# Better Lee Bounds

Vira Semenova

July, 29, 2021

# Introduction

Endogenous selection $\Rightarrow$ need bounds on ATE.

## Introduction

Endogenous selection $\Rightarrow$ need bounds on ATE.

Lee (2009): covariate set $\mathcal{X}$ map to a class of bounds on ATE

- any covariate subset $\mathcal{X}' \subseteq \mathcal{X}$ maps to a pair of valid bounds
- the full set $\mathcal{X}$ maps to sharp (the tightest possible) bounds

## Introduction

Endogenous selection $\Rightarrow$ need bounds on ATE.

Lee (2009): covariate set $\mathcal{X}$ map to a class of bounds on ATE

- any covariate subset $\mathcal{X}' \subseteq \mathcal{X}$ maps to a pair of valid bounds
- the full set $\mathcal{X}$ maps to sharp (the tightest possible) bounds

Sharp bounds are difficult to estimate, non-sharp bounds may not be very useful.

## Related literature

1. **Endogenous selection/bounds**: Heckman (1976), Heckman (1979), Manski (1989), Manski (1990), Frangakis and Rubin (2002), Angrist et al. (2002), Angrist et al. (2006),Lee (2009), Engberg et al, (2014), Feller et al. (2016), Angrist et al. (2013), Honore and Hu (2020), Mogstad et al. (2020a), Mogstad et al. (2020b), Kamat (2019)
2. **Orthogonal/debiased inference**: Newey (1994), Belloni et al. (2011), Chernozhukov et al. (2016), Belloni et al. (2016), Belloni et al. (2017), Chernozhukov et al. (2018)
3. **Machine learning for heterogenous treatment effects**: Athey and Imbens (2016), Wager and Athey (2018),Chernozhukov et al. (2017)
4. **Partial identification with convexity**: Beresteanu and Molinari (2008), Bontemps et al. (2012), Chandrasekhar et al. (2012)
5. **Monotonicity/LATE**: Imbens and Angrist (1994), Angrist and Imbens (1995), Kolesar (2013), Huber et al. (2015), Sloczynski (2021)
6. **Empirical Applications**: Lee (2009), Finkelstein et al. (2012)

## Outline

1. Overview of Lee (2009) bounds
2. Better Lee Bounds
3. JobCorps revisited

## Lee (2009): notation

Potential employment and wage outcomes

- $D = 1$ if subject wins a lottery
- $S(d) = 1$ if employed when $D = d$ for $d \in \{1, 0\}$
- $Y(d)$ wage when $D = d$ for $d \in \{1, 0\}$

## Lee (2009): notation

Potential employment and wage outcomes

- $D = 1$ if subject wins a lottery
- $S(d) = 1$ if employed when $D = d$ for $d \in \{1, 0\}$
- $Y(d)$ wage when $D = d$ for $d \in \{1, 0\}$

Observed data are $(X, D, S, S \cdot Y)$. $Y$ exists only if $S = 1$.

## Lee (2009): notation

Potential employment and wage outcomes

- $D = 1$ if subject wins a lottery
- $S(d) = 1$ if employed when $D = d$ for $d \in \{1, 0\}$
- $Y(d)$ wage when $D = d$ for $d \in \{1, 0\}$

Observed data are $(X, D, S, S \cdot Y)$. $Y$ exists only if $S = 1$.

Target population is the always-takers

$$S(1) = S(0) = 1 \quad \text{a.s.}$$

## Lee (2009): notation

Potential employment and wage outcomes

- $D = 1$ if subject wins a lottery
- $S(d) = 1$ if employed when $D = d$ for $d \in \{1, 0\}$
- $Y(d)$ wage when $D = d$ for $d \in \{1, 0\}$

Observed data are $(X, D, S, S \cdot Y)$. $Y$ exists only if $S = 1$.

Target population is the always-takers

$$S(1) = S(0) = 1 \quad \text{a.s.}$$

Target parameter

$$\text{ATE} = \mathbb{E}[Y(1) - Y(0) \mid S(1) = S(0) = 1]$$

Lee's monotonicity assumption

$$S(0) = 1 \Rightarrow S(1) = 1.$$

## Lee (2009): basic bounds

$\mathbb{E}[Y(0) \mid S(0) = 1]$ is point-identified. $\mathbb{E}[Y(1) \mid S(0) = 1]$ is not.

## Lee (2009): basic bounds

$\mathbb{E}[Y(0) \mid S(0) = 1]$ is point-identified. $\mathbb{E}[Y(1) \mid S(0) = 1]$ is not.

