Agnostic notes on regression adjustments to experimental data

Reexamining Freedman's critique

Winston Lin

7 June 2011

Outline

Outline

Outline

Outline

1. OLS adjustment and Freedman's critique
    - The practice of adjustment
    - Why adjust?
    - Freedman's critique
2. Toward an agnostic view
    - Overview of this paper
    - Lessons from survey sampling
3. Reexamining Freedman's complaints
    - Possible worsening of asymptotic precision
    - Inconsistent SE estimates
    - Finite-sample bias
4. Possible directions for further research

Outline

Example: A welfare reform experiment

	Average Earnings ($)		
	Treatment group	Control group	Difference
Year 1	2,470	1,550	920***
Year 2	3,416	2,233	1,183***
Year 3	3,562	2,552	1,010***

Riccio et al. (MDRC, 1994)

Example: A welfare reform experiment

	Average Earnings ($)		
	Treatment group	Control group	Difference
Year 1	2,470	1,550	920***
Year 2	3,416	2,233	1,183***
Year 3	3,562	2,552	1,010***

Riccio et al. (MDRC, 1994)

- 5,508 welfare recipients and applicants (Riverside County, CA)

Example: A welfare reform experiment

	Average Earnings ($)		
	Treatment group	Control group	Difference
Year 1	2,470	1,550	920***
Year 2	3,416	2,233	1,183***
Year 3	3,562	2,552	1,010***

Riccio et al. (MDRC, 1994)

- 5,508 welfare recipients and applicants (Riverside County, CA)

- Treatment group: Mandatory job search / basic education
- Control group: No mandate

Example: A welfare reform experiment

	Average Earnings ($)		
	Treatment group	Control group	Difference
Year 1	2,470	1,550	920***
Year 2	3,416	2,233	1,183***
Year 3	3,562	2,552	1,010***

- All the estimates are regression-adjusted.

Example: A welfare reform experiment

	Average Earnings ($)		
	Treatment group	Control group	Difference
Year 1	2,470	1,550	920***
Year 2	3,416	2,233	1,183***
Year 3	3,562	2,552	1,010***

- All the estimates are regression-adjusted.
- Adjustment is standard in the evaluation industry

Example: A welfare reform experiment

	Average Earnings ($)		
	Treatment group	Control group	Difference
Year 1	2,470	1,550	920***
Year 2	3,416	2,233	1,183***
Year 3	3,562	2,552	1,010***

- All the estimates are regression-adjusted.
- Adjustment is standard in the evaluation industry and common in academic publications.

The usual OLS adjustment

Y_i	Outcome
T_i	Treatment group dummy
X_i	Covariate(s) measured before random assignment

The usual OLS adjustment

Y_i	Outcome
T_i	Treatment group dummy
X_i	Covariate(s) measured before random assignment

OLS regression:

$$Y_i = \widehat{\alpha} + \widehat{\beta} \cdot X_i + \widehat{\gamma} T_i + \widehat{\epsilon}_i$$

The usual OLS adjustment

Y_i	Outcome
T_i	Treatment group dummy
X_i	Covariate(s) measured before random assignment

OLS regression:

$$Y_i \quad = \quad \widehat{\alpha} + \widehat{\beta} \cdot X_i + \widehat{\gamma}\, T_i + \widehat{\epsilon}_i$$

Regression-adjusted estimates:

Treatment group mean $\quad \widehat{\alpha} + \widehat{\beta} \cdot \overline{X} + \widehat{\gamma}$

The usual OLS adjustment

Y_i	Outcome
T_i	Treatment group dummy
X_i	Covariate(s) measured before random assignment

OLS regression:

$$Y_i \quad = \quad \widehat{\alpha} + \widehat{\beta} \cdot X_i + \widehat{\gamma} T_i + \widehat{\epsilon}_i$$

Regression-adjusted estimates:

Treatment group mean	$\widehat{\alpha} + \widehat{\beta} \cdot \overline{X} + \widehat{\gamma}$
Control group mean	$\widehat{\alpha} + \widehat{\beta} \cdot \overline{X}$

The usual OLS adjustment

Y_i	Outcome
T_i	Treatment group dummy
X_i	Covariate(s) measured before random assignment

OLS regression:

$$Y_i \quad = \quad \widehat{\alpha} + \widehat{\beta} \cdot X_i + \widehat{\gamma} T_i + \widehat{\epsilon}_i$$

Regression-adjusted estimates:

Treatment group mean	$\widehat{\alpha} + \widehat{\beta} \cdot \overline{X} + \widehat{\gamma}$
Control group mean	$\widehat{\alpha} + \widehat{\beta} \cdot \overline{X}$
Difference	$\widehat{\gamma}$

where $\overline{X}$ is the mean covariate value for the study population.

Outline

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing
- Precision improvement
  (Fisher 1932; Cochran 1957; Cox & McCullagh 1982)

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing
- Precision improvement
  (Fisher 1932; Cochran 1957; Cox & McCullagh 1982)

  Assumptions:

  - Regression model is correct

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing
- Precision improvement
  (Fisher 1932; Cochran 1957; Cox & McCullagh 1982)

  Assumptions:
  - Regression model is correct
  - Fixed covariate list (no fishing)

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing
- Precision improvement
  (Fisher 1932; Cochran 1957; Cox & McCullagh 1982)

  Assumptions:
  - Regression model is correct
  - Fixed covariate list (no fishing)
  - $K \ll N$

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing
- Precision improvement
  (Fisher 1932; Cochran 1957; Cox & McCullagh 1982)

  Assumptions:
  - Regression model is correct
  - Fixed covariate list (no fishing)
  - $K \ll N$
  - Outcome correlated with at least one covariate

Why adjust?

The unadjusted difference in means is an unbiased estimator of the average treatment effect.

So why do researchers use regression adjustment?

- Fishing
- Precision improvement

  (Fisher 1932; Cochran 1957; Cox & McCullagh 1982)

  Assumptions:

  - Regression model is correct
  - Fixed covariate list (no fishing)
  - $K \ll N$
  - Outcome correlated with at least one covariate

  Adjustment tends to reduce the variance of the estimated treatment effect.

  (This is the standard rationale.)

Why adjust?

- Conditional bias
    - Suppose that by chance, more disadvantaged people were assigned to the control group.

Why adjust?

- Conditional bias
    - Suppose that by chance, more disadvantaged people were assigned to the control group.
    - Dr. Pangloss: "That's OK. The difference in means is unbiased over all possible random assignments."

Why adjust?

- Conditional bias
  - Suppose that by chance, more disadvantaged people were assigned to the control group.
  - Dr. Pangloss: "That's OK. The difference in means is unbiased over all possible random assignments."
  - Arguably, inference should be conditional on a measure of covariate imbalance (Senn 1989; Cox & Reid 2000).

Why adjust?

- Conditional bias
  - Suppose that by chance, more disadvantaged people were assigned to the control group.
  - Dr. Pangloss: "That's OK. The difference in means is unbiased over all possible random assignments."
  - Arguably, inference should be conditional on a measure of covariate imbalance (Senn 1989; Cox & Reid 2000).
  - Adjustment may reduce conditional bias.

Why adjust?

- Conditional bias
  - Suppose that by chance, more disadvantaged people were assigned to the control group.
  - Dr. Pangloss: "That's OK. The difference in means is unbiased over all possible random assignments."
  - Arguably, inference should be conditional on a measure of covariate imbalance (Senn 1989; Cox & Reid 2000).
  - Adjustment may reduce conditional bias.
- Attrition or survey nonresponse bias

Why adjust?

- Conditional bias
    - Suppose that by chance, more disadvantaged people were assigned to the control group.
    - Dr. Pangloss: "That's OK. The difference in means is unbiased over all possible random assignments."
    - Arguably, inference should be conditional on a measure of covariate imbalance (Senn 1989; Cox & Reid 2000).
    - Adjustment may reduce conditional bias.
- Attrition or survey nonresponse bias
- Robustness check (Tukey 1991)

Outline

Freedman's critique of OLS adjustment

"On regression adjustments to experimental data" (*Adv. Appl. Math.*, 2008a)

"On regression adjustments in experiments with several treatments" (*Ann. Appl. Stat.*, 2008b)

Freedman's critique of OLS adjustment

"On regression adjustments to experimental data" (*Adv. Appl. Math.*, 2008a)

"On regression adjustments in experiments with several treatments" (*Ann. Appl. Stat.*, 2008b)

Freedman derives the asymptotic distribution of the OLS-adjusted estimator without assuming a regression model. He uses Neyman's (1923) model for randomization inference.

