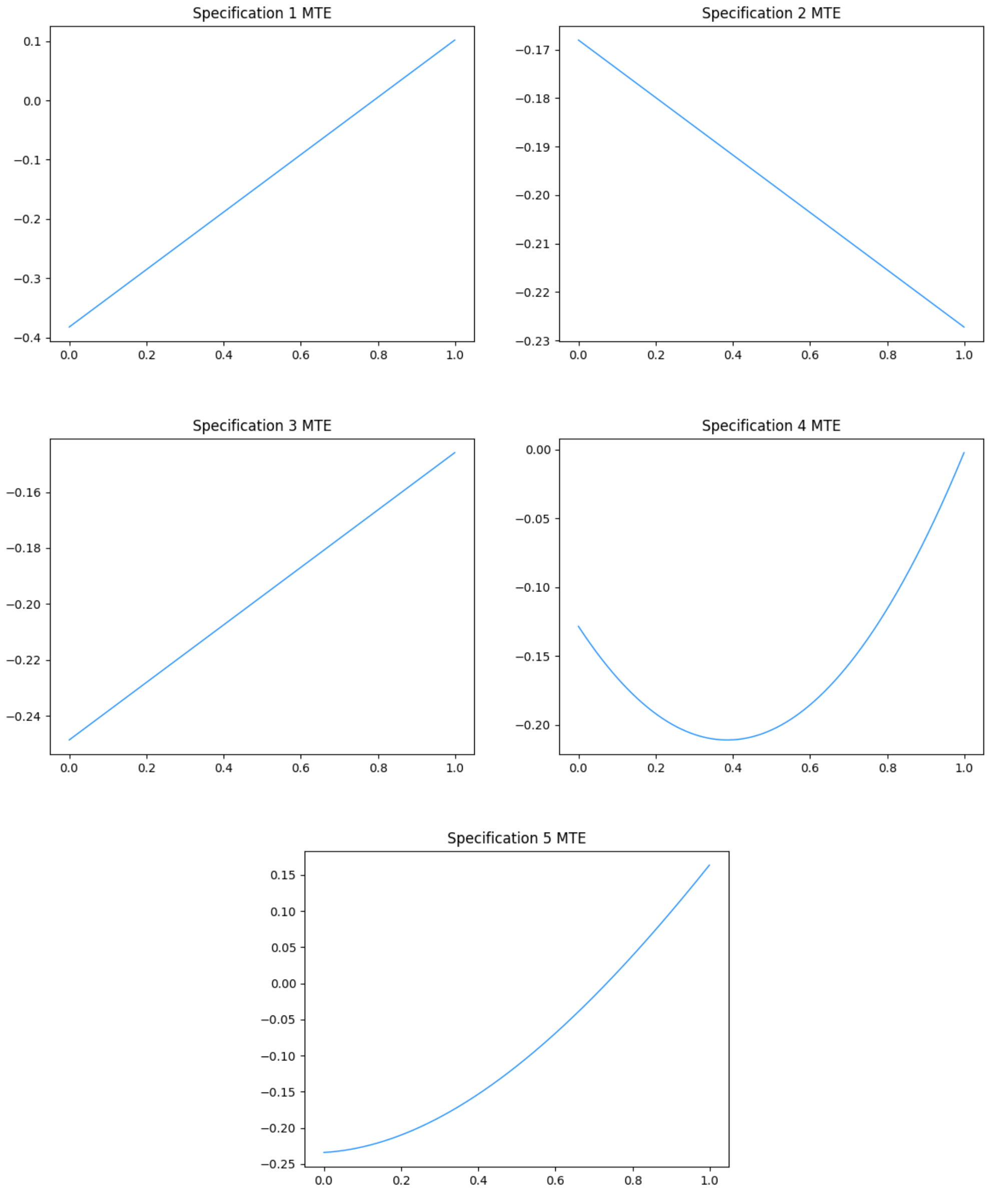

Figure 2: Z = *twins*

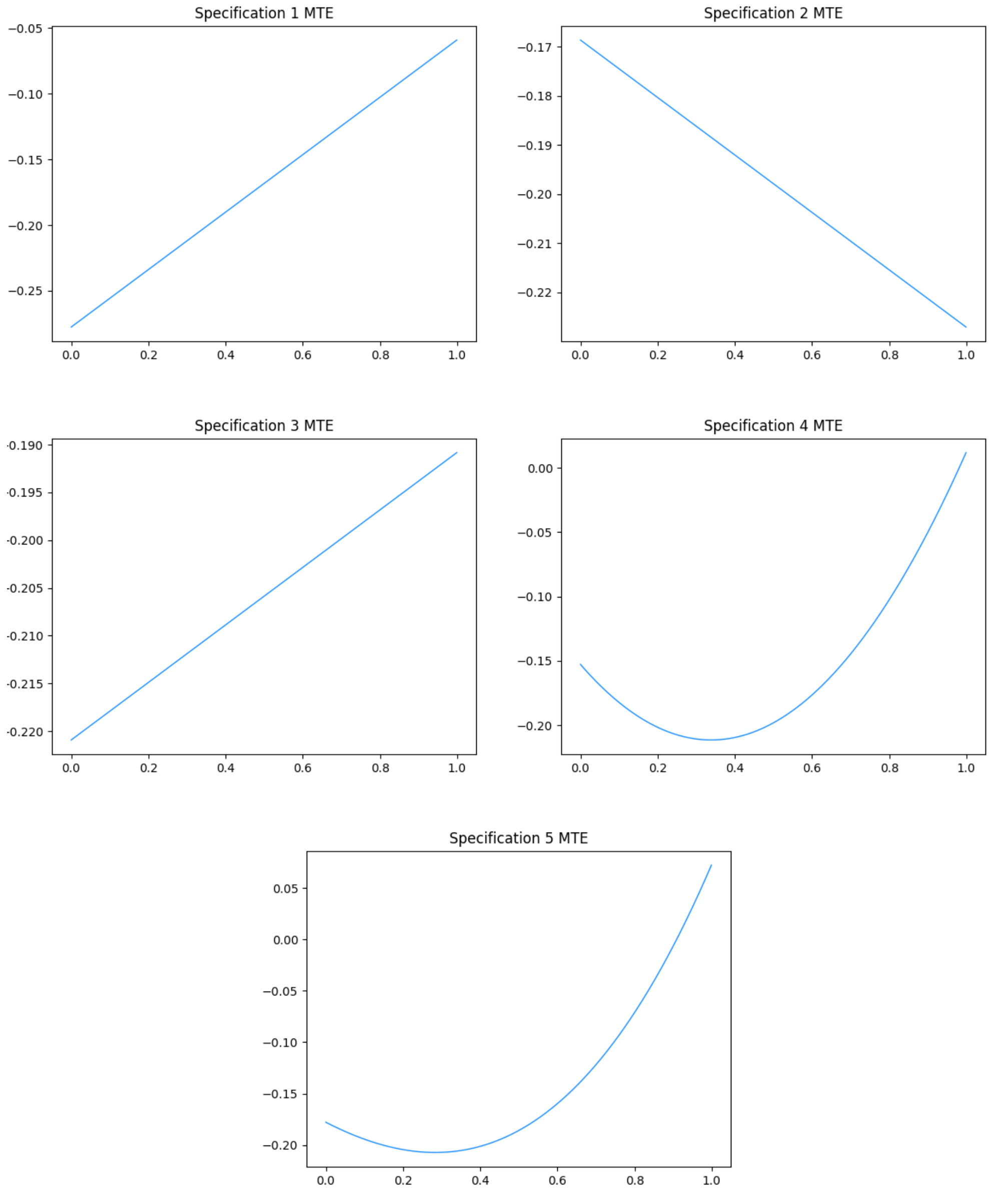

Figure 3: Z = *same sex, twins*

Problem 6

This procedure is just summarizing the paper but written out mostly for my own benefit. Computing the bounds, our objective is to solve

$$\underline{\beta}^* \equiv \inf_{m \in \mathcal{M}} \Gamma^*(m) \quad \text{subject to} \quad \Gamma_s(m) = \beta_s \quad \text{for all} \quad s \in \mathcal{S}$$
$$\overline{\beta}^* \equiv \sup_{m \in \mathcal{M}} \Gamma^*(m) \quad \text{subject to} \quad \Gamma_s(m) = \beta_s \quad \text{for all} \quad s \in \mathcal{S}$$

Working backwards from this we have

$$\Gamma^*(m) \equiv E\left[\int_0^1 m_0(u,X)\omega_0^*(u,X,Z)\mathrm{d}u\right] + E\left[\int_0^1 m_1(u,X)\omega_1^*(u,X,Z)\mathrm{d}u\right]$$

And for any IV-like specification s, we define

$$\Gamma_s(m) \equiv E\left[\int_0^1 m_0(u,X)\omega_{0s}(u,X,Z)\mathrm{d}u\right] + E\left[\int_0^1 m_1(u,X)\omega_{1s}(u,X,Z)\mathrm{d}u\right]$$

