Microeconometrics - Pset3

Jeanne Sorin

December 1, 2020

Problem 1

(a) We have

$$\begin{aligned}
ATU = E[Y(1) - Y(0)|D = 0] &= E[(Y(1) - Y(0)\frac{\mathbf{1}\{U > p(Z)\}}{P(D = 0)}] \\
&= E[E[Y(1) - Y(0)|Z, U]\frac{\mathbf{1}\{U > p(Z)\}}{P(D = 0)}] && LIE \\
&= E[m(u)\frac{\mathbf{1}\{U > p(Z)\}}{P(D = 0)}] && Y(1), Y(0), U \perp Z \\
&= E[\int_0^1 m(u)\frac{\mathbf{1}\{U > p(Z)\}}{P(D = 0)}du] && LIE \\
&= \int_0^1 m(u)\frac{P(p(z) < u}{P(D = 0)}du
\end{aligned}$$

(b) We have

$$\frac{E[Y|Z = z_1] - E[Y|Z = z_0]}{E[D|Z = z_1] - E[D|Z = z_0]}$$

Look at the numerator

$$\begin{aligned}
E[Y|Z = z_1] - E[Y|Z = z_0] =& E[Y(1)D + Y(0)(1 - D)|Z = z_1] - \\
& E[Y(1)D + Y(0)(1 - D)|Z = z_0] && \text{potential outcome def} \\
=& E[Y(1)D|Z = z_1] - E(Y(1)D|Z = z_0) + \\
& E(Y(0)(1 - D)|Z = z_1) - E(Y(0)(1 - D)|Z = z_0) \\
=& E[Y(1)|D(z_1) = 1, D(z_0) = 0] - \\
& E[Y(0)|D(z_1) = 1, D(z_0) = 0] \\
=& E[Y(1) - Y(0)|D(z_1) = 1, D(z_0) = 0] \\
=& E[Y(1) - Y(0)|p(z_0) < u < p(z_1)] && \text{definition of D} \\
=& \int_u \underbrace{E[Y(1) - Y(0)|U = u]}_{m(u)} \mathbf{1}[p(z_0) < u \leq p(z_1)]du && \text{Bayes Rule and cond inde of Z}
\end{aligned}$$

The denominator

$$\begin{aligned}
E[D|Z = z_1] - E[D|Z = z_0] &= P(D = 1|Z = z_1) - P(D = 1|Z = z_0) && \text{D is 0 or 1} \\
&= p(z_1) - p(z_0) && \text{definition of p}
\end{aligned}$$

Putting the two together gives us the desired result

$$E[Y|Z = z_1] - E[Y|Z = z_0] = \int_u \underbrace{E[Y(1) - Y(0)|U = u]}_{m(u)} \frac{\mathbf{1}[p(z_0) < u \leq p(z_1)]}{p(z_1) - p(z_0)}du$$

This estimator for the nonparametric Roy Model is equivalent to the LATE, i.e. the average MTE over the compliers, those who take the treatment under z_1 but not under z_0. As $z_1 \to z_0$, the LATE $\to$ the MTE. Depending on z, the LATE allows us to identify different parts of the MTE.

(c) We have

$$PRTE = \frac{E[Y^*] - E[Y]}{E[D^*] - E[D]}$$

The numerator

$$\begin{aligned}
E[Y] &= E[E(Y|P)] && LIE \\
&= \int_0^1 E[Y|P=m] f_p(m) dm \\
&\text{(Expectation is integral)} \\
&= \int_0^1 E[Y(1)\mathbf{1}(u \leq m) + Y(0)\mathbf{1}(u > m)|U=u] f_p(m) dm \\
&\text{(Potential outcomes def)} \\
&= \int_0^1 (E[Y(1)\mathbf{1}(u \leq p(z))|P=u] + E[Y(0)\mathbf{1}(u > p(z))|P=u]) f_p(u) du \\
&= \int_0^1 ((\int_0^1 \mathbf{1}(u \leq m) E[Y(1)|U=u] + \mathbf{1}(u > m) E[Y(0)|U=u]) du) f_p(m) dm \\
&\text{(conditional independence of P)} \\
&= \int_0^1 ((\underbrace{\int_0^1 \mathbf{1}(u > m) F_p(v) E[Y(1)|U=u] f_p(m) dm}_{[1-\lim_{v\uparrow u} F_p(v)]E(Y(1)|U=u)} + \underbrace{v\mathbf{1}(u < m) E[Y(0)|U=u]) f_p(m) dm}_{\lim_{v\uparrow u} F_p(v) \times E[Y(0)|U=u]}) du \text{ Fubini} \\
&= \int_0^1 ([1 - \lim_{v\uparrow u} F_p(v)] \times E(Y(1)|U=u) + \lim_{v\uparrow u} F_p(v) \times E[Y(0)|U=u]) du
\end{aligned}$$

Similarly for $E[Y^*]$.
The denominator

$$\begin{aligned}
E[D] &= E[D|Z] && \text{LIE and conditional independence of Z} \\
&= E(p(z)) && \text{U is uniform on [0,1]} \\
E[D^*] &= E(p(Z^*))
\end{aligned}$$

Combining the two and simplifying gives the desired result

$$\frac{E[Y^*] - E[Y]}{E[D^*] - E[D]} = \int_0^1 m(u) \frac{F_P^-(u) - F_{P^*}^-(u)}{E(p^*(Z^*)) - E(^(Z))} du$$

(d) We have

$$\begin{aligned}
E[s(D,Z)Y] &= \int_0^1 E[s(D,Z)Y|P=m]dF_p(m) && LIE \\
&= \int_0^1 [\int_0^1 \mathbf{1}(u \leq m)E[s(1,Z)|P=m]E[Y(1)|U=u]du + && \text{Conditional Index \&} \\
&\quad \int_0^1 \mathbf{1}(u > m)E[s(0,Z)|P=m]E[Y(0)|U=u]du]dF_p(m) && LIE \\
&= \int_0^1 [\int_0^1 \mathbf{1}(u > m)E[s(1,Z)|P=m]E[Y(1)|U=u]dF_p(m) + && \\
&\quad \int_0^1 \mathbf{1}(u \leq m)E[s(0,Z)|P=m]E[Y(0)|U=u]dF_p(m)]du && \text{Fubini} \\
&= \int_0^1 [E[s(1,Z)|P \geq v]m(1|v)(1-F_p^-(v) + E[s(1,Z)|P < v]m(0|v)(F_p^-(v)]dv && \text{From part c} \\
&= \int_0^1 m(1|u)(1-F_p^-(u)E[s(1,Z)|P \geq u]du + \int_0^1 m(0|u)F_p^-(u)E[s(0,Z)|P > u]du
\end{aligned}$$

(e) Construct an example under which the TSLS estimand is a weighted average of m(u) with weights that are negative for some values of u.
We can get negative weights if the instrument is the instruments are ordered in a way that is "at odds" with what we usuall do: Following Mogstad, Torgovitsky, Walter (2018), let's have 2 instruments (Z_1, Z_2) such that

$$P(D_{Z_1=1}=1, D_{Z_1=0}=0) = \pi_{1,compliers}$$
$$P(D_{Z_2=1}=1, D_{Z_2=0}=0) = \pi_{2,compliers}$$
$$\pi_{1,compliers} > \pi_{2,compliers}$$