Freedman's critique of OLS adjustment

"On regression adjustments to experimental data" (*Adv. Appl. Math.*, 2008a)

"On regression adjustments in experiments with several treatments" (*Ann. Appl. Stat.*, 2008b)

Freedman derives the asymptotic distribution of the OLS-adjusted estimator without assuming a regression model. He uses Neyman's (1923) model for randomization inference.

Finite population: The N subjects in the experiment

Freedman's critique of OLS adjustment

"On regression adjustments to experimental data" (*Adv. Appl. Math.*, 2008a)

"On regression adjustments in experiments with several treatments" (*Ann. Appl. Stat.*, 2008b)

Freedman derives the asymptotic distribution of the OLS-adjusted estimator without assuming a regression model. He uses Neyman's (1923) model for randomization inference.

Finite population: The N subjects in the experiment

Treatment group: Simple random sample

Freedman's critique of OLS adjustment

"On regression adjustments to experimental data"
(*Adv. Appl. Math.*, 2008a)

"On regression adjustments in experiments with
several treatments" (*Ann. Appl. Stat.*, 2008b)

Freedman derives the asymptotic distribution of the OLS-adjusted estimator without assuming a regression model. He uses Neyman's (1923) model for randomization inference.

Finite population: The N subjects in the experiment

Treatment group: Simple random sample
Control group: Everyone else

Neyman's model

Each subject i has two "potential outcomes":

Y_{1i}	Outcome that would occur if $T_i = 1$
Y_{0i}	Outcome that would occur if $T_i = 0$

Neyman's model

Each subject i has two "potential outcomes":

Y_{1i} Outcome that would occur if $T_i = 1$
Y_{0i} Outcome that would occur if $T_i = 0$

(Assume no interference between subjects.)

Neyman's model

Each subject i has two "potential outcomes":

Y_{1i}	Outcome that would occur if $T_i = 1$
Y_{0i}	Outcome that would occur if $T_i = 0$

(Assume no interference between subjects.)

Observed outcome: $Y_i \;=\; T_i Y_{1i} + (1 - T_i) Y_{0i}$.

Neyman's model

Each subject i has two "potential outcomes":

Y_{1i}	Outcome that would occur if $T_i = 1$
Y_{0i}	Outcome that would occur if $T_i = 0$

(Assume no interference between subjects.)

Observed outcome: $Y_i \;=\; T_i Y_{1i} + (1 - T_i) Y_{0i}$.

Estimand: Average treatment effect (ATE)

$$ATE \;\equiv\; \frac{1}{N} \sum_{i=1}^{N} (Y_{1i} - Y_{0i})$$

Neyman's model

T_i	Observed	Random

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random
X_i (covariate)	Observed	Fixed

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random
X_i (covariate)	Observed	Fixed

- Treatment effect $(Y_{1i} - Y_{0i})$ can vary with i

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random
X_i (covariate)	Observed	Fixed

- Treatment effect $(Y_{1i} - Y_{0i})$ can vary with i
- No assumptions about relationship between Y_i and X_i

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random
X_i (covariate)	Observed	Fixed

- Treatment effect $(Y_{1i} - Y_{0i})$ can vary with i
- No assumptions about relationship between Y_i and X_i
- No i.i.d. error term

Outline

OLS adjustment and Freedman's critique
- The practice of adjustment
- Why adjust?
- Freedman's critique

Toward an agnostic view
- Overview of this paper
- Lessons from survey sampling

Reexamining Freedman's complaints
- Asymptotic precision loss
- Inconsistent SE
- Finite-sample bias

Possible directions

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random
X_i (covariate)	Observed	Fixed

- Treatment effect $(Y_{1i} - Y_{0i})$ can vary with i
- No assumptions about relationship between Y_i and X_i
- No i.i.d. error term
- No imaginary superpopulation

Neyman's model

T_i	Observed	Random
Y_{1i}	Observed if $T_i = 1$	Fixed
Y_{0i}	Observed if $T_i = 0$	Fixed
Y_i	Observed	Random
X_i (covariate)	Observed	Fixed

- Treatment effect $(Y_{1i} - Y_{0i})$ can vary with i
- No assumptions about relationship between Y_i and X_i
- No i.i.d. error term
- No imaginary superpopulation
- Random assignment is the source of randomness

Freedman's conclusions

Freedman finds that under Neyman's model:

1. Adjustment can actually worsen asymptotic precision.

Freedman's conclusions

Freedman finds that under Neyman's model:

1. Adjustment can actually worsen asymptotic precision.
2. The conventional OLS standard error estimator is inconsistent.

Freedman's conclusions

Freedman finds that under Neyman's model:

1. Adjustment can actually worsen asymptotic precision.
2. The conventional OLS standard error estimator is inconsistent.
3. The adjusted estimator has a finite-sample bias of order $1/N$.

Freedman's conclusions

Freedman finds that under Neyman's model:

1. Adjustment can actually worsen asymptotic precision.
2. The conventional OLS standard error estimator is inconsistent.
3. The adjusted estimator has a finite-sample bias of order $1/N$.

"The reason for the breakdown is not hard to find: randomization does not justify the assumptions behind the OLS model."

Hidden bright side of Freedman's critique

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Adjustment hurts only under severe conditions

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Adjustment hurts only under severe conditions

- Suppose neither group has more than 75% of the subjects

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Adjustment hurts only under severe conditions

- Suppose neither group has more than 75% of the subjects
- Then for adjustment to hurt, X_i must covary more with the treatment effect than with the expected outcome.

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Adjustment hurts only under severe conditions

- Suppose neither group has more than 75% of the subjects
- Then for adjustment to hurt, X_i must covary more with the treatment effect than with the expected outcome.
- Not noted by Freedman, but follows from his asymptotic variance formula.

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Adjustment hurts only under severe conditions

- Suppose neither group has more than 75% of the subjects
- Then for adjustment to hurt, X_i must covary more with the treatment effect than with the expected outcome.
- Not noted by Freedman, but follows from his asymptotic variance formula.

SE estimation:

The conventional SE is consistent/conservative when allocation is 50–50

Hidden bright side of Freedman's critique

Asymptotic precision:

Adjustment can't hurt when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized

Adjustment hurts only under severe conditions

- Suppose neither group has more than 75% of the subjects
- Then for adjustment to hurt, X_i must covary more with the treatment effect than with the expected outcome.
- Not noted by Freedman, but follows from his asymptotic variance formula.

SE estimation:

The conventional SE is consistent/conservative when allocation is 50–50

- Noted by Freedman (2008a), but not emphasized
- The essential issue is the homoskedasticity assumption

The dark side has been emphasized

The dark side has been emphasized

"Random assignment does not justify any form of regression with covariates. If regression adjustments are introduced nevertheless, there is likely to be bias in any estimates of treatment effects and badly biased standard errors."

Berk et al. (2010), *Journal of Experimental Criminology*

Outline

Goal and formal results

Goal: More perspective and intuition

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?
- How strong should the warning label be?

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?
- How strong should the warning label be?

Two main formal results (under Freedman's assumptions):

1. Asymptotic distribution of "ANCOVA II"
   (OLS adjustment with treatment × covariate interactions)

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?
- How strong should the warning label be?

Two main formal results (under Freedman's assumptions):

1. Asymptotic distribution of "ANCOVA II"
   (OLS adjustment with treatment × covariate interactions)
   - ANCOVA II cannot hurt asymptotic precision

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?
- How strong should the warning label be?

Two main formal results (under Freedman's assumptions):

1. Asymptotic distribution of "ANCOVA II" (OLS adjustment with treatment × covariate interactions)
    - ANCOVA II cannot hurt asymptotic precision
    - Surprising, but not completely new

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?
- How strong should the warning label be?

