Non-existent outcomes in research on inequality

A causal approach

Ian Lundberg
Cornell Information Science
ilundberg@cornell.edu

Soonhong Cho
UCLA Political Science
soonhongcho@g.ucla.edu

Cornell Inequality
Discussion Group
16 Oct 2023

Replication
code here

Try our (beta) R package!
ilundberg.github.io/pstratreg

Idea

Outcomes that do not exist can hide inequality

Plan

- one concrete setting
- general methodological tools
- open questions

Parenthood reduces hourly wages for women

- Budig & England 2001; Gough & Noonan 2013

and increases wages for men

- Killewald 2013; Yu & Hara 2021

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

Maya

Maya

$$\underline{\text{if a mother}} - \underline{\text{if not}} = \underline{\text{effect}}$$

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

Average Observed

\$30

Maya is a Mother

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

Nancy is a Non-Mother

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

Mia is a Mother

if a mother	−	if not	=	effect
(no job)	−	(job)	=	−1
??	−	\$20	=	??

DROPPED

Nia is a Non-Mother

if a mother	−	if not	=	effect
(no job)	−	(job)	=	−1
??	−	\$20	=	??

SHOULD ALSO DROP

subgroup: always employed

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	=	??

SHOULD ALSO DROP

subgroup: always employed

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	=	??

subgroup: motherhood blocks employment

Goal Motherhood wage effect among the always-employed

↑

women who
would be employed if a mom
and
would be employed if not

Goal Motherhood wage effect among the always-employed

Difficulty Who are those people?

women who
would be employed if a mom
and
would be employed if not

Goal Motherhood wage effect among the $\underline{\text{always-employed}}$

Difficulty Who are those people?

↑

women who
would be employed if a mom
and
would be employed if not

	Mother?	Wage
Maya	Yes	\$30
Mia	Yes	??
Nancy	No	\$40
Nia	No	\$20

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Goal Motherhood wage effect among the always-employed

Difficulty Who are those people?

women who
would be employed if a mom
and
would be employed if not

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Bound.
Report the highest and lowest estimates consistent with the data

Estimated Effect of Motherhood

Upper Bound:

Lower Bound:

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Bound.
Report the highest and lowest estimates consistent with the data

Estimated Effect of Motherhood

Upper Bound: $+\$10$

Lower Bound:

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Bound.
Report the highest and lowest estimates consistent with the data

Estimated Effect of Motherhood

Upper Bound: $+$\$10

Lower Bound: $-$\$10

Goal Motherhood wage effect among the always-employed

women who
would be employed if a mom
and
would be employed if not

Difficulty Who are those people?

	Mother?	Wage
Maya	Yes	\$30
~~Mia~~	~~Yes~~	~~??~~
Nancy	No	\$40
Nia	No	\$20

Assumption.
Employed mothers would still be employed even if they had no children