And the IV-like estimand that is a feature of the observable data

$$\beta_s = E\left[\int_0^1 m_0(u,X)\omega_{0s}(u,X,Z)\mathrm{d}u\right] + E\left[\int_0^1 m_1(u,X)\omega_{1s}(u,X,Z)\mathrm{d}u\right]$$

$$\begin{aligned}&\text{where } \omega_{0s}(u,x,z) \equiv s(0,x,z)\mathbb{1}[u > p(x,z)]\\ &\text{and} \quad \omega_{1s}(u,x,z) \equiv s(1,x,z)\mathbb{1}[u \leq p(x,z)]\end{aligned}$$

the weights can be generated using

$$s(d,x,z) = \frac{z - E[Z]}{\mathrm{Cov}(D,Z)}$$

This IV-like estimand is not equal to the target parameter and is thus not an object of interest. Only some MTR functions are consistent with a given value of an IV-like estimand. We can use this intuition to construct bounds on β^*. Our goal is to characterize the values of β^* that could be generated by MTR functions that are elements of $\mathcal{M}$ and that could also deliver the collection of identified IV estimands. The admissible set of MTR pairs $m \in \mathcal{M}$ must have enough structure. Impose that $\mathcal{M}$ is a finite dimensional linear space. For every $m \equiv (m_1, m_1) \in \mathcal{M}$, there exists a finite dimensional vector $\theta \equiv (\theta_0, \theta_1) \in \mathbb{R}^{K_0+K_1}$ such that

$$m_d(u,x) = \sum_{k=0}^{K_d} \theta_{dk} b_{dk}(u,x) \quad \text{ for } d = 0,1$$

where $b_{dk}(u,x)$ are known basis functions. Substituting into the original equation of interest, we have

$$\begin{aligned}\Gamma^*(m) &= \sum_{d\in[0,1\}} \sum_{k=0}^{K_d} \theta_{dk} E\left[\int_0^1 b_{dk}(u,X)\omega_d^*(u,X,Z)\mathrm{d}u\right]\\ &\equiv \sum_{d\in\{0,1\}} \sum_{k=0}^{K_d} \theta_{dk}\gamma_{dk}^*, \quad \text{ where } \gamma_{dk}^* \equiv E\left[\int_0^1 b_{dk}(u,X)\omega_d^*(u,X,Z)\mathrm{d}u\right]\end{aligned}$$

Imposing this, we now can turn our objective into a linear program

$$\Gamma_s(m) = \sum_{d\in\{0,1]} \sum_{k=0}^{K_d} \theta_{dk}\gamma_{sdk}, \quad \text{ where } \gamma_{sdk} \equiv E\left[\int_0^1 b_{dk}(u,X)\omega_{ds}(u,X,Z)\mathrm{d}u\right]$$

And finally, how we get our bounds are

$$\vec{\beta}^* = \max_{\theta\in\Theta} \sum_{d\in\{0,1|} \sum_{k=0}^{K_d} \gamma_{dk}^*\theta_{dk} \quad \text{ subject to } \quad \sum_{d\in\{0,1|} \sum_{k=0}^{K_d} \gamma_{sdk}\theta_{dk} = \beta_s \text{ for all } s \in \mathcal{S}$$

$$\underline{\beta}^* = \min_{\theta\in\Theta} \sum_{d\in\{0,1|} \sum_{k=0}^{K_d} \gamma_{dk}^*\theta_{dk} \quad \text{ subject to } \quad \sum_{d\in\{0,1|} \sum_{k=0}^{K_d} \gamma_{sdk}\theta_{dk} = \beta_s \text{ for all } s \in \mathcal{S}$$

We get β_s from the IV-slope and $s(d,z)$ from the weight of the IV-slope. This is

$$\beta_s = \frac{Cov(Y,Z)}{Cov(D,Z)}$$
$$s(d,z) = \frac{z - E[Z]}{Cov(D,Z)}$$

And we also have bounds from the TSLS-slope estimator which we can derive from the table which gives the j-th component of the estimand. Deriving this we start from

$$\Pi \equiv E\left[\tilde{X}\widetilde{Z}'\right] E\left[\widetilde{Z}\widetilde{Z}'\right]^{-1}$$

where

$$\widetilde{Z} = \begin{bmatrix} \mathbb{1}\{Z=1\} \\ \mathbb{1}\{Z=2\} \\ \mathbb{1}\{Z=3\} \\ \mathbb{1}\{Z=4\} \end{bmatrix}'$$

and

$$\widetilde{X} = \begin{bmatrix} 1 & D \end{bmatrix}$$

From the j-th component we have from the paper that

$$\beta_s^j = e_j'(\Pi E[\widetilde{Z}\widetilde{X}])^{-1}\Pi E[\widetilde{Z}Y]$$
$$s(d,z)^j = e_j'(\Pi E[\widetilde{Z}\widetilde{X}])^{-1}\Pi \widetilde{Z}$$

