ECON 626: Applied Microeconomics

Lecture 10:

Attrition

Professors: Pamela Jakiela and Owen Ozier

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..."*

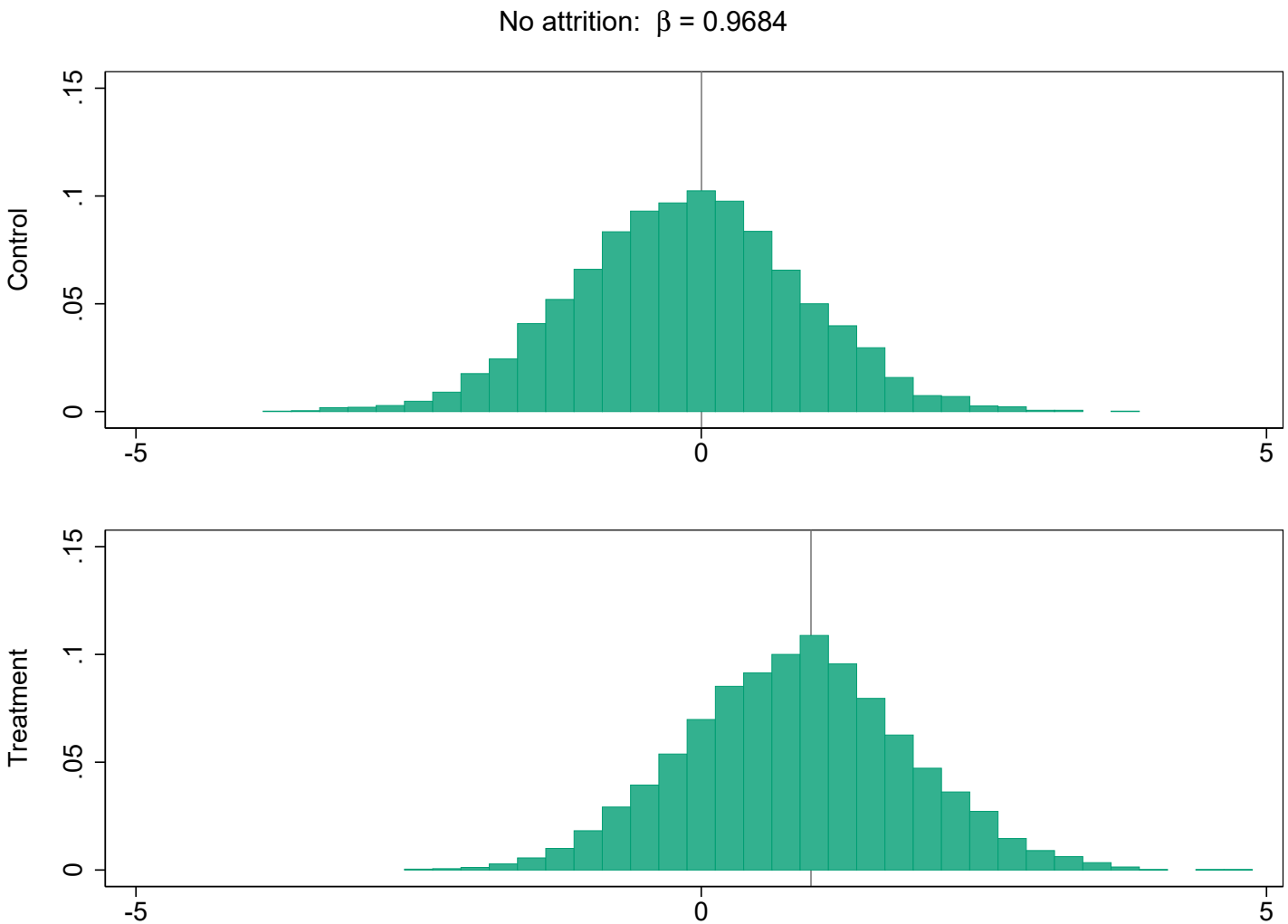
Motivation 1:

- 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?

