Covariance Adjustment in Randomized Experiments

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Motivation

- Treatment randomization $\rightsquigarrow$ unbiased causal estimates
- We adjust covariates for improved efficiency *before* randomization $\rightsquigarrow$ stratification
- In some cases, we cannot perform pre-randomization adjustment
- Danger of post-randomization covariance adjustment
  1. specification search, p-hacking $\rightsquigarrow$ pre-registration
  2. may lose some of the theoretical properties of randomization inference
- Are there principled ways of performing post-randomization covariance adjustment?
- What are the theoretical properties of such adjustments?

Building Intuition under the Linear Structural Model

- Let's assume the following LSM:

$$Y_i(t) \;=\; \alpha + \beta \cdot t + \boldsymbol{\gamma}^\top \mathbf{X}_i + \delta U_i + \epsilon_i \quad \text{where} \quad \mathbb{E}(\epsilon_i \mid \mathbf{X}_i, U_i) \;=\; 0$$

  and U_i is an unobserved pre-treatment variable correlated with $\mathbf{X}_i$
- Decomposition via projection:

$$U_i \;=\; \hat{\lambda} + \hat{\xi} T_i + \hat{\boldsymbol{\zeta}}^\top \mathbf{X}_i + \hat{\eta}_i$$

- Omitted variable bias formula:

$$\hat{\beta} \;\xrightarrow{p}\; \beta + \delta \cdot \xi, \quad \hat{\boldsymbol{\gamma}} \;\xrightarrow{p}\; \boldsymbol{\gamma} + \delta \cdot \boldsymbol{\zeta}$$

  where $(\xi, \boldsymbol{\delta}^\top, \boldsymbol{\zeta}^\top)$ are probability limits of $(\hat{\xi}, \hat{\boldsymbol{\delta}}^\top, \boldsymbol{\zeta}^\top)$
- Randomization of $T_i \rightsquigarrow \xi = 0$ though $\boldsymbol{\delta} \neq \mathbf{0}_K$
- Random assignment protects you against bias
- Does the result hold if we do not assume the LSM?

Covariance Adjustment via Regression

- Multiple linear regression model with centered covariates:

$$Y_i = \alpha + \beta T_i + \gamma^\top \widetilde{\mathbf{X}}_i + \epsilon_i \quad \text{for } i = 1, \ldots, n$$

  where $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \overline{\mathbf{X}}_n$
- Ordinary least squares estimator:

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \underset{(\alpha,\beta,\gamma)}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - \alpha - \beta T_i - \gamma^\top \widetilde{\mathbf{X}}_i)^2$$

  where the right hand side converges to,

$$\mathbb{E}\{(Y_i - \alpha - \beta T_i)^2\} + \mathbb{E}\{(\gamma^\top \widetilde{\mathbf{X}}_i)^2\} - 2 \cdot \underbrace{\mathbb{E}\{(Y_i - \alpha - \beta T_i)(\gamma^\top \widetilde{\mathbf{X}}_i)\}}_{=\mathbb{E}(Y_i \cdot \gamma^\top \widetilde{\mathbf{X}}_i)}$$

- Thus, $(\hat{\alpha}, \hat{\beta})$ converge to $(\mathbb{E}(Y_i(0)), \tau)$
- OLS estimator is consistent even if the model is incorrect

Efficiency Gain from Regression

- Asymptotic variance (Imbens and Rubin, Theorem 7.1):

$$\sqrt{n}(\hat{\beta} - \tau) \rightsquigarrow \mathcal{N}\left(0, \frac{\mathbb{E}\{(T_i - k)^2(Y_i - \alpha^* - \beta^* T_i - \gamma^{*\top}\widetilde{\mathbf{X}}_i)^2\}}{k^2(1-k)^2}\right)$$

  where $*$ represents a limiting value and $k = n_1/n$
- Consistently estimated by the heteroskedasticity-robust variance estimator regardless of model misspecification
- If the linear model is correct,

$$\mathbb{V}(\hat{\beta}) \approx \frac{\mathbb{E}\{\mathbb{V}(Y_i(1) \mid \widetilde{\mathbf{X}})\}}{n_1} + \frac{\mathbb{E}\{\mathbb{V}(Y_i(0) \mid \widetilde{\mathbf{X}})\}}{n_0} \leq \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}$$