Two main formal results (under Freedman's assumptions):

1. Asymptotic distribution of "ANCOVA II"
(OLS adjustment with treatment × covariate interactions)
    - ANCOVA II cannot hurt asymptotic precision
    - Surprising, but not completely new

2. Consistency/conservatism of Huber–White "sandwich" SE

Goal and formal results

Goal: More perspective and intuition

- *When* and *why* does adjustment do more harm than good?
- How strong should the warning label be?

Two main formal results (under Freedman's assumptions):

1. Asymptotic distribution of "ANCOVA II"
   (OLS adjustment with treatment × covariate interactions)
   - ANCOVA II cannot hurt asymptotic precision
   - Surprising, but not completely new

2. Consistency/conservatism of Huber–White "sandwich" SE
   - Not surprising at all

Perspective

Perspective

An "agnostic" way to study and teach regression:

Perspective

An "agnostic" way to study and teach regression:

- OLS has useful properties that don't depend on the model assumptions

Perspective

An "agnostic" way to study and teach regression:

- OLS has useful properties that don't depend on the model assumptions

  Convergence to the best linear predictor

Perspective

An "agnostic" way to study and teach regression:

- OLS has useful properties that don't depend on the model assumptions

  Convergence to the best linear predictor
  Consistent SE estimators (sandwich, jackknife)

Perspective

An "agnostic" way to study and teach regression:

- OLS has useful properties that don't depend on the model assumptions

  Convergence to the best linear predictor
  Consistent SE estimators (sandwich, jackknife)

- "Whether a regression specification is 'right' or 'wrong' … one can consider whether or not the population feature that [OLS] does consistently estimate is an interesting one."

  Goldberger (1991), *A Course in Econometrics*

Perspective

An "agnostic" way to study and teach regression:

- OLS has useful properties that don't depend on the model assumptions

  Convergence to the best linear predictor
  Consistent SE estimators (sandwich, jackknife)

- "Whether a regression specification is 'right' or 'wrong' ... one can consider whether or not the population feature that [OLS] does consistently estimate is an interesting one."

  Goldberger (1991), *A Course in Econometrics*

- Adopted by some econometricians

  White 1980a; Chamberlain 1982; Goldberger 1991; Angrist 1998; Angrist & Pischke 2009

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

- Regression estimators of population means (Cochran 1942)

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

- Regression estimators of population means (Cochran 1942)
- Very relevant to all of Freedman's issues

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

- Regression estimators of population means (Cochran 1942)
- Very relevant to all of Freedman's issues

From an agnostic perspective:

- Freedman's theorems are a major accomplishment.

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

- Regression estimators of population means (Cochran 1942)
- Very relevant to all of Freedman's issues

From an agnostic perspective:

- Freedman's theorems are a major accomplishment.
- Freedman's explanations oversimplify.

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

- Regression estimators of population means (Cochran 1942)
- Very relevant to all of Freedman's issues

From an agnostic perspective:

- Freedman's theorems are a major accomplishment.
- Freedman's explanations oversimplify.
  - "Since randomization does not justify the models, almost anything can happen." (Freedman 2008a, abstract)

Agnostic views of regression adjustment

Agnostic insights from survey sampling:

- Regression estimators of population means (Cochran 1942)
- Very relevant to all of Freedman's issues

From an agnostic perspective:

- Freedman's theorems are a major accomplishment.
- Freedman's explanations oversimplify.
    - "Since randomization does not justify the models, almost anything can happen." (Freedman 2008a, abstract)
    - We can give more specific explanations (and remedies) for each of Freedman's complaints.

Outline

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Estimand: $\overline{Y}_{pop}$

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Estimand: $\overline{Y}_{pop}$

$\overline{Y}_{sample}$ is unbiased, but it ignores the auxiliary info (X).

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Estimand: $\overline{Y}_{pop}$

$\overline{Y}_{sample}$ is unbiased, but it ignores the auxiliary info (X).

If $\overline{X}_{pop} > \overline{X}_{sample}$,

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Estimand: $\overline{Y}_{pop}$

$\overline{Y}_{sample}$ is unbiased, but it ignores the auxiliary info (X).

If $\overline{X}_{pop} > \overline{X}_{sample}$, then we expect $\overline{Y}_{pop} > \overline{Y}_{sample}$.

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Estimand: $\overline{Y}_{pop}$

$\overline{Y}_{sample}$ is unbiased, but it ignores the auxiliary info (X).

If $\overline{X}_{pop} > \overline{X}_{sample}$, then we expect $\overline{Y}_{pop} > \overline{Y}_{sample}$.

OLS regression estimator of $\overline{Y}_{pop}$:

Regression estimators in survey sampling

Example: Leaves on a plant (Watson 1937; Cochran 1977)

Y_i	Surface area	Measured for random sample
X_i	Mass	Measured for whole population

Estimand: $\overline{Y}_{pop}$

$\overline{Y}_{sample}$ is unbiased, but it ignores the auxiliary info (X).

If $\overline{X}_{pop} > \overline{X}_{sample}$, then we expect $\overline{Y}_{pop} > \overline{Y}_{sample}$.

OLS regression estimator of $\overline{Y}_{pop}$:

$$\widehat{\overline{Y}}_{OLS} \equiv \overline{Y}_{sample} + \widehat{\beta}_{OLS} \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

Consistency of regression estimators

Claim

Under simple random sampling,
$\widehat{\overline{Y}}_{OLS}$ is a consistent estimator of $\overline{Y}_{pop}$,
even if the regression model is false.

Consistency of regression estimators

Claim

Under simple random sampling,
$\widehat{\overline{Y}}_{OLS}$ is a consistent estimator of $\overline{Y}_{pop}$,
even if the regression model is false.

Sketch of Proof

$$\widehat{\overline{Y}}_{OLS} - \overline{Y}_{sample} \quad = \quad \widehat{\beta}_{OLS} \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

Consistency of regression estimators

Claim

Under simple random sampling,
$\widehat{\overline{Y}}_{OLS}$ is a consistent estimator of $\overline{Y}_{pop}$,
even if the regression model is false.

Sketch of Proof

$$\widehat{\overline{Y}}_{OLS} - \overline{Y}_{sample} \quad = \quad \widehat{\beta}_{OLS} \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

Under suitable regularity conditions:

1. $(\overline{X}_{pop} - \overline{X}_{sample}) \xrightarrow{p} 0$.
2. $\widehat{\beta}_{OLS}$ converges to a finite limit.
3. Therefore, $(\widehat{\overline{Y}}_{OLS} - \overline{Y}_{sample}) \xrightarrow{p} 0$.
4. $\overline{Y}_{sample}$ is consistent.
5. Therefore, $\widehat{\overline{Y}}_{OLS}$ is consistent.

Precision improvement in survey sampling

Claim

Under simple random sampling,
OLS adjustment of the estimated mean
cannot hurt asymptotic precision,
even if the regression model is false.

Precision improvement in survey sampling

Claim

Under simple random sampling,
OLS adjustment of the estimated mean
cannot hurt asymptotic precision,
even if the regression model is false.

Sketch of Proof (adapted from Cochran 1977)

First, imagine using a "fixed-slope regression estimator":

$$\widehat{\overline{Y}}_{fixedslope} \equiv \overline{Y}_{sample} + b \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

where b is a constant.

Precision improvement in survey sampling

Claim

Under simple random sampling,
OLS adjustment of the estimated mean
cannot hurt asymptotic precision,
even if the regression model is false.

Sketch of Proof (adapted from Cochran 1977)

First, imagine using a "fixed-slope regression estimator":

$$\widehat{\overline{Y}}_{fixedslope} \quad \equiv \quad \overline{Y}_{sample} + b \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

where b is a constant.

Note that $\overline{Y}_{sample}$ itself is a fixed-slope regression estimator (with $b = 0$).