Always-takers' share among the treated

$$p_0 = \frac{\Pr[S = 1|D = 0]}{\Pr[S = 1|D = 1]} \in (0, 1)$$

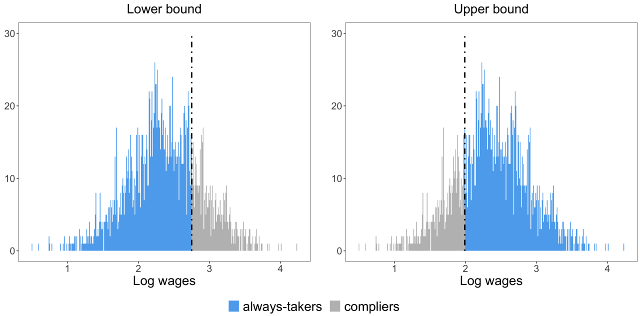
## Lee (2009): basic bounds

$\mathbb{E}[Y(0) \mid S(0) = 1]$ is point-identified. $\mathbb{E}[Y(1) \mid S(0) = 1]$ is not.

Always-takers' share among the treated

$$p_0 = \frac{\Pr[S = 1|D = 0]}{\Pr[S = 1|D = 1]} \in (0, 1)$$

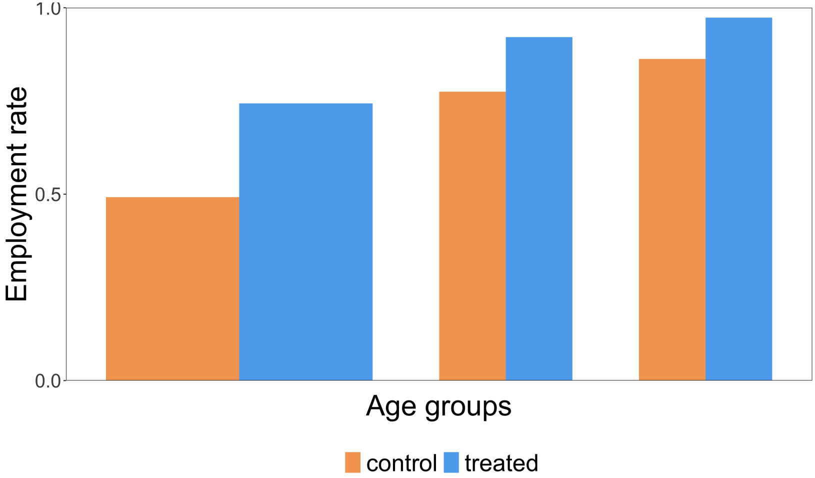
The borderline wage in the worst and the best case



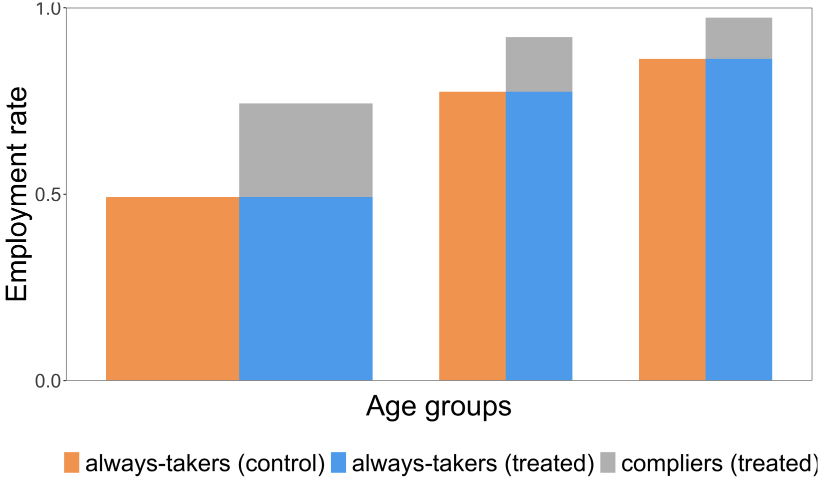

The ATE bounds are trimmed means

$$\widehat{\beta}^{\text{L}}_{\text{basic}} = 0.78 \text{ and } \widehat{\beta}^{\text{U}}_{\text{basic}} = 1.19$$