Bound.
Report the highest and lowest estimates consistent with the data

Estimated Effect of Motherhood

Upper Bound: $+\$10$

Lower Bound: $-\$10$

In practice,

- motherhood is not randomized
- identical people do not exist

Our contribution: Model-based estimates

Data: NLSY97

Parents
2,181 mothers
2,055 fathers

Pre $\longrightarrow$ **First Birth** $\longrightarrow$ **Post**

Pre → First Birth: $9+$ months

First Birth → Post: $12+$ months

Data: NLSY97

First Birth — **Pre** → (9+ months) → **First Birth** → (12+ months) → **Post**

Parents

2,181 mothers
2,055 fathers

Pre → 9+ months → First Birth → 12+ months → Post

Non-Parents

	person-periods	persons
women	20,578	2,796
men	25,921	3,439

Pre → 21+ months → Post: allowed to subsequently have a child

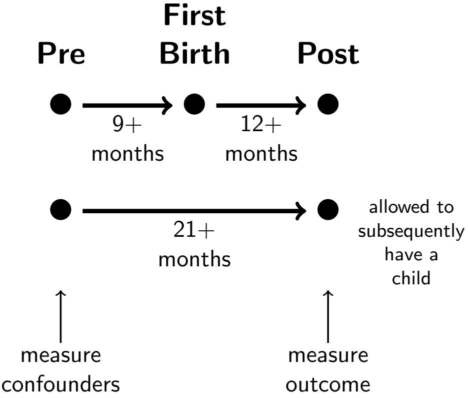
Data: NLSY97

Parents

2,181 mothers
2,055 fathers

Non-Parents

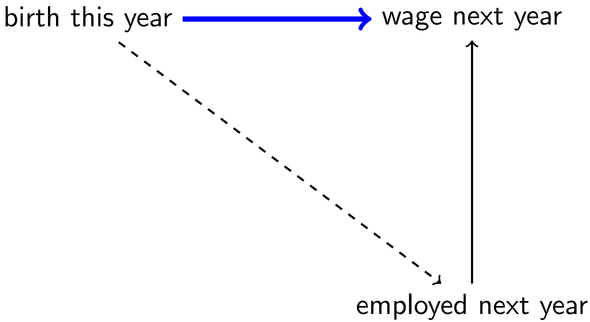
	person-periods	persons
women	20,578	2,796
men	25,921	3,439



