Noncompliance in Randomized Experiments

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

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
- Intention-to-Treat (ITT) analysis:
  - ITT effect can be estimated without bias
  - ITT analysis does not yield the treatment effect
- As-Treated analysis
  - comparison of the treated and untreated subjects
  - no benefit of randomization $\rightsquigarrow$ selection bias
- Can we estimate the treatment effect somehow?
- Encouragement design: randomize the encouragement to receive the treatment rather than the receipt of the treatment itself $\rightsquigarrow$ attractive to policy makers

Potential Outcomes Notation

- Randomized encouragement: $Z_i \in \{0, 1\}$
- Potential treatment variables: $(T_i(1), T_i(0))$
  1. $T_i(z) = 1$: would receive the treatment if $Z_i = z$
  2. $T_i(z) = 0$: would not receive the treatment if $Z_i = z$
- Observed treatment receipt indicator: $T_i = T_i(Z_i)$
- Observed and potential outcomes: $Y_i = Y_i(Z_i, T_i(Z_i))$
- Can be written as $Y_i = Y_i(Z_i)$

- No interference assumption for $T_i(Z_i)$ and $Y_i(Z_i, T_i)$
- Randomization of encouragement:

$$(Y_i(1), Y_i(0), T_i(1), T_i(0)) \perp\!\!\!\perp Z_i$$

- But $(Y_i(1), Y_i(0)) \not\perp\!\!\!\perp T_i \mid Z_i = z$

Principal Stratification (Angrist, et al. 1996. *J. Am. Stat. Assoc*)

- Four principal strata (latent types):
  - compliers $(T_i(1), T_i(0)) = (1, 0)$,
  - non-compliers $\left\{ \begin{array}{cc} \textit{always}-\textit{takers} & (T_i(1), T_i(0)) = (1, 1), \\ \textit{never}-\textit{takers} & (T_i(1), T_i(0)) = (0, 0), \\ \textit{defiers} & (T_i(1), T_i(0)) = (0, 1) \end{array} \right.$

- Observed and principal strata:

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

Instrumental Variables

- Assumptions:
  1. Randomized encouragement as an instrument for the treatment
  2. Monotonicity: No defiers

$$T_i(1) \geq T_i(0) \quad \text{for all } i.$$

  3. Exclusion restriction: Instrument (encouragement) affects outcome only through treatment

$$Y_i(1, t) \;=\; Y_i(0, t) \quad \text{for } t = 0, 1$$

  Zero ITT effect for always-takers and never-takers

- ITT effect decomposition:

$$\begin{aligned} \text{ITT} &= \text{ITT}_c \times \Pr(\text{compliers}) + \text{ITT}_a \times \Pr(\text{always} - \text{takers}) \\ &\quad + \text{ITT}_n \times \Pr(\text{never} - \text{takers}) \\ &= \text{ITT}_c \times \Pr(\text{compliers}) \end{aligned}$$

IV Estimand and Interpretation

- IV estimand:

$$\begin{aligned}
\text{ITT}_c &= \frac{\text{ITT}}{\Pr(\text{compliers})} \\
&= \frac{\mathbb{E}(Y_i \mid Z_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0)}{\mathbb{E}(T_i \mid Z_i = 1) - \mathbb{E}(T_i \mid Z_i = 0)} \\
&= \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)}
\end{aligned}$$

- $\text{ITT}_c =$ Complier Average Treatment Effect (CATE)
- Local Average Treatment Effect (LATE)
- CATE $\neq$ ATE unless ATE for noncompliers equals CATE
- Different encouragement (instrument) yields different compliers

Asymptotic Inference

- Wald estimator: $\widehat{\text{IV}}_{\text{Wald}} = \frac{\widehat{\text{Cov}(Y_i, Z_i)}}{\widehat{\text{Cov}(T_i, Z_i)}} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_T}$
- Identical to the two-stage least squares estimator:
  1. Regress T_i on Z_i and obtain fitted values $\widehat{T}_i$
  2. Regress Y_i on $\widehat{T}_i$

- Consistency: $\widehat{\text{IV}}_{\text{Wald}} \xrightarrow{p} \text{CATE} = \text{ITT}_c$
- Asymptotic variance via the Delta method:

