BGSE Development Economics Summer School

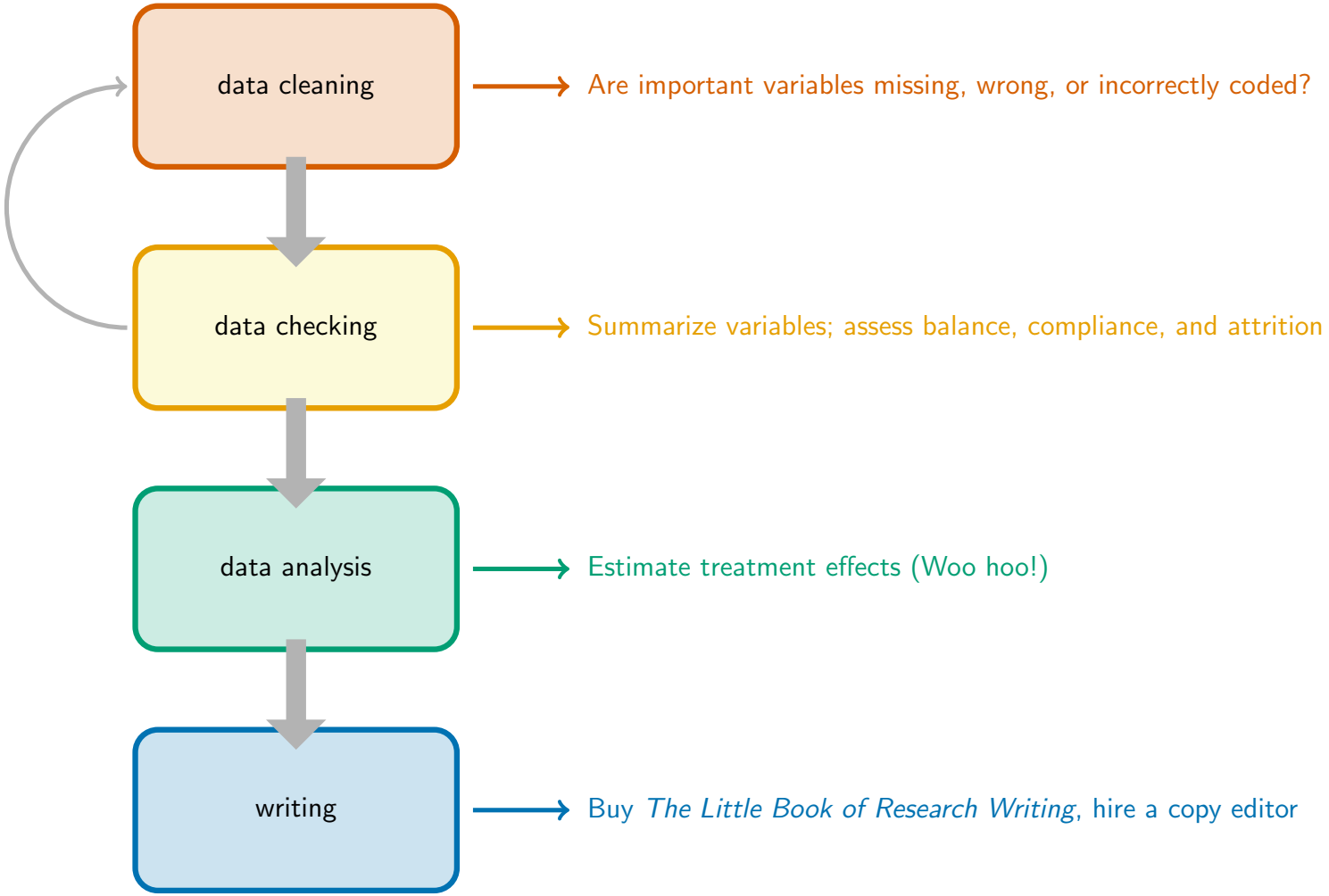
Day 3:

Analyzing Data from RCTs

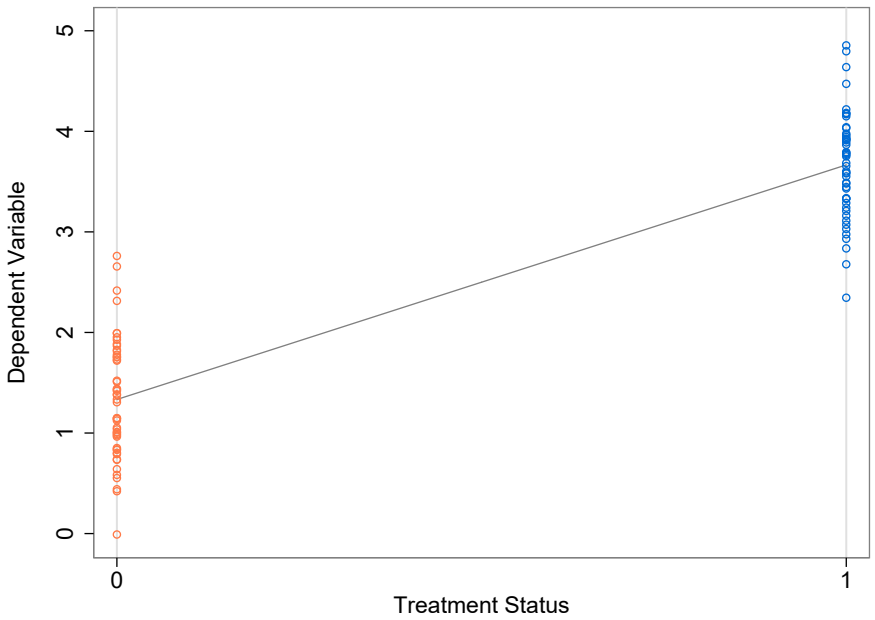
Professors: Pamela Jakiela and Owen Ozier

The Basics

Overview of the Research Process



Regression Analysis of Randomized Experiments



Simple regression framework for analyzing RCTs: $Y_i = \alpha + \beta D_i + \varepsilon_i$

- Treatment indicator $D_i = 0, 1 \Rightarrow$ only two possible values of $\hat{Y}_i$

Basic Regression Equation

More formally, we know that $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

What does this mean in an RCT (or any binary treatment) context?

Basic Regression Equation

Regress the outcome on... the treatment indicator and a constant

$$X_i = [D_i \quad 1]$$

Suppose pN observations have $D_i = 1$ and half have $D_i = 0$

$$\boldsymbol{X} = \begin{bmatrix} D_1 & 1 \\ D_2 & 1 \\ \dots & \dots \\ D_{(1-p)N} & 1 \\ D_{(1-p)N+1} & 1 \\ D_{(1-p)N+2} & 1 \\ \dots & \dots \\ D_N & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ \dots & \dots \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 1 \end{bmatrix} \; ; \; \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_{(1-p)N} \\ Y_{(1-p)N+1} \\ Y_{(1-p)N+2} \\ \dots \\ Y_N \end{bmatrix}$$

Basic Regression Equation

Set up $\boldsymbol{X}'\boldsymbol{X}$:

$$\begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ \dots & \dots \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} pN & pN \\ pN & N \end{bmatrix}$$

Equivalently, we can write:

$$\begin{bmatrix} pN & pN \\ pN & N \end{bmatrix} = pN \begin{bmatrix} 1 & 1 \\ 1 & 1/p \end{bmatrix}$$

Basic Regression Equation

Recall the formula for inverting a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Or, equivalently, we can write:

$$\left(pN \begin{bmatrix} 1 & 1 \\ 1 & 1/p \end{bmatrix} \right)^{-1} = \frac{1}{pN} \begin{bmatrix} 1 & 1 \\ 1 & 1/p \end{bmatrix}^{-1} = \frac{1}{pN} \left(\frac{p}{1-p} \right) \begin{bmatrix} 1/p & -1 \\ -1 & 1 \end{bmatrix}$$