Then, according to the paper, the sign of the weight on Z_2 compliers, i.e. the TSLS weights :

$$\text{sign}(\omega_{2c} = \mathbf{1}[\pi_{2,compliers} > 0] = \underbrace{\text{sign}[P(D_i=1|Z_{i,2}=1) - P(D_i=1|Z_{i2}=0)]}_{\text{potentially (-) if } P(D_i=1|Z_{i,2}=1)<P(D_i=1|Z_{i2}=0)]}$$

Problem 2

(a) We have, by definition

$$\begin{aligned} MTE &= E[Y(1) - Y(0)|U = u] \\ &= E[Y(1)|U = u] - E[Y(0)|U = u] \\ &= \alpha_1 - \alpha_0 + u(\beta_1 - \beta_0) && \text{linear MTR} \end{aligned}$$

For MTE to be constant, it must be that case that $\beta_1 = \beta_0$.
Besides, we have that

$$\begin{aligned} E[Y|D=1,Z=1] - E[Y|D=1,Z=0] &= E[Y(1)|D=1,Z=1] - E[Y(1)|D=1,Z=0] && \text{potential outcome} \\ &= E[Y(1)|u \le p(1)] - E[Y(1)|u \le p(0)] \\ &= \beta_1(E[u|u \le p(1)] - E[u|U \le (p(0))]) && \text{linear MTR} \\ &= \beta_1 \frac{p(1)}{2} - \beta_1 \frac{p(0)}{2} \\ &= \beta_1 \frac{p(1) - p(0)}{2} \end{aligned}$$

Similarly

$$\begin{aligned} E[Y|D=0,Z=1] - E[Y|D=0,Z=0] &= E[Y(0)|D=0,Z=1] - E[Y(0)|D=0,Z=0] && \text{potential outcome} \\ &= E[Y(0)|u > p(1)] - E[Y(0)|u > p(0)] \\ &= \beta_0(E[u|u > p(1)] - E[u|U > (p(0))]) && \text{linear MTR} \\ &= \beta_0 \frac{p(1) - p(0)}{2} \end{aligned}$$

Thus

$$\begin{aligned} &E[Y|D=1,Z=1] - E[Y|D=1,Z=0] = E[Y|D=0,Z=1] - E[Y|D=0,Z=0] \\ \Leftrightarrow &\beta_1 = \beta_0 \\ \Leftrightarrow &MTE(u) = \alpha_1 - \alpha_0 \end{aligned}$$

(b) The Wald estimate is

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

The numerator:

$$\begin{aligned} E[Y|Z=1] - E[Y|Z=0] = p(1)E[Y|Z=1,D=1] + (1-p(1))E[Y|Z=1,D=0] - \\ p(0)E[Y|Z=0,D=1] - (1-p(0))E[Y|Z=0,D=0] && LIE \end{aligned}$$

Replacing $E[Y|D=d,Z=z]$ by $E[Y|D=z,P=p(z)] = \alpha_d + \beta_d \frac{p(z)}{2}$:

$$\begin{aligned} E[Y|Z=1] - E[Y|Z=0] = &p(1)(\alpha_1 + \beta_1 \frac{p(1)}{2}) + (1-p(1))(\alpha_0 + \beta_0 \frac{p(1)}{2}) - \\ &p(0)(\alpha_1 + \beta_1 \frac{p(0)}{2}) + (1-p(0))(\alpha_0 + \beta_0 \frac{p(0)}{2}) \end{aligned}$$

The denominator

$$E[D|Z=1] - E[D|Z=0] = p(1) - p(0)$$

So the Wald estimate is

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

rearranging...

$$= (\alpha_1-\alpha_0)+(\beta_1-\beta_0)\frac{(p(1)+p(0))}{2}$$

Besides, the LATE with linear MTR is

$$\begin{aligned}
LATE &= E[Y(1)-Y(0)|D(z_1)>D(z_0)] \\
&= \int_0^1 m(u)\frac{\mathbf{1}(p(0)\leq u \leq p(1)}{p(1)-p(0)}du \\
&= \frac{1}{p(1)-p(0)}\int_{p(0)}^{p(1)}[(\alpha_1-\alpha_0)+(\beta_1-\beta_0)u]du \\
&= (\alpha_1-\alpha_0)+(\beta_1-\beta_0)\frac{(p(1)+p(0))}{2}
\end{aligned}$$

Evaluating the integral above at the bounds therefore yields the same expression for the LATE and the Wald estimate.

(c) In the previous part we show that in the simple case (binary instrument, linear MTR, no covariates), the Wald estimator and the LATE match. Note that the proof will generalize to non-linear MTRs because pulling out $\frac{1}{p(1)-p(0)}$ always allows us to express the numerator of the WALD as a weighted sum of MTEs, independent of $p(d)$, and therefore not relying on linearity.

Problem 3

(a) We have

$$\begin{aligned}
\pi &= P[M=1|R=c] \\
1-\pi &= 1-P[M=1|R=c] \\
1-\pi &= P[M=0|R=c] && M \in \{0,1\} \\
\pi &= 1-P[M=0|R=c] \\
\pi &= 1-\lim_{r \uparrow c} f_R(r|0) && f_{R|M}(r|O) \text{ continuous at } r=c
\end{aligned}$$

So π is identified.

(b) We want to find bounds on

$$\begin{aligned}
\delta &\equiv E[Y(1)-Y(0)|R=c, M=0] \\
&\equiv E[Y(1)|R=c, M=0] - \underbrace{E[Y(0)|R=c, M=0]}_{\text{continuous at } r=c} \\
&= E[Y(1)|R=c, M=0] - \underbrace{\lim_{r \uparrow c} E[Y|R=r]}_{\text{observed}}
\end{aligned}$$

We only need to get bounds on $E[Y(1)|R=c, M=0]$. From problem 1.2.c, we have

$$E[Y|R=c, Y \leq G^{-1}(1-\pi)] \leq E[Y(1)|R=c, M=0] \leq E[Y|R=c, Y \geq G^{-1}(\pi)]$$

Thus

$$E[Y|R=c, Y \leq G^{-1}(1-\pi)] - \lim_{r \uparrow c} E[Y|R=r] \leq \delta \leq E[Y|R=c, Y \geq G^{-1}(\pi)] - \lim_{r \uparrow c} E[Y|R=r]$$

(c) The assumptions and the result in part b) differ from the usual sharp regression discontinuity design framework because we are not given that the distribution of Y given r is continuous from the left at $r=c$, even though we are interested in the ATU. Thus, we don't know that $E[Y(1)|R=r, M=1]$ is continuous at $r=c$, so we need to adopt a more complicated approach to get bounds on δ : unlike in sharp RDD, δ is not point identified.