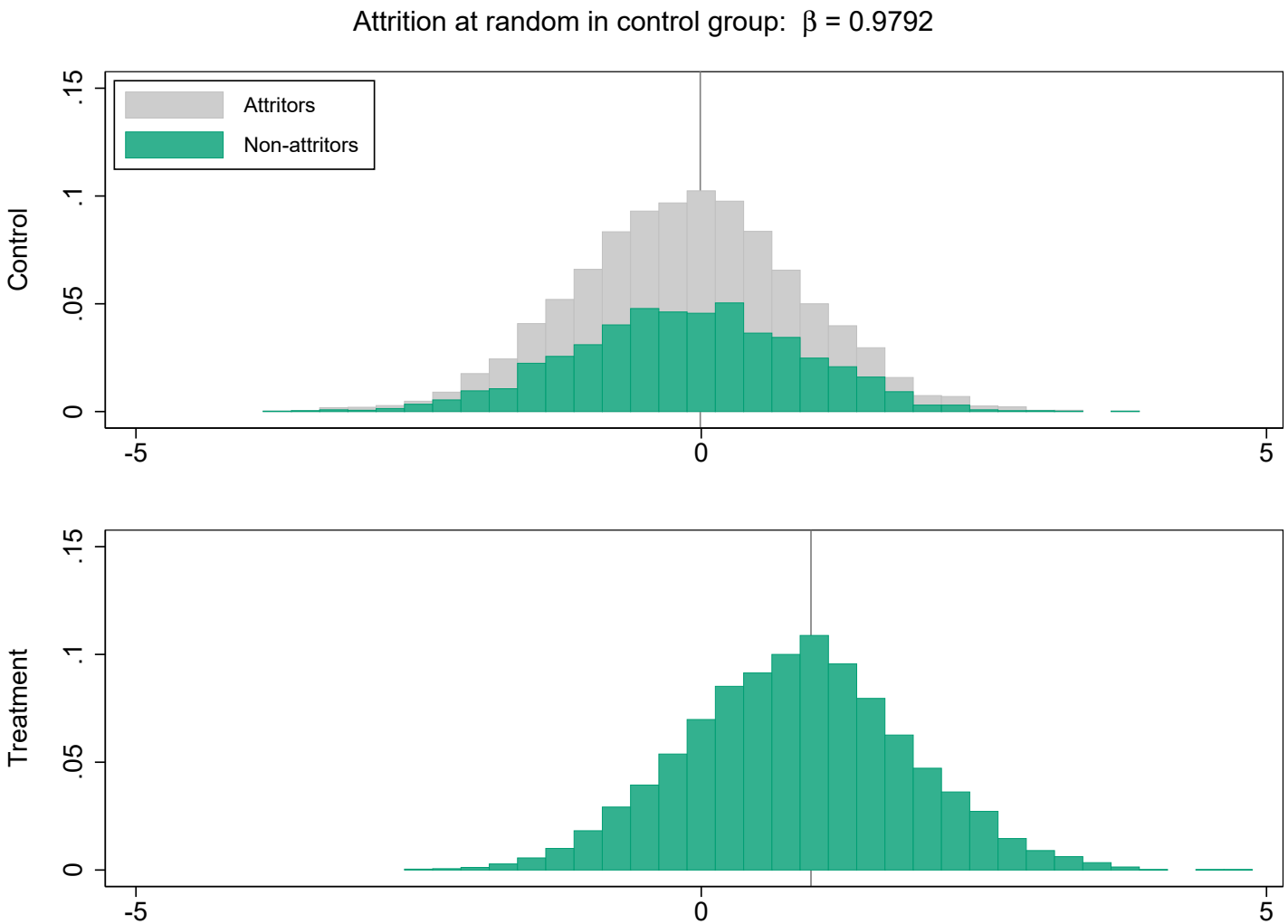
Motivation 2:

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

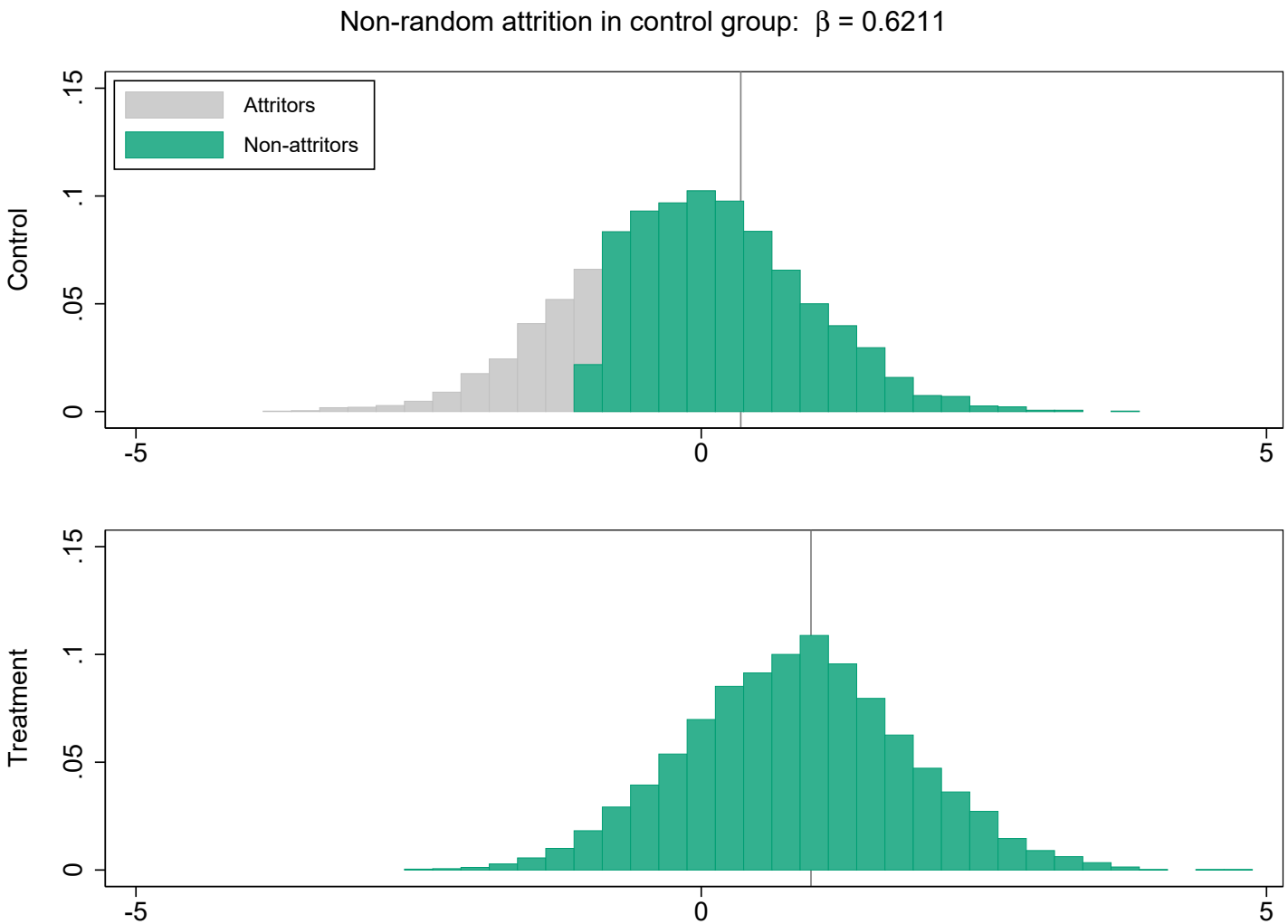
Attrition as Selection Bias: An Example