Basic Regression Equation

What about $\boldsymbol{X}'\boldsymbol{y}$?

$$\begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_{(1-p)N} \\ Y_{(1-p)N+1} \\ Y_{(1-p)N+2} \\ \dots \\ Y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=(1-p)N+1}^{N} Y_i \\ \\ \sum_{i=1}^{N} Y_i \end{bmatrix}$$

$$= \begin{bmatrix} \sum_T Y_i \\ \\ \sum_T Y_i + \sum_C Y_i \end{bmatrix} = \begin{bmatrix} pN\bar{Y}_T \\ \\ pN\bar{Y}_T + (1-p)N\bar{Y}_C \end{bmatrix}$$

Basic Regression Equation

We can now compute: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

$$\frac{1}{N(1-p)}\begin{bmatrix} 1/p & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} pN\bar{Y}_T \\ pN\bar{Y}_T + (1-p)N\bar{Y}_C \end{bmatrix}$$

$$= \frac{1}{N(1-p)}\begin{bmatrix} N\bar{Y}_T - pN\bar{Y}_T - (1-p)N\bar{Y}_C \\ -pN\bar{Y}_T + pN\bar{Y}_T + (1-p)N\bar{Y}_C \end{bmatrix}$$

$$= \begin{bmatrix} \bar{Y}_T - \bar{Y}_C \\ \bar{Y}_C \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \hat{\boldsymbol{\beta}}$$

Basic Regression Equation with Controls

More typical regression specification:

$$Y_{1,i} = \alpha + \beta D_i + \delta X_{1,i} + \gamma Y_{0,i} + \tau_{strata} + \varepsilon_i$$

We will typically want to include these controls:

- Dummies for randomization strata (κ_{strata})
- Baseline covariates that are not balanced across treatments*
- Baseline covariates that predict the outcome
  - Baseline values of outcome variables are (sometimes) most important

We do not want to include:

- Controls that could be impacted by treatment

“You Don’t Have to Take My Word For It”

American Economic Journal: Applied Economics

Vol. 11 No. 3 July 2019

Full Issue Download PDF
(AEA members only)

Find articles in this issue

☑ Title ☑ Abstract ☑ Author

All Classifications

Search

"You Don't Have to Take My Word For It"

American Economic Journal: Applied Economics 2019, 11(3): 128–154
https://doi.org/10.1257/app.20170226

Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program[†]

By Prashant Loyalka, Anna Popova, Guirong Li, and Zhaolei Shi[*]

Despite massive investments in teacher professional development (PD) programs in developing countries, there is little evidence on their effectiveness. We present results of a large-scale, randomized evaluation of a national PD program in China in which teachers were randomized to receive PD; PD plus follow-up; PD plus evaluation of the command of PD content; or no PD. Precise estimates indicate PD and associated interventions failed to improve teacher and student outcomes after one year. A detailed analysis of the causal chain shows teachers find PD content to be overly theoretical, and PD delivery too rote and passive, to be useful. (*JEL* I21, I28, J24, J45, O15, P36)

"You Don't Have to Take My Word For It"

We estimate the ATEs using the following ordinary least squares regression model:[22]

$$Y_{ij} = \alpha_0 + \alpha_1 D_j + X_{ij}\alpha + \tau_k + \varepsilon_{ij}, \tag{1}$$

where Y_{ij} is the outcome of interest measured at endline for student i in school j; D_j is one or more dummies indicating the treatment assignment of school j; X_{ij} is a vector of baseline control variables; and τ_k is a set of block fixed effects. In all specifications, X_{ij} includes the baseline value of the dependent variable whenever this is available. We also estimate treatment effects with an expanded set of baseline controls (we call these our "covariate-adjusted" regressions). For student-level outcomes, this expanded set of controls includes student age, student gender, parent educational attainment, a household asset index, class size, teacher gender, teacher age, teacher experience, teacher education level, a teacher certification dummy, a teacher major in math dummy, and teacher rank. For outcomes measured at the teacher level, student controls are omitted.

"You Don't Have to Take My Word For It"

Table 1—Impacts on Student Achievement (at Midline)

		(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Comparing PD as well as PD + Follow-up versus Control (left-out group)							
(1)	PD	−0.015	−0.035				
		(0.028)	(0.027)				
(2)	PD + Follow-up	0.000	−0.020				
		(0.031)	(0.030)				
(3)	Difference: PD + Follow-up − PD	0.015	0.015				
(4)	*p*-value: PD + Follow-up − PD	0.609	0.613				
(5)	Observations	15,987	15,713				
Panel B. Comparing PD + Evaluation versus PD (left-out group)							
(6)	PD + Evaluation			0.008	0.005		
				(0.029)	(0.028)		
(7)	Observations			10,725	10,483		
Panel C. Comparing PD + Evaluation versus Control (left-out group)							
(8)	PD + Evaluation					−0.003	−0.022
						(0.028)	(0.028)
(9)	Observations					10,967	10,774
(10)	Additional controls		X		X		X

Notes: Cluster-robust standard errors are in parentheses. Estimates are adjusted for student and teacher baseline covariates and block fixed effects. PD stands for professional development. According to the standard error estimates, none of the coefficients are statistically significant at even the 10 percent level. Of course, after adjusting *p*-values for multiple hypothesis testing (using the Free Step-Down Resampling Method of Westfall and Young 1993), the estimated coefficients remain statistically insignificant at the 10 percent level.

Variations on a Theme

- Treatment effect heterogeneity
  - Traditional approach: interact baseline covariate with treatment
  - What cool kids are doing: machine learning Davis & Heller (2017)
- Multiple treatments, cross-cutting designs
- Quantile regressions, distribution tests/regressions
  - The mean (in "average treatment effect") may not be of interest
  - Alternative statistics may yield greater power

Compliance with Treatment

How High Is Take-Up?

Even "free" programs are costly for participants, and take-up is often low

Intervention	Take-Up	Source
Business training	65%	McKenzie & Woodruff (2013)
Deworming medication	75%	Kremer & Miguel (2007)
Microfinance	13% – 31%	JPAL & IPA (2015)

Only people who do a program can be impacted by the program*

$\Rightarrow$ We might like to know how much a program impacted participants (it depends on our notion of treatment)

*Some restrictions apply

Imperfect Compliance

True model when outcomes are impacted by program participation (P_i):

$$Y_i = \alpha + \beta P_i + \varepsilon_i$$

- Program take-up is endogenous conditional on treatment
- Only those randomly assigned to treatment ($T_i = 1$) are eligible

We estimate standard regression specification:

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

What do we get?

Imperfect Compliance

Modifying our standard OLS equation, we get:

$$\hat{\beta} = E\left[Y_i | T_i = 1\right] - E\left[Y_i | T_i = 0\right]$$

$$= \alpha + \beta E\left[P_i | T_i = 1\right] + \varepsilon_i - \left(\alpha + \beta E\left[P_i | T_i = 0\right] + \varepsilon_i\right)$$

