A design perspective of instrumental variable method

Xinzhou Guo

HKUST

(Credited to Zhichao Jiang)

April 17, 2024

① Example : Encouragement trial

② Define Potential Outcome on Z :

Causal Quantity : Complier, ...

$E(Y(1) - Y(0) \mid \text{complier})$

identiability assumption : estimate

Instrumental variable method

- A general approach to identify causal effects with latent confounding
  1. polular in econometrics
  2. relies on the existence of an additional variable
  3. many controversial applications due to the strong assumptions

- Two perspectives
  - design perspective: encouragement design
  - econometric perspective: two-stage least squares

Encouragement Design

- Often, for ethical and logistical reasons, we cannot force all experimental units to follow the randomized treatment assignment
  1. some in the treatment group refuse to take the treatment
  2. some in the control group manage to receive the treatment

- Encouragement design: randomize the encouragement to receive the treatment rather than the receipt of the treatment itself

- Can we estimate the treatment effect (effect of the actually received treatment)?

Notation

- Randomized encouragement: $Z_i \in \{0, 1\}$
- SUTVA holds
- Potential treatment variables: $(D_i(1), D_i(0))$
  1. $D_i(z) = 1$: would receive the treatment if $Z_i = z$
  2. $D_i(z) = 0$: would not receive the treatment if $Z_i = z$
- Potential outcome: $Y_i(z)$ – why do we define on Z instead of D?
- Observed treatment receipt indicator: $D_i = D_i(Z_i)$ $= Z_i \cdot Y_i(1) + (1 - Z_i) Y_i(0)$
- Observed and potential outcomes: $Y_i = Y_i(Z_i)$ $= Z_i \cdot Y_i(1) + (1 - Z_i) Y_i(0)$

Effect of the treatment receipt?

- Randomization: $\{Y_i(1), Y_i(0), D_i(1), D_i(0)\} \perp Z_i$
- Intention-to-treat analysis:

$$\mathbb{E}\{Y_i(1) - Y_i(0)\} = \mathbb{E}(Y_i \mid Z_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0)$$

  - intention to treat effect: effect of the treatment assignment (not useful)
  - ITT analysis does not yield the treatment effect
- As-treated analysis: $\mathbb{E}(Y_i \mid D_i = 1) - \mathbb{E}(Y_i \mid D_i = 0)$
  - comparison of the treated and untreated subjects
  - no benefit of randomization $\rightsquigarrow$ selection bias
- Per-protocol analysis: $\mathbb{E}(Y_i \mid Z_i = 1, D_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0, D_i = 0)$
  - comparison of the treated and untreated subjects who follow the treatment assignment
- What's the definition of treatment effect?

$$E(Y_i(1) - Y_i(0) \mid \text{complier})$$
$$= E(Y_i(1) - Y_i(0) \mid D_i(1) = 1, D_i(0) = 0)$$

$$\underline{E(Y_i \mid Z_i = 1, D_i = 1)}$$
$$= E(Y_i(1) \mid \underline{Z_i = 1}, \underline{D_i(1) = 1})$$
$$= E(Y_i(1) \mid D_i(1) = 1)$$

$$E(Y_i(0) \mid D_i(0) = 0)$$

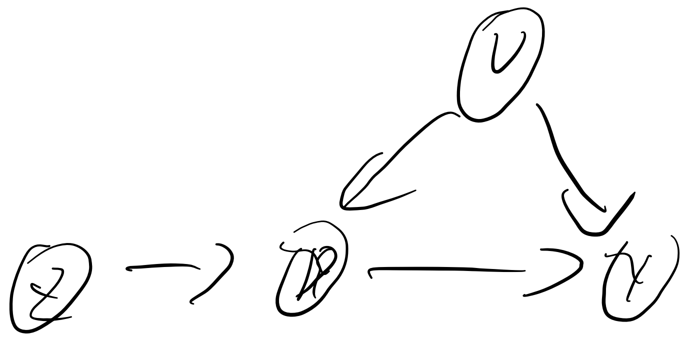
Compliance behavior

- Four principal strata (latent types):
  - compliers $(D_i(1), D_i(0)) = (1, 0)$,
  - non-compliers $\begin{cases} \text{always} - \text{takers} & (D_i(1), D_i(0)) = (1, 1), \\ \text{never} - \text{takers} & (D_i(1), D_i(0)) = (0, 0), \\ \text{defiers} & (D_i(1), D_i(0)) = (0, 1) \end{cases}$
  - denote the compliance behavior (a, n, c, d) by $U_i \rightsquigarrow D_i$ is a function of Z_i and U_i
- Observed strata and compliance behavior:

	$Z_i = 1$	$Z_i = 0$
$D_i = 1$	Complier/Always-taker	Defier/Always-taker
$D_i = 0$	Defier/Never-taker	Complier/Never-taker