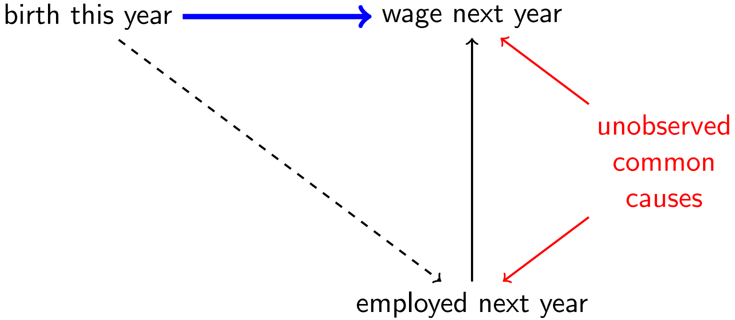
Causal assumptions

$$\text{birth this year} \longrightarrow \text{wage next year}$$

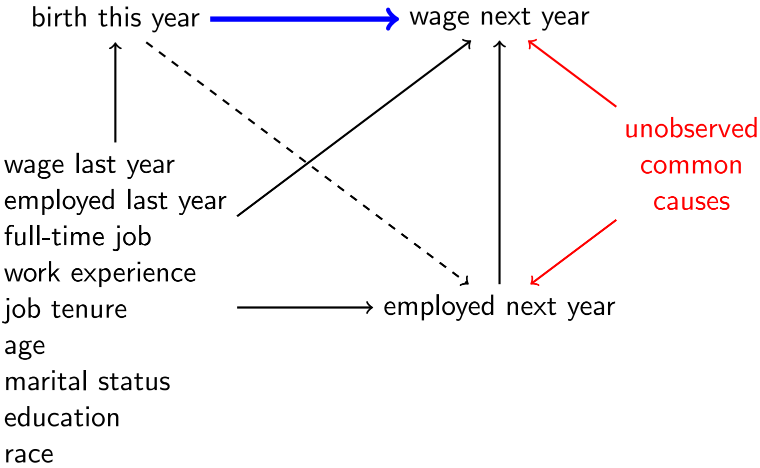
Causal assumptions



Causal assumptions



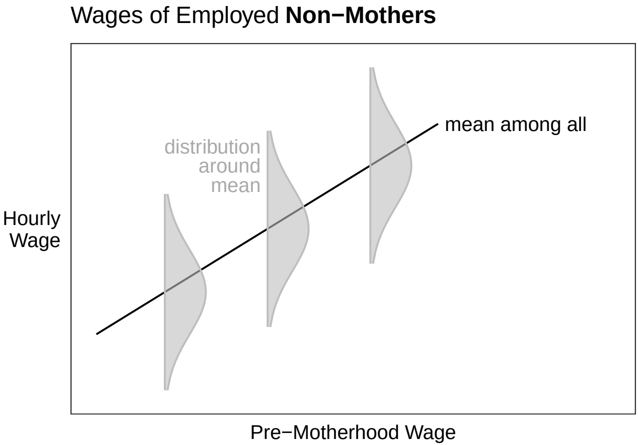
Causal assumptions



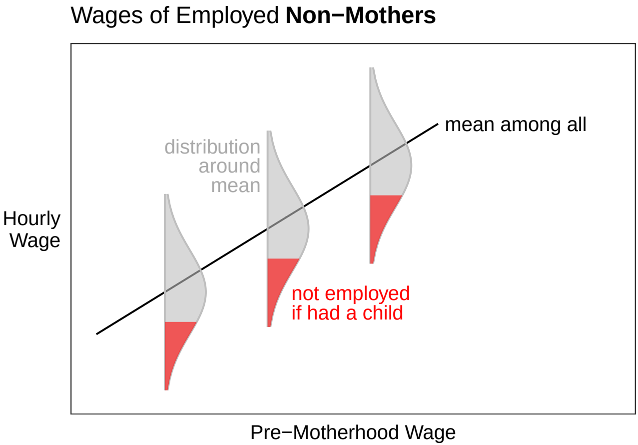
Modeling strategy

1. Logit of employment given motherhood and confounders
    - Estimate the proportion always-employed at each $\vec{X}, A$
2. OLS model of wage given motherhood and confounders
    - Predict wage under motherhood (no modifications)
    - Predict wage under no motherhood (with bounding)

