Non-existent outcomes in research on inequality: A causal approach

Ian Lundberg

UCLA Sociology and California Center for Population Research

Soonhong Cho

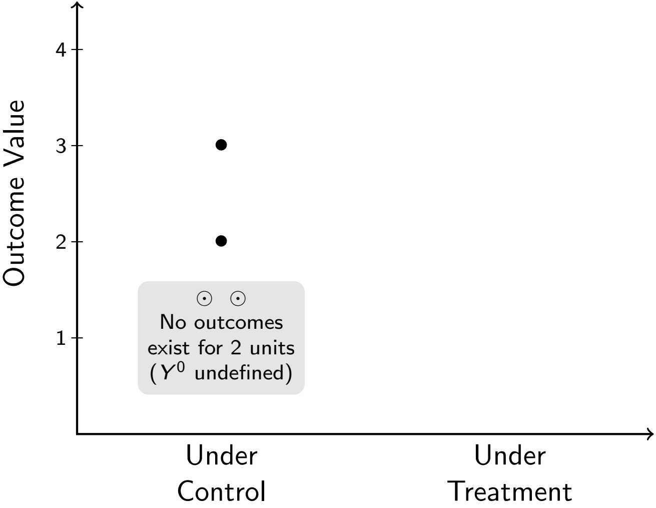
UCLA Political Science and California Center for Population Research

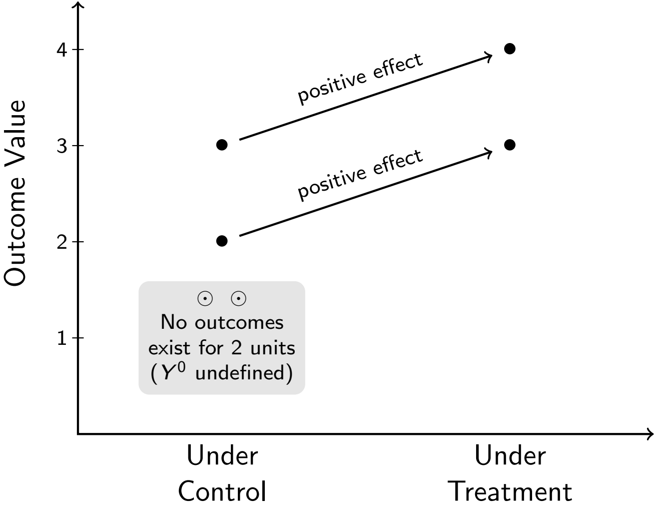
28 June 2025
University of Tokyo

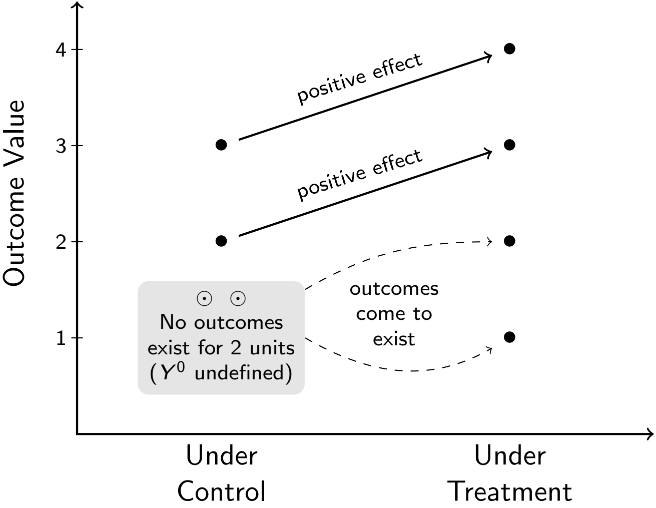
Research reported in this presentation was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number P2CHD041022.