Problem 4

We want to show that $E[Y(1)-Y(0)|R=r]$ is point identified for all $r \in [c_l, c_h]$:

$$\begin{aligned}
E[Y(1)-Y(0)|R=r] =& E[Y(1)-Y(0)|R=r,C=c_h]P(C=c_h|R=r)+ \\
& E[Y(1)-Y(0)|R=r,C=c_l]P(C=c_l|R=r) \qquad \text{LTP} \\
=& E[Y(1)-Y(0)|R=r,C=c_h] \underbrace{p_{h|r}}_{\text{observed}} + E[Y(1)-Y(0)|R=r,C=c_l] \underbrace{p_{l|r}}_{\text{observed}} \\
=& [g_1(r)-g_0(r)+\alpha_1-\alpha_0]p_{h|r} + [g_1(r)-g_0(r)]p_{l|r} \\
=& g_1(r)-g_0(r)+(\alpha_1-\alpha_0)p_{h|r}
\end{aligned}$$

Besides we can identify $g_1(r), g_0(r), \alpha_0, \alpha_1$ (LHS is the object to identify, RHS is observed):
Looking at the average of Y for low types (cutoffs at c_l) who don't take the treatment $D=0$ and with $R=r$:

$$g_0(r) = E[Y(0)|R=r,C=c_l] \tag{1}$$

Following a similar logic:

$$\begin{aligned}
g_0(r)+\alpha_0 &= E[Y(0)|R=r,C=c_h] \qquad (2) \\
\leftrightarrow \alpha_0 &= E[Y(0)|R=r,C=c_h] - E[Y(0)|R=r,C=c_l] \\
g_1(r) &= E[Y(1)|R=r,C=c_l] \qquad (3) \\
g_1(r)+\alpha_1 &= E[Y(1)|R=r,C=c_h] \qquad (4) \\
\leftrightarrow \alpha_1 &= E[Y(1)|R=r,C=c_h] - E[Y(1)|R=r,C=c_l]
\end{aligned}$$

Thus, substracting averages we see that we can point-identify α_0, α_1, $g_1(r), g_0(r)$ for each r. Note that $g_0(r)$ is identified for $r \in [c_l, c_h)$, so taking limits (and using the continuity) we also have that $g_0(r)$ is also identified at $r = c_h$. Same for $g_1(r)$ at $r = c_l$.

Problem 5

We have

- Y : worked for pay
- D : more than 2 children
- X : covariates (age, age at first birth, ages of the first two children in quarters, boy 1st, boy 2nd, black, hispanic, other race)
- Z the instrument
- $D = \mathbf{1}[U \leq p(X, Z)]$
- $m(d|u, x) = \mathbb{E}[Y(d)|U = u, X = x]$ the MTR

Mapping of the coefficients - by specification

Notice that coefficients from the regressions of Y on X and p (see code) must be slightly adjusted to recover the MTRs and the corresponding MTE.

There is a probably a way to code it such that you recover the MTR directly from a unique regression etc, but at least this is tractable.

Specification 1

$$
\begin{aligned}
E[Y|X = x, D = 1] &= E[Y(1)|X = x, U \leq p] \\
&= \frac{1}{p}\int_0^p E[Y(1)|X = x, U = u]du \\
&= \frac{1}{p}\int_0^p (\alpha_1 + \beta_1 u + \gamma_1' x)du \\
&= \frac{1}{p}[\alpha_1 u + \frac{\beta_1}{2}u^2 + \gamma_1' x u]_0^p \\
&= \underbrace{\alpha_1}_{\hat{\alpha_1}} + \underbrace{\frac{\beta_1}{2}}_{\hat{\beta_1}} p + \underbrace{\gamma_1'}_{\hat{\gamma_1'}} x
\end{aligned}
$$

Where the *hat* parameters designate those recover from the OLS regression $Y = \hat{\alpha_1} + \hat{\beta_1} p + \hat{\gamma_1}' x | D = 1$. Thus

$$
\begin{aligned}
\alpha_1 &= \hat{\alpha_1} \\
\beta_1 &= 2\hat{\beta_1} \\
\gamma_1 &= \hat{\gamma_1}
\end{aligned}
$$

Similarly

$$\begin{aligned}
E[Y|X=x,D=0] &= E[Y(0)|X=x,u>p] \\
&= \frac{1}{1-p}\int_p^1 E[Y(0)|X=x,U=u]du \\
&= \frac{1}{1-p}\int_p^1 (\alpha_0+\beta_0 u+\gamma_0' x)du \\
&= \frac{1}{1-p}[\alpha_0 u+\frac{\beta_0}{2}u^2+\gamma_0' xu]_p^1 \\
&= \frac{1}{1-p}[\alpha_0+\frac{\beta_0}{2}+\gamma_0' x-\alpha_0 p-\frac{\beta_0}{2}p^2-\gamma_0' xp] \\
&= \frac{1}{1-p}\alpha_0(1-p)+\frac{1}{1-p}\frac{\beta_0}{2}(1-p^2)+\frac{1}{1-p}\gamma_0' x(1-p) \\
&= \alpha_0+\frac{1+p}{2}\beta_0+\gamma_0' x \\
&= \underbrace{\alpha_0+\frac{1}{2}\beta_0}_{\hat{\alpha}_0}+p\underbrace{\frac{\beta_0}{2}}_{\hat{\beta}_0}+\underbrace{\gamma_0'}_{\hat{\gamma}_0'}x
\end{aligned}$$

Estimating the coefficients above from OLS conditioning on $D=0$, we will be able to recover the MTE_0 coefficients $\alpha_0,\beta_0,\gamma_0$:

$$\begin{aligned}
\alpha_0 &= \hat{\alpha_0}-\hat{\beta}_0 \\
\beta_0 &= 2\hat{\beta}_0 \\
\gamma_0 &= \hat{\gamma_0}
\end{aligned}$$

Specification 2

Because we want $\gamma_0=\gamma_1=\gamma$ we run the saturated regression (on the full dataset)

$$Y=\hat{\alpha}_0+\hat{\alpha}_1 D+\hat{\beta}_0 p+\hat{\beta}_1 pD+\hat{\gamma}' X$$

Thus, we recover the MTE as

$$MTE(u)=\hat{\alpha}_1+\hat{\beta}_1 u$$

Notice that $\hat{\gamma}' X$ cancels out as it is the same for $MTR(0)$ and $MTR(1)$ by design, and we evaluate X at the average of X.

Specification 3

As for specification 1

$$\begin{aligned}
E[Y|X=x,D=1] &= \frac{1}{p}\int_0^p [\alpha_1+\beta_1 u+\gamma' X+\delta_1' xu]du \\
&= \frac{1}{p}[\alpha_1 p+\frac{\beta_1}{2}p^2+\gamma' Xp+\frac{\delta_1'}{2}xp^2] \\
&= \underbrace{\alpha_1}_{\hat{\alpha_1}}+\underbrace{\frac{\beta_1}{2}}_{\hat{\beta_1}}p+\underbrace{\gamma'}_{\hat{\gamma_1}}X+\underbrace{\frac{\delta_1'}{2}}_{\hat{\delta_1'}}Xp
\end{aligned}$$