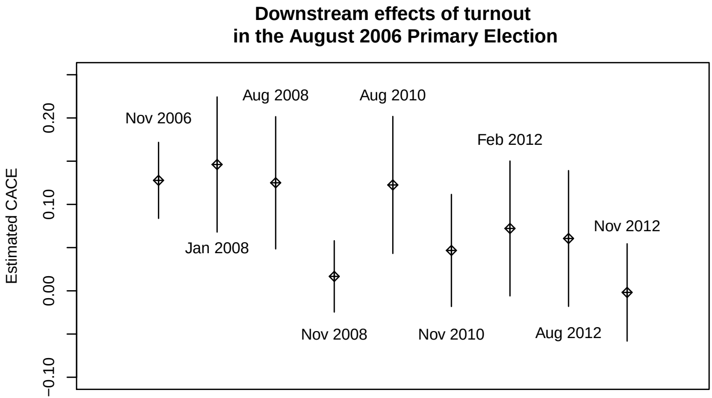
$$\begin{aligned}\mathbb{V}(\widehat{\text{IV}}_{\text{Wald}}) \quad \approx \quad & \frac{1}{\text{ITT}_T^4}\Big\{\text{ITT}_T^2\, \mathbb{V}(\widehat{\text{ITT}}_Y) + \text{ITT}_Y^2\, \mathbb{V}(\widehat{\text{ITT}}_T) \\ & -2\, \text{ITT}_Y\, \text{ITT}_T\, \text{Cov}(\widehat{\text{ITT}}_Y,\, \widehat{\text{ITT}}_T)\Big\}.\end{aligned}$$

Testing Habitual Voting (Coppock and Green. 2016. *Am. J. Political Sci.*)

- Settings (Revisit the Social Pressure Experiment):
  - Randomized encouragement to vote in the 2006 August primary
  - Treatment: turnout in the 2007 November municipal election
  - Outcome: turnout in the 2008 January party primary and subsequent elections
- Assumptions:
  1. Monotonicity: Being contacted by a canvasser would *never* discourage anyone from voting
  2. Exclusion restriction: being contacted by a canvasser in this election has no effect on turnout in the next election other than through turnout in this election
- CATE: Habitual voting for those who would vote if and only if they are contacted by a canvasser in this election

Downstream Effects

- Estimated proportion of principal strata:
  - compliers: est. $= 0.083$, s.e. $= 0.003$
  - always-takers: est. $= 0.311$, s.e. $= 0.001$
  - never-takers: est. $= 0.606$, s.e. $= 0.003$
- CATE:



Likelihood Inference

- Observed data $= \{Y_i, T_i, Z_i, \mathbf{X}_i\}_{i=1}^n$
- "Complete" data $= \{Y_i, T_i, Z_i, \mathbf{X}_i, C_i\}_{i=1}^n$

- The model:
  1. Compliance submodel: $f_\theta(C_i \mid \mathbf{X}_i)$

$$e.g., \quad \Pr(C_i = c \mid \mathbf{X}_i) \;=\; \frac{\exp(\mathbf{X}_i\gamma_c)}{1 + \exp(\mathbf{X}_i\gamma_c) + \exp(\mathbf{X}_i\gamma_a)}$$

  2. Outcome submodel: $g_\psi(Y_i \mid C_i, Z_i, \mathbf{X}_i)$

$$e.g., \quad Y_i \overset{\text{indep.}}{\sim} \mathcal{N}(\alpha_{C_i} + \beta \mathbf{1}\{C_i = c\} Z_i + \mathbf{X}_i^\top \delta,\ \sigma^2)$$

  Given C_i and Z_i, T_i is redundant

- The observed-data likelihood function:

$$\prod_{i=1}^{n} \{f_\theta(C_i = n \mid \mathbf{X}_i)\, g_\psi(Y_i \mid C_i = n, Z_i, \mathbf{X}_i)\}^{(1-T_i)Z_i}$$

$$\times \{f_\theta(C_i = a \mid \mathbf{X}_i)\, g_\psi(Y_i \mid C_i = a, Z_i, \mathbf{X}_i)\}^{T_i(1-Z_i)}$$

$$\times \left\{ \sum_{c' \in \{a,c\}} f_\theta(C_i = c' \mid \mathbf{X}_i)\, g_\psi(Y_i \mid C_i = c', Z_i, \mathbf{X}_i) \right\}^{T_i Z_i}$$

$$\times \left\{ \sum_{c' \in \{c,n\}} f_\theta(C_i = c' \mid \mathbf{X}_i)\, g_\psi(Y_i \mid C_i = c', Z_i, \mathbf{X}_i) \right\}^{(1-T_i)(1-Z_i)}$$