Some outcomes exist only for some people

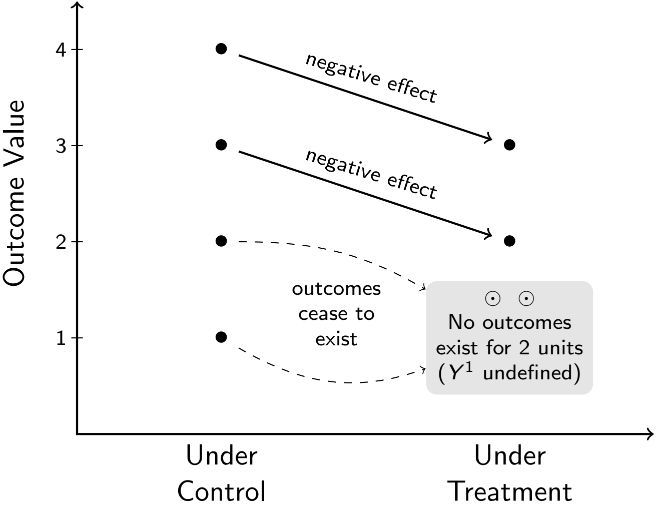
- only employed people have an hourly wage
- only married people can report marital satisfaction
- only people with children have descendants


Outcome Value
4
3
2
1
⊙ ⊙
No outcomes
exist for 2 units
(Y^0 undefined)
Under
Control
Under
Treatment


Outcome Value
4
3
2
1
positive effect
positive effect
⊙ ⊙
No outcomes exist for 2 units (Y^0 undefined)
Under Control
Under Treatment



Outcome Value
4
3
2
1
positive effect
positive effect
⊙ ⊙
No outcomes exist for 2 units (Y^0 undefined)
outcomes come to exist
Under Control
Under Treatment


Outcome Value
4
3
2
1
negative effect
negative effect
outcomes cease to exist
⊙ ⊙
No outcomes exist for 2 units (Y^1 undefined)
Under Control
Under Treatment

Our goals in this paper

When researchers drop missing outcomes,

- inequality can be obscured
- we may miss causal effects

We provide methods to study both:

- effects on outcome existence (employment)
- effects on outcome values (wage)

One concrete setting

Parenthood reduces hourly wages for women

(Budig & England 2001; Gough & Noonan 2013)

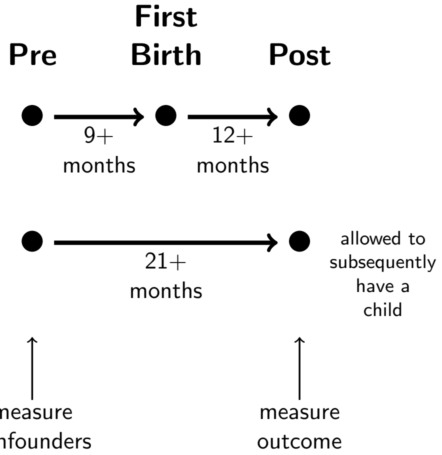
and increases wages for men

(Killewald 2013; Yu & Hara 2021)

The motherhood wage penalty may be disappearing over time

(Pal & Waldfogel 2016; Buchmann & McDaniel 2016; but see Jee et al. 2019)

Data: NLSY97



Parents

1,985 mothers
1,837 fathers

Non-Parents

	person-periods	persons
women	20,543	2,794
men	25,902	3,436

$$
\begin{aligned}
\text{log(Wage)} = \beta_0 &+ \beta_1(\text{Mother}) \\
&+ \beta_2(\text{Age}) \\
&+ \beta_3(\text{Married}) \\
&+ \beta_4(\text{Education}) \\
&+ \beta_5(\text{Work Experience}) \\
&+ \beta_6(\text{Full-Time}) \\
&+ \beta_7(\text{Tenure in Job}) \\
&+ \epsilon
\end{aligned}
$$

Maya

Maya

	if a mother	−	if not	=	effect
employment	_____	−	_____	=	_____
wage	_____	−	_____	=	_____

