Econ 2120: Final Review

Part 1: IV with Heterogeneous Treatment Effects

Ashesh Rambachan

Fall 2018

Outline

Outline

IV Model – Treatment Effect Heterogeneity

We'll now present the IV model with heterogeneous treatment effects. Then, we will place sufficient restrictions on the treatment effect heterogeneity so that the IV estimator will deliver an ATE.

Assume that treatment T_i is not randomly assigned but there is an instrument S_i that is randomly assigned and correlated with T_i.

For each individual i, there is a potential treatment function $T_i(\cdot)$ that can be evaluated at any $s \in \mathcal{S}$. $T_i(s)$ is the treatment realized for individual i at instrument level s. We observe

$$T_i = T_i(S_i).$$

For each individual i, there is a potential outcome function $Y_i(\cdot, \cdot)$ that can be evaluated at any level of the treatment and subsidy. We observe

$$Y_i(T_i, S_i)$$

IV Model – Treatment Effect Heterogeneity

We make two key assumptions:

(1) **Exclusion restriction**:

$$Y_i(t, s_1) = Y_i(t, s_2) \quad \forall s_1, s_2 \in \mathcal{S}.$$

(2) **Random assignment of instrument**:

$$\{\{Y_i(t), t \in \mathcal{T}\}, \{T_i(s), s \in \mathcal{S}\}\} \perp\!\!\!\perp S_i.$$

IV Model – Constant Treatment Effects

Suppose that $T_i \in \{0, 1\}$, $S_i \in \{0, 1\}$ and that the treatment effects are constant:

$$Y_i(1) = Y_i(0) + K$$

for all i. Then, the IV estimator identifies $ATE = K$. See the notes for the derivation.

Next time – we'll consider what the IV estimator identifies when we do not restrict the heterogeneity of the treatment effects.

Outline

IV Bounds on ATE

Under the IV assumptions, we can derive bounds on the ATE without treatment effect homogeneity assumptions. Suppose again that $Y \in [0, 1]$.

We have that

$$\begin{aligned} E[Y(1)] &= E[Y(1)|Z = z] \\ &= E[Y(1)|Z = z, T = 1]P(T = 1|Z = z) \\ &+ E[Y(0)|Z = z, T = 0]P(T = 0|Z = z) \\ &\leq E[Y(1)|Z = z, T = 1]P(T = 1|Z = z) + P(T = 0|Z = z) \end{aligned}$$

This holds for all $z \in \mathcal{Z}$. So, an upper-bound is

$$E[Y(1)] \leq \min_z E[Y(1)|Z = z, T = 1]P(T = 1|Z = z) + P(T = 0|Z = z).$$

Similarly, a lower bound is

$$\max_z E[Y(1)|Z = z, T = 1]P(T = 1|Z = z) \leq E[Y(1)].$$

Great Practice Q: Work out the IV bounds on ATE for a binary instrument, $Z = \{0, 1\}$ – this is a very simple calculation.

Outline

Selection Bias

Let $T_i \in \{0, 1\}$ and $S_i \in \mathcal{S}$. Define the potential outcome function, $Y(t, s)$ and impose the exclusion restriction

$$Y(t, s_1) \stackrel{d}{=} Y(t, s_2) \quad s_1, s_2 \in \mathcal{S}.$$

So, we write $Y(t)$ from now on.

Define

$$E[Y_i(1)] = \alpha_1, \; U_{i1} = Y_{i1} - E[Y_i(1)]$$
$$E[Y(0)] = \alpha_0, \; U_{i0} = Y_{i0} - E[Y_i(0)],$$

where $\beta = E[Y_i(1) - Y_i(0)]$. So, we write

$$Y_i(0) = \alpha_0 + U_{i0}$$
$$Y_i(1) = \alpha_1 + U_{i1}.$$

Selection Bias

Selection bias arises if T_i is correlated with the potential outcomes. We'll model this with a selection equation

$$T_i = 1\{V_i \leq g(S_i)\}, \;\; V_i \perp\!\!\!\perp S_i.$$

We assume that $S_i = (S_{i1}, \ldots, S_{iJ})'$ and assume $g : \mathbb{R}^J \to [0, 1]$. g is completely unrestricted and so, it is a normalization (observationally equivalent) to assume that

$$V_i | S_i = s \sim U[0, 1].$$

It's immediate that

$$P(T_i = 1 | S_i = s) = g(s).$$

This is known as a **threshold crossing model** or a **latent index model**. There are two key restrictions here:

(1) $V_i \perp\!\!\!\perp S_i$.

