02/15

$$Y = y_0 + (y_1 - y_0) D \qquad D \in \{0,1\}$$

$$Z \in \{0,1\}.$$

Exog (cond. on w): $(y_0, y_1, D_0, D_1) \perp\!\!\!\perp Z \mid W$

$$D = Z D_1 + (1-Z) D_0$$

Mon: $P(D_1 \geq D_0 \mid W) = 1$ for all realizations of W.

$$= D_0 + Z(D_1 - D_0)$$

Run $Y = \sum_{w \in W} \beta_{0,w} 1_{W=w} + \sum_{w \in W} \beta_{1,w} 1_{W=w} Z + U$

$$D = \sum_{w \in W} \pi_{0,w} 1_{W=w} + \sum_{w \in W} \pi_{1,w} Z 1_{W=w} + V.$$

$$\frac{\beta_{1,w}}{\pi_{1,w}} = \frac{Cov(Y, Z \mid W=w)}{Cov(D, Z \mid W=w)} = LATE(w)$$

$$= E(y_1 - y_0 \mid D_0 = 0, D_1 = 1, W=w).$$

In practice: $Y = \beta_0 + \beta_1 Z + W'\beta_2 + U$

$$D = \pi_0 + \pi_1 Z + W'\pi_2 + V$$

$$\beta_{IV} = \frac{E(Y\tilde{Z})}{E(D\tilde{Z})}, \text{ where } \tilde{Z} = Z - BLP(Z \mid W).$$

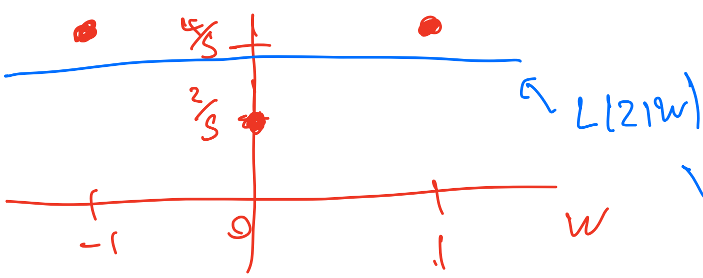
$$= E\left(LATE(w) \cdot \frac{[Cov(D, Z \mid W) + E(\tilde{Z} \mid W) E(Z \mid W) P(cp \mid W)]}{E(D\tilde{Z})} \right)$$

$$+ E\left(ATE(at, W) \cdot \frac{P(at \mid W)\, E(\tilde{Z} \mid W)}{E(D\tilde{Z})} \right).$$

Section 2.2 Blandhol et. al (2022):

$W = -1, 0, 1$ each with prob. $\frac{1}{3}$.

$$P(Z=1 \mid W) = E(Z \mid W) = \begin{cases} 4/5 & W \in \{-1, 1\} \quad \swarrow \text{ w.p. } \frac{2}{3} \\ 2/5 & W = 0 \quad \swarrow \text{ w.p. } \frac{1}{3} \end{cases}$$



$L(Z \mid W)$.

$$\frac{2}{3} \cdot \frac{4}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{2}{3}.$$

$$E(\tilde{Z} \mid W) = P(Z=1 \mid W) - \frac{2}{3} \leftarrow L(Z \mid W).$$

$$P(at \mid W) = \frac{1}{2} - \frac{|W|}{6} \qquad P(cp \mid W) = \frac{1}{6} + \frac{|W|}{6}$$

$$P(nt \mid W) = \frac{1}{3}$$

Simplify: $y_0 = 0$ so TE only depends on y_1.

Further assume $E(y_1 \mid cp, W) = \mu(cp)$

$E(y_1 \mid at, W) = \mu(at)$

$$\beta_{IV} = \frac{9}{7}\mu(cp) - \frac{2}{7}\mu(at).$$

Sloping downward in $\mu(at)$. Fix $\mu(cp) = \frac{1}{3}$

$$\frac{3}{7} - \frac{2}{7}\mu(at).$$

So when is $E(\tilde{Z} \mid W) \stackrel{?}{=} 0$

$$E(\tilde{Z} \mid W) = E(Z \mid W) - BLP(Z \mid W).$$

① $$D = \sum_{w \in W} 1(W=w)\pi_{0,w} + Z + V.$$

Regress $$Z = \sum_{w \in W} 1(W=w)\gamma_{0,w} + \tilde{Z}$$

Get $E(Z \mid W=w) = \gamma_{0,w}$, so $\tilde{Z} = Z - E(Z \mid W)$

$$E(\tilde{Z} \mid W) = E(Z \mid W) - E(Z \mid W) = 0.$$

② Z is completely randomly assigned (even independent of W).

$$E(\tilde{Z} \mid W) = E(Z \mid W) - BLP(Z \mid W)$$
$$= E(Z) - E(Z) = 0.$$

Multivalued Instruments

Can also generalize to multiple values of instrument, say different levels of a tuition subsidy $Z \in \{z_1, \ldots, z_K\}$.