Modeling strategy

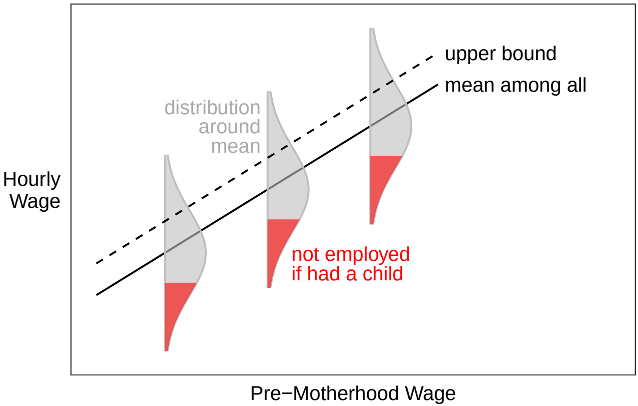


Modeling strategy

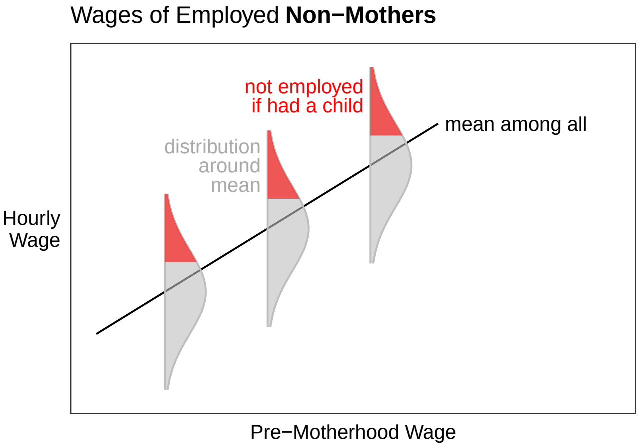


Modeling strategy

Wages of Employed **Non−Mothers**

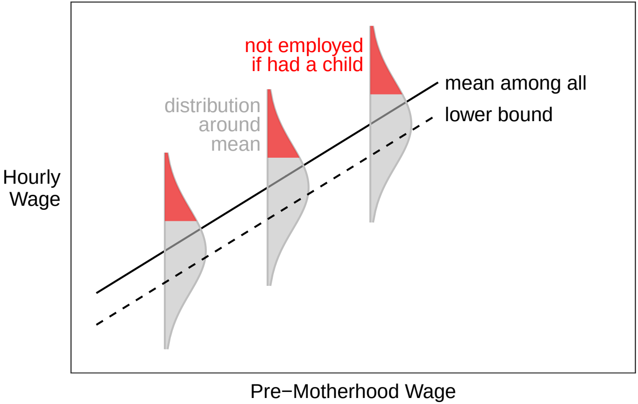


Modeling strategy

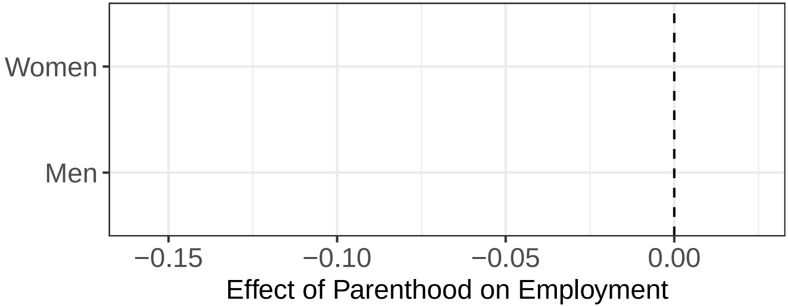


Modeling strategy

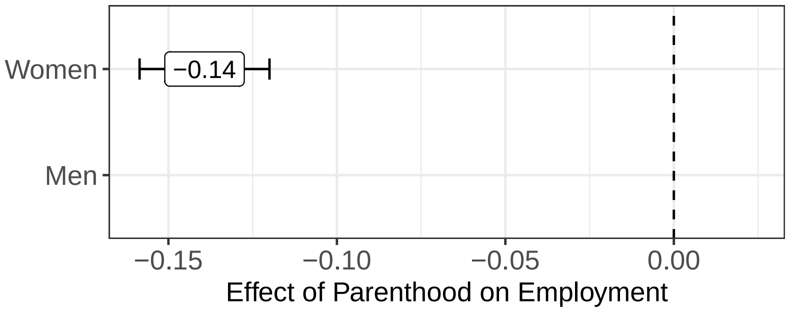
Wages of Employed **Non−Mothers**



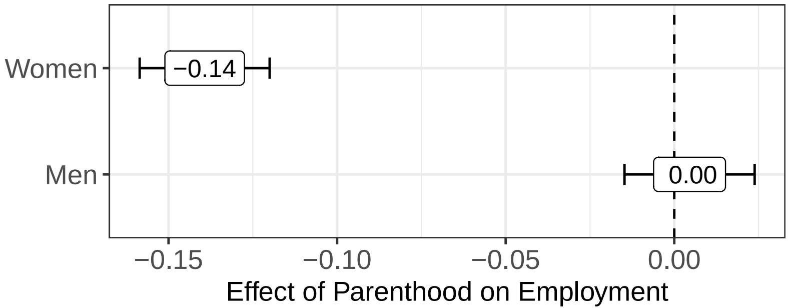
Results


Women
Men
−0.15
−0.10
−0.05
0.00
Effect of Parenthood on Employment

Results


Women
Men
−0.14
−0.15
−0.10
−0.05
0.00
Effect of Parenthood on Employment

Results


Women
Men
−0.14
0.00
−0.15
−0.10
−0.05
0.00
Effect of Parenthood on Employment

Results


Women
Men
High: $25 +
Mid: $15 − $25
Low: < $15
Not Employed
Wage Pre−Birth
−0.2
−0.1
0.0
0.1
Effect of Parenthood on Employment