- If the linear model is incorrect, the efficiency gain can be negative (Freedman. 2008. *Adv. Appl. Math.*)

Regression with Interactions

- Fully interacted model:

$$Y_i = \alpha + \beta T_i + \gamma^\top \widetilde{\mathbf{X}}_i + \delta^\top T_i \widetilde{\mathbf{X}}_i + \epsilon_i$$

- Imputation interpretation:

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\{T_i(Y_i - \widehat{Y_i(0)}) + (1 - T_i)(\widehat{Y_i(1)} - Y_i)\}$$

which is different from

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{Y_i(1)} - \widehat{Y_i(0)})$$

- $\hat{\beta}$ is consistent for the PATE and asymptotically normal
- Unlike the no interaction model, $\hat{\beta}$ is at least as efficient as the difference-in-means estimator (Linn. 2013. *Ann. Appl. Stat.*)

Covariance Adjustment with Permutation Test

- We could use the OLS estimator as a test statistic: see the California Alphabet Lottery example
- Rosenbaum suggests another procedure (Rosenbaum. 2002. *Stat. Sci.*)
  1. Under the sharp null hypothesis, obtain $Y_i(0)$ for all i.
  2. Regress $Y_i(0)$ on $\mathbf{X}_i$ (includes an intercept) and obtain the residuals $\hat{\epsilon}$
  3. Use Wilcoxon's rank sum statistic based on the residuals $S = \sum_{i=1}^{n} T_i R_i(\hat{\epsilon})$
- Alternatively, we can use the sum of residuals: $\sum_{i=1}^{n} T_i \hat{\epsilon}_i$
- Hodges-Lehmann estimator solves,

$$
\begin{aligned}
0 &= \mathbb{E}\left(\sum_{i=1}^{n} T_i \epsilon_i\right) = \sum_{i=1}^{n} T_i \hat{\epsilon}_i = \mathbf{T}^\top \mathbf{M}_{\mathbf{X}}(\mathbf{Y} - \hat{\tau}\mathbf{T}) \\
\Longleftrightarrow \quad \hat{\tau} &= (\mathbf{T}^\top \mathbf{M}_{\mathbf{X}} \mathbf{T})^{-1} (\mathbf{T}^\top \mathbf{M}_{\mathbf{X}}) \mathbf{Y} = \text{OLS estimator}
\end{aligned}
$$

Reducing Transphobia Revisited

- 3 month effect
- Dropping additional observations with missing covariates $n1 = 266$ and $n0 = 274$

- Neyman: est. $= 7.99$, s.e. $= 2.38$, 95% CI $= [3.33,\ 12.65]$
- Simple Regression: est. $= 7.99$, s.e. $= 2.37$, 2.38 (HC), 2.38 (HC2)

- Regression with covariates (age, gender, party ID, baseline thermometer rating)
    - no interaction: est. $= 4.28$, s.e. $= 1.64$, 1.64 (HC), 1.65 (HC2)
    - with interaction: est. $= 4.28$, s.e. $= 1.64$, 1.64 (HC), 1.65 (HC2)

Regression Adjustment in Stratified Designs

- Neyman's analysis of stratified designs $\rightsquigarrow$ aggregate within-strata estimates across strata
- Linear regression with fixed effects does within-strata comparison

$$
\begin{aligned}
Y_{ij} &= \alpha_j + \beta T_{ij} + \epsilon_{ij} \\
\hat{\beta} &\xrightarrow{p} \frac{\sum_{j=1}^{J} w_j \cdot k_j(1-k_j)\mathbb{E}(Y_{ij}(1) - Y_{ij}(0))}{\sum_{j=1}^{J} w_j \cdot k_j(1-k_j)} \neq \tau
\end{aligned}
$$

- $\hat{\beta}$ is consistent for PATE if:
  1. treatment assignment probability is identical across strata
  2. average treatment effect is identical across strata
- Weighted fixed effects with $1/k_j$ as regression weights equals the Neyman's estimator, removing the bias (Imai and Kim. 2016. *Working paper*)