Assumptions

- Randomization: $\{Y_i(1), Y_i(0), D_i(1), D_i(0)\} \perp Z_i$
- Monotonicity: $D_i(1) \geqslant D_i(0) \rightsquigarrow$ no defiers
- Exclusion restriction: $Y_i(1) = Y_i(0)$ for always takers and never-takers
- $\mathbb{E}\{D_i(1) - D_i(0)\} > 0 \rightsquigarrow$ there exists compliers



$$Y(1) = Y(Z=1)$$

$$Y(0) = Y(Z=0)$$

Monotonicity: implications

- Observed strata and compliance behavior under monotonicity

	$Z_i = 1$	$Z_i = 0$
$D_i = 1$	Complier/Always-taker	Always-taker
$D_i = 0$	Never-taker	Complier/Never-taker

-

$$
\begin{aligned}
\text{pr(never-taker)} &= \text{pr}\left(D_i = 0 \mid Z_i = 1\right) \\
\text{pr(always-taker)} &= \text{pr}\left(D_i = 1 \mid Z_i = 0\right) \\
\text{pr(complier)} &= \text{pr}\left(D_i = 1 \mid Z_i = 1\right) - \text{pr}\left(D_i = 1 \mid Z_i = 0\right)
\end{aligned}
$$

complier or always-taker

always-taker

Always-taker $E(Y(1) - Y(0) \mid D(1) = D(0) = 1)$

$= 0$

Exclusion restriction: implications

- ITT effect for units with different compliance behavior
  - always-taker: $\text{ACE}_a = \mathbb{E}\{Y_i(1) - Y_i(0) \mid D_i(1) = D_i(0) = 1\} = 0$
  - never-taker: $\text{ACE}_n = \mathbb{E}\{Y_i(1) - Y_i(0) \mid D_i(1) = D_i(0) = 0\} = 0$
  - complier: $\text{ACE}_c = \mathbb{E}\{Y_i(1) - Y_i(0) \mid D_i(1) = 1, D_i(0) = 0\}$
- ACE_c : Complier Average Causal Effect (CACE) or Local Average Treatment Effect (LATE)
  - for compliers: $Z_i = D_i$, ITT effect = treatment effect – we can estimate.

$$D(1) - D(0) = \begin{cases} 1 \\ 0 \end{cases}$$

$$\begin{aligned}
ACE_D &= E(D(1) - D(0)) \\
&= E\, I\{D(1) - D(0) = 1\} \\
&= P(D(1) - D(0) = 1) \\
&= P(D(1) = 1,\ D(0) = 0) = P(\text{complier})
\end{aligned}$$

Complier average causal effect

- ITT effect decomposition:

$$\begin{aligned}\mathrm{ACE}_Y = & \mathrm{ACE}_c \times \mathrm{pr}(\text{ compliers }) + \mathrm{ACE}_a \times \mathrm{pr}(\text{ always } - \text{ takers }) \\ & + \mathrm{ACE}_n \times \mathrm{pr}(\text{ never } - \text{ takers }) \\ = & \mathrm{ACE}_c \times \mathrm{pr}(\text{ compliers })\end{aligned}$$

- ITT effect (why?) on D : $\mathrm{ACE}_D = \mathrm{pr}\ (\text{compliers})$

- $\mathrm{ACE}_c = \mathrm{ACE}_Y / \mathrm{ACE}_D$

Identification

- ACE_Y and ACE_D are identifiable
- Identification formula:

$$\begin{aligned}\text{ACE}_c &= \frac{\mathbb{E}\left(Y_i \mid Z_i = 1\right) - \mathbb{E}\left(Y_i \mid Z_i = 0\right)}{\mathbb{E}\left(D_i \mid Z_i = 1\right) - \mathbb{E}\left(D_i \mid Z_i = 0\right)} \\ &= \frac{\text{Cov}\left(Y_i, Z_i\right)}{\text{Cov}\left(D_i, Z_i\right)}\end{aligned}$$

- Z acts as an IV for D – why?

Complier average causal effect

- CACE $= \mathbb{E}\{Y_i(1) - Y_i(0) \mid D_i(1) = 1, D_i(0) = 0\}$
  - average effect of encouragement for compliers
  - average treatment effect for compliers $(Z_i = D_i)$
- CACE $\neq$ treatment effect unless the treatment effect for non-compliers equals CACE
- Different encouragement yields different compliers

Alternative notation

$Y(z, d)$

$Y(1, 0)$?