And now aggregating that to all components we can see

$$\beta_s = (\Pi E[\widetilde{Z}\widetilde{X}])^{-1}\Pi E[\widetilde{Z}Y]$$
$$s(d,z) = (\Pi E[\widetilde{Z}\widetilde{X}])^{-1}\Pi \widetilde{Z})$$

We calculate β_s the same way that we would the IV estimand since we do not observe $E[\widetilde{Z}Y]$. The method for calculating the TSLS slope is equivalent to the IV slope, just different $s(d,z)$ weights.

Recall the IV-like estimand that is a feature of the observable data

$$\beta_s = E\left[\int_0^1 m_0(u,X)\omega_{0s}(u,X,Z)\mathrm{d}u\right] + E\left[\int_0^1 m_1(u,X)\omega_{1s}(u,X,Z)\mathrm{d}u\right]$$
$$\text{where } \omega_{0s}(u,x,z) \equiv s(0,x,z)\mathbb{1}[u > p(x,z)]$$
$$\text{and } \quad \omega_{1s}(u,x,z) \equiv s(1,x,z)\mathbb{1}[u \leq p(x,z)]$$

The two IV-like estimands give the constraints and the ATT gives the weights for the maximization/minimization problem.

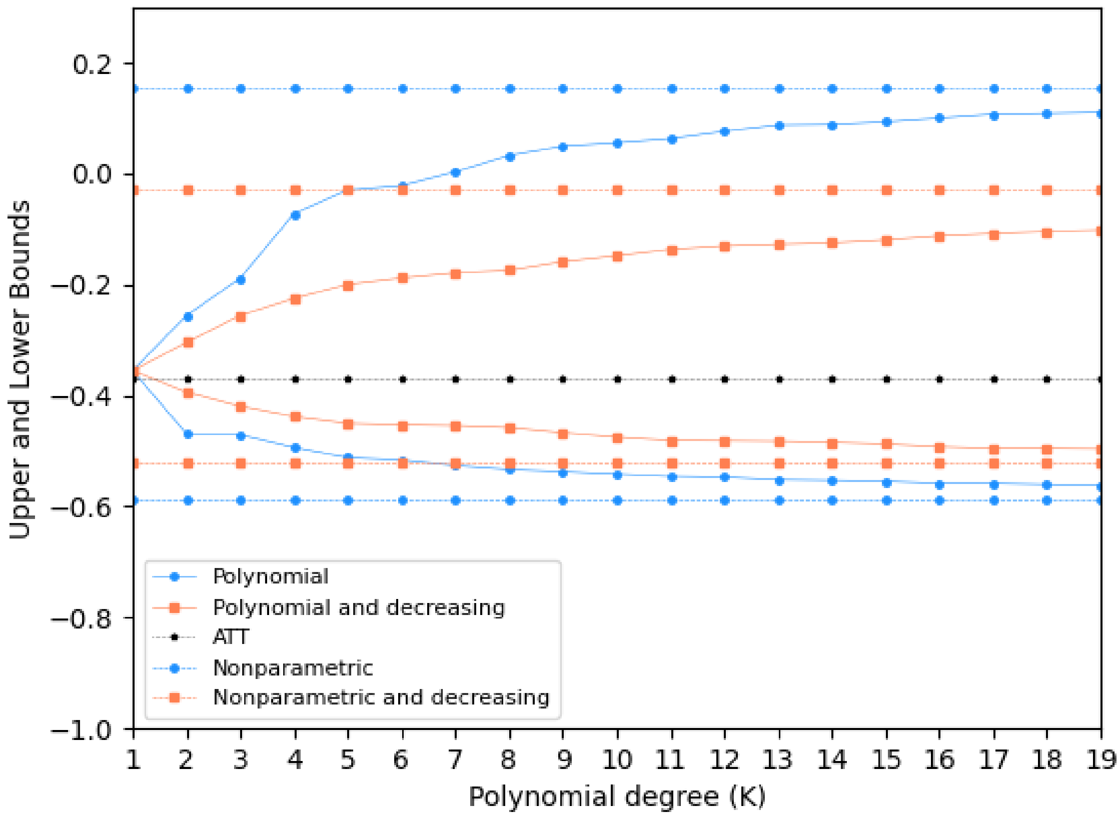

Figure 4: Figure 6 replication