Maya

	if a mother	−	if not	=	effect
employment		−		=	
wage		−		=	

Maya

	if a mother	−	if not	=	effect
employment		−		=	
wage		−		=	

Maya

	if a mother	−	if not	=	effect
employment		−		=	0
wage		−		=	

Maya

	if a mother	−	if not	=	effect
employment		−		=	0
wage	\$30	−		=	

Maya

	if a mother	−	if not	=	effect
employment		−		=	0
wage	\$30	−	\$40	=	

Maya

	if a mother	−	if not	=	effect
employment		−		=	0
wage	\$30	−	\$40	=	−\$10

Mia

	if a mother	−	if not	=	effect
employment		−		=	
wage		−		=	

Mia

	if a mother	−	if not	=	effect
employment		−		=	
wage		−		=	

Mia

	if a mother	−	if not	=	effect
employment		−		=	
wage		−		=	

Mia

	if a mother		if not		effect
employment		$-$		$=$	-1
wage		$-$		$=$	

Mia

	if a mother		if not		effect
employment		−		=	-1
wage	??	−		=	

Mia

	if a mother	−	if not	=	effect
employment		−		=	-1
wage	??	−	\$20	=	

Mia

	if a mother	−	if not	=	effect
employment		−		=	-1
wage	??	−	\$20	=	??

Mia

	if a mother		if not		effect
employment		$-$		$=$	-1
wage	??	$-$	\$20	$=$	??

Principal Stratification
Frangakis & Rubin 2002; Zhang & Rubin 2003
For an intro, see Miratrix et al. 2018

Maya

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Maya

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia is a Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia is a Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia is a Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

DROPPED

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia is a Mother

DROPPED

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Average Observed

\$30

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia is a Mother

DROPPED

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Average Observed

\$30

Average Observed

\$30

Maya is a Mother

if a mother	$-$	if not	$=$	effect
	$-$		$=$	0
\$30	$-$	\$40	$=$	$-$\$10

Nancy is a Non-Mother

if a mother	$-$	if not	$=$	effect
	$-$		$=$	0
\$30	$-$	\$40	$=$	$-$\$10

Mia is a Mother

if a mother	$-$	if not	$=$	effect
	$-$		$=$	-1
??	$-$	\$20	$=$	??

Nia is a Non-Mother

if a mother	$-$	if not	$=$	effect
	$-$		$=$	-1
??	$-$	\$20	$=$	??

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	−\$10

Mia is a Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

1) Average effect of motherhood on employment

Maya is a Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	**−\$10**

Nancy is a Non-Mother

if a mother	−	if not	=	effect
	−		=	0
\$30	−	\$40	=	**−\$10**

Mia is a Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

Nia is a Non-Mother

if a mother	−	if not	=	effect
	−		=	−1
??	−	\$20	=	??

1) Average effect of motherhood on employment
2) Wage effect among those employed regardless

Goal 1: Effect on outcome existence

Causal Estimand

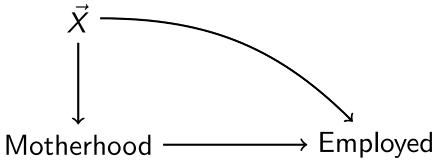
Define potential outcomes:
$S^1 =$ whether employed as a parent
$S^0 =$ whether employed as a non-parent

Causal estimand

$$\mathrm{E}(S^1 - S^0 \mid A = 1)$$

Goal 1: Effect on outcome existence

Causal Assumptions



Where $\vec{X}$ includes age, education, marital status, full-time employment, job tenure, work experience, and wage and employment each lagged by one year.