Precision improvement in survey sampling

Sketch of Proof (cont'd)

$$\widehat{\overline{Y}}_{fixed\,slope} \;=\; \overline{Y}_{sample} - b \cdot (\overline{X}_{sample} - \overline{X}_{pop})$$

is the sample mean of $Y_i - b \cdot (X_i - \overline{X}_{pop})$, so its variance is

Precision improvement in survey sampling

Sketch of Proof (cont'd)

$$\widehat{\overline{Y}}_{fixed\,slope} \;=\; \overline{Y}_{sample} - b \cdot (\overline{X}_{sample} - \overline{X}_{pop})$$

is the sample mean of $Y_i - b \cdot (X_i - \overline{X}_{pop})$, so its variance is

$$\frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{1}{N} \sum_{i=1}^{N} \left[(Y_i - \overline{Y}_{pop}) - b \cdot (X_i - \overline{X}_{pop}) \right]^2 .$$

Precision improvement in survey sampling

Sketch of Proof (cont'd)

$$\widehat{\overline{Y}}_{fixed\,slope} \;=\; \overline{Y}_{sample} - b \cdot (\overline{X}_{sample} - \overline{X}_{pop})$$

is the sample mean of $Y_i - b \cdot (X_i - \overline{X}_{pop})$, so its variance is

$$\frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{1}{N} \sum_{i=1}^{N} \left[(Y_i - \overline{Y}_{pop}) - b \cdot (X_i - \overline{X}_{pop}) \right]^2 .$$

What choice of b minimizes this variance?

Precision improvement in survey sampling

Sketch of Proof (cont'd)

$$\widehat{\overline{Y}}_{fixed\,slope} \;=\; \overline{Y}_{sample} - b\cdot(\overline{X}_{sample} - \overline{X}_{pop})$$

is the sample mean of $Y_i - b\cdot(X_i - \overline{X}_{pop})$, so its variance is

$$\frac{N-n}{N-1}\cdot\frac{1}{n}\cdot\frac{1}{N}\sum_{i=1}^{N}\left[(Y_i - \overline{Y}_{pop}) - b\cdot(X_i - \overline{X}_{pop})\right]^2 .$$

What choice of b minimizes this variance?
The "population least squares" slope, β_{PopLS}.

Precision improvement in survey sampling

Sketch of Proof (cont'd)

$$\widehat{\overline{Y}}_{fixed\,slope} = \overline{Y}_{sample} - b \cdot (\overline{X}_{sample} - \overline{X}_{pop})$$

is the sample mean of $Y_i - b \cdot (X_i - \overline{X}_{pop})$, so its variance is

$$\frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{1}{N} \sum_{i=1}^{N} \left[(Y_i - \overline{Y}_{pop}) - b \cdot (X_i - \overline{X}_{pop})\right]^2.$$

What choice of b minimizes this variance?
The "population least squares" slope, β_{PopLS}.
Call the resulting estimator $\widehat{\overline{Y}}_{PopLS}$.

Precision improvement in survey sampling

Sketch of Proof (cont'd)

$$\widehat{\overline{Y}}_{fixed\,slope} \;=\; \overline{Y}_{sample} - b\cdot(\overline{X}_{sample} - \overline{X}_{pop})$$

is the sample mean of $Y_i - b\cdot(X_i - \overline{X}_{pop})$, so its variance is

$$\frac{N-n}{N-1}\cdot\frac{1}{n}\cdot\frac{1}{N}\sum_{i=1}^{N}\left[(Y_i - \overline{Y}_{pop}) - b\cdot(X_i - \overline{X}_{pop})\right]^2 .$$

What choice of b minimizes this variance?
The "population least squares" slope, β_{PopLS}.
Call the resulting estimator $\widehat{\overline{Y}}_{PopLS}$.

$\widehat{\overline{Y}}_{PopLS}$ has lower variance than $\overline{Y}_{sample}$ if $\beta_{PopLS} \neq 0$.

Precision improvement in survey sampling

Sketch of Proof (cont'd)

Asymptotically, $\widehat{\overline{Y}}_{OLS}$ is as efficient as $\widehat{\overline{Y}}_{PopLS}$:

Precision improvement in survey sampling

Sketch of Proof (cont'd)

Asymptotically, $\widehat{\overline{Y}}_{OLS}$ is as efficient as $\widehat{\overline{Y}}_{PopLS}$:

$$\widehat{\overline{Y}}_{OLS} - \overline{Y}_{pop} = (\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop}) + (\widehat{\beta}_{OLS} - \beta_{PopLS}) \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

Precision improvement in survey sampling

Sketch of Proof (cont'd)

Asymptotically, $\widehat{\overline{Y}}_{OLS}$ is as efficient as $\widehat{\overline{Y}}_{PopLS}$:

$$\widehat{\overline{Y}}_{OLS} - \overline{Y}_{pop} = (\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop}) + (\widehat{\beta}_{OLS} - \beta_{PopLS}) \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

- $(\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop})$ is of order $1/\sqrt{n}$

Precision improvement in survey sampling

Sketch of Proof (cont'd)

Asymptotically, $\widehat{\overline{Y}}_{OLS}$ is as efficient as $\widehat{\overline{Y}}_{PopLS}$:

$$\widehat{\overline{Y}}_{OLS} - \overline{Y}_{pop} = (\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop}) + (\widehat{\beta}_{OLS} - \beta_{PopLS}) \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

- $(\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop})$ is of order $1/\sqrt{n}$
- $(\widehat{\beta}_{OLS} - \beta_{PopLS}) \cdot (\overline{X}_{pop} - \overline{X}_{sample})$ is of order $1/n$

Precision improvement in survey sampling

Sketch of Proof (cont'd)

Asymptotically, $\widehat{\overline{Y}}_{OLS}$ is as efficient as $\widehat{\overline{Y}}_{PopLS}$:

$$\widehat{\overline{Y}}_{OLS} - \overline{Y}_{pop} = (\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop}) + (\widehat{\beta}_{OLS} - \beta_{PopLS}) \cdot (\overline{X}_{pop} - \overline{X}_{sample})$$

- $(\widehat{\overline{Y}}_{PopLS} - \overline{Y}_{pop})$ is of order $1/\sqrt{n}$
- $(\widehat{\beta}_{OLS} - \beta_{PopLS}) \cdot (\overline{X}_{pop} - \overline{X}_{sample})$ is of order $1/n$

So for large enough n,

$$\mathrm{Var}(\widehat{\overline{Y}}_{OLS}) \approx \mathrm{Var}(\widehat{\overline{Y}}_{PopLS}) \leq \mathrm{Var}(\overline{Y}_{sample}).$$

■

Separated at birth?

	Survey sampling	Experiments

Separated at birth?

	Survey sampling	Experiments
Estimand	$\overline{Y}_{pop}$	$\overline{Y}_{1,pop} - \overline{Y}_{0,pop}$

Separated at birth?

	Survey sampling	Experiments
Estimand	$\overline{Y}_{pop}$	$\overline{Y}_{1,pop} - \overline{Y}_{0,pop}$
Outcome data	Y_i (sample)	Y_{1i} (treatment group) Y_{0i} (control group)

Separated at birth?

	Survey sampling	Experiments
Estimand	$\overline{Y}_{pop}$	$\overline{Y}_{1,pop} - \overline{Y}_{0,pop}$
Outcome data	Y_i (sample)	Y_{1i} (treatment group) Y_{0i} (control group)
Auxiliary data	X_i (population)	X_i (population)

Separated at birth?

	Survey sampling	Experiments
Estimand	$\overline{Y}_{pop}$	$\overline{Y}_{1,pop} - \overline{Y}_{0,pop}$
Outcome data	Y_i (sample)	Y_{1i} (treatment group) Y_{0i} (control group)
Auxiliary data	X_i (population)	X_i (population)
Cochran's classic paper	Sampling theory when the sampling-units are of unequal sizes *JASA* (1942)	Analysis of covariance: Its nature and uses *Biometrics* (1957)

Separated at birth?