So $\alpha_1,\beta_1,\gamma_1$ are recovered the same way as in specification 1. In addition we have

$$\delta_1=2*\hat{\delta_1}$$

Similarly for $MTR(0)$, we have α_0, β_0 recovered as in specification 1, and

$$\gamma_0 = \hat{\gamma_0} - \hat{\delta_0}$$
$$\delta_0 = 2\hat{\delta_0}$$

Specification 4

Following the same logic, $\alpha_1, \beta_{1,1}, \gamma_1$ are identified as in Specifications 1 and 3. Besides

$$\beta_{1,2} = 3 * \hat{\beta_{1,2}}$$

Moreover

$$\begin{aligned}
E[Y|D=0,X=x] &= \frac{1}{1-p}\int_p^1 [\alpha_0 + \beta_{0,1}u + \beta_{0,2}u^2 + \gamma_0' x]du \\
&= \frac{1}{1-p}[\alpha_0 u + \frac{\beta_{0,1}}{2}u^2 + \frac{\beta_{0,2}}{3}u^3 + \gamma_0' xu]_p^1 \\
&= \frac{1}{1-p}[\alpha_0(1-p) + \frac{\beta_{0,1}}{2}(1-p^2) + \frac{\beta_{0,2}}{3}(1-p^3) + \gamma_0' x(1-p)] \\
&= \alpha_0 + \frac{\beta_{0,1}}{2}(1+p) + \frac{\beta_{0,2}}{3}(1+p+p^2) + \gamma_0' x \\
&= \underbrace{\alpha_0 + \frac{\beta_{0,1}}{2} + \frac{\beta_{0,2}}{3}}_{\hat{\alpha}} + \underbrace{\frac{\beta_{0,1}}{2} + \frac{\beta_{0,2}}{3}}_{\hat{\beta_{0,1}}} + \underbrace{\frac{\beta_{0,2}}{3}}_{\hat{\beta_{0,2}}} p^2 + \underbrace{\gamma_0'}_{\hat{\gamma_0'}} x
\end{aligned}$$

Thus

$$\begin{aligned}
\beta_{0,2} &= 3\hat{\beta_{0,2}} \\
\beta_{0,1} &= 2(\hat{\beta_{0,1}} - \hat{\beta_{0,2}}) \\
\alpha_0 &= \hat{\alpha} - \hat{\beta_{0,1}} \\
\gamma_0' &= \hat{\gamma_0}'
\end{aligned}$$

Specification 5

Following the same logic, $\alpha_1, \beta_{1,1}, \beta_{1,2}, \gamma_1$ are identified as in Specifications 1, 3 and 4. Besides

$$\beta_{1,3} = 4 * \hat{\beta_{1,3}}$$

Finally (because the fun must end at some point):

$$\begin{aligned}
E[Y|D=0,X=x] &= \frac{1}{1-p}\int_p^1 [\alpha_0 + \beta_{0,1}u + \beta_{0,2}u^2 + \beta_{0,3}u^3 + \gamma_0' x]du \\
&= \frac{1}{1-p}[\alpha_0 u + (1/2)\beta_{0,1}u^2 + (1/3)\beta_{0,2}u^3 + (1/4)\beta_{0,3}u^4 + \gamma_0' xu]_p^1 \\
&= \frac{1}{1-p}[\alpha_0(1-p) + (1/2)\beta_{0,1}(1-p^2) + (1/3)\beta_{0,2}(1-p^3) + (1/4)\beta_{0,3}(1-p^4) + \gamma_0' x(1-p] \\
&= \alpha_0 + \frac{\beta_{0,1}}{2}(1+p) + \frac{\beta_{0,2}}{3}(1+p+p^2) + \frac{\beta_{0,3}}{4}(1+p+p^2+p^3) + \gamma_0' x \\
&= \underbrace{\alpha_0 + \frac{\beta_{0,1}}{2} + \frac{\beta_{0,2}}{3} + \frac{\beta_{0,3}}{4}}_{\hat{\alpha}} + \underbrace{\frac{\beta_{0,1}}{2} + \frac{\beta_{0,2}}{3} + \frac{\beta_{0,3}}{4}}_{\hat{\beta_{0,1}}} + \underbrace{\frac{\beta_{0,2}}{3} + \frac{\beta_{0,3}}{4}}_{\hat{\beta_{0,2}}} p^2 + \underbrace{\frac{\beta_{0,3}}{4}}_{\hat{\beta_{0,3}}} p^3 + \underbrace{\gamma_0'}_{\hat{\gamma_0'}} x
\end{aligned}$$

Thus

$$
\begin{aligned}
\beta_{0,3} &= 4\hat{\beta_{0,3}} \\
\beta_{0,2} &= 3(\hat{\beta_{0,2}} - \hat{\beta_{0,3}}) \\
\beta_{0,1} &= 2(\hat{\beta_{0,1}} - \hat{\beta_{0,2}}) \\
\alpha_0 &= \hat{\alpha_0} - \hat{\beta_{0,1}} \\
\gamma_0 &= \hat{\gamma_0}
\end{aligned}
$$

Structure of the Code

The enclosed code

- Computes the propensity score using logit
- Computes $MTR(1)$ and $MTR(0)$ for the different specifications (as detailed above)
- Computes the weights for the ATE, ATT, ATU, LATE for a pair (u, p).
- Recovers the parameter of interest by integrating (taking the average over) the weight*MTE.

- Contains a *routine* function that does the full work : for each specification
  1. Run the regression from above
  2. Recovers $MTR(1), MTR(0)$
  3. Compute the MTE for a grid of u, **evaluated at the average X**. Note that this is not exactly correct, as we don't necessarily expect the covariates to be uncorrelated with u (especially if we control for them!). Besides, for specification 3, Xs are interacted with u, so here we interact u with the average of X across the dataset. Of course, this is only an approximation.
  4. Compute the parameters of interest (ATE, ATT etc) following two different approaches
     - **Approach 1: Integrating the MTE with the appropriate weights**, evaluated at the average X. Notice that in the code, I compute the weights for p of the actual observations, but I could also compute the weights for p_{mean} where the propensity score is evaluated at the mean of X rather than the actual value of X observed in the data.
       While this takes advantage to the functional form of MTE to cover the whole interval $u \in [0, 1]$, this doesn't accound for the endogenous distribution of Xs.
     - **Approach 2: use the regression results to recover the unobserved potential outcomes** (ie $Y(0)$ for $D = 1$ etc) and then simply, from the data, get

$$
\begin{aligned}
ATE &= E[Y(1) - Y(0)] \\
ATT &= E[Y(1) - Y(0)|D = 1] \\
ATU &= E[Y(1) - Y(0)|D = 0]
\end{aligned}
$$

       This would have accounted for the endogenous distribution of Xs, but we would not have taken advantage of the fact that assuming a functional form for the $MTE(u)$ allows us to cover $u \in [0, 1]$, while our data doesn't have p covering the full interval[1].

  Because of the reasons mentioned above, I do not expect to find the same ATE, ATU, ATT for the two approaches. As you can see below, differences are quite large. It is unclear to me what, in my approach, drives such large differences (an error somewhere?)

The TSLS estimates presented below are very close to the ones in Table 11 of the paper.

[1] I don't have the results for approach 2, specification 2, because I'm running late.