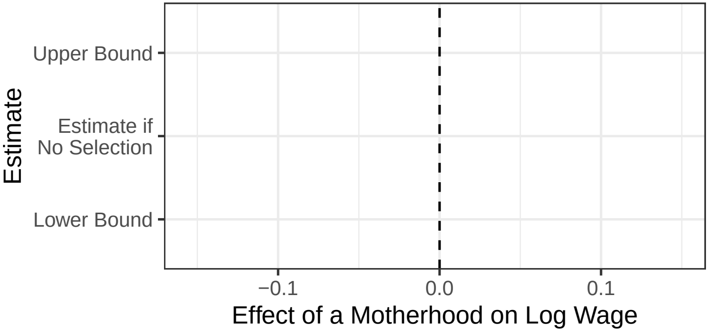
Results

Women
Men
High: $25 +
Mid: $15 − $25
Low: < $15
Not Employed
−0.10
−0.13
−0.16
−0.17
Wage Pre−Birth
−0.2
−0.1
0.0
0.1
Effect of Parenthood on Employment

Results

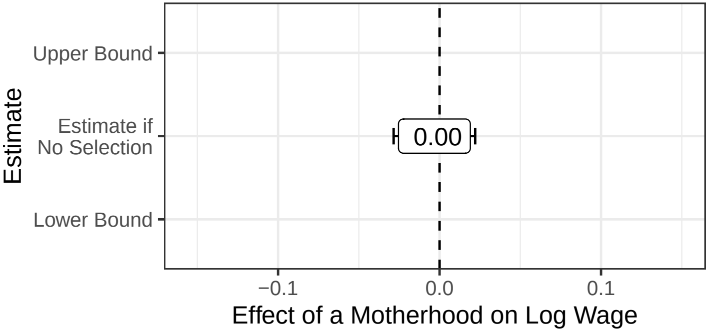
Women
Men
High: $25 +
Mid: $15 − $25
Low: < $15
Not Employed
−0.10
−0.13
−0.16
−0.17
0.00
−0.01
−0.02
0.06
Wage Pre−Birth
−0.2
−0.1
0.0
0.1
Effect of Parenthood on Employment

Results



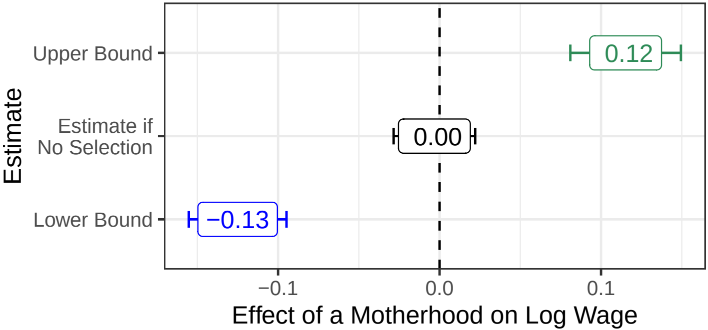
Upper Bound
Estimate if
No Selection
Lower Bound
Estimate
−0.1
0.0
0.1
Effect of a Motherhood on Log Wage

Results



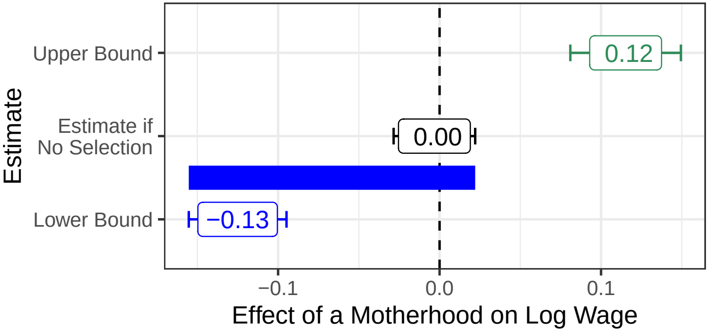
Upper Bound
Estimate if
No Selection
Lower Bound
Estimate
0.00
−0.1
0.0
0.1
Effect of a Motherhood on Log Wage

Results



Upper Bound
Estimate if
No Selection
Lower Bound
Estimate
0.12
0.00
−0.13
−0.1
0.0
0.1
Effect of a Motherhood on Log Wage