- Potential outcome: $Y_i(z, d)$
  - $Y_i = Y_i(Z_i) = Y_i(Z_i, D_i(Z_i))$
- Exclusion restriction: $Y_i(z, d) = Y_i(d)$
  - $\text{ACE}_Y = \mathbb{E}\{Y_i(D_i(1)) - Y_i(D_i(0))\}$
  - $\text{ACE}_c = \mathbb{E}\{Y_i(d = 1) - Y_i(d = 0) \mid U_i = c\}$

Inference

- Wald estimator: $\widehat{\text{CACE}}_{\text{Wald}} = \dfrac{\widehat{\text{ACE}}_Y}{\widehat{\text{ACE}}_D}$
- Consistency: $\widehat{\text{CACE}}_{\text{Wald}} \xrightarrow{p} \text{CACE} = \text{ACE}_c$
- Asymptotic variance via the Delta method:

$$\operatorname{var}\left(\widehat{\text{CACE}}_{\text{Wald}}\right) \approx \frac{1}{\text{ACE}_D^4}\left\{\text{ACE}_D^2 \operatorname{var}\left(\widehat{\text{ACE}}_Y\right) + \text{ACE}_Y^2 \operatorname{var}\left(\widehat{\text{ACE}}_D\right) - 2\text{ACE}_Y \text{ACE}_D \operatorname{cov}\left(\widehat{\text{ACE}}_Y, \widehat{\text{ACE}}_D\right)\right\}.$$

where

$$\operatorname{cov}\left(\widehat{\text{ACE}}_Y, \widehat{\text{ACE}}_D\right) = \frac{\operatorname{cov}\left(Y_i(1), D_i(1)\right)}{n_1} + \frac{\operatorname{Cov}\left(Y_i(0), D_i(0)\right)}{n_0}$$

$$\frac{1}{n_1}\sum_{Z_i=1}(Y_i - \bar{Y})(D_i - \bar{D})$$

$$\sqrt{n}\begin{pmatrix} \widehat{ACE}_Y - ACE_Y \\ \widehat{ACE}_D - ACE_D \end{pmatrix} \rightarrow N(0, \Sigma)$$

$$\sqrt{n}\left(\frac{\widehat{ACE}_Y}{\widehat{ACE}_D} - \frac{ACE_Y}{ACE_D} \right) \rightarrow ?$$

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \rightarrow N(0, \Sigma)$$

$$\sqrt{n}\left(g(\hat{\theta}_1, \hat{\theta}_2) - g(\theta_1, \theta_2) \right)$$

$$\rightarrow N(0, \nabla g^T(\theta_1, \theta_2)\, \Sigma\, \nabla g(\theta_1, \theta_2))$$

$$\widehat{CACE_{wald}} - CACE$$

$$= \frac{\widehat{ACE}_Y}{\widehat{ACE}_D} - \frac{ACE_Y}{ACE_D}$$

Variance calculation

$$\underbrace{\widehat{\text{CACE}}_{\text{Wald}}} - \text{CACE} \approx \frac{\widehat{\text{ACE}}_Y - \text{CACE} \cdot \widehat{\text{ACE}}_D}{\widehat{\text{ACE}}_D}$$

- $\widehat{\text{ACE}}_Y - \text{CACE} \cdot \widehat{\text{ACE}}_D$ is the difference in means estimator for the adjusted outcome $A_i = Y_i - \text{CACE} \cdot D_i$ – what is the interpretation?
- Variance estimation steps
  - obtain the adjusted outcome $A_i = Y_i - \widehat{\text{CACE}} \cdot D_i$
  - obtain the variance estimate based on the adjusted outcome

$$\widehat{V}_A = \frac{\widehat{\text{var}}\{A_i(1)\}}{n_1} + \frac{\widehat{\text{var}}\{A_i(0)\}}{n_0}$$

  - obtain the final variance estimate $\widehat{V}_A / \widehat{\text{ACE}}_D^2$

Weak instrumental variable

- $\widehat{\text{CACE}}_{\text{Wald}} = \frac{\widehat{\text{ACE}}_Y}{\widehat{\text{ACE}}_D}$ has poor properties when ACE_D is close to 0 – weak instrument
  - $\widehat{\text{CACE}}_{\text{Wald}}$ has finite sample bias
  - confidence interval has poor coverage rate

- How do we deal with a weak instrument (encouragement)?
  - testing: $H_0 : \text{CACE} = 0 \Longleftrightarrow H_0' : \text{ACE}_Y = 0$
  - confidence interval: Anderson-Rubin type confidence interval based on Fieller (1954)

Anderson-Rubin type confidence interval

H_0:

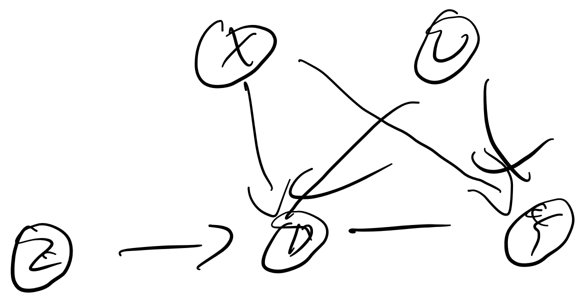
- Given the true value of CACE, $\mathbb{E}\left\{\widehat{\text{ACE}}_Y - \text{CACE} \cdot \widehat{\text{ACE}}_D\right\} = 0$
- Inverting hypothesis tests to obtain a confidence interval
- Define the adjusted outcome $A_i(t) = Y_i - tD_i$
  - ACE on $A_i(t)$: $\widehat{\text{ACE}}_A(t)$; variance estimator: $\widehat{V}_A(t)$
  - $\widehat{A}_i(t)$ and $\widehat{V}_A(t)$ are functions of t
  - $\mathbb{E}\left\{\widehat{\text{ACE}}_A(\text{CACE})\right\} = 0$ – test object and translated to $\widehat{\text{ACE}}_A(\text{CACE} = 0$
- $\text{pr}\left(\left|\frac{\widehat{\text{ACE}}_A(\text{CACE})}{\sqrt{\widehat{V}_A(\text{CACE})}}\right| \leqslant 1.96\right) = 0.95$
- Solve a quadratic inequality to obtain the confidence interval

$$\underbrace{(Y(1), Y(0), D(1) - D(0))}\perp \underline{Z}$$

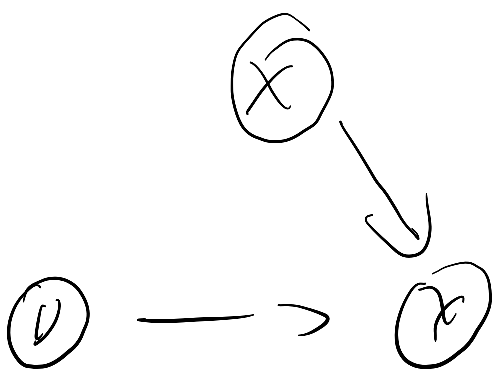
$$(Y(1) - tD(1), Y(0) - tD(0)) \perp Z$$

Evaluation of job training program

- A randomized field experiment investigating the efficacy of a job training intervention on unemployed workers
- Encouragement: Z; participation: D; job-search self-efficacy: Y
- Assumption:
  - Monotonicity: being encouraged would never discourage anyone from participating
  - Exclusion restriction: being encouraged has no effect on other than through participation ~~in the program~~
- CACE: effect of the encouragement (participation) on job-search self-efficacy for people who would participate if and only if they are encouraged
- ITT: est. $= 0.067$, s.e. $= 0.050, 95\%\text{CI} = [-0.031, 0.166]$
- CACE: est. $= 0.109$, s.e. $= 0.081, 95\%\text{CI} = [-0.050, 0.268]$



$Z \perp (D(1), D(0), Y(1), Y(0))$



$D \perp (Y(1), Y(0))$

$\hat{Y}_1 = \bar{Y}_1 - \bar{X}_1 \hat{r}_1$

$\hat{Y}_1 - \hat{Y}_0$

Covariate adjustment

$$\mathrm{CACE} = \frac{\mathrm{ACE}_Y}{\mathrm{ACE}_D}$$

- We can obtain the adjusted estimator for both the ACEs on D and Y

$$\widehat{\mathrm{ACE}}_{D,L} = \left\{\bar{D}(1) - \hat{\beta}_{D1}^{\top}\bar{X}(1)\right\} - \left\{\bar{D}(0) - \hat{\beta}_{D0}^{\top}\bar{X}(0)\right\}$$

$$\widehat{\mathrm{ACE}}_{Y,L} = \left\{\bar{Y}(1) - \hat{\beta}_{Y1}^{\top}\bar{X}(1)\right\} - \left\{\bar{Y}(0) - \hat{\beta}_{Y0}^{\top}\bar{X}(0)\right\}$$

- $\widehat{\mathrm{CACE}}_L = \frac{\widehat{\mathrm{ACE}}_{Y,L}}{\widehat{\mathrm{ACE}}_{D,L}}$
- Asymptotically more efficient than the simple ratio estimator

Variance calculation

$$\widehat{\mathrm{CACE}}_L - \mathrm{CACE} \approx \frac{\widehat{\mathrm{ACE}}_{Y,L} - \mathrm{CACE} \cdot \widehat{\mathrm{ACE}}_{D,L}}{\mathrm{ACE}_D}$$

- $\widehat{\mathrm{ACE}}_{Y,L} - \mathrm{CACE} \cdot \widehat{\mathrm{ACE}}_{D,L}$ is the difference in means estimator for the adjusted outcome

$$A_i = \begin{cases} \left(Y_i - \widehat{\beta}_{Y1}^{\top} X_i\right) - \mathrm{CACE} \cdot \left(D_i - \widehat{\beta}_{D1}^{\top} X_i\right) & \text{if } Z_i = 1 \\ \left(Y_i - \widehat{\beta}_{Y0}^{\top} X_i\right) - \mathrm{CACE} \cdot \left(D_i - \widehat{\beta}_{D0}^{\top} X_i\right) & \text{if } Z_i = 0 \end{cases}$$