Table 1: TSLS

Specification	TSLS Estimate
Same Sex	0.118
Twins	0.083
Both	0.113

Table 2: Treatment Effects - Same Sex Instrument

Specification	1	2	3	4	5
Approach 1					
ATE	-0.181	-0.197	-0.107	-0.042	-0.162
ATT	-0.289	-0.180	-0.141	0.100	0.012
ATU	-0.255	-0.194	-0.121	-0.106	-0.128
LATE	-0.226	-0.191	-0.187	-0.187	-0.187
Approach 2					
ATE	-0.189		-0.189	-0.189	-0.188
ATT	-0.196		-0.185	-0.186	-0.185
ATU	-0.184		-0.107	-0.042	-0.162

Table 3: Treatment Effects - Twins Instrument

Specification	1	2	3	4	5
Approach 1					
ATE	-0.283	-0.198	-0.156	0.011	-0.107
ATT	-0.526	-0.183	-0.208	0.172	-0.098
ATU	-0.255	-0.194	-0.113	-0.123	-0.123
LATE	-0.090	-0.205	-0.188	-0.187	-0.185
Approach 2					
ATE	-0.188		-0.190	-0.186	-0.182
ATT	-0.190		-0.187	-0.188	-0.188
ATU	-0.186		-0.156	0.011	-0.107

Table 4: Treatment Effects - Both Instruments

Specification	1	2	3	4	5
Approach 1					
ATE	-0.184	-0.198	-0.133	-0.030	-0.062
ATT	-0.293	-0.184	-0.148	0.086	0.013
ATU	-0.110	-0.207	-0.122	-0.107	-0.112
Approach 2					
ATE	-0.190		-0.189	-0.187	-0.186
ATT	-0.194		-0.190	-0.186	-0.184
ATI	-0.186		-0.187	-0.188	-0.188

Notice that I do not present *LATE* for the 2 instruments-case. If I had had more time, I could have investigated different *LATEs*, but I am running out of time. Potential parameters of interest could have been, with the first subscript designating the same sex instrument and the second designating the twins instrument: $LATE(z_{0,0} \rightarrow z_{0,1})$, $LATE(z_{0,0} \rightarrow z_{1,0})$, $LATE(z_{1,0} \rightarrow z_{1,1})$, $LATE(z_{0,1} \rightarrow z_{1,1})$. I'm not sure

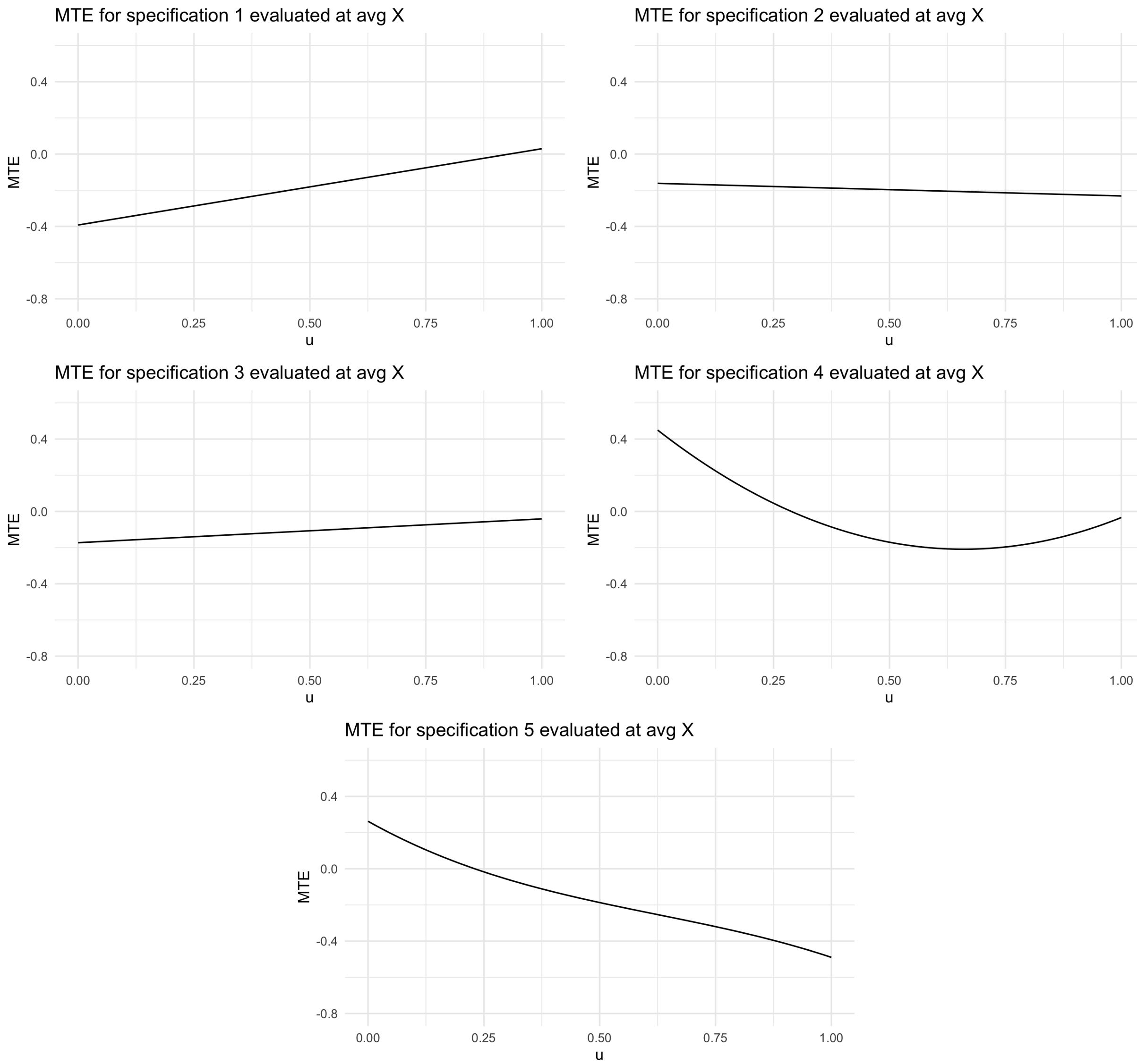

Figure 1: MTE(u) for Same Sex Instrument (evaluated at avg X)

what I would get out of computing $LATE(z_{0,1} \rightarrow z_{1,0})$.
In the case of two instruments, we then have that the TSLS estimate is a weighted sum of the two individual LATEs, where the weights depend on the relative strength of each instrument in the first stage.