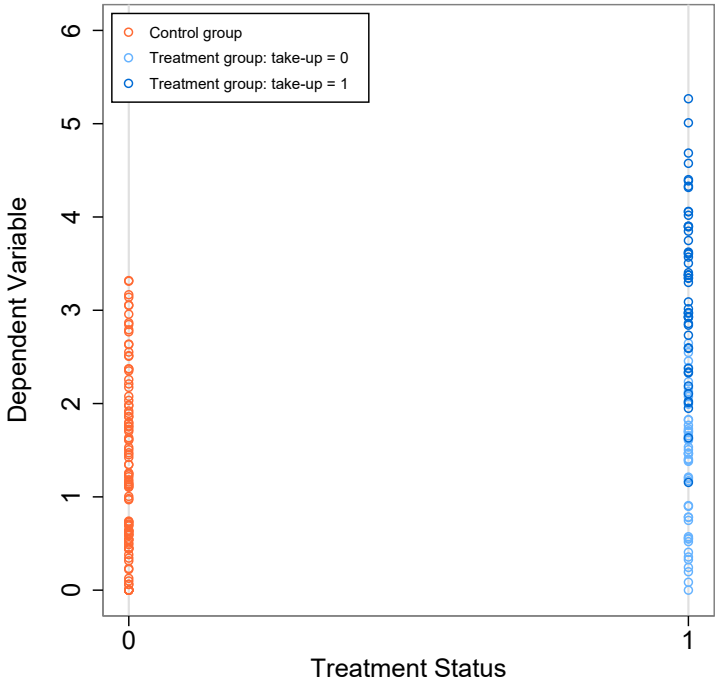
$$= \beta E\left[P_i | T_i = 1\right]$$

$$= \beta\lambda$$

where $\lambda < 1$ is the take-up rate in the treatment group

$\Rightarrow$ Low compliance scales down the estimated treatment effect

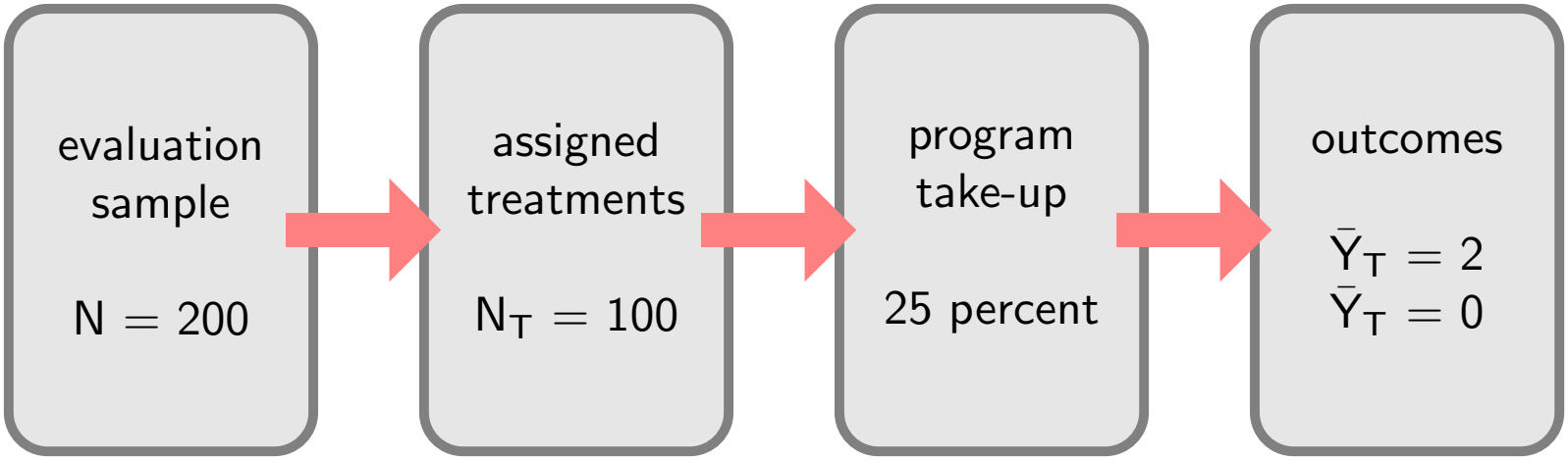
Treatment on the Treated



Your colleague suggests comparing the **compliers** to the control group

$\Rightarrow$ Is this a good idea?

Treatment on the Treated: A Thought Experiment



Questions:

- What was the average outcome among those who did the program?
- What does this suggest about the impact of treatment?

Treatment on the Treated: Intuition

The **treatment on the treated (TOT)** estimator:

$$\hat{\beta}_{tot} = \frac{E\left[Y_i | T_i = 1\right] - E\left[Y_i | T_i = 0\right]}{E\left[P_i | T_i = 1\right] - E\left[P_i | T_i = 0\right]}$$

Intuitively, the TOT scales up the ITT effect to reflect imperfect take-up

- Assumption: treatment only works through program take-up
  - Not always obvious whether this is true

Treatment on the Treated: Implementation

Estimated via two-stage least squares (2SLS):

$$Y_i = \alpha_1 + \beta_1 \hat{P}_i + \varepsilon_i \quad \text{[IV regression]}$$

$$P_i = \alpha_2 + \beta_2 T_i + \nu_i \quad \text{[first stage]}$$

Easy to implement using Stata's `ivregress 2sls` command

What Does Treatment on the Treated Measure?

$T = 0$	$T = 1$
always takers	always takers
compliers	compliers
never takers	never takers

TOT estimates local average treatment effect (LATE) on **compliers**

- Monotonicity assumption: there are no **defiers**
- When violated, TOT tells us about weighted difference between treatment effects on compliers and defiers... but it gets complicated

Alternative Experimental Designs

Q: When Is an RCT Not Just an RCT?

A: When SUTVA is violated! (Just kidding. Sort of.)

Exogenous variation from RCT can feed into creative research designs

- Deworming medication in Kenya (Miguel & Kremer 2004)
  - Cluster-randomized design to account for treatment spillovers
    - Exploit (random) variation in treatment status of children in neighboring schools to estimate spillovers from mass deworming
    - Use randomized phased-in design to identify (eventual) never-takers in comparison schools, compare to never-takers in treatment schools
  - Ozier (2018) uses same school-based deworming experiment to measure spillovers on younger siblings (under age 2 when "treated")
- Re-randomization to separate selection, treatment effects (Karlan & Zinman 2009; Leaver, Ozier, Serneels, & Zeitlin 2009)

Multiple Hypothesis Testing

Multiple Hypothesis Testing: The Problem

Consider testing 100 true null hypotheses — how many will rejected?

- What sort of ninny would test 100 hypotheses?
- Valid reasons for testing many hypotheses:
  - Studies often have 2 or 3 treatment arms (and rightly so!)
  - Difficult to predict which outcomes will be affected
    - Particularly true for secondary hypotheses/treatment effects
  - Different measures of the same outcome often available
  - Heterogeneity in treatment effects (across sub-samples)

How can we (credibly) test multiple hypotheses?

Bonferroni Corrections

Most conservative approach is the **Bonferroni method***

- Problem: you wish to test hypotheses $H_1, ...H_k$ using a test size of α
- Solution (of sorts): use a test size of α/k instead
  - **Family-wise error rate (FWER)**: probability of rejecting a false null
  - Bonferroni correction holds FWER below α
  - Bonferroni corrections are too conservative:
    - FWER ≈ 0.04877 when number of independent tests is large
    - Bonferroni corrections can be extremely conservative when tests are not independent (consider example of perfectly correlated tests)

Good news: if you are testing k hypotheses and a Bonferroni correction works (i.e. your results hold up), you don't need the rest of this lecture

*Purportedly developed by Olive Jean Dunn and not, ahem, Carlo Emilio Bonferroni

Stepdown Methods

Holm (1979) proposes a less conservative **stepdown method**:

0. Order k p-values from smallest to largest, $p_{(1)}, p_{(2)}, ..p_{(k)}$

1a. If $p_{(1)} > \alpha/k$, stop. Fail to reject all hypotheses

1b. Reject $H_{(1)}$ if $p_{(1)} < \alpha/k$. Proceed to Step 2.

2a. If $p_{(2)} > \alpha/(k-1)$, stop. Fail to reject all remaining hypotheses.

2b. Reject $H_{(2)}$ if $p_{(2)} < \alpha/(k-1)$. Proceed to Step 3.

...

j. Repeat as needed until you stop rejecting hypotheses because $p_{(j)} > \alpha/(k-(j-1))$ or all k hypotheses have been rejected

More good news: Romano & Wolf (JASA, 2005) state "This procedures holds under arbitrary dependence on the joint distribution of p-values."