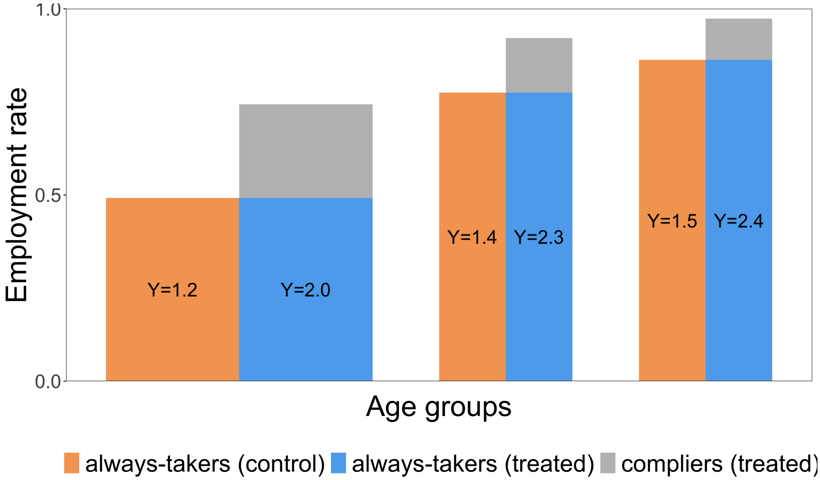
## Lee (2009): covariate-based bounds



## Lee (2009): covariate-based bounds, cont.

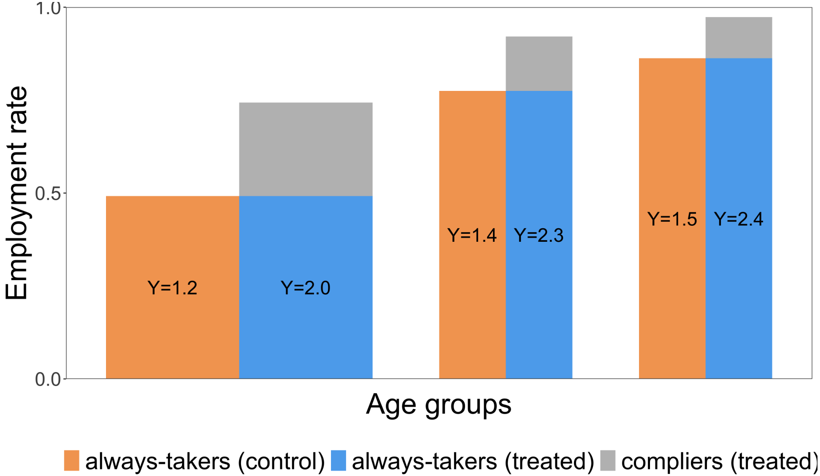


## Lee (2009): covariate-based bounds, cont.




lower bound $\widehat{\beta}_{\text{discrete}} = {}^1\!/_2(2.0 - 1.2) + {}^1\!/_4(2.3 - 1.4) + {}^1\!/_4(2.4 - 1.5) = 0.85$

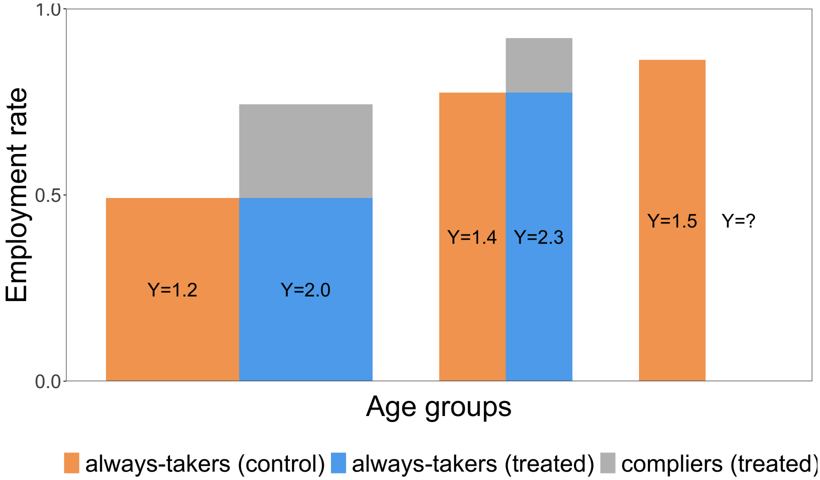
## Lee (2009): covariate-based bound, cont.




$$\widehat{\beta}_{\text{discrete}} = 1/2(2.0 - 1.2) + 1/4(2.3 - 1.4) + 1/4(2.4 - 1.5) = 0.85$$

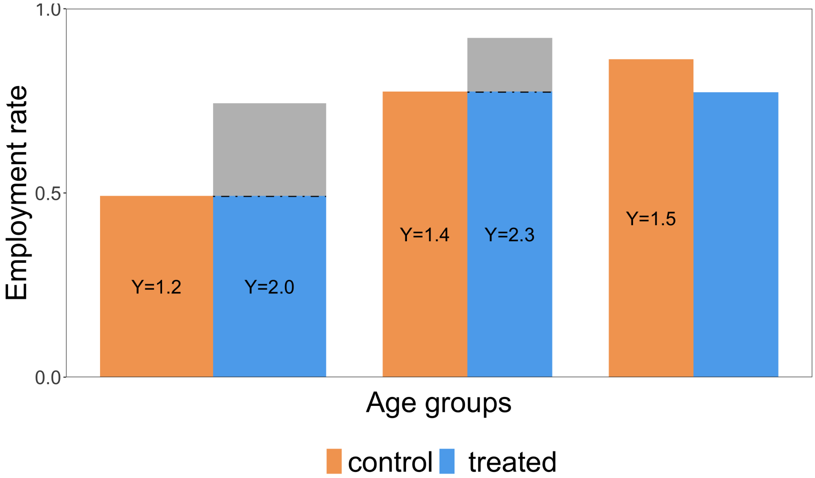
$$\widehat{\beta}_{\text{discrete}} = 0.85 > \widehat{\beta}_{\text{basic}} = 0.78$$

