Safe and Efficient Use of External Control Data under Model Misspecification for the Average Treatment Effect on the Treated with High-Dimensional Covariates

Chi-Shian Dai

University of Wisconsin-Madison
Department of Statistics

ENAR 3/28/2022

joint work with
Yang Ning and Jiwei Zhao

Motivation

- Estimate average treatment effect on treated group(ATT).
- Primary data set: Can be observational or experimental data.
  Ex: Randomized control trials.
- External controls: Observational data.

Data structure and Notations

		R	T	Y_1	Y_0	$\mathbf{X}$
Primary Study	1	1	1	✓		✓
	2	1	1	✓		✓
	$\vdots$	1	0		✓	✓
	n	1	0		✓	✓
External Control	$n+1$	0	0		✓	✓
	$\vdots$	0	0		✓	✓
	N	0	0		✓	✓

- $\mu(\mathbf{x}) = E(Y_0 \mid \mathbf{X} = \mathbf{x})$
- $p(\mathbf{x}) = E(T \mid \mathbf{X} = \mathbf{x}; R = 1)$
- $\pi(\mathbf{x}) = E(R \mid \mathbf{X} = \mathbf{x})$
- $\delta(\mathbf{x}) = p(\mathbf{x})\pi(\mathbf{x}) = E(T|\mathbf{X} = \mathbf{x})$

Challenge

- Safe: Make sure the external controls are helpful.
- High-dimensional Covariates: Model misspecification.
  Double Robustness:
  - Outcome models: $\mu(\mathbf{x})$
  - Propensity models: $p(\mathbf{x})$ and $\delta(\mathbf{x})$

Assumptions

- Consistency:
$$Y = Y_1 T + Y_0(1 - T)$$
- Ignorability:
$$\{R, T\} \perp \{Y_1, Y_0\} \mid \mathbf{X}$$
- Positivity: For each level of the covariate $\mathbf{X}$, the probability of receiving either treatment or control is greater than zero.

Naive Estimate(Without External Controls)

- Joint distrubution of $(Y, \mathbf{X}, T) \mid R = 1$

$$p(\mathbf{x})^t \{1 - p(\mathbf{x})\}^{(1-t)} f_1(y \mid \mathbf{x})^t f_0(y \mid \mathbf{x})^{1-t} g(\mathbf{x})$$

- Naive Influent function

$$\phi_{\mathrm{nv}}(\mu, p) = \frac{r}{\pi}\left[\frac{t\{y - \mu(\mathbf{x}) - \theta\}}{p} - \frac{(1-t)p(\mathbf{x})\{y - \mu(\mathbf{x})\}}{p\{1 - p(\mathbf{x})\}}\right]$$

- Naive Estimate for ATT

$$\widehat{\theta}_{\mathrm{nv}}(\mu, p) = \frac{\frac{1}{N}\sum_{i=1}^{N}[r_i t_i\{y_i - \mu(\mathbf{x}_i)\} - r_i(1-t_i)p(\mathbf{x})\{y_i - \mu(\mathbf{x})\}/\{1 - p(\mathbf{x})\}]}{\frac{1}{N}\sum_{i=1}^{N} r_i t_i}$$

- $\widehat{\theta}_{\mathrm{nv}}(\mu, p)$ is double robust for μ and p.

Efficient Estimate(With External Controls)

- Joint distrubution of $(Y, \mathbf{X}, R, T)$

$$p(\mathbf{x})^{rt}\{1-p(\mathbf{x})\}^{r(1-t)} f_1(y \mid \mathbf{x})^{rt} f_0(y \mid \mathbf{x})^{1-rt} g(\mathbf{x})^r h(\mathbf{x})^{1-r} \pi(\mathbf{x})^r (1-\pi)^{1-r}$$

- Efficient Influent Function

$$\phi_{\mathrm{eff}}(\mu, \delta) = \frac{rt\{y-\mu(\mathbf{x})-\theta\}}{\pi p} - \frac{(1-rt)\delta(\mathbf{x})\{y-\mu(\mathbf{x})\}}{\pi p\{1-\delta(\mathbf{x})\}}$$

- Efficient estimate for ATT

$$\widehat{\theta}_{\mathrm{eff}}(\mu, \delta) = \frac{\frac{1}{N}\sum_{i=1}^{N}[r_i t_i\{y_i-\mu(\mathbf{x})\}-(1-r_i t_i)\delta(\mathbf{x})\{y_i-\mu(\mathbf{x})\}/\{1-\delta(\mathbf{x})\}]}{\frac{1}{N}\sum_{i=1}^{N} r_i t_i}$$

- $\widehat{\theta}_{\mathrm{eff}}(\mu, \delta)$ is double robust for μ and δ.