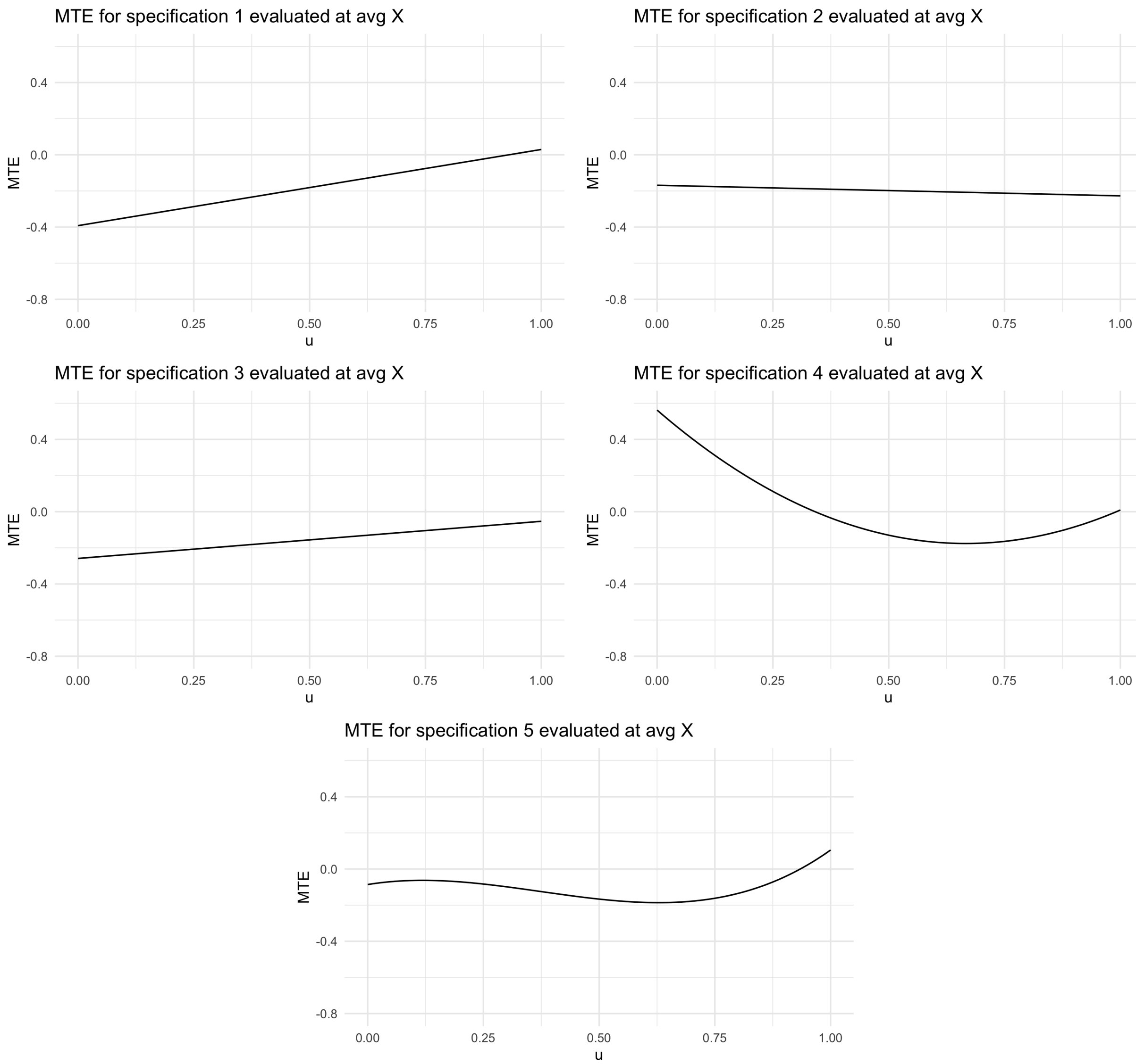

Figure 2: MTE(u) for Twins Instrument (evaluated at avg X)

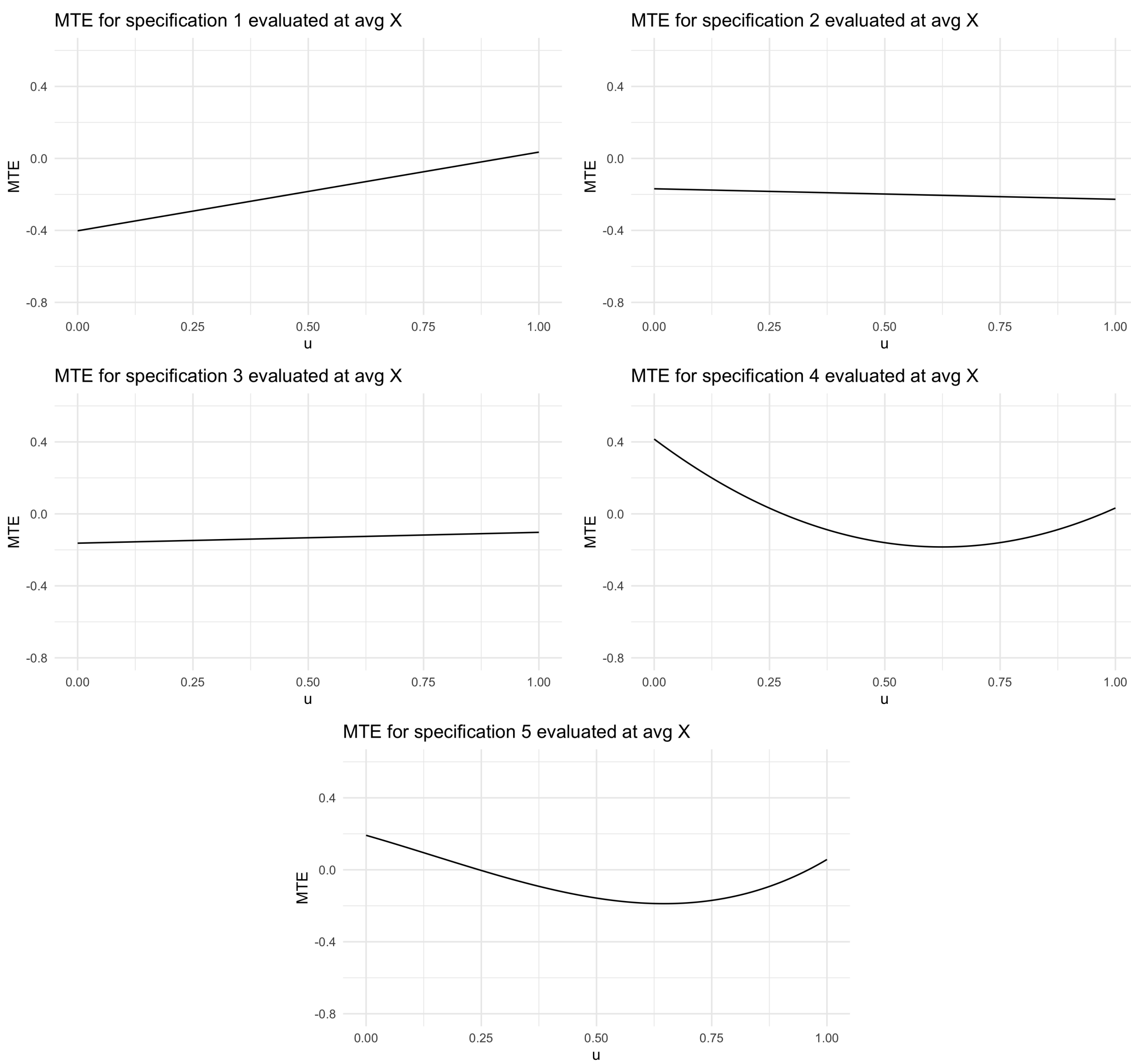

Figure 3: MTE(u) for Twins and Same Sex Instrument (evaluated at avg X)

Problem 6

The Setup

Extending Heckman and Vytlacil (2005) and Mogstad et al. (2017, proposition 1), the paper shows that we can write any parameter as (abstracting from the Xs, as they are not relevant for this problem):

$$\Gamma(m) \equiv \mathbb{E}[\int_0^1 m_0(u)\omega_0(u,Z)du] + \mathbb{E}[\int_0^1 m_1(u)\omega_1(u,Z)du]$$

Where $m \in \mathbf{M}$. Replacing $\mathbf{M}$ by a finite dimensional linear space (Mogstad et al. (2017)):

$$m_d(u) = \sum_{k=0}^{K_d} \theta_{dk} b_{dk}(u) \qquad d = 0,1$$
$$b_{dk}(u) \text{ a known basis function}$$