- Variance estimation steps
  - obtain the adjusted outcome
  - obtain the variance estimate based on the adjusted outcome

$$\widehat{V}_{A,L} = \frac{\widehat{\mathrm{var}}\{A_i(1)\}}{n_1} + \frac{\widehat{\mathrm{var}}\{A_i(0)\}}{n_0}$$

  - obtain the final variance estimate $\widehat{V}_A / \widehat{\mathrm{ACE}}_{D,L}^2$

Evaluation of a job training program

- Do covariate adjustment for the ITT effects
  - center the covariates (without intercept)
  - regress Y on X for both groups to obtain $\widehat{\beta}_{Y1}$ and $\widehat{\beta}_{Y0}$
  - regress D on X for both groups to obtain $\widehat{\beta}_{D1}$ and $\widehat{\beta}_{D0}$
- Calculate the ITT effects: $\widehat{\text{ACE}}_{D,L} = 0.617, \widehat{\text{ACE}}_{Y,L} = 0.059$
- Calculate the CACE: $\widehat{\text{CACE}}_L = \widehat{\text{ACE}}_{Y,L} / \widehat{\text{ACE}}_{D,L} = 0.096$
- Obtain the adjusted outcome

$$A_i = \begin{cases} \left(Y_i - \widehat{\beta}_{Y1}^\top X_i\right) - \widehat{\text{CACE}} \cdot \left(D_i - \widehat{\beta}_{D1}^\top X_i\right) & \text{if } Z_i = 1 \\ \left(Y_i - \widehat{\beta}_{Y0}^\top X_i\right) - \widehat{\text{CACE}} \cdot \left(D_i - \widehat{\beta}_{D0}^\top X_i\right) & \text{if } Z_i = 0 \end{cases}$$

- Calculate the variance estimate $\widehat{V}_{A,L} / \widehat{\text{ACE}}_{D,L}^2 = 0.006$
-

$$\text{CACE: est.} = 0.109, \text{ s.e.} = 0.081, 95\%\text{CI} = [-0.050, 0.268]$$

$$\text{CACE (covs): est.} = 0.096, \text{ s.e.} = 0.080, 95\%\text{CI} = [-0.061, 0.252]$$

Necessary condition for identification

- Are exclusion restriction and monotonicity necessary?
- Necessary condition for identification: number of observed frequencies is larger than the number of parameters
  - observed data: $\mathrm{pr}\left(Z_i, D_i, Y_i\right) \rightsquigarrow 7$ observed frequencies
  - compliance behavior: $U_i = a, n, c$
  - parameters: $\mathrm{pr}(Y, Z, U) = \mathrm{pr}(Z)\,\mathrm{pr}(U \mid Z)\,\mathrm{pr}(Y \mid Z, U)$
  - number of parameters $= \begin{cases} 12 & \text{no asm.} \\ 9 & \text{with mon.} \\ 10 & \text{with ex.} \\ 7 & \text{with both} \end{cases}$
- Without exclusion restriction or monotonicity, CACE is not identifiable

$$E(Y(z) \mid U)$$

$$= E(Y \mid Z = z, U = u)$$

$$E(Y \mid Z = 1, U = c)$$

$$= \frac{E(DY \mid Z = 1) - E(DY \mid Z = 0)}{P(\text{complier})}$$

Identification for the joint distribution

- Proportions of principal strata

$$\pi_a = \mathrm{pr}(D = 1 \mid Z = 0), \pi_n = \mathrm{pr}(D = 0 \mid Z = 1), \pi_c = 1 - \pi_a - \pi_n$$

-

$$\mathbb{E}\{Y(z) \mid U\} = \mathbb{E}(Y \mid Z = z, U) \text{ for } U = a, n$$

exclusion ⇒

$$\mathbb{E}\{Y(1) \mid U = a\} = \mathbb{E}\{Y(0) \mid U = a\} = \mathbb{E}(Y \mid Z = 0, D = 1)$$

$$\mathbb{E}\{Y(1) \mid U = n\} = \mathbb{E}\{Y(0) \mid U = n\} = \mathbb{E}(Y \mid Z = 1, D = 0)$$

-

$$\mathbb{E}\{Y(z) \mid U\} = \mathbb{E}(Y \mid Z = z, U) \text{ for } U = c$$

$$\mathbb{E}(Y \mid z = 1, U = c) = \pi_c^{-1}\{\mathbb{E}(DY \mid Z = 1) - \mathbb{E}(DY \mid Z = 0)\}$$

$$\mathbb{E}(Y \mid z = 0, U = c) = \pi_c^{-1}\{\mathbb{E}((1 - D)Y \mid Z = 0) - \mathbb{E}((1 - D)Y \mid Z = 1)\}$$

- Can replace Y with $\mathbf{1}(Y = y)$ to obtain $\mathrm{pr}(Y = y \mid Z = z, U)$

Testable conditions for IV assumptions

How do we test the assumptions?