$\widehat{\theta}_{\text{eff}}$ might be worse than $\widehat{\theta}_{\text{nv}}$

Propensity models $\delta(\mathbf{x})$ and $p(\mathbf{x})$ are correct:

$\widehat{\theta}_{\text{eff}}$ is more efficient than $\widehat{\theta}_{\text{nv}}$.

Only outcome model $\mu(\mathbf{x})$ is correct:

$\widehat{\theta}_{\text{eff}}$ might be less efficient than $\widehat{\theta}_{\text{nv}}$.

Proposed Method

- Models Aggregation:
  - Influent function

$$\phi_a = a\phi_{\mathrm{eff}} + (1-a)\phi_{\mathrm{nv}} = \phi_{\mathrm{nv}} + a(\phi_{\mathrm{eff}} - \phi_{\mathrm{nv}}),$$

  - ATT estimate

$$\widehat{\theta}_a = a\widehat{\theta}_{\mathrm{eff}} + (1-a)\widehat{\theta}_{\mathrm{nv}}.$$

- Find a such that

$$\begin{aligned} \widehat{a} &= \arg\min_a \tilde{E}\{\phi_{\mathrm{nv}} + a(\phi_{\mathrm{eff}} - \phi_{\mathrm{nv}})\}^2 \\ &= \arg\min_a \widehat{SD}\{a\widehat{\theta}_{\mathrm{eff}} + (1-a)\widehat{\theta}_{\mathrm{nv}}\} \\ &= \frac{\tilde{E}[\phi_{\mathrm{nv}}(\widehat{\mu},\widehat{p})\{\phi_{\mathrm{nv}}(\widehat{\mu},\widehat{p}) - \phi_{\mathrm{eff}}(\widehat{\mu},\widehat{\delta})\}]}{\tilde{E}\{\phi_{\mathrm{nv}}(\widehat{\mu},\widehat{p}) - \phi_{\mathrm{eff}}(\widehat{\mu},\widehat{\delta})\}^2} \end{aligned}$$

Safety of $\widehat{\theta}_{\widehat{a}}$

Safety of $\widehat{\theta}_{\widehat{a}}$

- Propensity models $\delta(\mathbf{x})$ and $p(\mathbf{x})$ are correct: $SD(\widehat{\theta}_{\widehat{a}}) = SD(\widehat{\theta}_{\mathrm{eff}}) \leq SD(\widehat{\theta}_{\mathrm{nv}})$
- Only outcome model $\mu(\mathbf{x})$ is correct: $SD(\widehat{\theta}_{\widehat{a}}) \leq SD(\widehat{\theta}_{\mathrm{nv}})$

Estimate of Nuisance Models

- $\mu_{\text{nv}}(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\alpha}_{\text{nv}}$, $\mu_{\text{eff}}(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\alpha}_{\text{eff}}$, $p(\mathbf{x}) = \text{expit}(\mathbf{x}^\top \boldsymbol{\beta})$, $\delta(\mathbf{x}) = \text{expit}(\mathbf{x}^\top \gamma)$.
- Consider the influent function of θ_a: $\phi_a(\boldsymbol{\alpha}_{\text{eff}}, \boldsymbol{\alpha}_{\text{nv}}, \gamma, \boldsymbol{\beta}) = a\phi_{\text{eff}}(\boldsymbol{\alpha}_{\text{eff}}, \gamma) + (1-a)\phi_{\text{nv}}(\boldsymbol{\alpha}_{\text{nv}}, \boldsymbol{\beta})$
- Find $\boldsymbol{\alpha}_{\text{eff}}$, $\boldsymbol{\alpha}_{\text{nv}}$, γ, $\boldsymbol{\beta}$ such that $\tilde{E}\{\nabla\phi_a(\boldsymbol{\alpha}_{\text{eff}}, \boldsymbol{\alpha}_{\text{nv}}, \gamma, \boldsymbol{\beta})\} = 0$, which is equivalent to