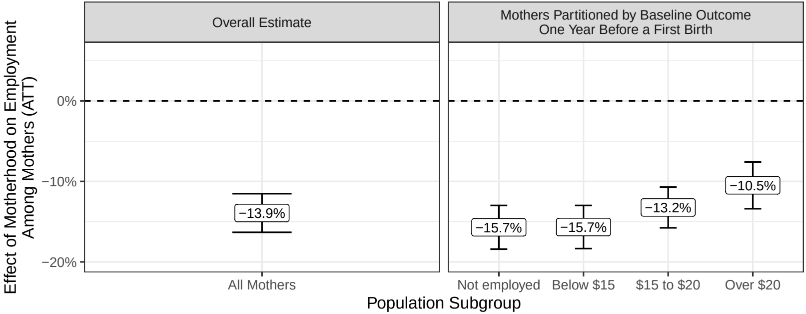
Goal 1: Effect on outcome existence

Estimation Strategy

- Regress Y on treatment and confounders
- Predict for all mothers average

$$\frac{1}{n_{\text{Mothers}}}\sum\left(\hat{Y}_i^1 - \hat{Y}_i^0\right)$$

Results for **mothers**



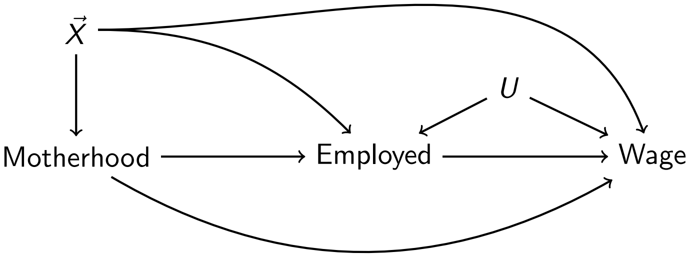
Goal 2: Effect on outcome value

Causal Estimand

$$\tau = \mathrm{E}\Bigg(\underbrace{Y^1 - Y^0}_{\substack{\text{Effect of Treatment} \\ \text{on Outcome Value}}} \quad \Big| \quad \underbrace{A = 1}_{\substack{\text{Among} \\ \text{the Treated}}} \; , \quad \underbrace{S^1 = S^0 = 1}_{\substack{\text{Whose Outcome} \\ \text{Exists Regardless} \\ \text{of Treatment}}} \Bigg) \tag{1}$$

Effect of motherhood on wage
among mothers who would be employed
as a parent or as a non-parent

Causal assumption: Exchangeability



Causal assumption: Monotonicity

$$S_i^1 \leq S_i^0 \quad \text{for all } i \qquad \text{(negative monotonicity)} \tag{2}$$

A woman who would be employed as a non-mother ($S_i^0 = 1$) would also be employed as a mother ($S_i^1 = 1$)

Causal assumption: Mean dominance

$$\begin{aligned}&\text{E}(Y^0 \mid \vec{X} = \vec{x}, S^0 = S^1 = 1) \geq \\ &\text{E}(Y^0 \mid \vec{X} = \vec{x}, S^0 = 1, S^1 = 0) \qquad \forall \vec{x}\end{aligned}$$

Women employed regardless of motherhood
would have higher mean wages as non-mothers
than women employed only if they were non-mothers

Identification: Making use of the assumptions

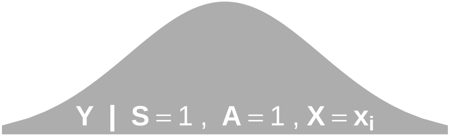
Recall that our goal is:

$$\begin{pmatrix}\text{mean factual}\\ \text{wage}\end{pmatrix} \quad - \quad \begin{pmatrix}\text{mean counterfactual wage}\\ \text{as non-mothers}\end{pmatrix}$$

among mothers employed regardless of motherhood.

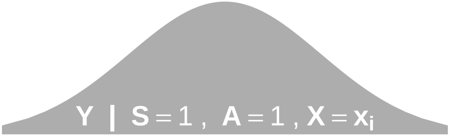
Identification: Mean factual wage as mothers

We observe mothers with a wage distribution



Identification: Mean factual wage as mothers

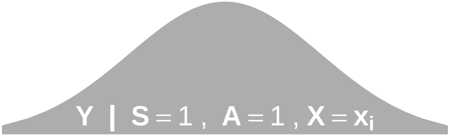
We observe mothers with a wage distribution



By assumption (monotonicity),
all would also be employed as non-mothers

Identification: Mean factual wage as mothers

We observe mothers with a wage distribution



By assumption (monotonicity),
all would also be employed as non-mothers

Estimate $\mathrm{E}(Y^1 \mid A = 1, S^0 = S^1 = 1, \vec{X})$ by the mean.