Many potential treatments: $D(z_1), D(z_2), \ldots, D(z_K)$.

Exogeneity : $y_0, y_1, D(z_1), \dots, D(z_K) \perp\!\!\!\perp Z$

Monotonicity : $P(D(z_1) \leq D(z_2) \leq \dots \leq D(z_K)) = 1$.

Now have multiple types of complies:

Those who would take care with subsidy $\geq z_j$ but not otherwise are denoted cp_j. At extreme ends are always takers and never-takers.

$$P(D=0 \mid Z=z_K) = P(nt) \,; \quad P(D=1 \mid Z=z_1) = P(at).$$

$$P(D=1 \mid Z=z_2) = P(D(z_2)=1 \mid Z=z_2)$$

= Proportion of always takes and complies with lowest level of subsidy.

$$P(D=1 \mid Z=z_2) - P(D=1 \mid Z=z_1)$$ = Proportion of complies with lowest level of subsidy.

$\vdots$

$$P(D=1 \mid Z=z_K) - P(D=1 \mid Z=z_{K-1})$$ = Proportion of complies with highest level of subsidy.

$$P(at) + \sum_{j=2}^{K} \left[P(D=1 \mid Z=z_j) - P(D=1 \mid Z=z_{j-1}) \right]$$

$$= P(D=1 \mid Z=z_K)$$

Now add $P(D=0 \mid Z=z_K)$ $P(nt)$ get 1.

Now there are several ways to choose instruments.

1. Include Z linearly.
2. Split Z into indicators: $Z^* = \{1(Z=z_1), \ldots, 1(Z=z_K)\}$.

We could do the previous analysis keeping only 2 instrument values: z_j, z_{j-1}. Monotonicity holds for these 2 values,

So $$\beta_{IV} = \frac{E(Y \mid Z=z_j) - E(Y \mid Z=z_{j-1})}{E(D \mid Z=z_j) - E(D \mid Z=z_{j-1})}$$

$$= LATE(cp_j)$$

cp_j - those who take treatment when $Z=z_j$ but not z_{j-1}.

Now: Combine all instrument values and find β_{IV} for a general choice of scalar instrument: $c'Z^*$

$$Z^* = (1(Z=z_1), \ldots, 1(Z=z_K))',$$

(Suppose $c = [z_1, z_2, \ldots, z_K]'$. Then $c'Z^* = Z$.

$$c'Z^* = \sum_{j=1}^{K} z_j 1(Z=z_j))$$

For a general choice of instrument $c'Z^*$:

$$\beta_{IV} = \frac{Cov(Y, c'Z^*)}{Cov(D, c'Z^*)}$$

RF: $Y = \beta_0 + \beta_1 (c'Z^*) + U$

FS: $D = \pi_0 + \pi_1 (c'Z^*) + V$

Note: If we instead specify the first stage as
$D = Z^{*\prime} c + V; \; E(Z^* V) = 0$, coefficient we end up
with is $c = [E(D|Z=z_1), \ldots, E(D|Z=z_K)]$, so the instrument
is effectively $Z^{*\prime} c = E(D|Z) = P(D=1|Z)$ which
is the propensity score at Z.

$$\mathrm{Cov}(Y, c' Z^*)$$

$$= E(Y(c'Z^* - E(c'Z^*)))$$

$$= \sum_{j=1}^{K} E(Y|Z=z_j)(c_j - E(c'Z^*)) P(Z=z_j)$$

Law of total expectation

$$E(Y|Z=z_j) = E(Y|Z=z_1) + \sum_{l=2}^{j} E(Y|Z=z_l) - E(Y|Z=z_{l-1})$$

$$\mathrm{LATE}(cp_j) = \frac{E(Y|Z=z_j) - E(Y|Z=z_{j-1})}{\underbrace{P(D=1|Z=z_j)}_{p(z_j)} - \underbrace{P(D=1|Z=z_{j-1})}_{p(z_{j-1})}}$$

$$E(Y|Z=z_j) = E(Y|Z=z_1) + \sum_{l=2}^{j} \mathrm{LATE}(cp_l)(p(z_l) - p(z_{l-1}))$$