(2) V_i, S_i enter in an additively separable way.

Selection Bias

Let's see how selection bias arises. We have

$$\begin{aligned} Y_i &= Y_i(T_i) \\ &= Y_i(0) + T_i\{Y_i(1) - Y_i(0)\} \\ &= \alpha_0 + T_i\beta + \{U_{i0} + T_i(U_{i1} - U_{i0})\}. \end{aligned}$$

Then,

$$E[Y_i|T_i] = \alpha_0 + T_i\beta + E[U_{i0}|T_i] + T_iE[U_{i1} - U_{i0}|T_i]$$

and the predictive effect is

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = \beta + E[U_{i1}|T_i = 1] - E[U_{i0}|T_i = 0].$$

Selection bias arises if T_i is correlated with either U_{i1}, U_{i0}.

Random assignment of the instrument

Suppose that

$$\{Y_i(0), Y_i(1), V_i\} \perp\!\!\!\perp S_i.$$

Consider the reduced form expectations. First,

$$E[T_i|S_i = s] = P(T_i = 1|S_i = s) = g(s).$$

Next,

$$\begin{aligned} E[Y_i|S_i = s] &= E[\alpha_0 + T_i\beta + \{U_{i0} + T_i(U_{i1} - U_{i0})\}] \\ &= \alpha_0 + g(s)\beta + E[T_i(U_{i1} - U_{i0})|S_i = s], \end{aligned}$$

Random assignment of the instrument

Continuing, we have that

$$\begin{aligned} E[T_i(U_{i1} - U_{i0})|S_i = s] &= E[E[T_i(U_{i1} - U_{i0})|S_i = s, V_i]|S_i = s] \\ &= E[1\{V_i \leq g(s)\}E[U_{i1} - U_{i0}|V_i]|S_i = s] \\ &= \int_0^1 1\{v \leq g(s)\}E[U_{i1} - U_{i0}|V_i = v]dv \\ &= \int_0^{g(s)} E[U_{i1} - U_{i0}|V_i = v]dv \end{aligned}$$

So, under random assignment of the instrument, we have that

$$E[T_i|S_i = s] = g(s)$$

$$E[Y_i|S_i = s] = \alpha_0 + g(s)\beta + \int_0^{g(s)} E[U_{i1} - U_{i0}|V_i = v]dv.$$

Marginal treatment effects

Suppose that $E[U_{i1} - U_{i0}|V_i] \neq 0$ – that is, the latent V_i is correlated with the individual gains from treatment. In other words, there is selection into the treatment based on individual gains. The **marginal treatment effect** is defined as

$$\begin{aligned} MTE(v) &= E[Y_i(1) - Y_i(0)|V_i] \\ &= \beta + E[U_{i1} - U_{i0}|V_i = v]. \end{aligned}$$

This is the ATE for the subpopulation with $V_i = v$. This sub-population is **marginal** meaning that their treatment status will change for a small change in the subsidy that has $g(s)$ go from just below to just above v.

By iterated expectations, we can relate MTE to ATE:

$$ATE = E[E[Y_i(1) - Y_i(0)|V_i]] = E[MTE(V_i)].$$

Marginal treatment effects

Recall: In the simple linear IV case, the Wald estimator returned the ratio of the reduced form slopes. So, let's look at something similar here.

Claim:

$$\frac{\frac{\partial E[Y_i|S_i=s]}{\partial s_j}}{\frac{\partial E[T_i|S_i=s]}{\partial s_j}} = MTE(g(s)) \quad j = 1, \ldots, J.$$

Proof: We have

$$\frac{\partial E[Y_i|S_i = s]}{\partial s_j} = \frac{\partial g(s)}{\partial s_j}\beta + E[U_{i1} - U_{i0}|V_i = g(s)]\frac{\partial g(s)}{\partial s_j}$$

$$\frac{\partial E[T_i|S_i = s]}{\partial s_j} = \frac{\partial g(s)}{\partial s_j}.$$

Result is immediate.