Identification: Counterfactual wage as non-mothers

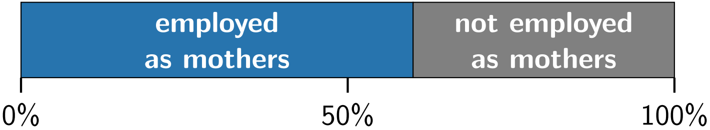
Identification: Counterfactual wage as non-mothers

We observe wages among non-mothers

Identification: Counterfactual wage as non-mothers

We observe wages
among non-mothers

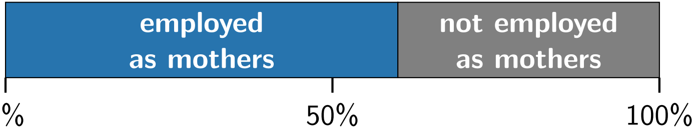
But some of them would not be employed as mothers



Identification: Counterfactual wage as non-mothers

We observe wages
among non-mothers

But some of them would not be employed as mothers



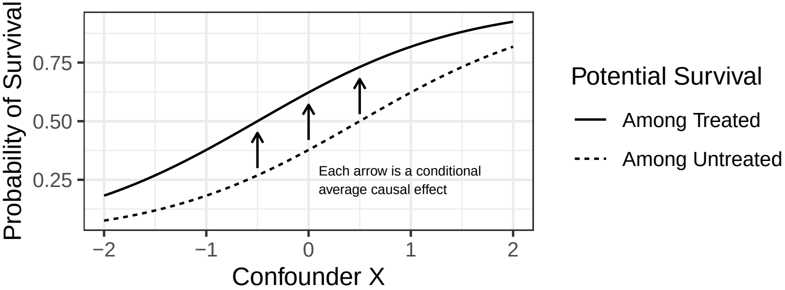
Two extreme possibilities:

Lower Bound:
Mean of Blue Region

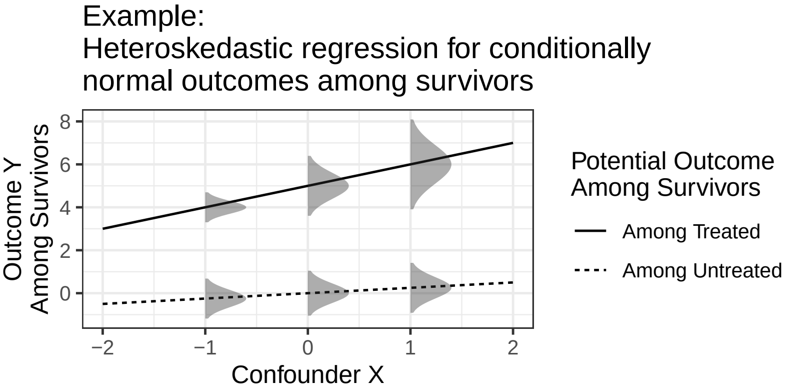
Upper Bound:
Mean of Blue Region

Estimation by regression and simulation

Example:
Logistic regression for survival



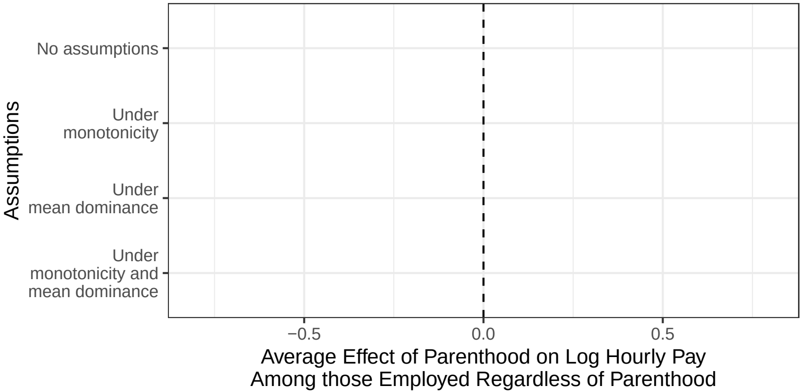
Estimation by regression and simulation



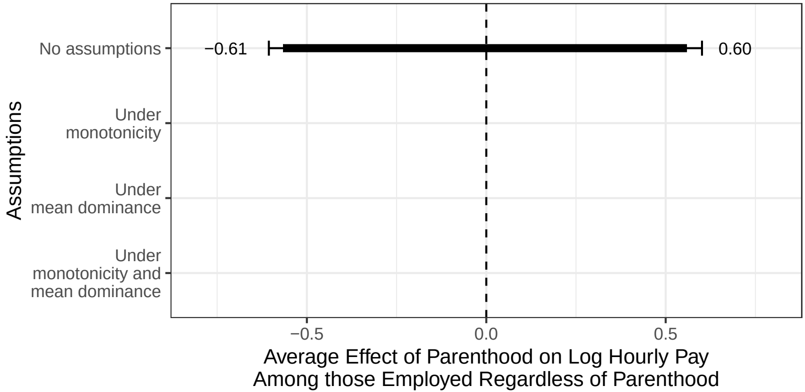
Review of our strategy

1. Define the causal estimands
2. Make causal assumptions (e.g., DAG + monotonicity)
3. Model Y among those with outcomes
4. Simulate a distribution for counterfactual estimation
5. Create bounds by considering extremes
   - Target subgroup are the **highest** paid
   - Target subgroup are the **lowest** paid