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

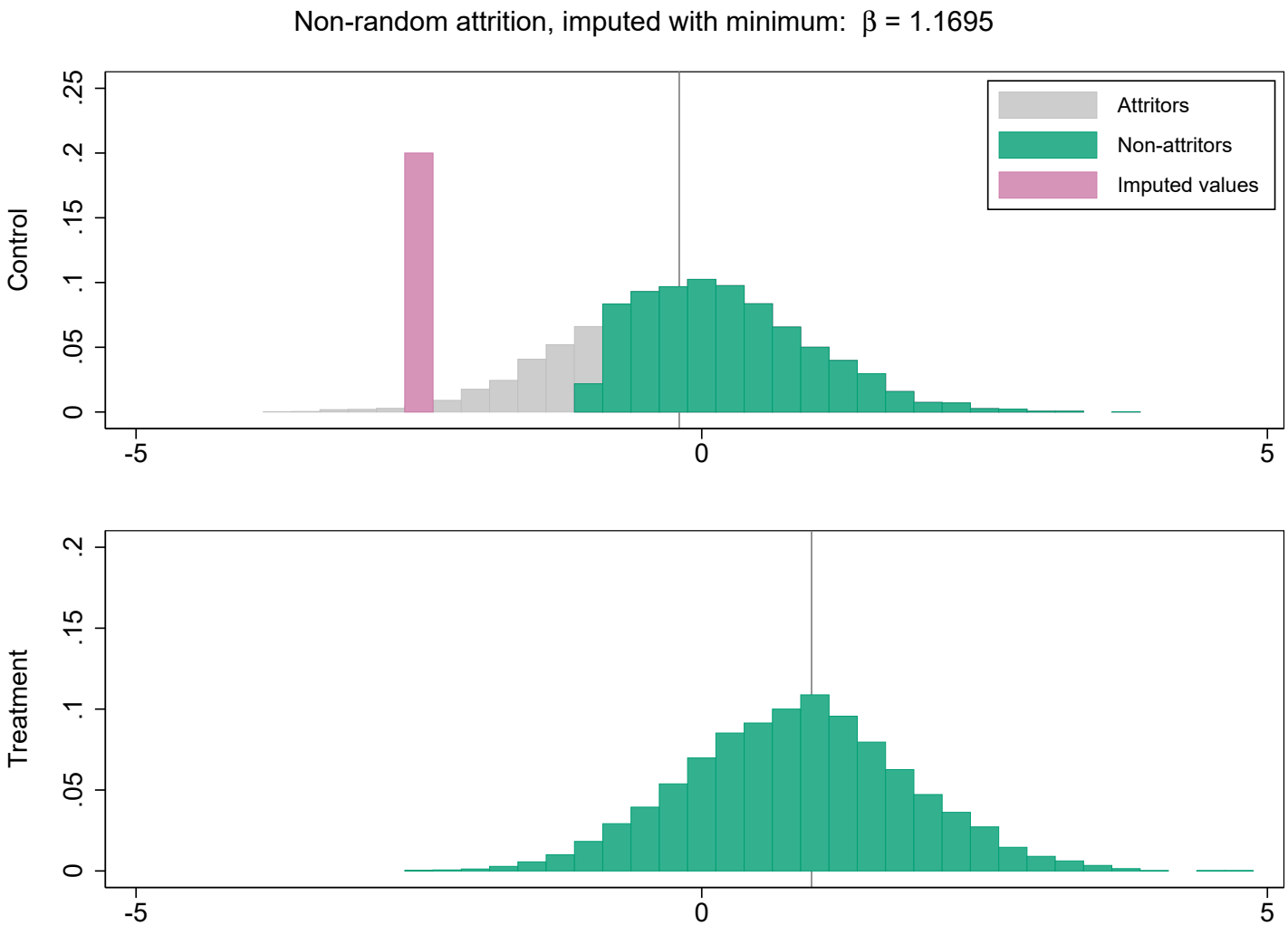
Approach 1: implement Heckman two-step correction for selection