Thus

$$\Gamma(m) \equiv \sum_{k=0}^{K_1} \theta_{1,k} \underbrace{\mathbb{E}[\int_0^1 b_{1k}(u)\omega_1(u,Z)du]}_{\equiv \gamma_1} + \sum_{k=0}^{K_0} \theta_{0,k} \underbrace{\mathbb{E}[\int_0^1 b_{0k}(u)\omega_0(u,Z)du]}_{\equiv \gamma_0}$$

In order to find the bounds on ω, we tackle the following maximization (minimization) problem:

$$\max_{\{\theta_d,k\}_{d=0,1,k=0:K}} \Gamma_{ATT}(m) = \sum_{k=0}^{K} \theta_{1,k} \underbrace{\mathbb{E}[\int_0^1 b_{1,k}(u)\omega_{1,ATT}(u,Z)du]}_{\gamma_{1,k,ATT}} + \sum_{k=0}^{K} \theta_{0,k} \underbrace{\mathbb{E}[\int_0^1 b_{0,k}(u)\omega_{0,ATT}(u,Z)du]}_{\gamma_{0,k,ATT}}$$

$$\text{s.t. } (1)\beta_{IVSlope} = \sum_{k=0}^{K} \theta_{1,k}\mathbb{E}[\int_0^1 b_{1,k}(u)\omega_{1,IVSlope}(u,Z)du] + \sum_{k=0}^{K} \theta_{0,k}\mathbb{E}[\int_0^1 b_{0,k}(u)\omega_{0,IVSlope}(u,Z)du]$$
$$(2)\beta_{TSLS} = \sum_{k=0}^{K} \theta_{1,k}\mathbb{E}[\int_0^1 b_{1,k}(u)\omega_{1,TSLS}(u,Z)du] + \sum_{k=0}^{K} \theta_{0,k}\mathbb{E}[\int_0^1 b_{0,k}(u)\omega_{0,TSLS}(u,Z)du]$$

Where

$$\beta_{IVSlope} = \int_0^1 [\underbrace{m_1(u) - m_0(u)}_{\text{known DGP, or from estimation}}]\omega_{IVSlope}(u,Z)du$$
$$\beta_{TSLS} = \int_0^1 [m_1(u) - m_0(u)]\omega_{TSLS}(u,Z)du$$

The objects of interest

We are given

$$m_1(u) = 0.35 - 0.3u - 0.05u^2$$
$$m_0(u) = 0.9 - 1.1u + 0.3u^2$$
$$Z \in \{1,2,3,4\}, P(Z=z) = 1/4 \forall z \in Z$$
$$P(D=1|Z=z) \equiv p(z); p(1) = 0.12; p(2) = 0.29; p(3) = 0.48; p(4) = 0.78$$

We need to compute the following objects

1. $\mathbb{E}(Z), \mathbb{E}(D), cov(Z, D)$

$$\begin{aligned} \mathbb{E}(Z) &= \frac{1+2+3+4}{4} = 2.5 \\ \mathbb{E}(D) &= \frac{p(1)+p(2)+p(3)+p(4)}{4} = \frac{0.12+0.29+0.48+0.78}{4} = 0.4175 \\ Cov(D,Z) &= \mathbb{E}(DZ) - \mathbb{E}(D)\mathbb{E}(Z) = 1.315 - 1.044 = 0.2712 \end{aligned}$$

2. $\omega_{1,IVSlope}, \omega_{0,IVSlope}$
   As for ω_{TSLS} and ω_{ATT}, we have that $\omega_{IVSlope}$ can only take 4 strictly positive values as we have 4 instruments.
   We have

$$\begin{aligned} \omega_{1,IVSlope}(u) &= s_{IVSlope}\mathbf{1}[u \leq p(z)] \\ &= \frac{z - E(z)}{Cov(D,Z)}\mathbf{1}[u \leq p(z)] \\ \omega_{0,IVSlope}(u) &= \frac{z - E(z)}{Cov(D,Z)}\mathbf{1}[u > p(z)] \end{aligned}$$

   Thus

$$\begin{aligned} \omega_{1,IVSlope}(u|u <= p1) =& 0.25\left(\frac{1-E(z)}{Cov(D,Z)} + \frac{2-E(z)}{Cov(D,Z)} + \frac{3-E(z)}{Cov(D,Z)} + \frac{4-E(z)}{Cov(D,Z)}\right) \\ \omega_{1,IVSlope}(u|p1 < u <= p2) =& 0.25\left(\frac{2-E(z)}{Cov(D,Z)} + \frac{3-E(z)}{Cov(D,Z)} + \frac{4-E(z)}{Cov(D,Z)}\right) \\ & etc \\ \omega_{0,IVSlope}(u|u <= p1) &= 0 \\ \omega_{0,IVSlope}(u|p1 < u <= p2) &= 0.25\left(\frac{1-E(z)}{Cov(D,Z)}\right) \\ \omega_{0,IVSlope}(u|p2 < u <= p3) &= 0.25\left(\frac{1-E(z)}{Cov(D,Z)} + \frac{2-E(z)}{Cov(D,Z)}\right) \\ & etc \end{aligned}$$

3. $\omega_{1,TSLS}, \omega_{0,TSLS}$
   Following the similar logic

$$\begin{aligned} \omega_{1,TSLS} &= s_{TSLS}\mathbf{1}[u \leq p(z)] \\ \omega_{0,TSLS} &= s_{TSLS}\mathbf{0}[u > p(z)] \end{aligned}$$

   Where

$$\begin{aligned} s_{TSLS}[z] &= (\Pi * \tilde{Z})^{-1} * \Pi * e_4 \\ e_4 &= diag(1,4) \\ \Pi &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ p1 & p2 & p3 & p4 \end{bmatrix} \\ \tilde{Z} &= \frac{1}{4}\Pi' \end{aligned}$$

4. $\omega_{1,ATT}, \omega_{0,ATT}$

$$\omega_{1,ATT}(u) = \mathbf{1}[u \le p1]\frac{1}{E(D)} + \mathbf{1}[p1 < u \le p2]\frac{1-0.25}{E(D)} + \mathbf{1}[p2 < u \le p3]\frac{1-2*0.25}{E(D)} + \mathbf{1}[p3 < u \le p4]\frac{1-3*0.25}{E(D)}$$
$$\omega_{0,ATT}(u) = -\omega_{1,ATT}(u)$$

5. $\beta_{IVSlope}$ (true value)
Notice that because we have 4 ordered instruments, in equal share, we can rewrite (and this is true in general, and used all along the code)

$$\begin{aligned}\beta_{IVSlope} &= \int_0^1 [m_1(u) - m_0(u)]\omega_{IVSlope}(u, Z)du \\ &= \int_0^{p1} [m_1(u) - m_0(u)]\omega_{IVSlope}(p1)du + \int_{p1}^{p2} [m_1(u) - m_0(u)]\omega_{IVSlope}(p2)du + \\ &\quad \int_{p2}^{p3} [m_1(u) - m_0(u)]\omega_{IVSlope}(p3)du + \int_{p3}^{p4} [m_1(u) - m_0(u)]\omega_{IVSlope}(p4)du\end{aligned}$$

Notice that $\omega_{IVSlope}(u, p1)$ doesn't depend on u for $u \in [0, p1]$, so we can get rid of the dependent on u:

$$\begin{aligned}\beta_{IVSlope} =& \omega_{IVSlope}(p1)\int_0^{p1} [m_1(u) - m_0(u)]du + \omega_{IVSlope}(p2)\int_{p1}^{p2} [m_1(u) - m_0(u)]du + \\ & \omega_{IVSlope}(p3)\int_{p2}^{p3} [m_1(u) - m_0(u)]du + \omega_{IVSlope}(p4)\int_{p3}^{p4} [m_1(u) - m_0(u)]du\end{aligned}$$