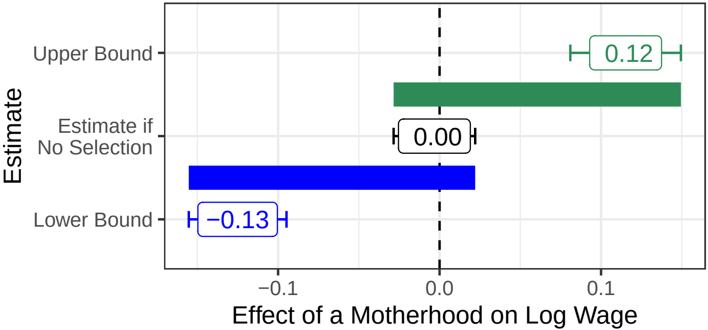
Results



If you believe the **lowest**-earning

non-mothers would stop paid employment if they had a child

Results



If you believe the **lowest**-earning or **highest**-earning

non-mothers would stop paid employment if they had a child

What we know that we did not know before

What we know that we did not know before

We knew

in recent years, motherhood only weakly predicts pay*

What we know that we did not know before

We knew

in recent years, motherhood only weakly predicts pay*

*among the employed

What we know that we did not know before

We knew
in recent years, motherhood only weakly predicts pay*

*among the employed

This fact is consistent with two stories

What we know that we did not know before

We knew
in recent years, motherhood only weakly predicts pay*

*among the employed

This fact is consistent with two stories

1. motherhood's causal effect on pay is small

What we know that we did not know before

We knew

in recent years, motherhood only weakly predicts pay*

*among the employed

This fact is consistent with two stories

1. motherhood's causal effect on pay is small

or

2. employed non-mothers the wrong comparison population

What we know that we did not know before

We knew
in recent years, motherhood only weakly predicts pay*

*among the employed

This fact is consistent with two stories

1. motherhood's causal effect on pay is small

or

2. employed non-mothers the wrong comparison population
   - lowest-earning non-mothers might stop paid work with a child

What we know that we did not know before

We know how to think
about outcomes that
don’t exist

What we know that we did not know before

We know how to think about outcomes that **don't exist**

 $-$  $= -1$

?? $-$ \$25 $=$??

What we know that we did not know before

We know how to think about outcomes that **don't exist**

 $-$  $= -1$

?? $-$ \$25 $=$??

in the labor market

— some people are not employed

What we know that we did not know before

We know how to think about outcomes that **don't exist**

 $-$  $= -1$

?? $-$ \$25 $=$??

in the labor market

— some people are not employed

in intergenerational mobility

— some people have no kids

What we know that we did not know before

We know how to think about outcomes that **don't exist**

 $-$  $= -1$

$?? \quad - \quad \$25 \; = \; ??$

in the labor market
— some people are not employed

in intergenerational mobility
— some people have no kids

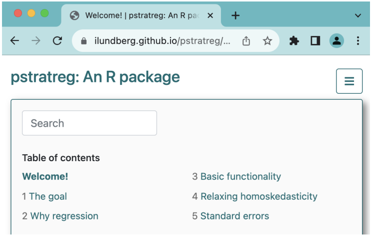
in assortative mating
— some people have no spouse

What we know that we did not know before

We know how to think about outcomes that **don't exist**

 $-$  $= -1$

$?? \quad - \quad \$25 = ??$



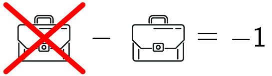
in the labor market
— some people are not employed

in intergenerational mobility
— some people have no kids

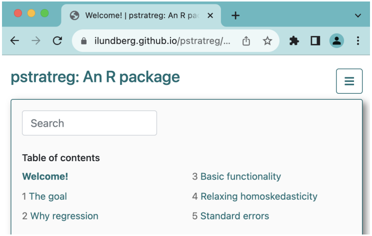
in assortative mating
— some people have no spouse