## Lee (2009): degenerate cell example




$$\widehat{\beta}_{\text{discrete}} = {}^1\!/_2(2.0 - 1.2) + {}^1\!/_4(2.3 - 1.4) + {}^1\!/_4(? - 1.5) = ?$$

## Lee (2009): heterogeneous monotonicity example




$$\widehat{\beta}_{\text{discrete}} = {}^1\!/_2(2.0 - 1.2) + {}^1\!/_4(2.3 - 1.4) + ? = ?$$

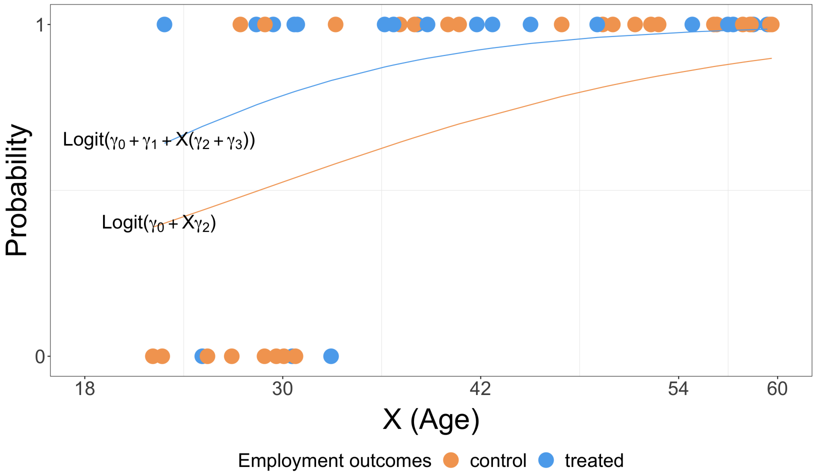
# Outline

1. Overview of Lee (2009) bounds
2. Better Lee Bounds
   - degenerate cell problem
   - monotonicity
3. JobCorps revisited

## Better Lee Bounds: the probability of employment

employment probability

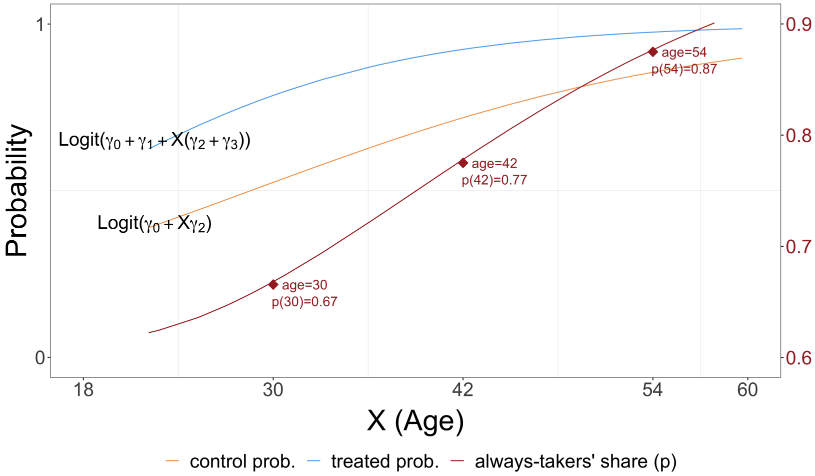
$$\Pr(S = 1 \mid D, X) = \text{Logit}(\gamma_0 + D \cdot \gamma_1 + X \cdot \gamma_2 + D \cdot X \cdot \gamma_3)$$



## Better Lee Bounds: the always-takers' share among the treated

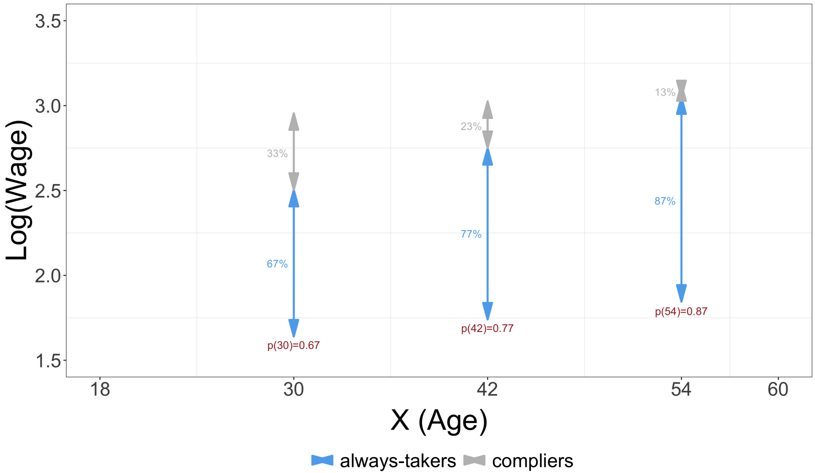
always-takers' share among the treated

$$p(X) = \frac{\text{Logit}(\gamma_0 + X \cdot \gamma_2)}{\text{Logit}(\gamma_0 + \gamma_1 + X \cdot (\gamma_2 + \gamma_3))}$$