	Survey sampling	Experiments
Estimand	$\overline{Y}_{pop}$	$\overline{Y}_{1,pop} - \overline{Y}_{0,pop}$
Outcome data	Y_i (sample)	Y_{1i} (treatment group) Y_{0i} (control group)
Auxiliary data	X_i (population)	X_i (population)
Cochran's classic paper	Sampling theory when the sampling-units are of unequal sizes *JASA* (1942)	Analysis of covariance: Its nature and uses *Biometrics* (1957)
Theory	Agnostic	Linear model

What the analogy suggests

1. Estimate $\overline{Y}_{1,pop}$

What the analogy suggests

1. Estimate $\overline{Y}_{1,pop}$
   - Regress Y_i on X_i in treatment group $\longrightarrow$ $\widehat{\beta}_{treat}$

What the analogy suggests

① Estimate $\overline{Y}_{1,pop}$

- Regress Y_i on X_i in treatment group $\longrightarrow \widehat{\beta}_{treat}$
- $\widehat{\overline{Y}}_{1,OLS} = \overline{Y}_{treat} + \widehat{\beta}_{treat} \cdot (\overline{X}_{pop} - \overline{X}_{treat})$

What the analogy suggests

1. Estimate $\overline{Y}_{1,pop}$
   - Regress Y_i on X_i in treatment group $\longrightarrow$ $\widehat{\beta}_{treat}$
   - $\widehat{\overline{Y}}_{1,OLS} \; = \; \overline{Y}_{treat} + \widehat{\beta}_{treat} \cdot (\overline{X}_{pop} - \overline{X}_{treat})$
2. Estimate $\overline{Y}_{0,pop}$
   - Regress Y_i on X_i in control group $\longrightarrow$ $\widehat{\beta}_{control}$
   - $\widehat{\overline{Y}}_{0,OLS} \; = \; \overline{Y}_{control} + \widehat{\beta}_{control} \cdot (\overline{X}_{pop} - \overline{X}_{control})$

What the analogy suggests

1. Estimate $\overline{Y}_{1,pop}$
   - Regress Y_i on X_i in treatment group $\longrightarrow \widehat{\beta}_{treat}$
   - $\widehat{\overline{Y}}_{1,OLS} = \overline{Y}_{treat} + \widehat{\beta}_{treat} \cdot (\overline{X}_{pop} - \overline{X}_{treat})$
2. Estimate $\overline{Y}_{0,pop}$
   - Regress Y_i on X_i in control group $\longrightarrow \widehat{\beta}_{control}$
   - $\widehat{\overline{Y}}_{0,OLS} = \overline{Y}_{control} + \widehat{\beta}_{control} \cdot (\overline{X}_{pop} - \overline{X}_{control})$
3. Take the difference

$$\widehat{ATE}_{ANCOVA\ II} = \widehat{\overline{Y}}_{1,OLS} - \widehat{\overline{Y}}_{0,OLS}$$

What the analogy suggests

1. Estimate $\overline{Y}_{1,pop}$
    - Regress Y_i on X_i in treatment group $\longrightarrow \widehat{\beta}_{treat}$
    - $\widehat{\overline{Y}}_{1,OLS} = \overline{Y}_{treat} + \widehat{\beta}_{treat} \cdot (\overline{X}_{pop} - \overline{X}_{treat})$
2. Estimate $\overline{Y}_{0,pop}$
    - Regress Y_i on X_i in control group $\longrightarrow \widehat{\beta}_{control}$
    - $\widehat{\overline{Y}}_{0,OLS} = \overline{Y}_{control} + \widehat{\beta}_{control} \cdot (\overline{X}_{pop} - \overline{X}_{control})$
3. Take the difference

$$\widehat{ATE}_{ANCOVA\ II} = \widehat{\overline{Y}}_{1,OLS} - \widehat{\overline{Y}}_{0,OLS}$$

- Equivalent to regressing Y_i on T_i, X_i, and $T_i \cdot (X_i - \overline{X}_{pop})$

What the analogy suggests

① Estimate $\overline{Y}_{1,pop}$

- Regress Y_i on X_i in treatment group $\longrightarrow \widehat{\beta}_{treat}$
- $\widehat{\overline{Y}}_{1,OLS} = \overline{Y}_{treat} + \widehat{\beta}_{treat} \cdot (\overline{X}_{pop} - \overline{X}_{treat})$

② Estimate $\overline{Y}_{0,pop}$

- Regress Y_i on X_i in control group $\longrightarrow \widehat{\beta}_{control}$
- $\widehat{\overline{Y}}_{0,OLS} = \overline{Y}_{control} + \widehat{\beta}_{control} \cdot (\overline{X}_{pop} - \overline{X}_{control})$

③ Take the difference

$$\widehat{ATE}_{ANCOVA\ II} = \widehat{\overline{Y}}_{1,OLS} - \widehat{\overline{Y}}_{0,OLS}$$

- Equivalent to regressing Y_i on T_i, X_i, and $T_i \cdot (X_i - \overline{X}_{pop})$
- Similar to a well-known nonexperimental method

  Educational statistics: Peters–Belson (Cochran 1969)
  Labor economics: Oaxaca–Blinder (Kline 2011)

Outline

1. OLS adjustment and Freedman's critique
    - The practice of adjustment
    - Why adjust?
    - Freedman's critique
2. Toward an agnostic view
    - Overview of this paper
    - Lessons from survey sampling
3. Reexamining Freedman's complaints
    - Possible worsening of asymptotic precision
    - Inconsistent SE estimates
    - Finite-sample bias
4. Possible directions for further research

Finite-population asymptotics

Finite-population asymptotics

Infinite-population asymptotics:

- Sample size $n \to \infty$
- Population and estimand don't change

Finite-population asymptotics

Infinite-population asymptotics:

- Sample size $n \to \infty$
- Population and estimand don't change
- Regularity conditions are about the population:

$$E(X_i^4) < \infty$$

Finite-population asymptotics

Infinite-population asymptotics:

- Sample size $n \to \infty$
- Population and estimand don't change
- Regularity conditions are about the population:

$$E(X_i^4) < \infty$$

Finite-population asymptotics:

- Population size $N \to \infty$

Finite-population asymptotics

Infinite-population asymptotics:

- Sample size $n \to \infty$
- Population and estimand don't change
- Regularity conditions are about the population:

  $$E(X_i^4) < \infty$$

Finite-population asymptotics:

- Population size $N \to \infty$
- Regularity conditions are about an imaginary infinite sequence of populations:

Finite-population asymptotics

Infinite-population asymptotics:

- Sample size $n \to \infty$
- Population and estimand don't change
- Regularity conditions are about the population:

$$E(X_i^4) \; < \; \infty$$

Finite-population asymptotics:

- Population size $N \to \infty$
- Regularity conditions are about an imaginary infinite sequence of populations:

$$\frac{1}{N}\sum_{i=1}^{N} X_{i,N}^4 \; < \; L \qquad \text{for } N = 1, 2, \ldots$$

Freedman's regularity conditions

For simplicity, assume a single covariate X_i.
Results generalize to multiple covariates.

Freedman's regularity conditions

For simplicity, assume a single covariate X_i.
Results generalize to multiple covariates.

Condition 1

Y_{1i}, Y_{0i}, and X_i have bounded fourth moments.
For example, there exists $L < \infty$ such that

$$\frac{1}{N}\sum_{i=1}^{N} X_i^4 \; < \; L \qquad \text{for } N = 1, 2, \ldots .$$

Freedman's regularity conditions

Condition 2

The population means, variances, and covariances of Y_{1i}, Y_{0i}, and X_i converge to finite limits.
The limits of the variances are positive.

Freedman's regularity conditions

Condition 2

The population means, variances, and covariances of Y_{1i}, Y_{0i}, and X_i converge to finite limits.
The limits of the variances are positive.

Condition 3

Both the treatment group and the control group are of order N in size:

Freedman's regularity conditions

Condition 2

The population means, variances, and covariances of Y_{1i}, Y_{0i}, and X_i converge to finite limits.
The limits of the variances are positive.

Condition 3

Both the treatment group and the control group are of order N in size:

Let N_T denote the treatment group size. Then

$$\lim_{N \to \infty} \frac{N_T}{N} = p \qquad \text{where } 0 < p < 1.$$

Central Limit Theorem

Freedman's finite-population CLT for experiments

Under Conditions 1–3,

$$\sqrt{N}\left(\begin{bmatrix} \overline{Y}_{1,treat} \\ \overline{Y}_{0,control} \\ \overline{X}_{treat} \\ \overline{X}_{control} \end{bmatrix} - \begin{bmatrix} \overline{Y}_{1,pop} \\ \overline{Y}_{0,pop} \\ \overline{X}_{pop} \\ \overline{X}_{pop} \end{bmatrix}\right) \xrightarrow{d} \text{Normal}\ (0, V)$$

where the elements of V follow the pattern on the next slide.