6. β_{TSLS} (true value)
Similarly as above

$$\begin{aligned}\beta_{TSLS} =& \omega_{TSLS}(p1)\int_0^{p1} [m_1(u) - m_0(u)]du + \omega_{TSLS}(p2)\int_{p1}^{p2} [m_1(u) - m_0(u)]du + \\ & \omega_{TSLS}(p3)\int_{p2}^{p3} [m_1(u) - m_0(u)]du + \omega_{TSLS}(p4)\int_{p3}^{p4} [m_1(u) - m_0(u)]du\end{aligned}$$

7. β_{ATT} (true value): same as above with ω_{ATT}

8. **Basis Function for Bernstein Polynomial**
Using Mogstal, Santos & Torgovitsky (2018) Bernstein Polynomial formula

$$b_{k,K}(u) = \sum_{i=k}^{K}(-1)^{i-k}\left(\begin{array}{c}K\\ i\end{array}\right)\left(\begin{array}{c}i\\ k\end{array}\right)u^i$$

9. **Basis Function for Constant Splines**
Following my understanding of Mogstad, Santos and Torgovitsky (2018, ECMA) (p 1602), because we have a discrete instrument that can take 4 different values, we split our $u \in [0, 1]$ interval into $4+1 = 5$ constant splines.
Besides, given the set up of the code, we set the spline boundaries at the propensity scores.

10. **The monotonicity constraints**
For both $d = 0, 1$ we want to ensure $\theta_{d,0} > \theta_{d,1} > \theta_{d,2} > ... > \theta_{d,K}$. To do so, we add, in addition to constraints (1) and (2), $K * 2$ constraints that $\theta_{d,i} - \theta_{d,i+1} > 0$.

The code's structure

Because the instrument takes discrete values $\{1, 2, 3, 4\}$ with equal probability, it appears to be equivalent to set up the problem (1) as averaging over the bins defined by the propensity scores and (2) as formally integrating over all u. Therefore, following discussions with a few members of the study group[2], I go for the first approach. Note that this would need to be thought through a little more if the instrument was to take different values with different probabilities.

The code goes on to compute (for each propensity score bin)

- The bases functions (see discussion above)
  - Bernstein Polynomial
  - Constant Splines
- The $s(d, z)$ for the constraints
  - $s_{IVslope}(d, z)$
  - $s_{TSLS}(d, z)$
- The weights
  - $\omega_{1,ATT} = -\omega_{0,ATT}$
  - $\omega_{1,IVSlope} = -\omega_{0,IVSlope}$
  - $\omega_{1,TSLS} = -\omega_{0,TSLS}$
- The $\gamma(.)$ functions (functions of $s(d, z), \omega_d$ but not θ)
- The true $\beta_{IVslope}$ and β_{TSLS} values (needed for the RHS of the constraints) by integrating the MTE over the 4 propensity score bins (because $p(z) = 1/4 \forall z$ this is the same as integrating over u) multiplied by the corresponding weights
- The routine *beautiful* that uses the gurobi package. For a given $K = K_1 = K_0$ (parametric case), it finds the $\{\theta_{d,k}\}_{d=0,1;k=0:K}$ that maximizes (minimizes) $\Gamma_{ATT}(m)$ such that $\Gamma_{IVSlope}(m) = \beta_{IVslope}$ and $\Gamma_{TSLS}(m) = \beta_{TSLS}$ to get the upper bound (lower bound) on β_{ATT}.
- Finally, I make a beautiful graph reproducing Figure 6 of the paper, with more colors.

Refer to the enclosed code for more details.

[2] I want to especially thank Sunny for the trick

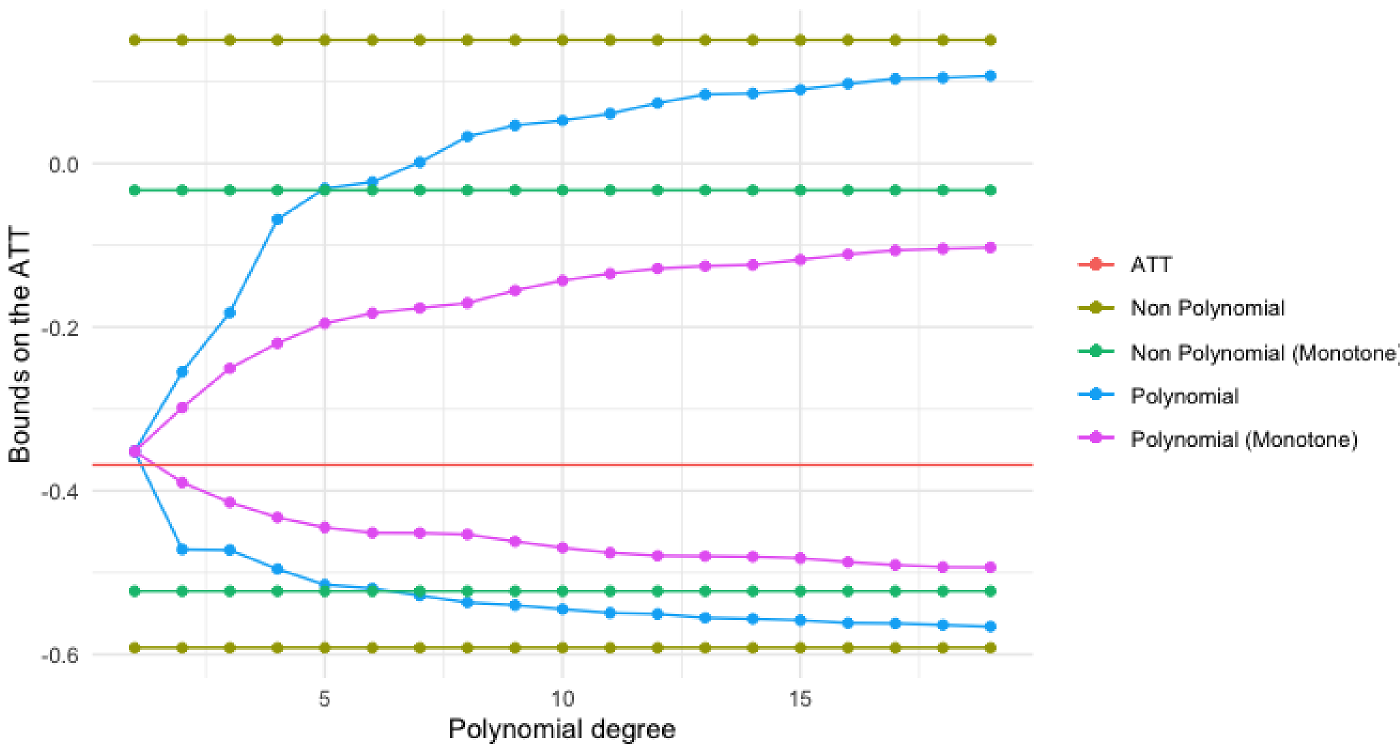

Figure 4: Reproduction of Figure 6 of Mogstad and Torgovitsky (2018)