## Better Lee Bounds: the worst case for the always-takers

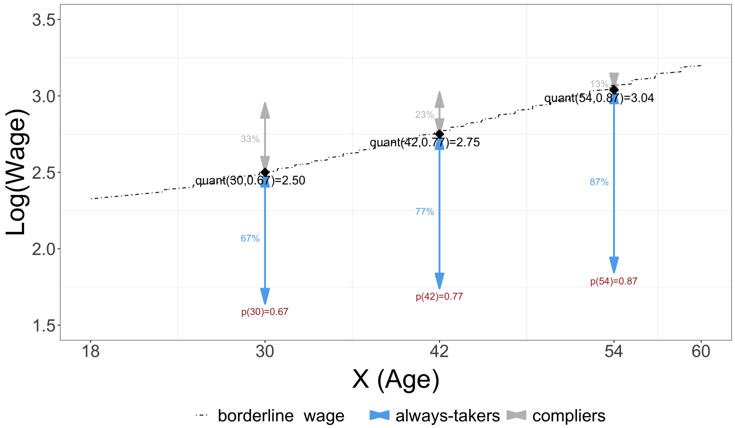
worst case: the always-takers' wages are below the compliers' wages for every age



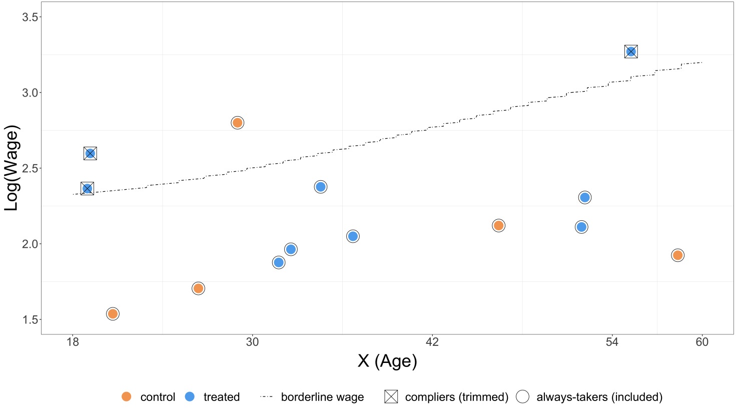
# Better Lee Bounds: the borderline always-takers' wage

quant($X, u$) is $u$-quantile of the treated wages

the borderline wage is quant($X, p(X)$) for each $X$



# Better Lee Bounds: the trimmed wage sample




$$\widehat{\beta}_{\text{better}} = \bar{Y}_{\text{treated}} - \bar{Y}_{\text{control}} = 0.83$$

$$\widehat{\beta}_{\text{better}} = 0.83 \approx \widehat{\beta}_{\text{discrete}} = 0.85$$

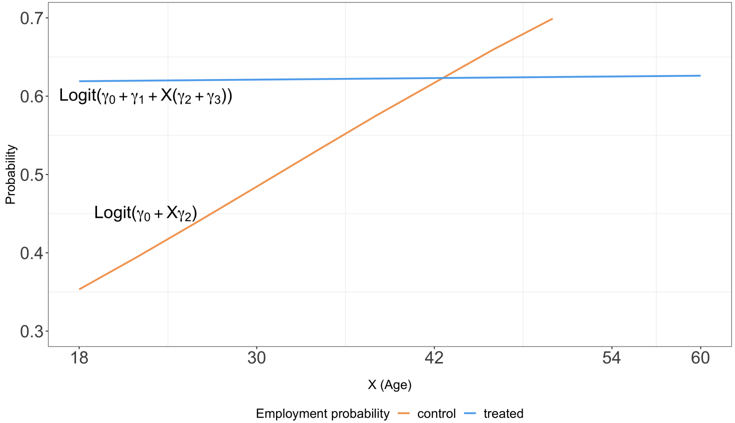
# Outline

1. Overview of Lee (2009) bounds
2. Better Lee Bounds
   - degenerate cell problem
   - monotonicity
3. JobCorps revisited

# Better Lee Bounds: conditional monotonicity

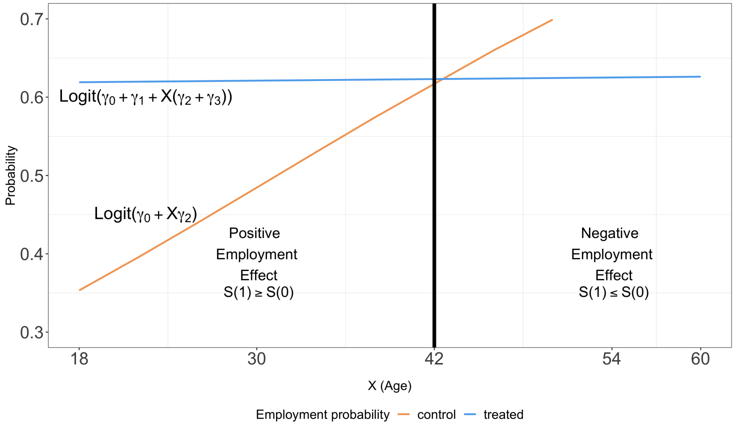
employment probability

$$s(D, X) = \Pr(S = 1 \mid D, X) = \text{Logit}(\gamma_0 + D \cdot \gamma_1 + X \cdot \gamma_2 + D \cdot X \cdot \gamma_3)$$