- Assumptions imply conditions that can be used to falsify the assumptions
- Testable conditions for exclusion restriction and monotonicity

$$
\begin{cases}
\operatorname{pr}(Y_i = y, D_i = 1 \mid Z_i = 1) \geqslant \operatorname{pr}(Y_i = y, D_i = 1 \mid Z_i = 0) \\
\operatorname{pr}(Y_i = y, D_i = 0 \mid Z_i = 1) \leqslant \operatorname{pr}(Y_i = y, D_i = 0 \mid Z_i = 0)
\end{cases}
$$

- Statistical test for testable conditions

$$
\mathbb{E}(Q \mid Z = 1) - \mathbb{E}(Q \mid Z = 0) \geqslant 0,
$$

where $Q = D \cdot \mathbf{1}(Y = y), (1 - D) \cdot \mathbf{1}(Y = y)$

$E(D \cdot 1(Y=y) \mid Z=1)$

$$P(Y=y, D=1 \mid Z=1)$$

$$= P(Y(1)=y, D(1)=1 \mid Z=1)$$

$$= P(Y(1)=y, D(1)=1).$$

$$P(Y=y, D=1 \mid Z=0)$$

$$= P(Y(0)=y, D(0)=1)$$

$$P(Y(1)=y, D(1)=1)$$

$$\overset{?}{\geq} P(Y(0)=y, D(0)=1)$$

$$\Rightarrow P(Y(1)=y \mid D(1)=1)\, P(D(1)=1)$$

$$\overset{?}{\geq} P(Y(0)=y \mid D(0)=1)\, P(D(0)=1)$$

Examples

- Investigators et al. (2014) assess the effectiveness of the emergency endovascular versus the open surgical repair strategies for patients with a clinical diagnosis of ruptured aortic aneurism
- Instrument $Z = 1$: endovascular strategy; treatment D : treatment received; outcome $Y = 0$: alive
- Statistical tests for testable conditions all pass
- CACE: est. $= 0.131; 95\%$ CI: $(-0.036, 0.298)$

Examples

- In Hirano et al. (2000), physicians are randomly selected to receive a letter encouraging them to inoculate patients at risk for flu
- Instrument $Z = 1$: encouragement; treatment D : flu shot; outcome $Y = 0$: flu-shot related hospitalization
- Statistical tests for testable conditions all pass, however,

$$\mathbb{E}\{D(1 - Y) \mid Z = 1\} < \mathbb{E}\{D(1 - Y) \mid Z = 0\}$$

- CACE: est. $= 0.116; 95\%$ CI: $(-0.061, 0.293)$
- $\mathbb{E}(Y \mid Z = 1, U = c) = 1.004$

$= E(Y(1) \mid U = c)$

CACE in observational studies

- Conditional independence and exclusion restriction may be more plausible after conditioning on covariates $\mathbf{X}_i$
  - $Z_i \perp \{Y_i(1), Y_i(0), D_i(1), D_i(0)\} \mid \mathbf{X}_i$
  - $Y_i(1) = Y_i(0)$ for $U = a, n$
- Within subgroup defined by $\mathbf{X}$, estimate

$$\mathrm{CACE}(\mathbf{x}) = \mathbb{E}\left\{Y_i(1) - Y_i(0) \mid D_i(1) = 1, D_i(0) = 0, \mathbf{X}_i = \mathbf{x}\right\}$$

- Estimate $\mathrm{CACE}(\mathbf{x})$ for each $\mathbf{x}$ and then average over $\mathbf{X}$

$$\mathrm{CACE} = \mathbb{E}\left\{\mathrm{CACE}\left(\mathbf{X}_i\right) \mid D_i(1) = 1, D_i(0) = 0\right\},$$

  which requires the estimation of $\mathrm{pr}\left(\mathbf{X}_i \mid D_i(1) = 1, D_i(0) = 0\right)$
- For continuous $\mathbf{X}_i$, we need to model $\mathrm{CACE}(\mathbf{x})$, which is hard to estimate because the model is for compliers only

D, Y, X

$Y(D=1)$

$Y(D=0)$

$$\Rightarrow \quad E\, g(Y(1), X) = E\,\frac{D\,g(Y, X)}{e(X)}$$

if $D = Z$

$$CACE = E(Y(1) - Y(0))$$

Weighting method for CACE

We can do weighting for compliers.