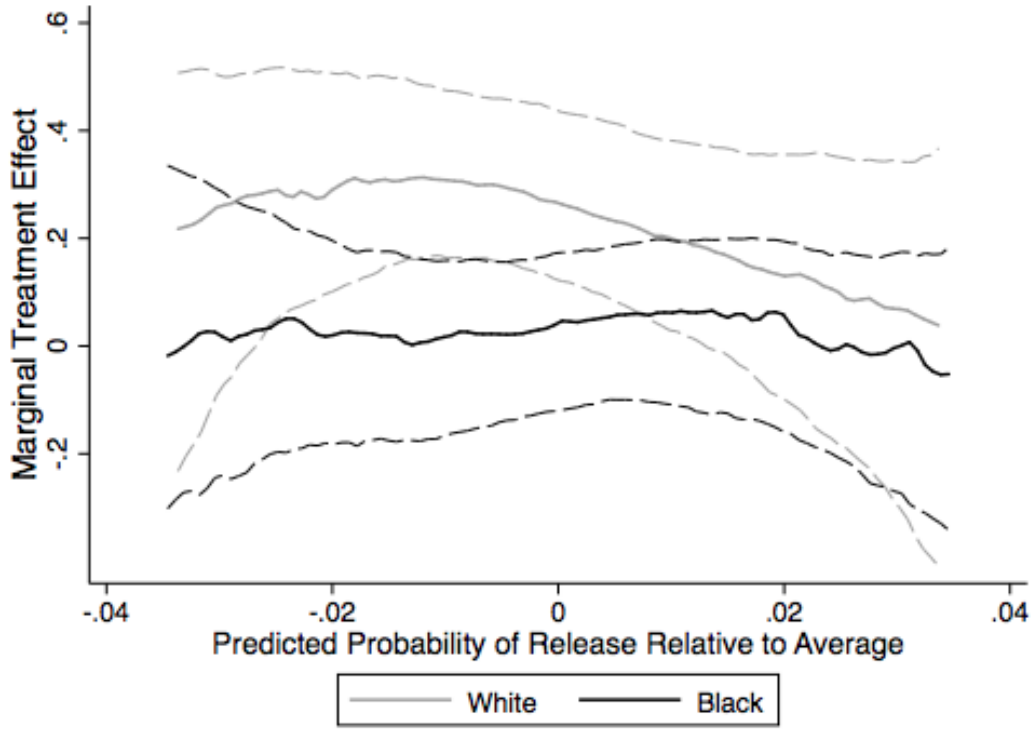
Marginal treatment effects

Consider the case where S_i consists of a single, continuous variable.

To apply these results, we need to construct flexible approximations of the CEFs. We can do this with least squares using flexible basis functions such as polynomials or some other series. For T_i, we can nest this inside a logit/probit and use MLE for estimation.

Arnold, Dobbie & Yang (2018): Figure (2)

Figure 2: Marginal Treatment Effects



Note: This figure reports the marginal treatment effects (MTEs) of pre-trial release on pre-trial rearrest separately by race. To estimate each MTE, we first estimate the predicted probability of release using only judge leniency. We then estimate the relationship between the predicted probability of release and rearrest prior to disposition using a local quadratic estimator (bandwidth = 0.030). Finally, we use the numerical derivative of the local quadratic estimator to calculate the MTE at each point in the distribution. Standard errors are computed using 500 bootstrap replications clustered at the judge-by-shift level. See the text for additional details.

LATE

Now suppose that the instrument only takes two values, $s = a, s = b$.

For example, suppose that in an RCT, you only randomize ITT (e.g. Moving to Opportunity).

Suppose that the instrument is relevant: $E[T_i|S_i = a] \neq E[T_i|S_i = b]$ and WLOG, assume $g(a) < g(b)$.

