To adjust or not to adjust? Estimating the average treatment effect in randomized experiments with missing covariates

Peng Ding

UC Berkeley, Statistics

February 17, 2022

joint work with Anqi Zhao from National University of Singapore

https://arxiv.org/abs/2108.00152

Randomized experiments and covariate adjustment

- Gold standard for unbiased estimation of treatment effects
  - simple OLS works: $Y_i \sim Z_i$ with outcome Y and treatment Z
- A large literature on covariate adjustment to improve efficiency
  - Fisher (1935): $Y_i \sim Z_i + x_i$ with covariates x
  - Lin (2013): $Y_i \sim Z_i + x_i + Z_i x_i$ with centered covariates
  - Lin (2013) is generally better than Fisher (1935) asymptotically
- EHW robust SE is a convenient approximation to the true SE
- Design-based theory (Neyman 1923, 1934; Freedman 2008; Lin 2013)
  - parameter: $\tau = n^{-1} \sum_{i=1}^{n} \{Y_i(1) - Y_i(0)\}$
  - Z_i's are random permutation of 1s and 0s
  - conditional on all potential outcomes and covariates
  - no outcome modeling assumption

Theory is nice but practice can be complicated...

- Proper covariate adjustment promises asymptotic efficiency gain
- Practitioners may not want to use it or can often misuse it
  - for many reasons
- This talk focuses on a major complication

  missingness in covariates

Is missingness in covariates a real problem?

- Duflo et al (2011 AER) field experiment in Western Kenya
- Effect of free delivery of fertilizer on fertilizer use
- 204 received treatment and 673 received control
- 27 covariates: education, previous fertilizer use, gender, income, etc
- 7 covariates have missing values
- $\approx 20\%$ units have some missing covariates

Missingness patterns in Duflo et al (2011 AER)

only for the 7 covariates subject to missingness: 1 for missing and 0 for observed

missingness pattern	sample size
0000000	716
0000100	59
0001000	1
0010011	2
0100000	19
0100100	1
1000000	1
1011111	71
1111111	7

The current default covariate adjustment

- With missing x, default OLS functions in R and Stata drop units
- Called the *complete-case (cc) analysis*
- $Y_i \sim Z_i$ uses all units
- $Y_i \sim Z_i + x_i$ or $Y_i \sim Z_i + x_i + Z_i x_i$ drops units with any missingness
- Question: Is covariate adjustment based on the cc analysis really better than the simple difference in means?

How to deal with missing x in randomized experiments?

- Modeling x? impute x? multiple imputation? or even fancier methods? (Little and Rubin 2002 book on missing data)
- Goal of this talk:
  - adjust for x but do not want to use complicated methods
  - recommend easy-to-implement methods: better to still use OLS
  - stronger guarantees than naive methods, without modeling assumptions
- Before going to technical details, state our final recommendation:
  - *missingness-indicator method (mim)*
  - impute missing covariate values all by 0
  - augment imputed covariates by the associated missingness indicators
  - use the augmented covariates in Fisher (1935) or Lin (2013)
  - report the EHW robust SE

Start from a simple yet reasonable scenario

- Let M_i be the missingness indicator vector corresponding to x_i
- Assume

$$M_i(1) = M_i(0) \quad (i = 1, \ldots, n)$$

- M is not affected by the treatment
- Reasonable in experiments with x collected before treatment
- M is effectively a covariate vector: this is key for later discussion
- M can be related to x and even $Y(1)$, $Y(0)$
- Allows for *missing not at random* in the sense of Rubin (1976)

Method 1: complete-case (cc) analysis

- Default if run OLS or other statistical models
- Loss of efficiency if many units miss at least some covariate values
- Problematic because the complete cases may not represent all units
- Default OLS $Y_i \sim Z_i + x_i$ or $Y_i \sim Z_i + x_i + Z_i x_i$ can be biased
- A good reason to avoid covariate adjustment due to this complexity
- Although cc analysis is widely used, we strongly discourage using it!

Method 2: complete-covariate (ccov) analysis

- Dropping units seems inferior as in cc analysis
- Adjust only for covariates that are completely observed for all units
- Always ensures efficiency gain with at least one predictive covariate
- Theory for ccov analysis is very simple: follows from Fisher and Lin
- Use ccov analysis as a benchmark in our discussion
- Reduces to unadjusted estimator if all covariates have missing values
- Can be inferior in efficiency if most covariates have missing values

Method 3: single imputation (imp)

- Impute missing values of x_{ij} by c_j ($j = 1, \ldots, p$)
  - c_j's can be fixed numbers, e.g., 0's
  - c_j's can even be data-dependent, e.g., mean of observed x_{ij}'s
  - theory only requires c_j's have finite limits
  - can view the imputed covariates $x_i^{\text{imp}}(c)$ as "pseudo covariates"
- Use $x_i^{\text{imp}}(c)$ in covariate adjustment
- Asymptotically better than ccov analysis in efficiency
- Theory depends on c_j's; can optimize over them
- Do not discuss complex imputation schemes: generally suboptimal
- Multiple imputation seems overkill since EHW robust SE works

Method 4: missingness-indicator method (mim)

- Impute the missing covariates all by 0's: imputed covariates x_i^{imp}
- View x_i^{imp} as well as M_i as "pseudo covariates"
- Use (x_i^{imp}, M_i) in covariate adjustment
- Report the EHW robust SE
- The choice of 0 for the imputation is not restrictive
  - point estimator and SE are invariant to choice of c_j's (numeric fact)
  - true for both Fisher and Lin
  - true only if the missingness indicators are included in OLS

Our recommendation: mim

- Uses all covariates and all units
- Always improves efficiency over unadjusted, ccov, and imp analyses
- No dependence on the imputed values for the missing covariates
- Very easy to implement via OLS + EHW robust SE
- No need to model the missing data mechanism and covariates
- Works even if the missing mechanism depends on missing covariates
- Works even if the linear outcome model is wrong

The mim is not new at all!