Theorem (Abadie Kappa)

Suppose that the assumptions for CACE hold conditional on $\mathbf{X}_i$. Define

$$\kappa_{0i} = (1-D)\frac{(1-Z) - \mathrm{pr}(Z=0 \mid \mathbf{X}_i)}{\mathrm{pr}(Z=0 \mid \mathbf{X}_i)\,\mathrm{pr}(Z=1 \mid \mathbf{X}_i)},$$

$$\kappa_{1i} = D\frac{Z - \mathrm{pr}(Z=1 \mid \mathbf{X}_i)}{\mathrm{pr}(Z=0 \mid \mathbf{X}_i)\,\mathrm{pr}(Z=1 \mid \mathbf{X}_i)}.$$

Then

$$\mathbb{E}\{g(Y_i(0), \mathbf{X}_i) \mid U_i = c\} = \frac{1}{\mathrm{pr}\{U_i = c\}}\mathbb{E}\{\kappa_{0i} g(Y, X)\}$$

$$\mathbb{E}\{g(Y_i(1), \mathbf{X}_i) \mid U_i = c\} = \frac{1}{\mathrm{pr}\{U_i = c\}}\mathbb{E}\{\kappa_{1} g(Y, X)\}$$

$$Z\left(\frac{1}{(1-e(X_i))(e(X_i))} - \frac{1}{1-e(X_i)}\right)$$

$$= \frac{Z}{e(X_i)}$$

$$E[4DZY - 2DY] - \cdots$$

$$= E(2D(2Z-1)Y) - E(2(1-D)(1-2Z)Y)$$

$$= E\, 2Y(2Z-1)(D+1-D)$$

$$= \frac{1}{2}\{E\, 2Y(1) - E\, 2Y(0)\}$$

Abadie's procedure for estimating CACE

- Two-step procedure
  - step 1: fit a model for $\text{pr}(Z_i = 1 \mid \mathbf{X}_i)$, and calculate κ_{0i} and κ_{1i} for each unit
  - step 2: apply Abadie's formula

- Abadie's procedure simplifies to Wald estimator without covariates

Model-assisted weighting method

- We may want to assume some parametric models for $Y(z)$, e.g. $\mathbb{E}\{Y_i(z) \mid \mathbf{X}_i, D_i(1)=1, D_i(0)=0\} = \beta_{z0} + \beta_{z,X}\mathbf{X}_i$ estimation is not direct because the model is for compliers only
- Least squares for compliers:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\mathbb{E}\left[\{Y_i(z) - (\beta_{z0} + \beta_{z,X}\mathbf{X}_i)\}^2 \mid D_i(1)=1, D_i(0)=0\right]$$

$$\equiv\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\mathbb{E}\left[\kappa_{zi}\{Y_i - (\beta_{z0} + \beta_{z,X}\mathbf{X}_i)\}^2\right]$$

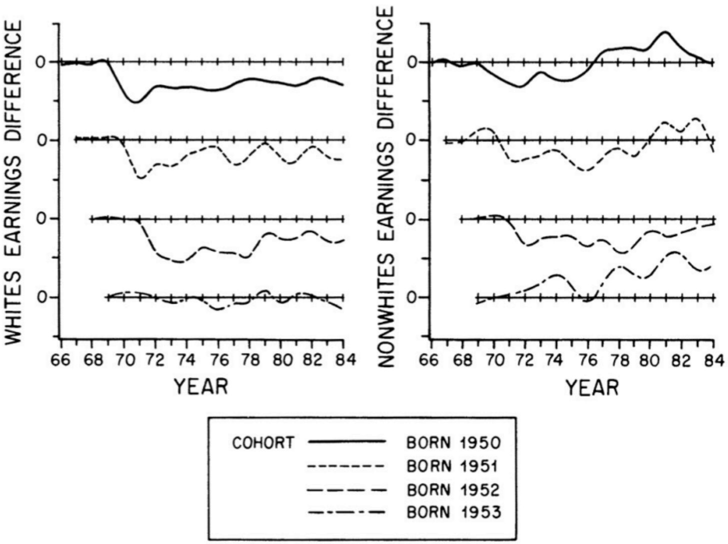
  why does this work?
- Abadie's procedure
  - step 1: fit a model for $\operatorname{pr}(Z_i = 1 \mid \mathbf{X}_i)$, and calculate κ 's for each unit
  - step 2: weighted least squares estimation

Effect of veteran status on earnings

- There were five draft lotteries during the Vietnam War period. In each lottery, priority for induction was determined by a Random Sequence Number (RSN) from 1-365 that was assigned to birthdates in the cohort being drafted
- Men were called for induction by RSN up to a ceiling determined by the Defense Department, and only men with lottery numbers below the ceiling could have been drafted
- Draft lottery RSNs were randomly assigned in a televised drawing held a few months before men reaching draft age were to be called
- Draft-eligibility ceilings were announced later in the year, once Defense Department manpower needs were known
- Subsequent selection from the draft-eligible pool was based on a number of criteria: physical examination and a mental aptitude test

IV setup

- IV Z_i : draft-eligibility; treatment D_i : veteran status; outcome Y_i earnings in $1981-1984$
- Assumptions hold?



Notes: The figure plots the difference in FICA taxable earnings by draft-eligibility status for the four cohorts born 1950–53. Each tick on the vertical axis represents $500 real (1978) dollars.