# Better Lee Bounds: conditional monotonicity, cont.

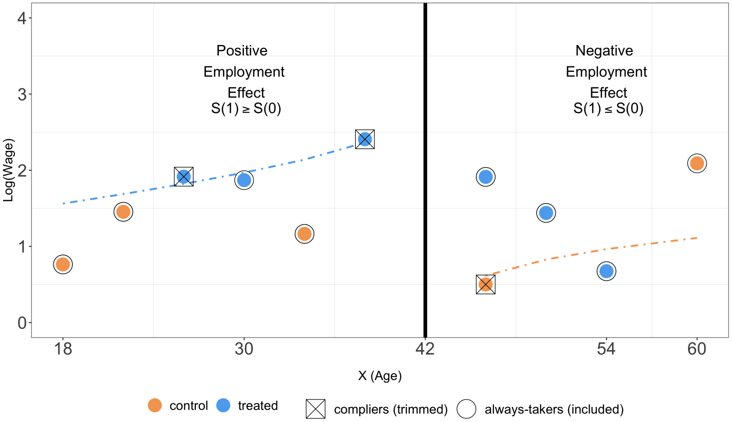
$$X_{\text{pos}} = \{X : s(0,X) < s(1,X)\} \text{ and } X_{\text{neg}} = \{X : s(0,X) > s(1,X)\}$$



# Better Lee Bounds: anatomy of bounds under conditional monotonicity

the always-takers' share is $p(X) = s(0, X)/s(1, X)$ if $X \in X_{\text{pos}}$ and $1/p(X)$ otherwise

$$\text{the borderline wage is } \begin{cases} \text{quant}_1(X, p(X)) & X \in X_{\text{pos}} \\ \text{quant}_0(X, 1 - 1/p(X)) & X \in X_{\text{neg}} \end{cases}$$



## Better Lee Bounds: conditional monotonicity, cont.

- Conditional monotonicity as in Kolesar (2013)

  $$\text{either } \Pr(S(1) \geq S(0) \mid X) = 1 \text{ or } \Pr(S(1) \leq S(0) \mid X) = 1 \quad \text{a.s.}$$

  - weakest form of monotonicity assumption. It is untestable
- This paper: assumes that subjects are correctly classified into $X_{\text{pos}}$ and $X_{\text{neg}}$
- Future work: allow for incorrect classification as in (Andrews, Kitagawa, McCloskey, 2018)
- Future work: cond. monotonicity induces a smaller distortion than the unconditional one

## Better Lee Bounds: many covariates

Sparsity: few (out of many) covariates are relevant for employment and wage

- logistic and quantile series $\Rightarrow$ logistic and quantile LASSO
- automated penalty choice as in (Belloni et al (ECMA, 2017))
- bias correction to account for regularization

## Better Lee Bounds: many covariates

Sparsity: few (out of many) covariates are relevant for employment and wage

- logistic and quantile series ⇒ logistic and quantile LASSO
- automated penalty choice as in (Belloni et al (ECMA, 2017))
- bias correction to account for regularization

No sparsity: agnostic approach (Chernozhukov, Demirer, Duflo, Fernandez-Val, 2017)

- bounds width is proportional to first-stage $R^2$ ⇒ rank covariates by explained variance!
- sharpness is not guaranteed, but tighter in practice

# Outline

1. Overview of Lee (2009) bounds
2. Better Lee Bounds
3. JobCorps revisited

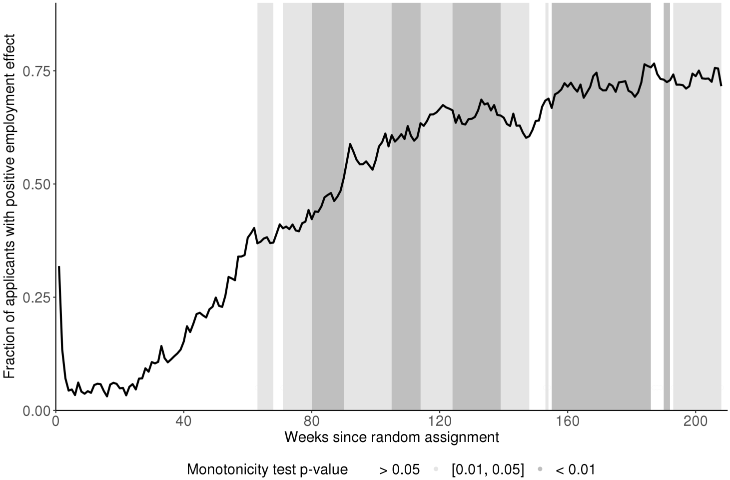
## JobCorps: overview of the program

- JobCorps is a training program that helps youth ages 16 through 24 to get a better job, make more money, and gain control over their lives (U.S. Department of Labor, 2005b).