Examine 1st term in $\mathrm{Cov}(Y, c'Z^*)$ using $E(Y|Z=z_1)$.

$$= \sum^{K} E(Y|Z=z_1)(c_j - E(c'Z^*)) P(Z=z_j)$$

$$= E(Y|Z=z_1) \underbrace{\sum_{j=1}^{k} \left(c_j - E(c'Z^*)\right) P(Z=z_j)}_{= \sum_{j=1}^{k} c_j P(Z=z_j) - E(c'Z^*) = 0}.$$

$$\sum_{j=1}^{k} c_j P(Z=z_j) = E(c'Z^*)$$

So
$$\mathrm{Cov}(Y, c'Z^*) = \sum_{j=1}^{k} \sum_{l=2}^{j} \left(LATE(cp_l)\right)\left(p(z_l) - p(z_{l-1})\right)\left(c_j - E(c'Z^*)\right) P(Z=z_j)$$

Want to write this as a weighted average of LATEs, so find a way to swap order of summation:

$\mathrm{Cov}(Y, c'Z^*) =$

$$\sum_{j=1}^{k} \sum_{l=2}^{k} \mathbb{1}(l \le j) \left(LATE(cp_l)\right)\left(p(z_l) - p(z_{l-1})\right)\left(c_j - E(c'Z^*)\right) P(Z=z_j)$$

$$= \sum_{l=2}^{k} LATE(cp_l)\left(p(z_l) - p(z_{l-1})\right)\left(\sum_{j=\cancel{1}l}^{k} \cancel{\mathbb{1}(l \le j)}\left(c_j - E(c'Z^*)\right) P(Z=z_j)\right)$$

$$= \sum_{l=2}^{k} \left[\left(p(z_l) - p(z_{l-1})\right) \sum_{j=l}^{k} \left(c_j - E(c'Z^*)\right) P(Z=z_j)\right] LATE(cp_l)$$

So $$\beta_{IV} = \sum_{l=2}^{k} \left[\frac{\left(p(z_l) - p(z_{l-1})\right) \sum_{j=l}^{k} \left(c_j - E(c'Z^*)\right) P(Z = z_j)}{Cov(D, c'Z^*)} \right] LATE(c_{p_l})$$

Sign of weights? $p(z_l) - p(z_{l-1}) \geq 0$ by Monotonicity.

$$\sum_{j=l}^{k} \left(c_j - E(c'Z^*)\right) P(Z = z_j) = E\left(c'Z^* 1(Z \geq z_l)\right) - E(c'Z^*) P(Z \geq z_l)$$

$$= \left(E(c'Z^* | Z \geq z_l) - E(c'Z^*)\right) P(Z \geq z_l)$$

$$\geq 0$$ if $c'Z^*$ is larger when Z is larger.

This would hold when including Z linearly. The denominator will also be positive in this case (exercise).

Note: Weights depend on magnitude of Z if Z is included linearly. This isn't desirable if it means that the LATE for those who comply only at the largest subsidy value is highly weighted when transitions from, say, \$300 to \$500 induce more individuals to change their treatment choice than a shift from \$500 to \$800 would. Using 2SLS with Z^* (as noted above) in the first stage leads to $Z^{*\prime}c = p(Z)$, which avoids this issue, but is then susceptible to many instruments bias.

Multiple Instruments

Suppose some individuals are offered a subsidy and others are offered the opportunity to attend the course remotely. Individuals are offered neither, one or both: $Z \in \{(0,0), (1,0), (0,1), (1,1)\}$.

Now monotonicity is much harder to justify. How to order $(1,0)$, $(0,1)$? An American studying Spanish with a teacher in Spain may be more concerned about whether the class is offered remotely than the cost, but a student who can attend in-person may be more likely to do so with a subsidy: Not all agents respond to the instruments the same way, just as not all agents have the same response to the treatment itself.

Abadie's Kappa Abadie (2003). 02/17

Methods for finding ATEs so far:

① Estimate $\beta_{IV}^{(w)} = \dfrac{Cov(Y, Z \mid W)}{Cov(D, Z \mid W)}$ = LATE (W).
at each W.
(curse of dimensionality)

② Reduce dimension by including covariates linearly: Avoid many instruments bias + curse of dimensionality but lose ability to easily interpret β_{IV}.

③ Abadie's K: Note that
$$E(y_1 - y_0 \mid W, D_1 > D_0) \quad (= LATE(w))$$

Strong Exogeneity $= E(Y \mid W, D=1, D_1 > D_0) - E(Y \mid W, D=0, D_1 > D_0)$.