Stepdown Methods

More complicated/powerful bootstrap-based stepdown methods exist

- Examples: Westfall & Young (1993), Romano & Wolf (2005)
- These procedures exploit additional assumptions to increase power (so you don't need them if simpler methods "work" in your setting)
- They are also more computationally-intensive, often including phrases like "efficient computation" or "computationally feasible"
- Approaches generally use some form of stepdown structure
  - At each step, "accept"/reject decisions use empirical distribution of bootstrapped p-values associated with not-yet-rejected hypotheses

Controlling the False Discovery Rate

Anderson (JASA, 2008): "[Family-wise error rate] adjustments become increasingly severe as the number of tests grows — it is inherent in controlling the probability of making a single false rejection."

- Alternative is to tolerate some small number of false positives

The **false discovery rate**: expected proportion of rejections that are Type I errors (i.e. where null was true and should not have been rejected)

- FWER and FDR are identical under the null (all rejections are errors)
- When some null hypotheses are false, FDR adjustments can be less stringent than FWER adjustments (because FDR $<$ FWER)

Controlling the False Discovery Rate

Benjamini & Hochberg (1995) propose an approach to FDR control:

1. Order k p-values from smallest to largest, $p_1, p_2, ..., p_j, ..., p_k$, where j indicates the rank of the p-value for a specific hypothesis
2. Rejecting all p-values with $p_j < qj/k$ yields an expected FDR no higher than q when p-values are independent or positively correlated

All of the procedures discussed so far modify test sizes ("accept"/reject)

- We often want an adjusted p-value, not a yes/no decision

Anderson (2008) proposed intuitive approach to calculating BH q-values:

- Rescale p-values by number of hypotheses / p-value rank
- Adjust for non-monotonicity

Multiple Test Corrections: Example

p-value	Bonferroni	Holm (0.05)	Anderson
0.001	0.005	0.005	0.005
0.002	0.010	0.008	0.005
0.040	0.200	0.120	0.05125
0.041	0.205	–	0.05125
0.099	0.495	–	0.099

Multiple Hypothesis Testing: Summary

Try to avoid testing a large number of hypotheses

- Aggregate your main outcomes into indices (when appropriate)
- Consider pre-specifying "surprising" relationships

Try a simple approach to p-value adjustment (when appropriate)

- Rescale p-values à la Bonferroni or Anderson (2008)
- User-written Stata program `rwolf` (when appropriate)

Be suspicious of (your own and others') p-values near significance cutoffs

Attrition

Attrition as Selection Bias

Angrist and Pishke (2008):

> *"The goal of most empirical economic research is to overcome selection bias, and therefore to say something about the causal effect..."*

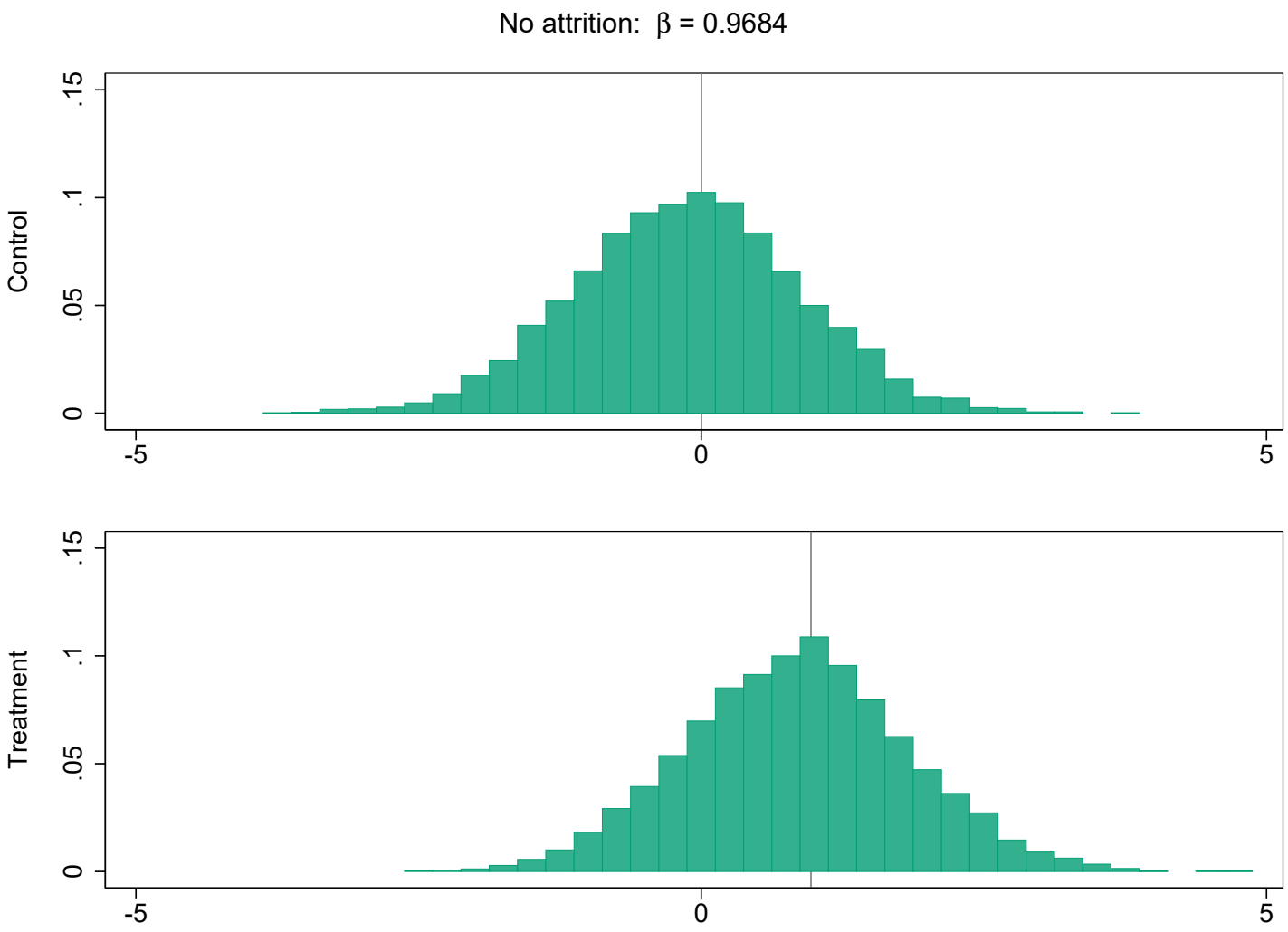
Motivation 1:

- What do we do when an RCT should identify the effect of interest, but there is attrition from the sample (i.e. missing endline data)?
- What if that attrition is differential across arms?