What we know that we did not know before

We know how to think about outcomes that **don't exist**



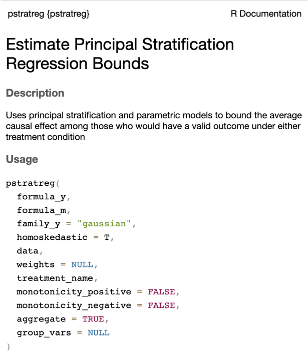
?? − \$25 = ??



in the labor market
— some people are not employed

in intergenerational mobility
— some people have no kids

in assortative mating
— some people have no spouse



pstratreg {pstratreg} R Documentation

Estimate Principal Stratification Regression Bounds

Description

Uses principal stratification and parametric models to bound the average causal effect among those who would have a valid outcome under either treatment condition

Usage

```
pstratreg(
  formula_y,
  formula_m,
  family_y = "gaussian",
  homoskedastic = T,
  data,
  weights = NULL,
  treatment_name,
  monotonicity_positive = FALSE,
  monotonicity_negative = FALSE,
  aggregate = TRUE,
  group_vars = NULL
)
```

Appendix

Sample restrictions: Persons

	Mothers	Non-Mothers	Fathers	Non-Fathers
All	2,999	3,746	2,662	4,200
18+ at pre-period	2,315	2,832	2,273	3,522
Observed in windows*	2,098	2,830	2,029	3,517
Non-missing wage if employed	1,985	2,794	1,837	3,436

* Requirement is

- parents: observed at
  - pre-birth: 3 years to 9 months before
  - post-birth: 1–3 years after
- for non-parents, two observations
  - at least 1 year + 9 months apart
  - no more than 6 years apart

Sample restrictions: Person-Periods

	Mothers	Non-Mothers	Fathers	Non-Fathers
All	2,999	31,510	2,662	39,325
18+ at pre-period	2,315	21,743	2,273	28,159
Observed in windows*	2,098	21,704	2,029	29,135
Non-missing wage if employed	1,985	20,543	1,837	25,902

* Requirement is

- parents: observed at
  - pre-birth: 3 years to 9 months before
  - post-birth: 1–3 years after
- for non-parents, two observations
  - at least 1 year + 9 months apart
  - no more than 6 years apart

Variable definitions

- A treatment: first birth occurs
- Y outcome: log hourly wage in 2022 dollars
  - including tips, overtime bonuses
  - includes non-hourly workers
  - for the job current at interview date
- $\vec{X}$ confounders
  - log(wage in pre-period)
  - employed in pre-period
  - full time (35+ hours) in pre-period current job
  - log(years of full-time work experience + 1), from weekly arrays
  - log(years of tenure in current job + 1), from job roster
  - age + age squared
  - marital status (married with spouse present, cohabiting, other)
  - education (less than high school, high school degree, 2-year college degree, 4-year college degree)
  - race (Hispanic, non-Hispanic Black, other)

Statistical models: Mediator (employment)

A indicates parenthood
$M = 1$ indicates employment
$\vec{X}$ are confounders

$$\text{logit}\left(P(M = 1 \mid \vec{X}, A)\right) = \alpha + \vec{X}'\vec{\gamma} + A\left(\beta + \vec{X}'\vec{\eta}\right)$$

Conditional average effect on mediator (employment)

$$\hat{P}(M^1 = 1 \mid \vec{X}, A) = \text{logit}^{-1}\left(\hat{\alpha} + \vec{X}'\hat{\vec{\gamma}} + 1 \times \left(\hat{\beta} + \vec{X}'\hat{\vec{\eta}}\right)\right)$$
$$\hat{P}(M^0 = 1 \mid \vec{X}, A) = \text{logit}^{-1}\left(\hat{\alpha} + \vec{X}'\hat{\vec{\gamma}} + 0 \times \left(\hat{\beta} + \vec{X}'\hat{\vec{\eta}}\right)\right)$$

Conditional Average Effect = Top − Bottom

Statistical models: Outcome (wage)

A indicates parenthood. $M = 1$ indicates employment.
Y is log hourly wage. $\vec{X}$ are confounders.