So... Try to estimate $E(Y \mid W, D, D_1 > D_0)$

Problem: Don't know who is a complier, but we can find the complier treatment effect using an appropriate weighting scheme.

① Impose restriction: $E(Y \mid W, D, D_1 > D_0) = h(D, W, \theta_0)$
where $\theta_0 \in \Theta \rightarrow$ Compact subset of $\mathbb{R}^d$.
$\{h(\cdot, \cdot, \theta) : \theta \in \Theta\}$ comprises a parametric class of functions (h is known).

Let $g(Y, D, W) = (Y - E(Y \mid D, W, D_1 > D_0))^2$

$$= (Y - h(D, W, \theta_0))^2 .$$

θ_0 satisfies $\theta_0 \in \underset{\theta}{\text{argmin}}\ E(g(Y, D, W) \mid D_1 > D_0)$

Conditional expectation for compliers.

$$E(K g(Y, D, W))$$

$$= E(K(Y - h(D, W, \theta))^2)$$

where $K = 1 - \frac{D(1-Z)}{P(Z=0 \mid W)} - \frac{(1-D)Z}{P(Z=1 \mid W)}$.

$$\text{Min} \sum_{i=1}^{n} \hat{K}_i (y_i - h(D_i, W_i, \theta))^2 .$$

Probs. need to be estimated

$y = x'\beta + u$ $\quad \sum (y_i - x_i'\beta)^2$

$y = h(x, \beta) + u$ $\quad \min_{\beta} \sum (y_i - h(x_i, \beta))^2$ Non-linear least squares.

"Extremum Estimation" - finding minimizer (extreme value) of some

$\beta_0 \in \arg\min_{\beta} E[(y - h(x, \beta))^2]$. criterion function.

$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$.

"Asymptotic normality of extremum estimators".

<u>Treatment choice</u>

Suppose utility from choosing d is $U(d) - u(d \mid Z, W)$.

Choose $d = 1 \iff U(1) - u(1 \mid Z, W) \geq U(0) - u(0 \mid Z, W)$.

$$D = \mathbb{1}(\underbrace{U(0) - U(1)}_{U} \leq \underbrace{u(0 \mid Z, W) - u(1 \mid Z, W)}_{u(Z, W)}).$$

is called the <u>selection equation</u>. "Threshold crossing model".

① U is cont. distributed conditional on $W = w$.

② $(y_0, y_1, U) \perp\!\!\!\perp Z \mid W$.

- <u>Roy model</u> is composed of selection equation and these assumptions. $U(1), U(0)$ may be related to y_1 and y_0

eg. $U(d)$ is predicted wage after attending/not computing class
$Y(d)$ is actual wage after attending/not.

Selection equation gives us potential treatments:

$$D(z) = \mathbb{1}(U \leq u(z,W)).$$

- Vytlacil (2002) showed this formulation is equivalent to "LATE" model described last week. $D(z)$ generated by the Roy model satisfy strong exogeneity and monotonicity. Conversely, if SE and Mon. hold in "LATE" model, $\exists$ U, $u(\cdot,\cdot)$ s.t. (1) and (2) hold with $D = \mathbb{1}(U \leq u(Z,W))$.

Vytlacil

Vytlacil equivalence Theorem.

$\Rightarrow$: $D(z) = \mathbb{1}(U \leq u(z,W))$ depends on U and W.

Since $\underset{(2)}{U \perp\!\!\!\perp Z | W}$, $D(z) \perp\!\!\!\perp Z | W$, so

$$D_1, D_0 \perp\!\!\!\perp Z | W$$

$$(Y_1, Y_0, D_1, D_0) \perp\!\!\!\perp Z | W.$$

So strong exogeneity in "LATE" formulation holds.

Conditional on $W = w$: $D(0) = \mathbb{1}(U \leq u(0,w))$

$$\leq \mathbb{1}(U \leq u(1,w)) = D(1)$$

$\Leftrightarrow$ $u(1,w) \geq u(0,w)$.

So either $D(1) \geq D(0)$ or $D(0) \geq D(1)$ w.p. 1.

At each $W = w$, monotonicity holds (but order might

$\Leftarrow$: See Vytlacil (2002)

change across values of W).