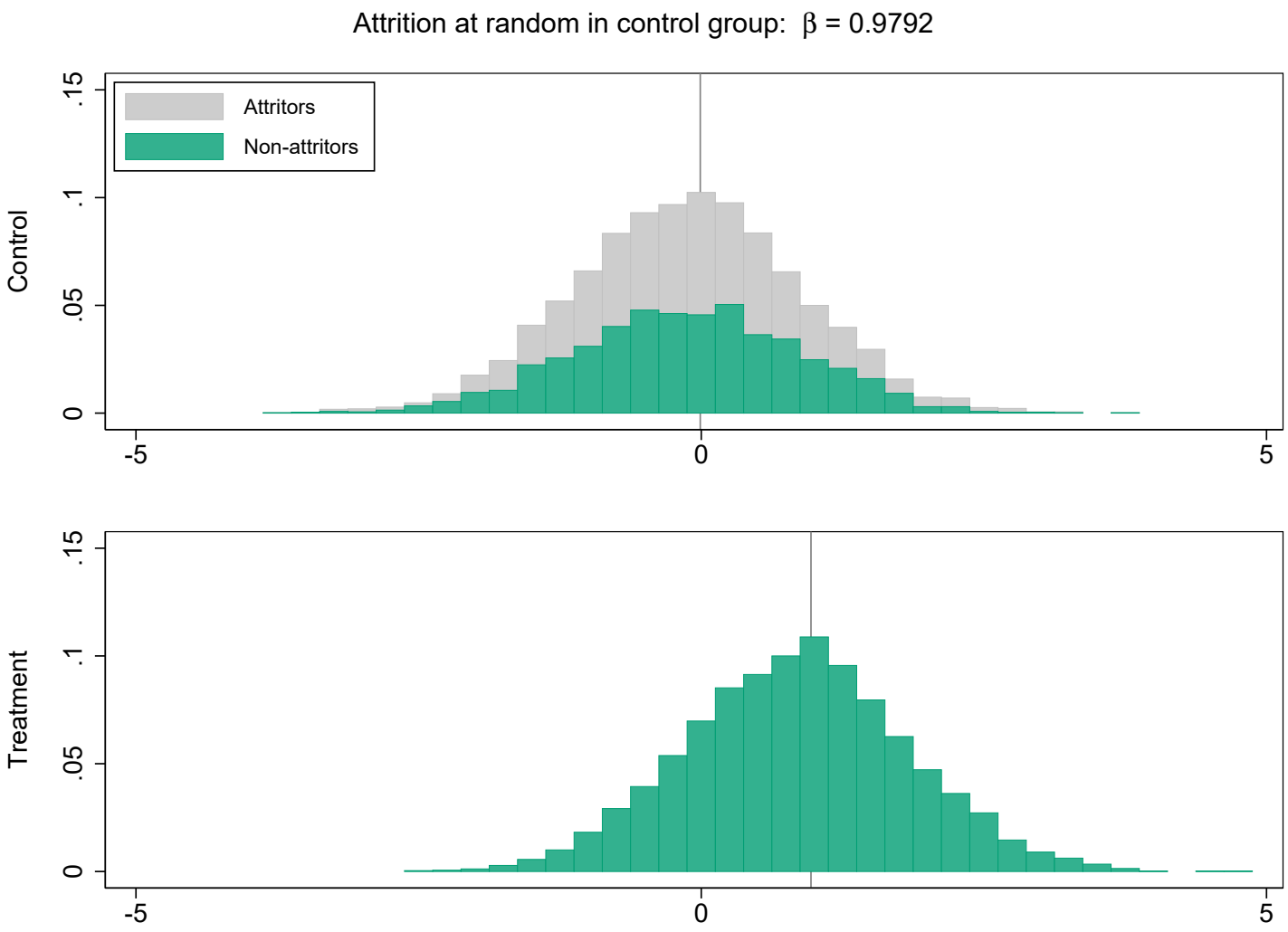
Motivation 2:

- What can we do when outcomes (e.g. profits) are not always observed and are more likely to be observed in treatment group?

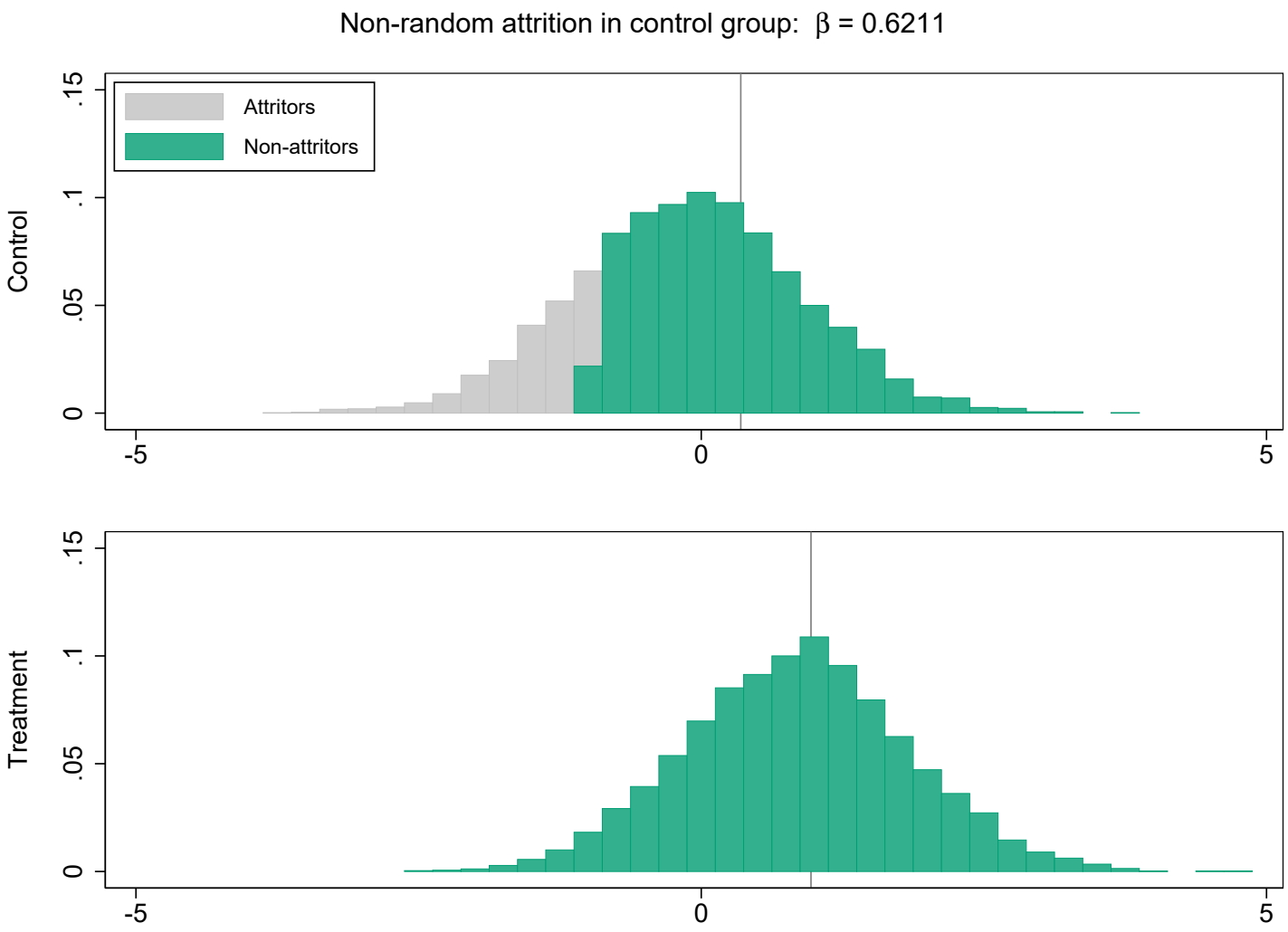
Attrition as Selection Bias: An Example



Random Attrition Is OK



Non-Random Attrition Is a Problem



Non-Random Attrition Is a Problem

We want to know if business training increases micro-enterprise profits

- We only observe profits (Y) for business that still exist ($Z \geq 0$)

The true model of profits is given by:

$$
\begin{aligned}
Y^* &= \beta D + \delta_1 + U \\
Z^* &= \gamma D + \delta_2 + V \\
Y &= \mathbb{1}[Z^* \geq 0]
\end{aligned}
$$

Standard approach to estimating treatment effects yields:

$$
\begin{aligned}
\hat{\beta}_{ITT} &= E[Y|D=1] - E[Y|D=0] \\
&= \beta + \underbrace{E[U|D=1, V \geq -\delta_2 - \gamma] - E[U|D=0, V \geq -\delta_2]}_{\text{selection bias if U and V are not independent}}
\end{aligned}
$$