Central Limit Theorem

Freedman's finite-population CLT (cont'd)

$$N \, \mathrm{Var}(\overline{Y}_{1,treat}) \quad \rightarrow \quad \frac{1-p}{p} \lim_{N \to \infty} \sigma^2_{Y_1}$$

Central Limit Theorem

Freedman's finite-population CLT (cont'd)

$$N\ \mathrm{Var}(\overline{Y}_{1,treat}) \quad \rightarrow \quad \frac{1-p}{p} \lim_{N\rightarrow\infty} \sigma^2_{Y_1}$$

$$N\ \mathrm{Var}(\overline{Y}_{0,control}) \quad \rightarrow \quad \frac{p}{1-p} \lim_{N\rightarrow\infty} \sigma^2_{Y_0}$$

Central Limit Theorem

Freedman's finite-population CLT (cont'd)

$$
\begin{aligned}
N\ \mathrm{Var}(\overline{Y}_{1,treat}) &\rightarrow \frac{1-p}{p}\lim_{N\to\infty}\sigma^2_{Y_1}\\
N\ \mathrm{Var}(\overline{Y}_{0,control}) &\rightarrow \frac{p}{1-p}\lim_{N\to\infty}\sigma^2_{Y_0}\\
N\ \mathrm{Cov}(\overline{Y}_{1,treat},\overline{Y}_{0,control}) &\rightarrow -\lim_{N\to\infty}\sigma_{Y_1,Y_0}
\end{aligned}
$$

Central Limit Theorem

Freedman's finite-population CLT (cont'd)

$$
\begin{aligned}
N\ \mathrm{Var}(\overline{Y}_{1,treat}) &\rightarrow \frac{1-p}{p} \lim_{N\to\infty} \sigma^2_{Y_1} \\
N\ \mathrm{Var}(\overline{Y}_{0,control}) &\rightarrow \frac{p}{1-p} \lim_{N\to\infty} \sigma^2_{Y_0} \\
N\ \mathrm{Cov}(\overline{Y}_{1,treat}, \overline{Y}_{0,control}) &\rightarrow -\lim_{N\to\infty} \sigma_{Y_1,Y_0} \\
N\ \mathrm{Cov}(\overline{X}_{treat}, \overline{Y}_{1,treat}) &\rightarrow \frac{1-p}{p} \lim_{N\to\infty} \sigma_{X,Y_1}
\end{aligned}
$$

etc.

Some notation for my theorems

Limits of population LS slopes:

$$\beta_1 \quad \equiv \quad \lim_{N\to\infty} \frac{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})(Y_{1i} - \overline{Y}_{1,pop})}{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})^2}$$

Some notation for my theorems

Limits of population LS slopes:

$$\beta_1 \equiv \lim_{N\to\infty} \frac{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})(Y_{1i} - \overline{Y}_{1,pop})}{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})^2}$$

$$\beta_0 \equiv \lim_{N\to\infty} \frac{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})(Y_{0i} - \overline{Y}_{0,pop})}{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})^2}$$

Some notation for my theorems

Limits of population LS slopes:

$$\beta_1 \equiv \lim_{N\to\infty} \frac{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})(Y_{1i} - \overline{Y}_{1,pop})}{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})^2}$$

$$\beta_0 \equiv \lim_{N\to\infty} \frac{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})(Y_{0i} - \overline{Y}_{0,pop})}{\sum_{i=1}^{N}(X_i - \overline{X}_{pop})^2}$$

Potential outcomes, minus variation predicted by X_i:

$$R_{1i} \equiv Y_{1i} - \beta_1 \cdot (X_i - \overline{X}_{pop})$$

Some notation for my theorems

Limits of population LS slopes:

$$\beta_1 \equiv \lim_{N\to\infty} \frac{\sum_{i=1}^N (X_i - \overline{X}_{pop})(Y_{1i} - \overline{Y}_{1,pop})}{\sum_{i=1}^N (X_i - \overline{X}_{pop})^2}$$

$$\beta_0 \equiv \lim_{N\to\infty} \frac{\sum_{i=1}^N (X_i - \overline{X}_{pop})(Y_{0i} - \overline{Y}_{0,pop})}{\sum_{i=1}^N (X_i - \overline{X}_{pop})^2}$$

Potential outcomes, minus variation predicted by X_i:

$$R_{1i} \equiv Y_{1i} - \beta_1 \cdot (X_i - \overline{X}_{pop})$$

$$R_{0i} \equiv Y_{0i} - \beta_0 \cdot (X_i - \overline{X}_{pop})$$

Asymptotic distribution of ANCOVA II

Theorem 1

Under Conditions 1–3,

$$\sqrt{N}\ \left(\widehat{ATE}_{ANCOVA\ II} - ATE\right)\ \xrightarrow{d}\ \text{Normal}\ (0, v)$$

where

$$v\ =\ \frac{1-p}{p}\lim_{N\to\infty}\sigma^2_{R_1} + \frac{p}{1-p}\lim_{N\to\infty}\sigma^2_{R_0} + 2\lim_{N\to\infty}\sigma_{R_1,R_0}.$$

ANCOVA II cannot hurt asymptotic precision

Corollary 1.1

Asymptotically, ANCOVA II is at least as efficient as the difference in means,

ANCOVA II cannot hurt asymptotic precision

Corollary 1.1

Asymptotically, ANCOVA II is at least as efficient as the difference in means, and strictly more efficient unless

$$(1-p)\beta_1 + p\beta_0 \quad = \quad 0.$$

ANCOVA II cannot hurt asymptotic precision

Corollary 1.1

Asymptotically, ANCOVA II is at least as efficient as the difference in means, and strictly more efficient unless

$$(1-p)\beta_1 + p\beta_0 \quad = \quad 0.$$

Corollary 1.2

Asymptotically, ANCOVA II is at least as efficient as ANCOVA I (OLS adjustment without interactions),

ANCOVA II cannot hurt asymptotic precision

Corollary 1.1

Asymptotically, ANCOVA II is at least as efficient as the difference in means, and strictly more efficient unless

$$(1 - p)\beta_1 + p\beta_0 \quad = \quad 0.$$

Corollary 1.2

Asymptotically, ANCOVA II is at least as efficient as ANCOVA I (OLS adjustment without interactions), and strictly more efficient unless either $\beta_1 = \beta_0$ or $p = 0.5$.

Outline of proof of Theorem 1

1. Imagine a “fixed slope” estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0

Outline of proof of Theorem 1

1. Imagine a "fixed slope" estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$

Outline of proof of Theorem 1

1. Imagine a "fixed slope" estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$
3. Check that R_{1i} and R_{0i} satisfy the regularity conditions (use Jensen's inequality)
4. Apply Freedman's CLT to get the asymptotic distribution

Outline of proof of Theorem 1

1. Imagine a "fixed slope" estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$
3. Check that R_{1i} and R_{0i} satisfy the regularity conditions (use Jensen's inequality)
4. Apply Freedman's CLT to get the asymptotic distribution
5. Show that $\sqrt{N}(\widehat{ATE}_{ANCOVA\ II} - \widehat{ATE}_{Ideal\ ANCOVA\ II}) \xrightarrow{p} 0$

Outline of proof of Theorem 1

1. Imagine a "fixed slope" estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$
3. Check that R_{1i} and R_{0i} satisfy the regularity conditions (use Jensen's inequality)
4. Apply Freedman's CLT to get the asymptotic distribution
5. Show that $\sqrt{N}(\widehat{ATE}_{ANCOVA\ II} - \widehat{ATE}_{Ideal\ ANCOVA\ II}) \xrightarrow{p} 0$
   - This is $\sqrt{N}(\widehat{\beta}_{treat} - \beta_1) \cdot (\overline{X}_{pop} - \overline{X}_{treat})$ minus the analogous term for the control group