- An old yet not so popular literature: e.g. Cohen and Cohen (1975)
- Used a lot in observational studies, especially for matching: Rosenbaum and Rubin (1984, Appendix B), Rosenbaum (2009, page 241), Hainmueller and Hangartner (2013), Fogarty et al (2016), etc
- Problematic in non-randomized studies: Greenland and Finkle (1995), Doners et al (2006), Yang, Wang and Ding (2019), etc
- Randomization justifies mim though!
- Some versions recommended also by White and Thompson (2005), Carpenter and Kenward (2007), and Gerber and Green (2012)
- We provide the design-based theory for mim

Method 5: missingness-pattern (mp) method

- Missingness-pattern = combination of the missingness indicators
- Example with 2 missing covariates:

missingness pattern (M_i)	x_{i1}	x_{i2}	number of units
$(0,0)$	obs	obs	$N_{(00)}$
$(0,1)$	obs	mis	$N_{(01)}$
$(1,0)$	mis	obs	$N_{(10)}$
$(1,1)$	mis	mis	$N_{(11)}$

Method 5: missingness-pattern (mp) method

- It is also an intuitive method: just use whatever covariates we have!
- Propose the mp method:
  - stratify the units based on their covariates missingness patterns
  - use all the available covariates within each missingness pattern
  - weighted average of the estimators across missingness patterns
- The idea might not be entirely new either: Wilks (1932), Matthai (1951), Rosenbaum and Rubin (1984 Appendix B)
- We have not seen its use in covariate adjustment in randomized experiments with missing covariates

Illustrating the mp method with 2 missing covariates

- Fit additive regression for each missingness pattern:
  - regress Y_i on $(1, Z_i, x_{i1}, x_{i2})$ over $\{i : M_i = (0,0)\}$ to obtain $\hat{\tau}_{F,(0,0)}$
  - regress Y_i on $(1, Z_i, x_{i1})$ over $\{i : M_i = (0,1)\}$ to obtain $\hat{\tau}_{F,(0,1)}$
  - regress Y_i on $(1, Z_i, x_{i2})$ over $\{i : M_i = (1,0)\}$ to obtain $\hat{\tau}_{F,(1,0)}$
  - regress Y_i on $(1, Z_i, \emptyset)$ over $\{i : M_i = (1,1)\}$ to obtain $\hat{\tau}_{F,(1,1)}$
- Weighted average with $\rho_{(0,0)} = N_{(0,0)}/N$, etc:

$$\hat{\tau}_F^{\text{mp}} = \rho_{(0,0)}\hat{\tau}_{F,(0,0)} + \rho_{(0,1)}\hat{\tau}_{F,(0,1)} + \rho_{(1,0)}\hat{\tau}_{F,(1,0)} + \rho_{(1,1)}\hat{\tau}_{F,(1,1)}$$

- Can also obtain $\hat{\tau}_L^{\text{mp}}$ analogously

Comments on the mp method

- Somewhat more transparent, without explicit imputation
- Use all available covariate information for all units
- Much more demanding in sample size within each missingness pattern
- Not applicable in the motivating Duflo et al (2011 AER) example
- Potentially useful in other examples

Properties of the mp method

- Post-stratification estimators with covariate adjustment within each missingness pattern (Miratrix, Sekhon and Yu 2013) – conditional
- Effectively it uses $x_i^{\text{imp}}, M_{i1}, \ldots, M_{iJ}$ and their full interactions as "pseudo covariates", up to collinearity adjustment – unconditional
- Can be conveniently implemented by a single OLS
- EHW robust SE is a convenient approximation to the true SE
- Asymptotically more efficient than mim recommended before

Summary of the methods

- Complete-case (cc) analysis: not recommended
- Complete-covariate (ccov) analysis: benchmark
- Single imputation (imp): OK; but not invariant, not efficient
- Missingness-indicator method (mim): recommended
- Missingness-pattern (mp) method: can improve mim with more data

More comparisons

- Efficiency comparison based on Fisher (1935) is tricky with treatment effect heterogeneity (Freedman 2008)
- Focus on covariate adjustment based on Lin (2013)
- Ordering by asymptotic efficiency:

  $$mp > mim > imp > ccov > unadjusted$$

- The ordering is intuitive based on the amount of covariate information in the adjustment: the more the better

Discussion of other methods

- mp uses $x_i^{\text{imp}}, M_{i1}, \ldots, M_{iJ}$ and their full interactions
- mim uses $x_i^{\text{imp}}, M_{i1}, \ldots, M_{iJ}$ without any interactions
- Many intermediate choices of "pseudo covariates"
- Other methods use $x_i^{\text{imp}}(c), f(M_{i1}, \ldots, M_{iJ})$ as "pseudo covariates"
  - lose the invariance with respect to c
  - less demanding for sample size than mim
  - Rummel (1970): "missingness count" $\sum_{j=1}^{J}(1 - M_{ij})$ in factor analysis
- From unadjusted to mp estimators, there is a range of estimators
  - future direction: data-dependent choice of model specifications, e.g., combined with lasso (Bloniarz et al. 2015)