Before we looked at a ratio of derivatives. Here the analogue is a ratio of differences

$$\begin{aligned}\frac{E[Y_i|S_i = b] = E[Y_i|S_i = a]}{E[T_i|S_i = b] = E[T_i|S_i = a]} &= \beta + \frac{1}{g(b) - g(a)} \int_{g(a)}^{g(b)} E[U_{i1} - U_{i0}|V_i = v]a \\ &= \beta + E[U_{i1} - U_{i0}|g(a) \leq V_i \leq g(b)] \\ &= E[Y_i(1) - Y_i(0)|g(a) \leq V_i \leq g(b)] \\ &= LATE.\end{aligned}$$

LATE and Monotonicity

Imbens & Angrist (1994): The ratio of differences is the Wald Estimator for a binary instrument. Going to switch and set $b = 1, a = 0$. We have

$$\begin{aligned} Cov(S_i, Y_i) &= E[S_i Y_i] - E[S_i]E[Y_i] \\ &= E[Y_i|S_i = 1]E[S_i] \\ &\quad - (E[Y_i|S_i = 1]E[S_i] + E[Y_i|S_i = 0]E[1 - S_i])E[S_i] \\ &= (E[Y_i|S_i = 1] - E[Y_i|S_i = 0])E[S_i]E[1 - S_i]. \end{aligned}$$

Next, by a similar argument,

$$Cov(S_i, T_i) = (E[T_i|S_i = 1] - E[T_i|S_i = 0])E[S_i]E[1 - S_i].$$

So,

$$\frac{Cov(S_i, Y_i)}{Cov(S_i, T_i)} = \frac{E[Y_i|S_i = 1] - E[Y_i|S_i = 0]}{E[T_i|S_i = 1] - E[T_i|S_i = 0]}.$$

LATE and Monotonicity

Recall: We also defined a potential treatment function, $T_i(s)$ and the instrument being randomly assigned now means

$$\{Y_i(0), Y_i(1), T_i(), T_i(1)\} \perp\!\!\!\perp S_i.$$

The instrument also satisfies the exclusion restriction. Consider the numerator of the Wald estimator. We have

$$\begin{aligned}
&E[Y_i|S_i = 1] - E[Y_i|S_i = 0] \\
&\quad = E[Y_i(1)T_i(1) + Y_i(0)(1 - T_i(1))] \\
&\quad - E[Y_i(1)T_i(0) + Y_i(0)(1 - T_i(0))] \\
&\quad = E[(T_i(1) - T_i(0))(Y_i(1) - Y_i(0))] \\
&\quad = P(T_i(1) - T_i(0) = 1)E[(Y_i(1) - Y_i(0))|T_i(1) - T_i(0) = 1] \\
&\quad - P(T_i(1) - T_i(0) = -1)E[(Y_i(1) - Y_i(0))|T_i(1) - T_i(0) = -1
\end{aligned}$$

LATE and Monotonicity

Next, we have that

$$E[T_i|S_i = 1] - E[T_i|S_i = 0] = E[T_i(1) - T_i(0)]$$
$$= P(T_i - T_i(0) = 1) - P(T_i - T_i(0) = -1).$$

So, even if $Y_i(1) - Y_i(0) > 0$ w.p. 1, we can have that

$$\frac{E[Y_i|S_i = 1] - E[Y_i|S_i = 0]}{E[T_i|S_i = 1] - E[T_i|S_i = 0]} < 0.$$

Without additional assumptions, the Wald Estimator does not identify an average treatment effect for any sub-population!

LATE and Monotonicity

Imbens & Angrist (1994): Introduce monotonicity assumption.

$$P(T_i(1) \geq T_i(0)) = 1 \text{ or } P(T_i(1) \leq T_i(0)) = 1.$$

In other words, "one-sided non-compliance." In other other words, "no defiers" – population only consists of

Never-takers: $T_i(0) = T_i(1) = 0$

Always-takers: $T_i(1) = T_i(0) = 1$

Compliers: $T_i(1) = 1, T_i(0) = 0$

The Wald Estimator identifies an ATE for the compliers.

Under monotonicity, the Wald Estimator simplifies to

$$\begin{aligned}\frac{E[Y_i|S_i = 1] - E[Y_i|S_i = 0]}{E[T_i|S_i = 1] - E[T_i|S_i = 0]} &= E[Y_i(1) - Y_i(0)|\, T_i(1) - T_i(0) = 1] \\ &= LATE.\end{aligned}$$

Monotonicity and Choice Model

Vytlacil (2002): Monotonicity is **equivalent** to the threshold crossing model we presented at the beginning.

> The keys in the choice model are: (1) Additive separability of S_i, V_i and (2) $V_i \perp\!\!\!\perp S_i$.
>
> See lecture notes for proof.

In other words, monotonicity is equivalent to assuming a particular choice model for selection into treatment. If you're uncomfortable with the choice model, you are uncomfortable with monotonicity and vice versa.