$$\tilde{E}\left[\{RT - (1-RT)\exp(\mathbf{X}^{\mathrm{T}}\gamma)\}\mathbf{X}\right] = \mathbf{0}, \quad (1)$$

$$\tilde{E}\left[\{RT - R(1-T)\exp(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})\}\mathbf{X}\right] = \mathbf{0}, \quad (2)$$

$$\tilde{E}\{R(1-T)(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\alpha}_{\text{nv}})\exp(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{X}\} = \mathbf{0}, \quad (3)$$

$$\tilde{E}\{(1-RT)(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\alpha}_{\text{eff}})\exp(\mathbf{X}^{\mathrm{T}}\gamma)\mathbf{X}\} = \mathbf{0}. \quad (4)$$

- High-dimensionality: Consider the corresponded optimization form for (1)-(4) and add L1 penalties.

Algorithm

1: Using the primary study data only, compute $\widehat{\boldsymbol{\beta}}$ in (2) and $\widehat{\boldsymbol{\alpha}}_{\rm nv}$ in (3), then compute $\widehat{\theta}_{\rm nv}(\widehat{\boldsymbol{\alpha}}_{\rm nv}, \widehat{\boldsymbol{\beta}})$.

2: Using all the available data, compute $\widehat{\boldsymbol{\gamma}}$ in (1) and $\widehat{\boldsymbol{\alpha}}_{\rm eff}$ in (4), then compute $\widehat{\theta}_{\rm eff}(\widehat{\boldsymbol{\alpha}}_{\rm eff}, \widehat{\boldsymbol{\gamma}})$

3: Compute $\widehat{a}$ with

$$\widehat{a} = \frac{\frac{1}{N}\sum_{i=1}^{N}\{\phi_{\rm nv}(\widehat{\boldsymbol{\alpha}}_{\rm nv}, \widehat{\boldsymbol{\beta}}) - \phi_{\rm eff}(\widehat{\boldsymbol{\alpha}}_{\rm eff}, \widehat{\boldsymbol{\gamma}})\}\phi_{\rm nv}(\widehat{\boldsymbol{\alpha}}_{\rm nv}, \widehat{\boldsymbol{\beta}})}{\frac{1}{N}\sum_{i=1}^{N}\{\phi_{\rm nv}(\widehat{\boldsymbol{\alpha}}_{\rm nv}, \widehat{\boldsymbol{\beta}}) - \phi_{\rm eff}(\widehat{\boldsymbol{\alpha}}_{\rm eff}, \widehat{\boldsymbol{\gamma}})\}^2}.$$

4: Compute $\widehat{\theta}_{\widehat{a}} = \widehat{a}\, \widehat{\theta}_{\rm eff}(\widehat{\boldsymbol{\alpha}}_{\rm eff}, \widehat{\boldsymbol{\gamma}}) + (1 - \widehat{a})\widehat{\theta}_{\rm nv}(\widehat{\boldsymbol{\alpha}}_{\rm nv}, \widehat{\boldsymbol{\beta}})$.

Real Data Application

The data are drawn from LaLonde (1986).

- Goal: Evaluations of Job Training Programs.
- Population: No income in 1975.
- Response: [Log(Income+1) in 1978] - [Log(Income+1) in 1975].
- Primary Data: An experimental dataset from National Supported Work program. Sample size is 289. Number of treated group is 111.
- External Controls: An observational dataset from Current Population Survey. Sample size is 554.
- Covariates: Education, Age, Race, Marital Status ...
  Dimension: $d = 106$.

Results

Method	ATT	SD	p.value
$\widehat{\theta}_{\mathrm{nv}}$	0.78	0.49	0.11
$\widehat{\theta}_{\mathrm{eff}}$	0.909	0.42	0.0313
$\widehat{\theta}_{\widehat{a}}$	0.911	0.42	0.0308

$\widehat{a}$=1.02.

Acknowledgement

Jiwei Zhao

Yang Ning

Jun Shao

Thank you

Thank you

References

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, pages 604–620.