- Drawback: requires an instrument for selection into sample

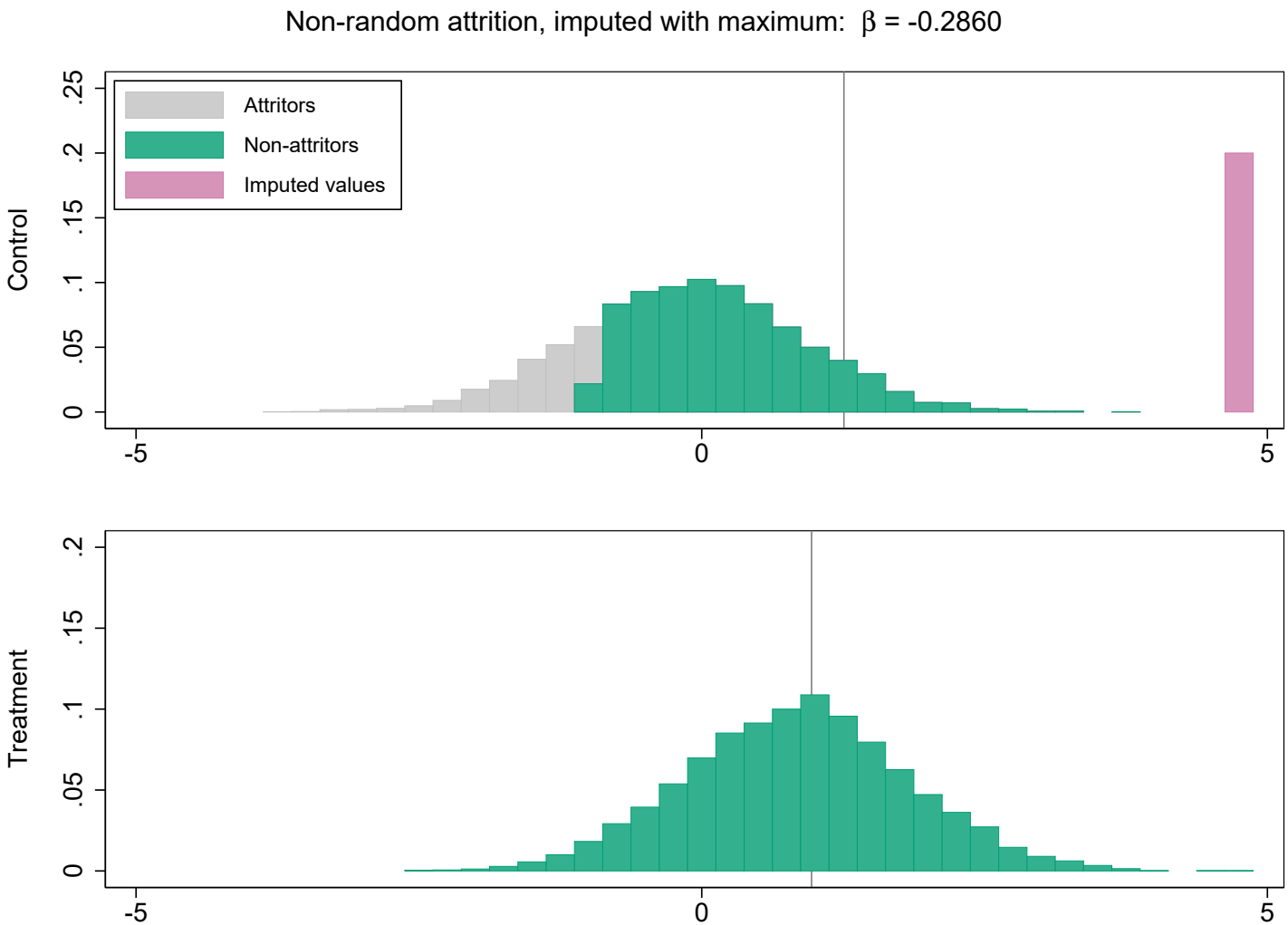
Approach 2: implement Manski bounds (Horowitz and Manski 2000)