Approaches to Selection Bias from Attrition

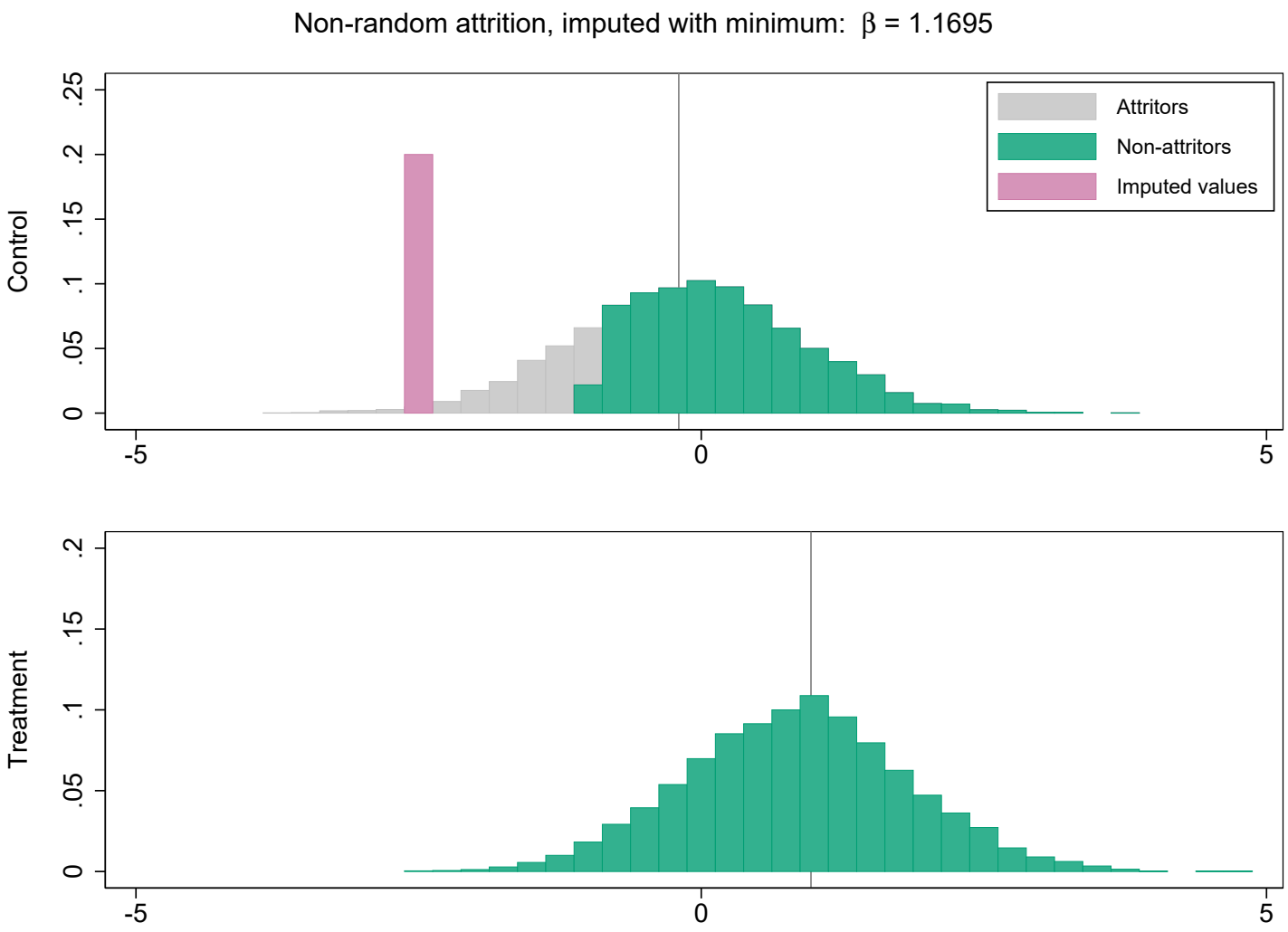
Approach 1: implement Heckman two-step correction for selection

- Drawback: requires an instrument for selection into sample

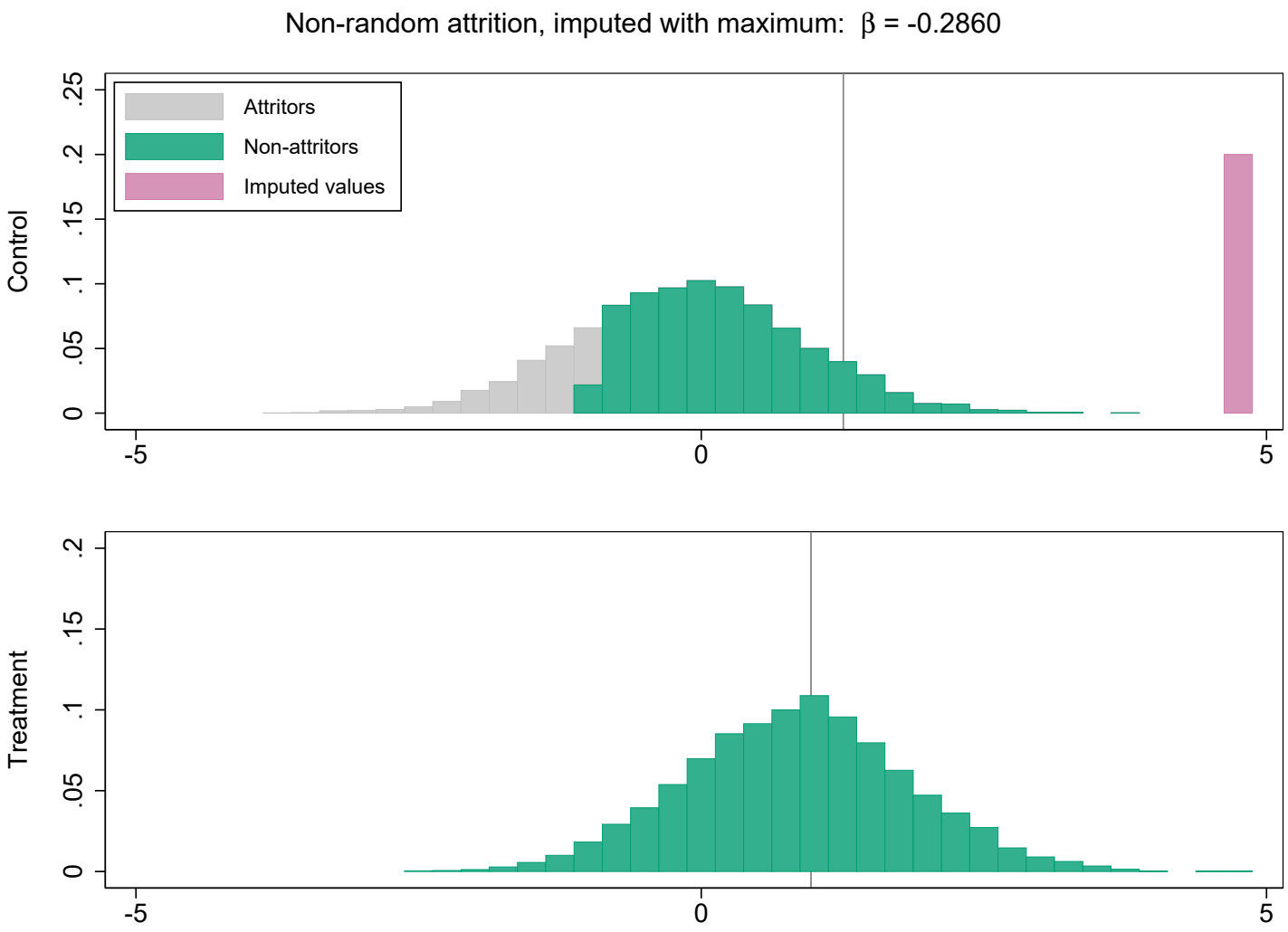
Approach 2: implement Manski bounds (Horowitz and Manski 2000)