Post-stratification

- Weighted fixed effects estimator $\rightsquigarrow$ stratification estimator after the treatment is assigned
- A major difference in inference: the number of treated within each strata is random
- Inference conditions on the event that each strata has at least one treated and one control unit (Miratrix et al. 2013. *J. Royal Stat. Soc. Ser. B*)
  1. Unbiased for SATE
  2. Variance for SATE

$$\frac{1}{n}\sum_{j=1}^{J} w_j \left\{ \mathbb{E}\left(\frac{n_{0j}}{n_{1j}}\right) S_{1j}^2 + \mathbb{E}\left(\frac{n_{1j}}{n_{0j}}\right) S_{0j}^2 + 2S_{01j} \right\}$$

- Blocking before or after randomization?
  - pre-randomization blocking is typically more efficient than post-stratification but the difference is small
  - post-stratification is prone to p-hacking

STAR Project Revisited

- Pre-randomization stratification by schools
- Covariates to be considered: gender, race, birth year

- Effect of kindergarden class size on high school graduation:
  1. Neyman: est. $= 0.018$, s.e. $= 0.017$
  2. Linear regression with strata fixed effects
     - No covariate: est. $= 0.004$, s.e. $= 0.016$, 0.016 (HC), 0.016 (HC2)
     - With covariates (gender, birth year, race):
       est. $= -0.005$, s.e. $= 0.016$, 0.016 (HC), 0.016 (HC2)

- Effect of kindergarden class size on 8th grade reading score
  1. Neyman: est. $= 2.76$, se. $= 1.73$
  2. Linear regression with strata fixed effects
     - No covariate: est. $= 2.58$, s.e. $= 1.72$, 1.69 (HC), 1.72 (HC2)
     - With covariates: est. $= 2.45$, s.e. $= 1.70$, 1.68 (HC), 1.71 (HC2)

Matched-Pair Designs

- Linear regression with pair fixed effects

$$Y_{ij} \;=\; \alpha_j + \beta T_{ij} + \epsilon_{ij} \quad \text{where } T_{1j} + T_{2j} = 1 \text{ for all } j$$

  - identical to the average pairwise difference estimator
  - homoskedastic (or HC2) variance estimator is also identical
  - covariates can be added, equivalent to the first-difference model

$$(Y_{1j} - Y_{2j}) \;=\; \beta \underbrace{(T_{1j} - T_{2j})}_{=W_j} + \gamma^\top (\mathbf{X}_{1j} - \mathbf{X}_{2j}) + (\epsilon_{1j} - \epsilon_{2j})$$

- An alternative model:

$$(Y_{1j} - Y_{2j}) \;=\; \beta W_j + \gamma^\top (\mathbf{X}_{1j} + \mathbf{X}_{2j}) + \eta_j$$

- Both models yield conservative variances but they are smaller than the standard variance (Fogarty. 2018. *Biometrika*)

Seguro Popular Revisited

- Outcome: health cluster-level Seguro Popular enrollment rates
- Cluster-level covariates: average assets, education, urban/rural

- Neyman: est. $= 0.374$, s.e. $= 0.036$

- Linear regression with fixed effects:
  - No covariates: est. $= 0.374$, s.e. $= 0.036$
  - With covariates: est. $= 0.372$, s.e. $= 0.035$

- First difference regression:
  - No covariates: est. $= 0.374$, s.e. $= 0.036$
  - With covariates: est. $= 0.390$, s.e. $= 0.035$

Summary

- Covariance adjustment can be used to improve efficiency in randomized experiments
- Under various experimental designs, linear regression models are useful methods for this purpose
- Randomization of treatment assignment protects researchers from misspecification
  - independence between treatment and covariates
  - linear regression estimators are often consistent even when the model is incorrect
- Danger of p-hacking $\rightsquigarrow$ pre-randomization blocking is preferred whenever possible
- Suggested readings:
  - IMBENS AND RUBIN, Chapters 7, 9, and 10
  - ANGRIST AND PISCHKE, Chapters 3 and 8