- 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

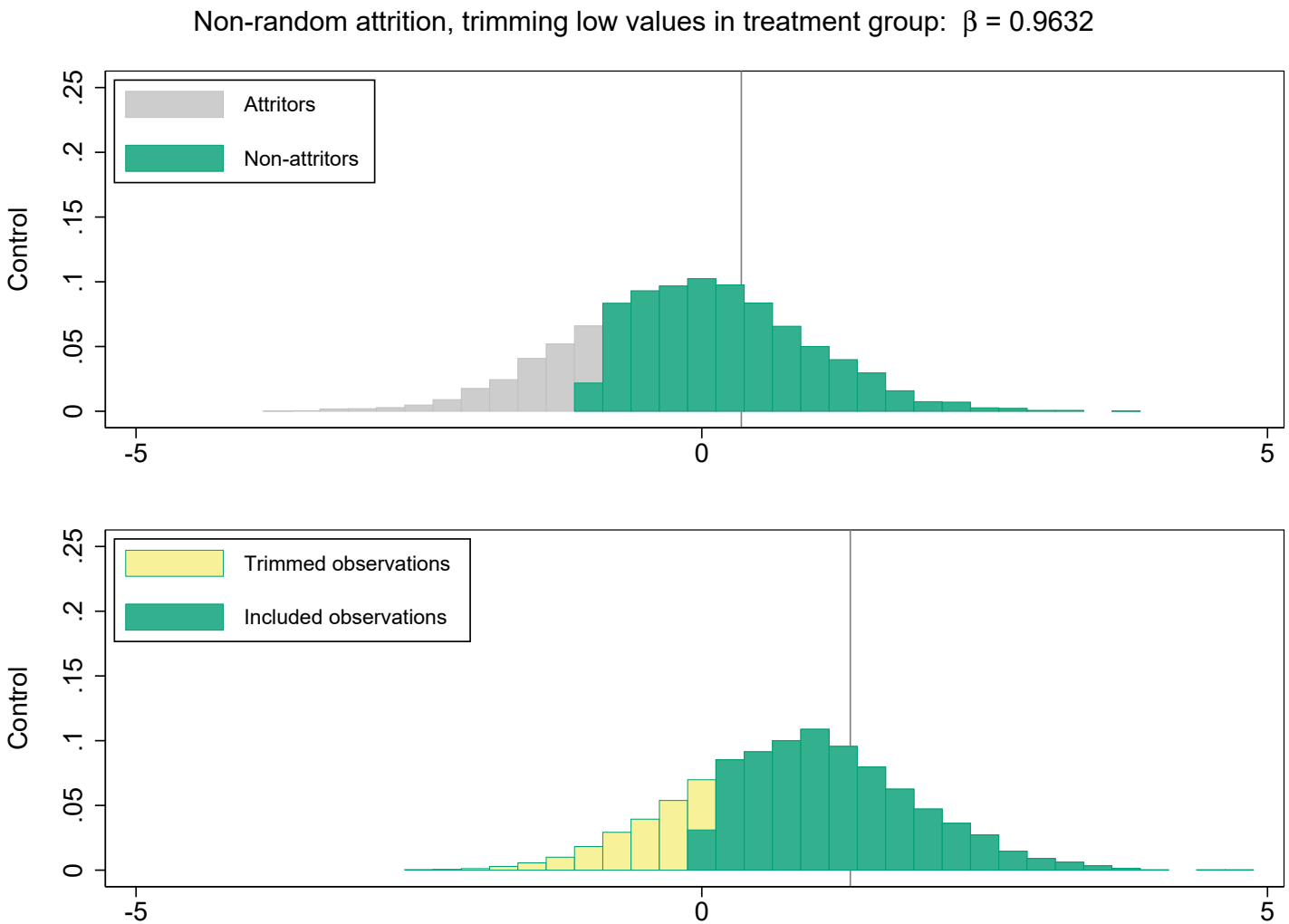
$$LB = E[Y|D = 1, S = 1, Y \leq y_{1-p_0}] - E[Y|D = 0, S = 1]$$