- The complete-data likelihood function:

$$\prod_{i=1}^{n} \prod_{c' \in \{a,c,n\}} \{f_\theta(C_i = c' \mid \mathbf{X}_i) g_\psi(Y_i \mid C_i = c', Z_i, X_i)\}^{C_i = c'}$$

Optimization Using the *EM* Algorithm

- The Expectation and Maximization algorithm by Dempster, Laird, and Rubin
- Useful for maximizing the likelihood function with missing data
- Goal: maximize the observed-data log-likelihood, $l_n(\theta \mid Y_{obs})$
- The *EM* algorithm: Repeat the following steps until convergence
  1. *E*-step: Compute

     $$Q(\theta \mid \theta^{(t)}) \equiv \mathbb{E}\{l_n(\theta \mid Y_{obs}, Y_{mis}) \mid Y_{obs}, \theta^{(t)}\}$$

     where $l_n(\theta \mid Y_{obs}, Y_{mis})$ is the complete-data log-likelihood
  2. *M*-step: Find

     $$\theta^{(t+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q(\theta \mid \theta^{(t)})$$

- The *ECM* algorithm: *M*-step replaced with multiple conditional maximization steps

Monotone Convergence Property

- The observed-data likelihood increases each step:

$$l_n(\theta^{(t+1)} \mid Y_{obs}) \ \geq \ l_n(\theta^{(t)} \mid Y_{obs})$$

- Sketch of Proof:
  1. $l_n(\theta \mid Y_{obs}) = \log f(Y_{obs}, Y_{mis} \mid \theta) - \log f(Y_{mis} \mid Y_{obs}, \theta)$
  2. Taking the expectation w.r.t. $f(Y_{mis} \mid Y_{obs}, \theta^{(t)})$

$$l_n(\theta \mid Y_{obs}) \ = \ Q(\theta \mid \theta^{(t)}) - \int \log f(Y_{mis} \mid Y_{obs}, \theta) f(Y_{mis} \mid Y_{obs}, \theta^{(t)}) dY_{mis}$$

  3. Finally,

$$\begin{aligned}
& l_n(\theta^{(t+1)} \mid Y_{obs}) - l_n(\theta^{(t)} \mid Y_{obs}) \\
= \ & Q(\theta^{(t+1)} \mid \theta^{(t)}) - Q(\theta^{(t)} \mid \theta^{(t)}) \\
& + \int \log \frac{f(Y_{mis} \mid Y_{obs}, \theta^{(t)})}{f(Y_{mis} \mid Y_{obs}, \theta^{(t+1)})} f(Y_{mis} \mid Y_{obs}, \theta^{(t)}) dY_{mis} \\
\geq \ & 0
\end{aligned}$$

- Stable, no derivative required

Application to the Complier Average Causal Effect

1. E-step:

$$\begin{aligned} w_i(c^*)^{(t+1)} &= \Pr(C_i = c^* \mid \mathbf{Y}, \mathbf{Z}, \mathbf{X}) \\ &= \frac{f_{\theta^{(t)}}(C_i = c^* \mid \mathbf{X}_i) g_{\psi^{(t)}}(Y_i \mid C_i = c^*, Z_i, \mathbf{X}_i)}{\sum_{c' \in \{a,c,n\}} f(C_i = c' \mid \mathbf{X}_i) g_{\psi^{(t)}}(Y_i \mid C_i = c', Z_i, \mathbf{X}_i)} \end{aligned}$$

   for $c^* \in \{c, a, n\}$.

2. M-step:

$$\begin{aligned} \theta^{(t+1)} &= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{c' \in \{a,c,n\}} w(c')^{(t+1)} \log f_\theta(C_i = c' \mid \mathbf{X}_i) \\ \psi^{(t+1)} &= \underset{\psi}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{c' \in \{a,c,n\}} w(c')^{(t+1)} \log g_\psi(Y_i \mid C_i = c', Z_i, \mathbf{X}_i) \end{aligned}$$

Summary

- Noncompliance in randomized experiments
- ITT vs. CATE (LATE) $\rightsquigarrow$ additional assumptions are required
    1. randomization of instrument
    2. monotonicity
    3. exclusion restriction
- Traditional instrumental variables $\rightsquigarrow$ ignoring heterogeneity

- Tradeoff between internal and external validity:
    - compliers vs. noncompliers
    - compliers as latent group defined by an instrument

- Suggested readings:
    - IMBENS AND RUBIN, Chapters 23 and 24
    - ANGRIST AND PISCKE, Chapter 4