## JobCorps: overview of the program

- JobCorps is a training program that helps youth ages 16 through 24 to get a better job, make more money, and gain control over their lives (U.S. Department of Labor, 2005b).
- $N = 9,145$ applicants – the sample sample as in Lee (2009)
- $p = 5,177$ baseline covariates $=$ demographics, reasons for joining JobCorps, medical, arrest, and drug use records, wage history
  - Lee (2009): 28 demographic covariates
- Lee (2009): week 90 wage effect is $[0.048, 0.049]$

## JobCorps: monotonicity failure demonstration




week 90: JobCorps helps (hurts) employment for 50 % of subjects

Lee's week 90 estimates $= [0.048, 0.049]$

## JobCorps: Better Lee Bounds on week 90 wage effect

| | (1) series | (2) lasso | (3) union | (4) agnostic |
|---|---|---|---|---|
| est. bounds | [-0.005, 0.091] | [0.040, 0.046] | [0.041, 0.059] | [0.041, 0.043] |
| 95 % CI | (-0.05, 0.135) | (0.001, 0.078) | (-0.02, 0.112) | (-0.02, 0.101) |
| # emplmnt covs | 28 | 5 177 (9) | 15 | 12 |
| # wage covs | 28 | 421 (6) | 15 | 12 |

1. 28 Lee's covs
2. 5, 177 = all covs, (9) = emplpyment equation, (6) = wage equation
3. 15 = union of employment and wage covs selected in Column 2
4. 12 covs selected on the sample Lee excluded

# JobCorps: Better Lee Bounds on week 90 wage effect

| | (1) series | (2) lasso | (3) union | (4) agnostic |
|---|---|---|---|---|
| est. bounds | [-0.005, 0.091] | [0.040, 0.046] | [0.041, 0.059] | [0.041, 0.043] |
| 95 % CI | (-0.05, 0.135) | (0.001, 0.078) | (-0.02, 0.112) | (-0.02, 0.101) |
| # emplmnt covs | 28 | 5 177 (9) | 15 | 12 |
| # wage covs | 28 | 421 (6) | 15 | 12 |

$1^{st}$ **stage** :

28 Lee's covs = age, race, education, parental educ., income, earnings at baseline

logit Employed $D$ $(X_1 - X_{28})$ $D * (X_1 - X_{28})$

qreg LogWage $D$ $(X_1 - X_{28})$ $D * (X_1 - X_{28})$ if Employed $==1$

## JobCorps: Better Lee Bounds on week 90 wage effect

| | (1) series | (2) lasso | (3) union | (4) agnostic |
|---|---|---|---|---|
| est. bounds | [-0.005, 0.091] | [0.040, 0.046] | [0.041, 0.059] | [0.041, 0.043] |
| 95 % CI | (-0.05, 0.135) | (0.001, 0.078) | (-0.02, 0.112) | (-0.02, 0.101) |
| # emplmnt covs | 28 | 5 177 (9) | 15 | 12 |
| # wage covs | 28 | 421 (6) | 15 | 12 |

$1^{st}$ **stage** :

5177=28 covs + reasons for joining JobCorps, medical, arrest, and drug use records, wage history

post lasso logit Employed $D$ $(X_1 - X_{5177})$ $D * (X_1 - X_{5177})$

post lasso qreg LogWage $D$ $(X_1 - X_{421})$ $D * (X_1 - X_{421})$ if Employed $== 1$

## JobCorps: Better Lee Bounds on week 90 wage effect

| | (1) series | (2) lasso | (3) union | (4) agnostic |
|---|---|---|---|---|
| est. bounds | [-0.005, 0.091] | [0.040, 0.046] | [0.041, 0.059] | [0.041, 0.043] |
| 95 % CI | (-0.05, 0.135) | (0.001, 0.078) | (-0.02, 0.112) | (-0.02, 0.101) |
| # emplmnt covs | 28 | 5 177 (9) | 15 | 12 |
| # wage covs | 28 | 421 (6) | 15 | 12 |

$1^{st}$ **stage** :

$$\mathcal{X} = 15 = 9 \text{ emplmnt} + 6 \text{ wage covs}$$

logit Employed $D$ $(X_1 - X_{15})$ $D * (X_1 - X_{15})$

qreg LogWage $D$ $(X_1 - X_{15})$ $D * (X_1 - X_{15})$ if Employed $== 1$

## JobCorps: Better Lee Bounds on week 90 wage effect

| | (1) series | (2) lasso | (3) union | (4) agnostic |
|---|---|---|---|---|
| est. bounds | [-0.005, 0.091] | [0.040, 0.046] | [0.041, 0.059] | [0.041, 0.043] |
| 95 % CI | (-0.05, 0.135) | (0.001, 0.078) | (-0.02, 0.112) | (-0.02, 0.101) |
| # emplmnt covs | 28 | 5 177 (9) | 15 | 12 |
| # wage covs | 28 | 421 (6) | 15 | 12 |