- Makes no assumptions besides bounded support for the outcome
  - What is the worst-case scenario for missing observations?
- Replaces missing values with maximum or minimum in the support
- Drawback: results may be uninformative (i.e. CIs may be wide)
  - Manksi bounds still serve as a useful benchmark
  - May work well with certain (e.g. binary) outcomes

Manski Upper Bound: Attrition from Control Group



Manski Lower Bound: Attrition from Control Group



Bounds Under Monotonicity

Approach 3: Lee (2009) derives bounds under monotonicity assumption

> *"treatment... can only affect sample selection in 'one direction' "*

Monotonicity allows us to ignore those who attrit from both arms

- Bounded support not required (not imputing missing values)
- Throw away highest/lowest values from less-attritted study arm
- Identifies the average treatment effect for never-attriters

Bounds Under Monotonicity

Each individual characterized by $(Y_1^*, Y_0^*, S_1^*, S_0^*)$:

- Y_1^*, Y_0^* are potential outcomes
- S_1^*, S_0^* are potential outcomes for attrition
  - Observed in sample when $S = S_1^* D + S_0^*(1 - D) = 1$
  - Never-attritors: $S_1^* = S_0^* = 1$
  - Marginal types: $S_1^* = 1$ and $S_0^* = 0$
  - This assumes treatment reduces attrition, but it can go either way (but not both ways as the same time under monotonicity)

Bounds Under Monotonicity

Recall our simple example:

$$\begin{aligned} E[Y|D=0] &= E[Y^*|D=0, Z^* \geq 0] \\ &= \delta_1 + E[U|D=0, V \geq -\delta_2] \end{aligned}$$

$$\begin{aligned} E[Y|D=1] &= E[Y^*|D=1, Z^* \geq 0] \\ &= \delta_1 + \beta + E[U|D=1, V \geq -\delta_2 - \gamma] \end{aligned}$$

We need to know $E[U|D=1, V \geq -\delta_2]$ to identify treatment effect β

- Notice that those with $V \geq -\delta_2$ are never-attritors
- Those with $-\delta_2 - \gamma \leq V < -\delta_2$ only attrit from control group

Bounds Under Monotonicity

$E[Y|D=1, Z^* \geq 0]$ is a weighted average:

$$= (1-p)\underbrace{E[Y^*|D=1, V \geq -\delta_2]}_{\text{outcome among never-attrittors}} + p\underbrace{E[Y^*|D=1, -\delta_2 - \gamma \leq V < -\delta_2]}_{\text{outcome among marginal types}}$$

where $p = Pr[-\delta_2 - \gamma \leq V < -\delta_2]/Pr[V \geq -\delta_2 - \gamma]$

Throwing out p observations allows us to bound treatment effect:

> *"We cannot identify which observations are inframarginal and which are marginal. But the 'worst-case' scenario is that the smallest p values of Y belong to the marginal group.*

Lee Bounds in Theory

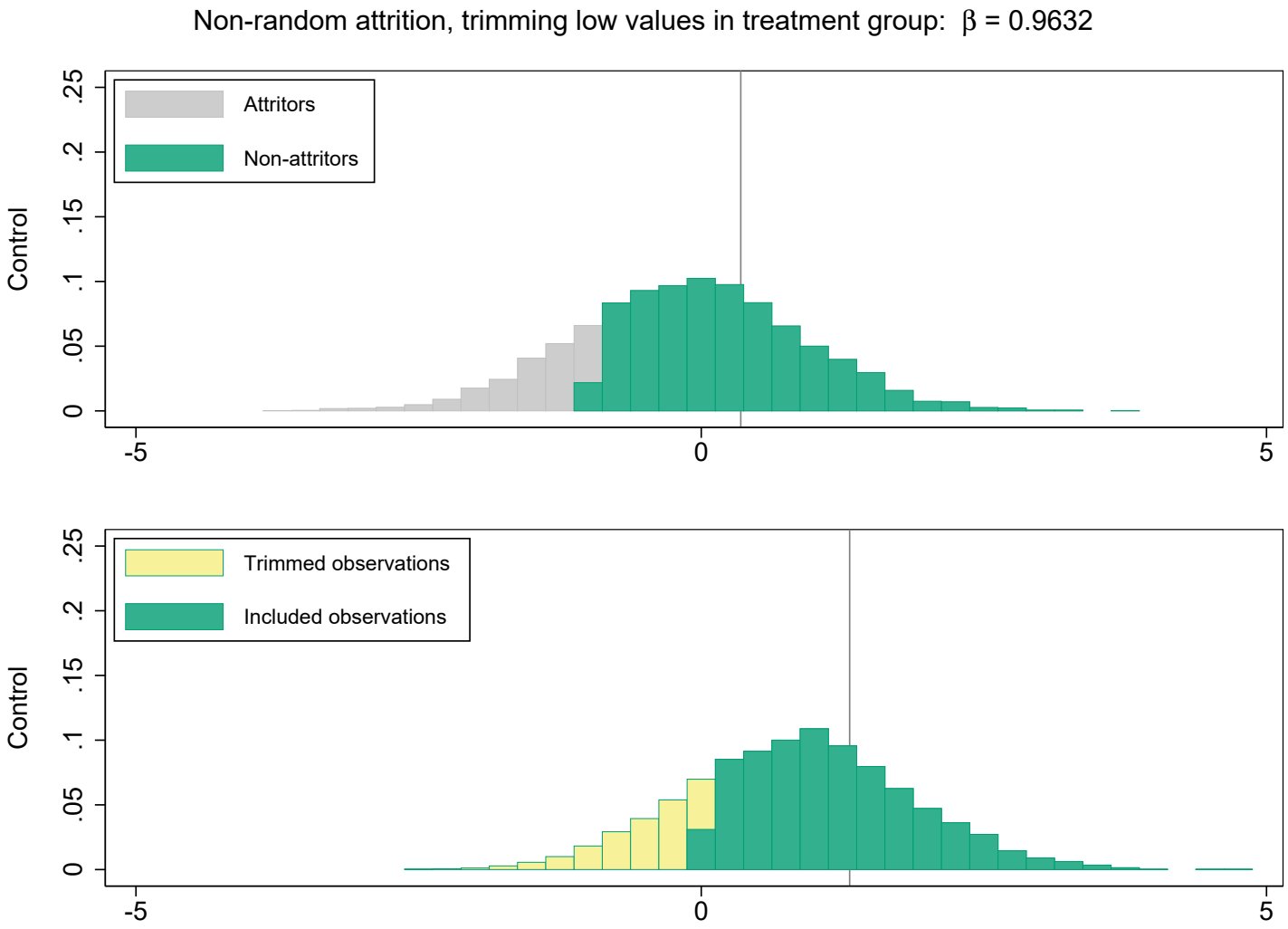
$$LB = E[Y|D = 1, S = 1, Y \leq y_{1-p_0}] - E[Y|D = 0, S = 1]$$