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

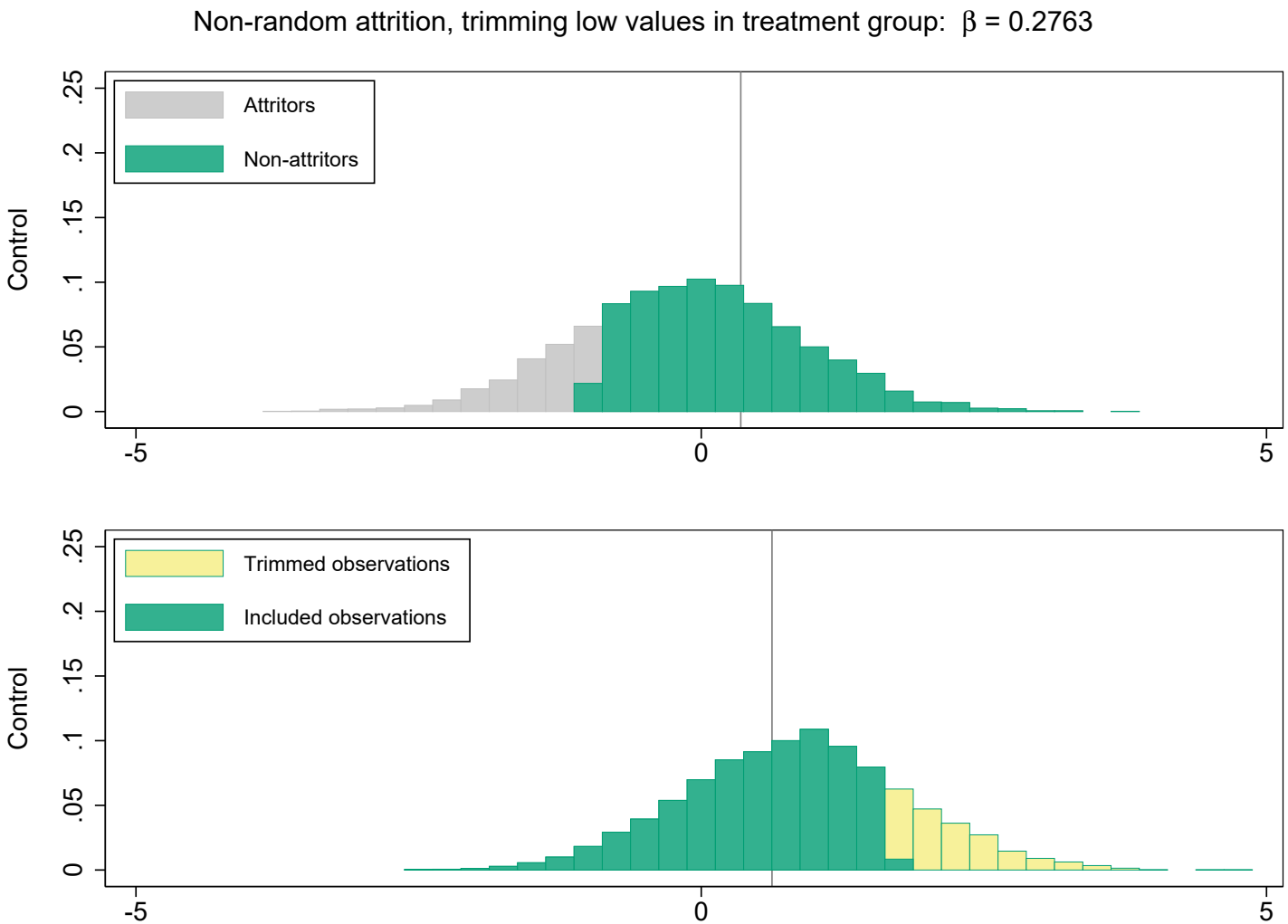
$$y_q = G^{-1}(q) \text{ where } G \text{ is the CDF of } Y \text{ conditional on } D = 1, S = 1$$

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

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.