$1^{st}$ **stage** :

12 covs explaining most variation in wage

logit Employed $D$ $(X_1 - X_{12})$ $D * (X_1 - X_{12})$

qreg LogWage $D$ $(X_1 - X_{12})$ $D * (X_1 - X_{12})$ if Employed $== 1$

## Final thoughts

**Today**: Lee bounds $\rightarrow$ Better Lee Bounds

- ✓ discrete bounds $\rightarrow$ smooth interpolations
- ✓ monotonicity $\rightarrow$ conditional monotonicity

## Final thoughts

**Today**: Lee bounds $\rightarrow$ Better Lee Bounds

✓ discrete bounds $\rightarrow$ smooth interpolations

✓ monotonicity $\rightarrow$ conditional monotonicity

**This paper**:

+ unknown propensity score (e.g., quasi-experiments)
+ non-compliance
+ multiple outcomes / short panels
+ achieved nearly point-identification in Finkelstein et al. (2012) and Angrist et al, (2002)

## Future work

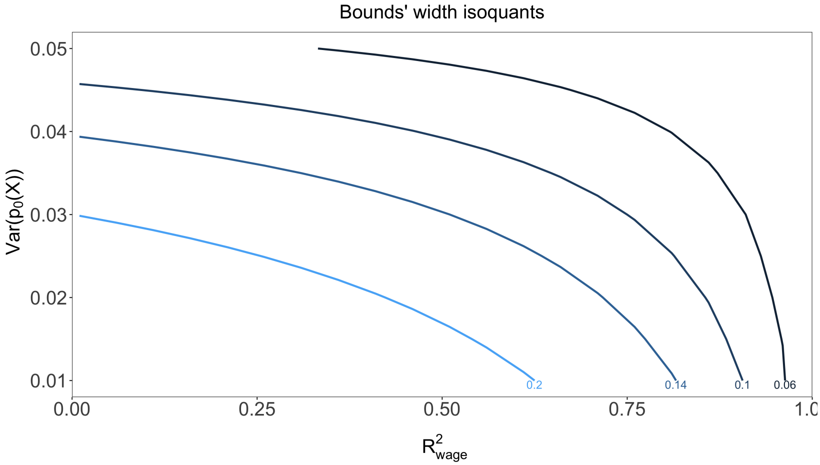
**This paper**: relax monotonicity

- point-identified $\Rightarrow$ bounded always-takers' share
- justify “better" in identification sense: conditional monotonicity induces a smaller distortion than the unconditional one

**Other bounds types**

- sharp bounds $\neq$ tightest CI!
- strong partial ID $<$ very tight bounds $<$ point ID

# Intuition for bounds' width



Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *The American Economic Review*, 92(5):1535–1558.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.

Angrist, J. D., Pathak, P. A., and Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics*, 5(4):1–27.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7460.

Belloni, A., Chernozhukov, V., et al. (2011). $\ell_1$-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130.

Belloni, A., Chernozhukov, V., Fernandez-Val, I., and Hansen, C. (2017).

Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85:233–298.

Belloni, A., Chernozhukov, V., and Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34(4):606–619.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.

Bontemps, C., Magnac, T., and Maurin, E. (2012). Set identified linear models. *Econometrica*, 80:1129–1155.

Chandrasekhar, A., Chernozhukov, V., Molinari, F., and Schrimpf, P. (2012). Inference for best linear approximations to set identified functions. *arXiv e-prints*, page arXiv:1212.5627.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21:C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2017). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. *arXiv e-prints*, page arXiv:1712.04802.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2016). Locally Robust Semiparametric Estimation. *arXiv e-prints*, page arXiv:1608.00033.

Feller, A., Greif, E., Ho, N., Miratrix, L., and Pillai, N. (2016). Weak separation in mixture models and implications for principal stratification.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J., Allen, H., Baicker, K., and Group, O. H. S. (2012). The oregon health insurance experiment: Evidence from the first year. *Quarterly Journal of Economics*, 127(3):1057–1106.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4):475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Honore, B. and Hu, L. (2020). Selection without exclusion. *Econometrica*, 88(88):1007–1029.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Kolesar, M. (2013). Estimation in an instrumental variable model with treatment effect heterogeneity.

Lee, D. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Manski, C. F. (1989). Anatomy of the selection problem. *The Journal of Human Resources*, 24(3):343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.

Mogstad, M., Torgovitsky, A., and Walters, C. (2020a). The causal interpretation of two-stage least squares with multiple instrumental variables.

Mogstad, M., Torgovitsky, A., and Walters, C. (2020b). Policy evaluation with multiple instrumental variables.

Newey, W. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):245–271.

Sloczynski, T. (2021). When should we (not) interpret linear iv estimands as late?

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal Journal of the American Statistical Association*, 113:1228–1242.