FIGURE 2. THE DIFFERENCE IN EARNINGS BY DRAFT-ELIGIBILITY STATUS

Wald estimates

TABLE 3—WALD ESTIMATES

Cohort	Year	Draft-Eligibility Effects in Current $: FICA Earnings (1)	Adjusted FICA Earnings (2)	Total W-2 Earnings (3)	$\hat{p}^e - \hat{p}^n$ (4)	Service Effect in 1978 $ (5)
1950	1981	−435.8	−487.8	−589.6	0.159	−2,195.8
		(210.5)	(237.6)	(299.4)	(0.040)	(1,069.5)
	1982	−320.2	−396.1	−305.5		−1,678.3
		(235.8)	(281.7)	(345.4)		(1,193.6)
	1983	−349.5	−450.1	−512.9		−1,795.6
		(261.6)	(302.0)	(441.2)		(1,204.8)
	1984	−484.3	−638.7	−1,143.3		−2,517.7
		(286.8)	(336.5)	(492.2)		(1,326.5)
1951	1981	−358.3	−428.7	−71.6	0.136	−2,261.3
		(203.6)	(224.5)	(423.4)	(0.043)	(1,184.2)
	1982	−117.3	−278.5	−72.7		−1,386.6
		(229.1)	(264.1)	(372.1)		(1,312.1)
	1983	−314.0	−452.2	−896.5		-2,181.8
		(253.2)	(289.2)	(426.3)		(1,395.3)
	1984	−398.4	−573.3	−809.1		−2,647.9
		(279.2)	(331.1)	(380.9)		(1,529.2)
1952	1981	−342.8	−392.6	−440.5	0.105	−2,502.3
		(206.8)	(228.6)	(265.0)	(0.050)	(1,556.7)
	1982	−235.1	−255.2	−514.7		−1,626.5
		(232.3)	(264.5)	(296.5)		(1,685.8)
	1983	−437.7	−500.0	−915.7		−3,103.5
		(257.5)	(294.7)	(395.2)		(1,829.2)
	1984	−436.0	−560.0	−767.2		−3,323.8
		(281.9)	(330.1)	(376.0)		(1,959.3)

Physical activity and weight after buying a car

Physical activity and weight following car ownership in Beijing, China: quasi-experimental cross sectional study

Michael L Anderson,[1] Fangwen Lu,[2] Jun Yang[3]

- In January 2011, to deal with the problem of congestion, Beijing capped the number of new vehicles allowed at 240000 each year and introduced a vehicle permit (license plate) lottery
- After that date, only residents who entered and won the lottery were entitled to a license plate.
- The lottery was drawn monthly, and winners had to purchase a car within six months of winning. By mid-2012 the probability of winning fell below 2% a month

Effect of winning the lottery

- IV: winning the lottery; treatment: buying a car; outcome: weekly transit rides, minute daily walking/bicycling, weight
- IV assumptions satisfied?

Table 3 | Age stratified associations between winning the lottery and transit use, activity, and weight

	Time since winning (95% CI)		
Dependent variables	**0.1 years (minimum)**	**2.6 years (average)**	**5.1 years (maximum)**
Individuals aged ≥40			
Weekly transit rides	−2.18 (−4.13 to −0.24)	−2.1 (−3.35 to −0.85)	−2.02 (−5.16 to 1.12)
Minutes daily walking/bicycling	12.1 (−4.66 to 28.86)	−2.59 (−12.12 to 6.94)	−17.29 (−36.52 to 1.95)
Weight (kg)	1.29 (−5.07 to 7.65)	3.24 (−0.31 to 6.8)	5.2 (−2.59 to 12.99)
Individuals aged ≥50			
Weekly transit rides	−2.88 (−5.57 to −0.19)	−1.9 (−3.61 to −0.18)	−0.91 (−5.45 to 3.63)
Minutes daily walking/bicycling	27.4 (−0.28 to 55.08)	−1.19 (−13.76 to 11.38)	−29.78 (−54.08 to −5.49)
Weight (kg)	−1 (−8.4 to 6.4)	4.67 (0.04 to 9.31)	10.34 (0.49 to 20.19)

$D(1) = 0$

$D(0) = 1$

Summary

- ITT vs. CACE $\rightsquigarrow$ additional assumptions are required
  1. randomization of instrument
  2. monotonicity
  3. exclusion restriction

- Problems of external validity:
  - compliers vs. non-compliers
  - compliers as latent group defined by an instrument

- Exclusion restriction and monotonicity imply testable conditions

Suggested readings

- Non-compliance
  - Angrist, Imbens & Rubin, "Identification of Causal Effects Using Instrumental Variables"
  - Imbens and Rubin, Chapters 23 and 24
  - Angrist and Pischke, Chapter 4
- Weighting method for CACE
  - Abadie (2003). "Semi-parametric instrumental variable estimation of treatment response models"