Outline of proof of Theorem 1

1. Imagine a “fixed slope” estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$
3. Check that R_{1i} and R_{0i} satisfy the regularity conditions (use Jensen's inequality)
4. Apply Freedman's CLT to get the asymptotic distribution
5. Show that $\sqrt{N}(\widehat{ATE}_{ANCOVA\ II} - \widehat{ATE}_{Ideal\ ANCOVA\ II}) \xrightarrow{p} 0$
    - This is $\sqrt{N}(\widehat{\beta}_{treat} - \beta_1) \cdot (\overline{X}_{pop} - \overline{X}_{treat})$ minus the analogous term for the control group
    - Show $\widehat{\beta}_{treat} \xrightarrow{p} \beta_1$ (use Chebyshev and Cauchy-Schwarz to prove a WLLN)

Outline of proof of Theorem 1

1. Imagine a "fixed slope" estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$
3. Check that R_{1i} and R_{0i} satisfy the regularity conditions (use Jensen's inequality)
4. Apply Freedman's CLT to get the asymptotic distribution
5. Show that $\sqrt{N}(\widehat{ATE}_{ANCOVA\ II} - \widehat{ATE}_{Ideal\ ANCOVA\ II}) \xrightarrow{p} 0$
   - This is $\sqrt{N}(\widehat{\beta}_{treat} - \beta_1) \cdot (\overline{X}_{pop} - \overline{X}_{treat})$ minus the analogous term for the control group
   - Show $\widehat{\beta}_{treat} \xrightarrow{p} \beta_1$ (use Chebyshev and Cauchy-Schwarz to prove a WLLN)
   - By Freedman's CLT, $\sqrt{N}(\overline{X}_{pop} - \overline{X}_{treat})$ is of order 1

Outline of proof of Theorem 1

1. Imagine a "fixed slope" estimator $\widehat{ATE}_{Ideal\ ANCOVA\ II}$, using β_1 and β_0
2. That equals $\overline{R}_{1,treat} - \overline{R}_{0,control}$
3. Check that R_{1i} and R_{0i} satisfy the regularity conditions (use Jensen's inequality)
4. Apply Freedman's CLT to get the asymptotic distribution
5. Show that $\sqrt{N}(\widehat{ATE}_{ANCOVA\ II} - \widehat{ATE}_{Ideal\ ANCOVA\ II}) \xrightarrow{p} 0$
    - This is $\sqrt{N}(\widehat{\beta}_{treat} - \beta_1) \cdot (\overline{X}_{pop} - \overline{X}_{treat})$ minus the analogous term for the control group
    - Show $\widehat{\beta}_{treat} \xrightarrow{p} \beta_1$ (use Chebyshev and Cauchy-Schwarz to prove a WLLN)
    - By Freedman's CLT, $\sqrt{N}(\overline{X}_{pop} - \overline{X}_{treat})$ is of order 1
    - So $\sqrt{N}(\widehat{\beta}_{treat} - \beta_1) \cdot (\overline{X}_{pop} - \overline{X}_{treat}) \xrightarrow{p} 0$

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
  - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
  - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
  - Treatment group has 75% or 90% of the subjects

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
  - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
  - Treatment group has 75% or 90% of the subjects
- Outcome-covariate relationships are far from linear

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
    - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
    - Treatment group has 75% or 90% of the subjects
- Outcome-covariate relationships are far from linear

Population size: $N = 1,000$

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
  - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
  - Treatment group has 75% or 90% of the subjects
- Outcome-covariate relationships are far from linear

Population size: $N = 1,000$

Scenario	N_T	Y_{1i}	Y_{0i}
A	900	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
  - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
  - Treatment group has 75% or 90% of the subjects
- Outcome-covariate relationships are far from linear

Population size: $N = 1,000$

Scenario	N_T	Y_{1i}	Y_{0i}
A	900	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$
B	750	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$
C	750	$e^{X_i/2} - e^{X_i} + \nu_i$	$e^{X_i/2} + e^{X_i} + \epsilon_i$
D	750	$e^{X_i/2} + \nu_i$	$e^{X_i/2} - 2e^{X_i} + \epsilon_i$

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
  - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
  - Treatment group has 75% or 90% of the subjects
- Outcome-covariate relationships are far from linear

Population size: $N = 1,000$

Scenario	N_T	Y_{1i}	Y_{0i}
A	900	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$
B	750	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$
C	750	$e^{X_i/2} - e^{X_i} + \nu_i$	$e^{X_i/2} + e^{X_i} + \epsilon_i$
D	750	$e^{X_i/2} + \nu_i$	$e^{X_i/2} - 2e^{X_i} + \epsilon_i$

$X_i, \nu_i, \epsilon_i \sim$ Normal $(0, 1)$, fixed across replications.

Monte Carlo simulations: Setup

Illustrates efficiency of ANCOVA II in scenarios where:

- ANCOVA I hurts precision or doesn't help much
    - X_i covaries more with $Y_{1i} - Y_{0i}$ than with Y_{1i} or Y_{0i}
    - Treatment group has 75% or 90% of the subjects
- Outcome-covariate relationships are far from linear

Population size: $N = 1,000$

Scenario	N_T	Y_{1i}	Y_{0i}
A	900	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$
B	750	$e^{X_i/2} + e^{X_i} + \nu_i$	$e^{X_i/2} - e^{X_i} + \epsilon_i$
C	750	$e^{X_i/2} - e^{X_i} + \nu_i$	$e^{X_i/2} + e^{X_i} + \epsilon_i$
D	750	$e^{X_i/2} + \nu_i$	$e^{X_i/2} - 2e^{X_i} + \epsilon_i$

$X_i, \nu_i, \epsilon_i \sim$ Normal $(0, 1)$, fixed across replications.

10,000 replications of random assignment.

Monte Carlo simulations: Results

Scenario	Monte Carlo SD		
	Difference in means	ANCOVA I	ANCOVA II
A	0.168	0.323	0.144
B	0.073	0.144	0.071
C	0.141	0.158	0.096
D	0.221	0.207	0.166

Comments on asymptotic precision

- Related work:
  - Yang & Tsiatis (2001)
    Tsiatis, Davidian, Zhang, & Lu (2008)

Comments on asymptotic precision

- Related work:
  - Yang & Tsiatis (2001)
    Tsiatis, Davidian, Zhang, & Lu (2008)
  - Schochet (2010)

Comments on asymptotic precision

- Related work:
  - Yang & Tsiatis (2001)
    Tsiatis, Davidian, Zhang, & Lu (2008)
  - Schochet (2010)
- I'm not advocating ANCOVA II

Comments on asymptotic precision

- Related work:
    - Yang & Tsiatis (2001)
      Tsiatis, Davidian, Zhang, & Lu (2008)
    - Schochet (2010)
- I'm not advocating ANCOVA II
- But the result sheds light on Freedman's warning that adjustment can hurt asymptotic precision.

Comments on asymptotic precision

- Related work:
  - Yang & Tsiatis (2001)
    Tsiatis, Davidian, Zhang, & Lu (2008)
  - Schochet (2010)
- I'm not advocating ANCOVA II
- But the result sheds light on Freedman's warning that adjustment can hurt asymptotic precision.

  The essential problem is omission of
  treatment × covariate interactions,
  not the linear model.

Outline

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
    - Don't need homoskedasticity or even linearity

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
    - Don't need homoskedasticity or even linearity
- Many independent discoverers
    - Sample analog of Huber's (1967) asymptotic variance
    - Eicker 1967; Fuller 1975; Hinkley 1977; White 1980ab

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
    - Don't need homoskedasticity or even linearity
- Many independent discoverers
    - Sample analog of Huber's (1967) asymptotic variance
    - Eicker 1967; Fuller 1975; Hinkley 1977; White 1980ab
- Downward bias and high variance in small samples

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
    - Don't need homoskedasticity or even linearity
- Many independent discoverers
    - Sample analog of Huber's (1967) asymptotic variance
    - Eicker 1967; Fuller 1975; Hinkley 1977; White 1980ab
- Downward bias and high variance in small samples
    - Bias corrections exist, but still have high variance

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
    - Don't need homoskedasticity or even linearity
- Many independent discoverers
    - Sample analog of Huber's (1967) asymptotic variance
    - Eicker 1967; Fuller 1975; Hinkley 1977; White 1980ab
- Downward bias and high variance in small samples
    - Bias corrections exist, but still have high variance
    - "HC2" (MacKinnon & White 1985) or "$J(1)$ jackknife" (Wu 1986)

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
  - Don't need homoskedasticity or even linearity
- Many independent discoverers
  - Sample analog of Huber's (1967) asymptotic variance
  - Eicker 1967; Fuller 1975; Hinkley 1977; White 1980ab
- Downward bias and high variance in small samples
  - Bias corrections exist, but still have high variance
  - "HC2" (MacKinnon & White 1985) or "$J(1)$ jackknife" (Wu 1986) yields the familiar

$$\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_0^2}{n_0}$$

for the difference in means.