Marginal Treatment Effects

Consider ATEs for a particular value of U.

$$E(y_1 - y_0 \mid U = u) \qquad U|W \sim F(\cdot \mid w).$$

$$p(z,w) = P(D=1 \mid Z=z, W=w)$$
$$= P(U \leq u(z,w) \mid Z=z, W=w)$$

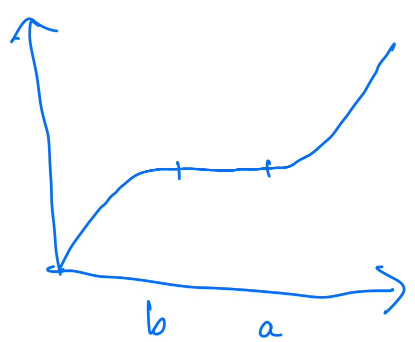
$$U \perp\!\!\!\perp Z \mid W \quad = P(U \leq u(z,w) \mid W=w).$$
$$= F(u(z,w) \mid W=w).$$

$$D = 1(U \leq u(Z,W))$$

$$= 1(\underbrace{F(U|W)}_{R} \leq \underbrace{F(u(Z,W)|W)}_{p(Z,W)})$$

w.p. 1 since U is continuous and CDFs are weakly increasing.

$$a \leq b \iff F(a) \leq F(b).$$



Note: R has a uniform dist.

$$P(R \leq r \mid W=w) =$$

$$P(\underset{\uparrow}{F(U|W=w)} \leq r \mid W=w) = r.$$

CDF of U applied to U has a uniform distribution (probability integral transform).

R is independent of W. Since R is a function of U, W, so since $U \perp\!\!\!\perp Z \mid W$, R is independent of $Z \mid W$, so $R \perp\!\!\!\perp (Z, W)$.

$D = \mathbb{1}\left(R \leq p(Z,W)\right)$.

Drop covariates and set $Z \in \{0, 1\}$. Suppose $p(1) \geq p(0)$.

$R < p(0) \rightarrow$ Always taker

$p(0) < R \leq p(1) \rightarrow$ Complier

$R > p(1) \rightarrow$ Never Taker.

"Marginal" Treatment effect: $E(y_1 - y_0 \mid R = r, W = w) = m(r, w)$.

$\hookrightarrow$ $R = p(1) \rightarrow$ these individuals are indifferent between taking the class and not when they receive a subsidy to do so (on the margin)

The MTE can be used to represent ATE, ATT and LATE as an integral of m, weighted by different functions.

Eg: $\text{ATE} = E\big(E(y_1 - y_0 \mid R, W) \big)$

$$= E\big(E(E(y_1 - y_0 \mid R, W) \mid W) \big)$$

$$= E\big(E(m(R, W) \mid W) \big)$$

$(R \perp\!\!\!\perp W)$ $$= E\left(\int_0^1 m(r, W)\, dr \right)$$

$$\text{ATT} = E(y_1 - y_0 \mid D = 1)$$

$$= E\left(\frac{(y_1 - y_0) D}{P(D=1)} \right)$$

$D = \mathbb{1}(R \leq p(Z, W))$.

$$= E\left(\frac{E(D\, E(y_1 - y_0 \mid Z, R, W) \mid W)}{P(D=1)} \right)$$

$$= E\left(\frac{E(D\, E(y_1 - y_0 \mid R, W) \mid W)}{P(D=1)} \right)$$

$(y_1, y_0) \perp\!\!\!\perp Z \mid W$,

$R \perp\!\!\!\perp Z \mid W$.

$$= E\left(\int_0^1 \frac{\mathbb{1}(r \leq p(Z, W))}{P(D=1)}\, m(r, W)\, dr \right)$$

$$LATE(w) = E(y_1 - y_0 \mid \underbrace{D_1 > D_0, W = w}_{\Leftrightarrow\ p(0,w) \le R \le p(1,w)})$$

$$= E\left(E\left(y_1 - y_0 \mid \cancel{D_1 > D_0}, R, W=w \right) \mid D_1 > D_0, W = w \right)$$

Given R, W, $\{D_1 > D_0\}$ is known already.

$$= E\left(m(R,w) \mid D_1 > D_0, W=w \right)$$

$$= E\left(\frac{m(R,w)\, \mathbb{1}(p(0,w) \le R \le p(1,w))}{P(p(0,w) \le R \le p(1,w) \mid W=w)} \mid W=w \right)$$

$$(R \perp W) = \int_0^1 m(r,w) \frac{\mathbb{1}(p(0,w) \le r \le p(1,w))}{p(1,w) - p(0,w)}\, dr$$

In all cases, the weights can be estimated, so if we know $m(r,w)$ then we know the average effect. Different estimands require knowledge of integrals of m over different regions. With the LATE, it would suffice to know m on the region $p(0,w) \le r \le p(1,w)$ for each w.