$$UP = E[Y|D = 1, S = 1, Y \geq y_{p_0}] - E[Y|D = 0, S = 1]$$

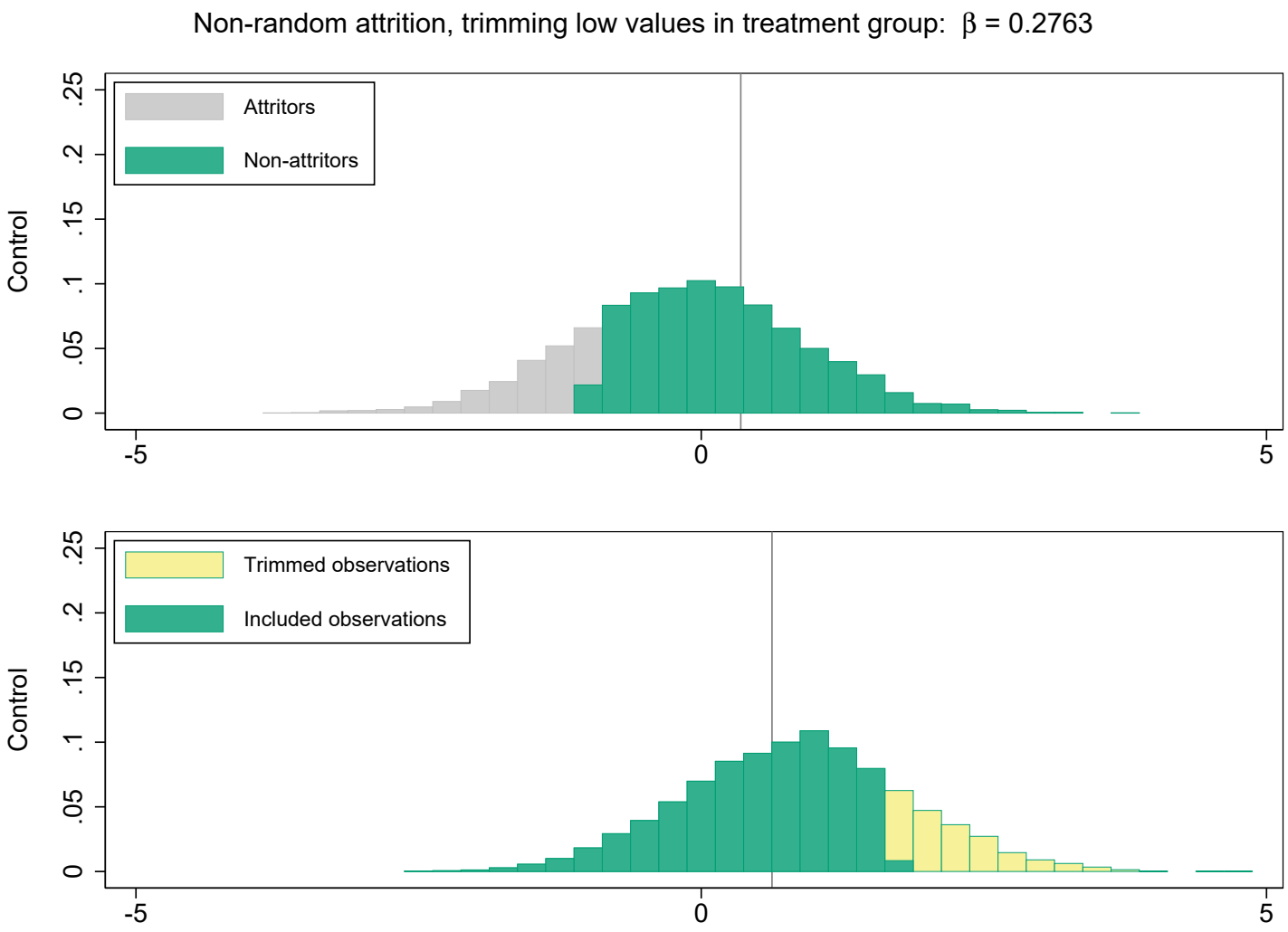
$y_q = G^{-1}(q)$ where G is the CDF of Y conditional on $D = 1, S = 1$

$$p_o = \frac{Pr[S = 1|D = 1] - Pr[S = 1|D = 0]}{Pr[S = 1|D = 1]}$$

Lee (Upper) Bounds in Practice



Lee (Lower) Bounds in Practice



Lee Bounds in Practice

Table IV: Bounds on Treatment Effects for ln(wage) in Week 208 using Trimming Procedure

Control	(i)	Number of Observations	3599	Control Standard Error	
	(ii)	Proportion Non-missing	0.566	Std. Error	0.0082
	(iii)	Mean ln(wage) for employed	1.997		
				Treatment UB Standard Error	
Treatment	(iv)	Number of Observations	5546	Component 1	0.0053
	(v)	Proportion Non-missing	0.607	Component 2	0.0021
	(vi)	Mean ln(wage) for employed	2.031	Component 3	0.0083
				Total	0.0100
		$p = [(v)-(ii)]/(v)$	0.068		
	(vii)	pth quantile	1.636	Treatment LB Standard Error	
	(viii)	Trimmed Mean: $E[Y\|Y>y_p]$	2.090	Component 1	0.0058
				Component 2	0.0037
	(ix)	(1-p)th quantile	2.768	Component 3	0.0144
	(x)	Trimmed Mean: $E[Y\|Y<y_{1-p}]$	1.978	Total	0.0159

Lee Bounds in Practice: Confidence Intervals

For the entire interval, you can do better than:

$$\left[\widehat{\Delta^{LB}} - 1.96\frac{\widehat{\sigma_{LB}}}{\sqrt{n}}, \widehat{\Delta^{UB}} + 1.96\frac{\widehat{\sigma_{UB}}}{\sqrt{n}}\right]$$

Instead (Imbens and Manski 2004), use:

$$\left[\widehat{\Delta^{LB}} - \bar{C}_n\frac{\widehat{\sigma_{LB}}}{\sqrt{n}}, \widehat{\Delta^{UB}} + \bar{C}_n\frac{\widehat{\sigma_{UB}}}{\sqrt{n}}\right]$$

where $\bar{C}_n$ satisfies:

$$\Phi\left(\bar{C}_n + \sqrt{n}\frac{\widehat{\Delta^{UB}} - \widehat{\Delta^{LB}}}{\max(\widehat{\sigma_{LB}}, \widehat{\sigma_{UB}})}\right) - \Phi\left(-\bar{C}_n\right) = 0.95$$

Lee Bounds in Practice: Covariates

Estimating Lee bounds within bins narrows bounds

- The **tightened** bounds are averages over $X = x$ bins
- ITT effects are also weighted across bins
- If attrition is concentrated in specific cells, we can limit bounding exercise to the component of average where attrition actually occurs

Lee Bounds in Practice: `leebounds` in Stata

Title

leebounds — Lee (2009) treatment-effect bounds

Syntax

leebounds *depvar treatvar* [*if*] [*in*] [*weight*] [, *options*]

depvar specifies the outcome variable.

treatvar specifies a binary variable, indicating receipt of treatment. Estimating the effect of *treatvar* on *depvar* is subject of the empirical analysis. The (alphanumerically) larger value of *treatvar* is assumed to indicate treatment.

options	Description
select(*varname*)	selection indicator
tight(*varlist*)	covariates for tightened bounds
cieffect	compute confidence interval for treatment effect
vce(**analytic**|**bootstrap**)	compute analytic or bootstrapped standard errors; default is **vce(analytic)**
level(#)	set confidence level; default is **level(95)**

pweights, **fweights**, and **iweights** are allowed; see weight. Observations with negative weight are skipped for any weight type.
bootstrap is allowed; see prefix.

Description

leebounds computes treatment-effect bounds for samples with nonrandom sample selection or attrition, as proposed by Lee (2009). The lower and upper bound correspond to extreme assumptions about the missing information that are consistent with the observed data. As opposed to parametric approaches to correcting for sample-selection bias, such as the classical Heckman (1979) estimator, Lee (2009) bounds rest on very few assumptions, that is, random assignment of treatment and monotonicity. Monotonicity means that the treatment status affects selection in just one direction. That is, receiving a treatment makes selection either more or less likely for any observation. In technical terms, the approach rests on a trimming procedure. Either from below or from above, the group (treatment, control) that suffers less from sample attrition is trimmed at the quantile of the outcome variable that corresponds to the share of excess observations in this group. Calculating group differentials in mean outcome yields the lower and the upper bound, respectively, for the treatment effect depending on whether trimming is from below or above.