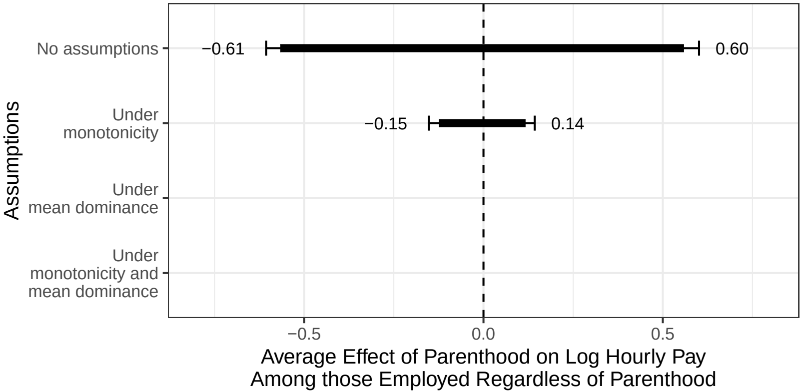
Results: Effect of motherhood on wages



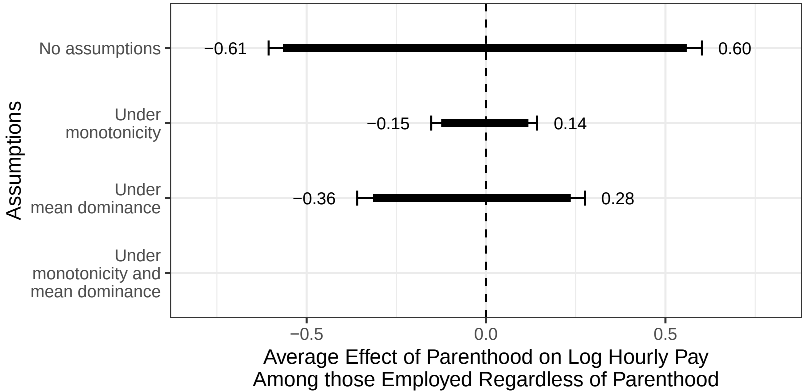
Results: Effect of motherhood on wages



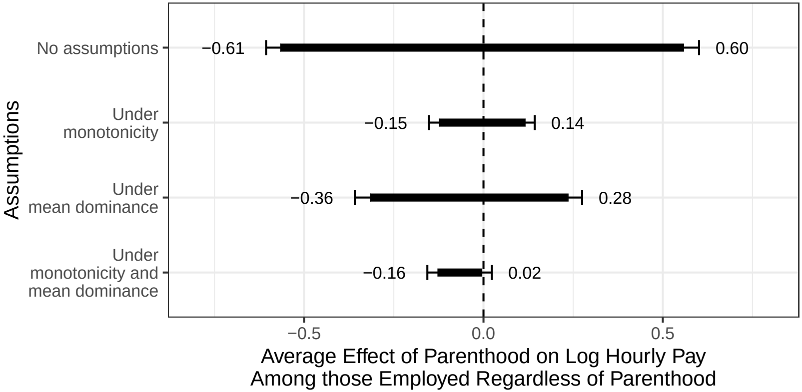
Results: Effect of motherhood on wages



Results: Effect of motherhood on wages



Results: Effect of motherhood on wages



Discussion: Building on past work

Discussion: Building on past work

When you can't point-identify,
make assumptions to set-identify (Manski 1995, 2003)

Discussion: Building on past work

When you can't point-identify,
make assumptions to set-identify (Manski 1995, 2003)

Principal stratification already did this (Frangakis & Rubin 2002)

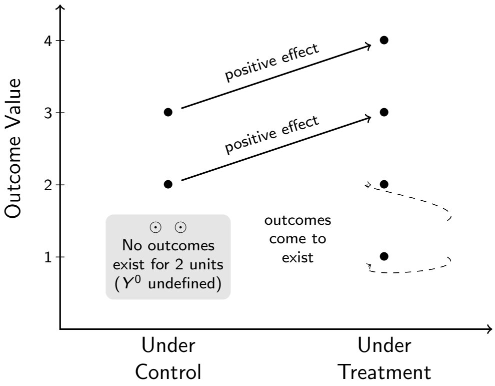
Discussion: Building on past work

When you can't point-identify,
make assumptions to set-identify (Manski 1995, 2003)

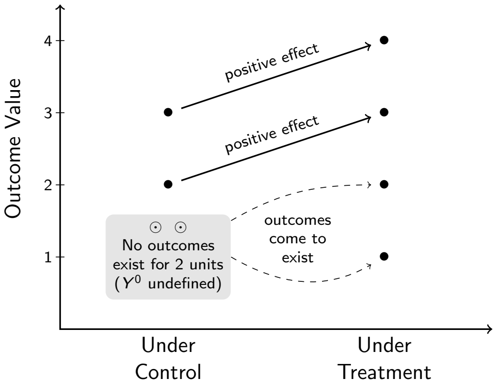
Principal stratification already did this (Frangakis & Rubin 2002)

Our new piece:
Regression + simulation for observational studies
with measured confounders

In social stratification,
many causes of outcome values
may also shape outcome existence



In social stratification,
many causes of outcome values
may also shape outcome existence



Thanks!

Ian Lundberg
ianlundberg@ucla.edu

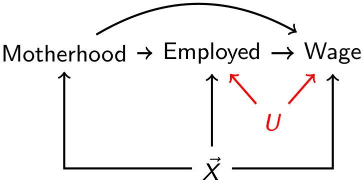
Soonhong Cho
tnsehdtm@gmail.com

Draft of paper
tinyurl.com/nonexistentoutcomes

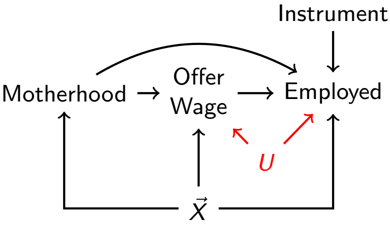
R package
ilundberg.github.io/pstratreg

Appendix Slides

Differences from Heckman selection model



Our DAG



Heckman DAG