The sandwich estimator

Conventional OLS variance estimator: $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$

Sandwich estimator: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\epsilon}_i^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

- Consistent under infinite- or finite-population sampling
    - Don't need homoskedasticity or even linearity
- Many independent discoverers
    - Sample analog of Huber's (1967) asymptotic variance
    - Eicker 1967; Fuller 1975; Hinkley 1977; White 1980ab
- Downward bias and high variance in small samples
    - Bias corrections exist, but still have high variance
    - "HC2" (MacKinnon & White 1985) or "$J(1)$ jackknife" (Wu 1986) yields the familiar

$$\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_0^2}{n_0}$$

for the difference in means. In randomized experiments, this is unbiased or conservatively biased (Neyman 1923).

The sandwich SE is asymptotically conservative in randomized experiments

Theorem 2

Let $\widehat{w}$ denote the sandwich variance estimator for $\widehat{ATE}_{ANCOVA\ II}$.

The sandwich SE is asymptotically conservative in randomized experiments

Theorem 2

Let $\widehat{w}$ denote the sandwich variance estimator for $\widehat{ATE}_{ANCOVA\ II}$. Under Conditions 1–3,

$$N\ \widehat{w} \quad \xrightarrow{p} \quad \frac{1}{p}\lim_{N\to\infty}\sigma^2_{R_1} + \frac{1}{1-p}\lim_{N\to\infty}\sigma^2_{R_0}$$

The sandwich SE is asymptotically conservative in randomized experiments

Theorem 2

Let $\widehat{w}$ denote the sandwich variance estimator for $\widehat{ATE}_{ANCOVA\ II}$. Under Conditions 1–3,

$$\begin{aligned} N\ \widehat{w} &\overset{p}{\to} \frac{1}{p}\lim_{N\to\infty}\sigma^2_{R_1} + \frac{1}{1-p}\lim_{N\to\infty}\sigma^2_{R_0} \\ &= \text{Avar}\left(\sqrt{N}\left[\widehat{ATE}_{ANCOVA\ II} - ATE\right]\right) + \\ &\quad \lim_{N\to\infty}\sigma^2_{R_1-R_0} \end{aligned}$$

The sandwich SE is asymptotically conservative in randomized experiments

Theorem 2

Let $\widehat{w}$ denote the sandwich variance estimator for $\widehat{ATE}_{ANCOVA\ II}$. Under Conditions 1–3,

$$\begin{aligned} N\ \widehat{w} \quad &\xrightarrow{p} \quad \frac{1}{p}\lim_{N\to\infty}\sigma^2_{R_1} + \frac{1}{1-p}\lim_{N\to\infty}\sigma^2_{R_0} \\ &= \quad \text{Avar}\left(\sqrt{N}\left[\widehat{ATE}_{ANCOVA\ II} - ATE\right]\right) \ + \\ &\quad \lim_{N\to\infty}\sigma^2_{R_1-R_0} \end{aligned}$$

Related result

A similar result holds for ANCOVA I.

Comments on SE estimation

These results, together with asymptotic normality, imply asymptotically valid confidence intervals.

Comments on SE estimation

These results, together with asymptotic normality, imply asymptotically valid confidence intervals.

Inference with small N is on shakier ground.
But that's a problem for the difference in means as well.

Comments on SE estimation

These results, together with asymptotic normality, imply asymptotically valid confidence intervals.

Inference with small N is on shakier ground.
But that's a problem for the difference in means as well.

Freedman's critique conflates two independent choices:

- To adjust, or not to adjust?
- To assume homoskedasticity, or not?

Outline

Notes on small-sample bias

Unconditional unbiasedness is overrated

Notes on small-sample bias

Unconditional unbiasedness is overrated

- MSE (or another loss function), not bias
- Adjustment may reduce conditional bias

Notes on small-sample bias

Unconditional unbiasedness is overrated

- MSE (or another loss function), not bias
- Adjustment may reduce conditional bias

The leading term in the bias reflects omitted squared covariates (Cochran 1942, 1977)

Notes on small-sample bias

Unconditional unbiasedness is overrated

- MSE (or another loss function), not bias
- Adjustment may reduce conditional bias

The leading term in the bias reflects omitted squared covariates (Cochran 1942, 1977)

With a saturated model, ANCOVA II is unbiased

- Equivalent to post-stratification (subclassification and weighting)

Notes on small-sample bias

Unconditional unbiasedness is overrated

- MSE (or another loss function), not bias
- Adjustment may reduce conditional bias

The leading term in the bias reflects omitted squared covariates (Cochran 1942, 1977)

With a saturated model, ANCOVA II is unbiased

- Equivalent to post-stratification (subclassification and weighting)
- Miratrix (2011) gives exact variance

Possible directions for further research

Permutation inference with covariate adjustment (Rosenbaum 2002)

Possible directions for further research

Permutation inference with covariate adjustment (Rosenbaum 2002)

- Some see this as a remedy for Freedman's complaints 2011)
- Actually doesn't address any of Freedman's issues
- Assumes constant treatment effect
  and thus homoskedasticity

Possible directions for further research

Permutation inference with covariate adjustment (Rosenbaum 2002)

- Some see this as a remedy for Freedman's complaints 2011)
- Actually doesn't address any of Freedman's issues
- Assumes constant treatment effect
  and thus homoskedasticity
- Janssen (1997): Look at permutation distribution of unequal-variance *t*-statistic

Possible directions for further research

Permutation inference with covariate adjustment (Rosenbaum 2002)

- Some see this as a remedy for Freedman's complaints 2011)
- Actually doesn't address any of Freedman's issues
- Assumes constant treatment effect
  and thus homoskedasticity
- Janssen (1997): Look at permutation distribution of unequal-variance t-statistic

Cluster-randomized experiments

- Multilevel models allow between- and within-cluster slopes to differ

Possible directions for further research

Permutation inference with covariate adjustment (Rosenbaum 2002)

- Some see this as a remedy for Freedman's complaints 2011)
- Actually doesn't address any of Freedman's issues
- Assumes constant treatment effect
  and thus homoskedasticity
- Janssen (1997): Look at permutation distribution of unequal-variance t-statistic

Cluster-randomized experiments

- Multilevel models allow between- and within-cluster slopes to differ
- Gains from adjustment may be especially high (Raudenbush 1997)

Possible directions for further research

Permutation inference with covariate adjustment (Rosenbaum 2002)

- Some see this as a remedy for Freedman's complaints 2011)
- Actually doesn't address any of Freedman's issues
- Assumes constant treatment effect
  and thus homoskedasticity
- Janssen (1997): Look at permutation distribution of unequal-variance t-statistic

Cluster-randomized experiments

- Multilevel models allow between- and within-cluster slopes to differ
- Gains from adjustment may be especially high (Raudenbush 1997)
- But the number of clusters may often be too small to rely on asymptotics

Acknowledgments

I would like to thank:

- Jas Sekhon for very valuable advice and encouragement

Acknowledgments

I would like to thank:

- Jas Sekhon for very valuable advice and encouragement
- Peter Aronow, Gus Baker, Erin Hartman, Danny Hidalgo, Alex Mayer, Luke Miratrix, Cyrus Samii, Dylan Small, and seminar participants at Berkeley for helpful comments

Acknowledgments

I would like to thank:

- Jas Sekhon for very valuable advice and encouragement
- Peter Aronow, Gus Baker, Erin Hartman, Danny Hidalgo, Alex Mayer, Luke Miratrix, Cyrus Samii, Dylan Small, and seminar participants at Berkeley for helpful comments
- David Freedman for his generous advice and help earlier in my education

All shortcomings of this work are solely my responsibility.