$$Y \mid \vec{X}, A, M = 1 \sim \text{Normal}\left(\mu(\vec{X}, A), \sigma^2(\vec{X}, A)\right)$$
$$\mu\left(\vec{X}, A\right) = \lambda + \vec{X}'\vec{\nu} + A\left(\tau + \vec{X}'\vec{\delta}\right)$$
$$\log\left[\sigma^2\left(\vec{X}, A\right)\right] = \xi + \vec{X}'\vec{\psi} + A\omega$$

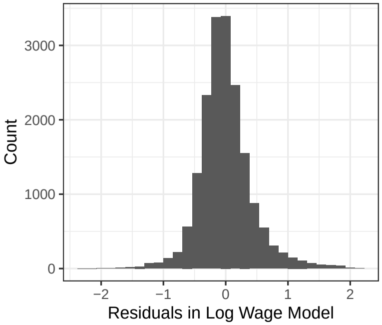
Estimate by

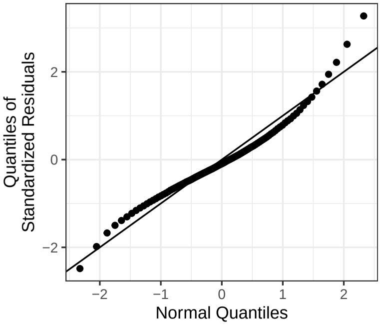
- for $\mu(\vec{X}, A)$, estimate by OLS with Y as the outcome
- then we calculate the residual $\hat{\epsilon} = Y - \hat{Y}$
- for $\sigma^2(\vec{X}, A)$, estimate by Gamma GLM with a log link with $\hat{\epsilon}^2$ as the outcome (Western & Bloome 2009)

Statistical models: Effects on wage

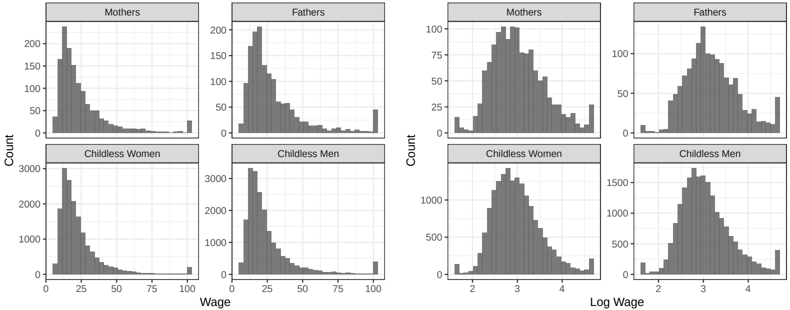
1. At the $\vec{X}$ of each unit, simulate
   - $Y^1 \sim \text{Normal}\left(\mu(\vec{X}, A = 1), \sigma^2(\vec{X}, A = 1)\right)$
   - $Y^0 \sim \text{Normal}\left(\mu(\vec{X}, A = 0), \sigma^2(\vec{X}, A = 0)\right)$
2. Remove the upper (lower) portion at this $\vec{X}$ value estimated to be the not-always-employed
3. Estimate the mean of the simulated values
4. Repeat for every unit

Normality of residuals





Histogram: Wage



Alternative: Utility functions

Alternative solution: Code the unemployed people with a wage

- if unemployed, then code with minimum wage

This works if you believe a utility function

$$U = \begin{cases} \log(\text{min wage}) & \text{if not employed} \\ \log(\text{wage}) & \text{if employed} \end{cases}$$

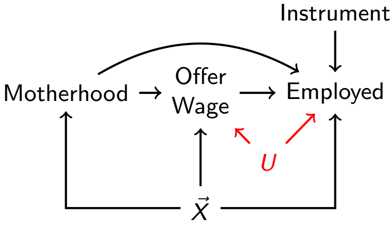
But if it is just a convenience

- then it does not bound effects on Y
- you might miss the real story
  - if half the sample was not employed, would you do this?

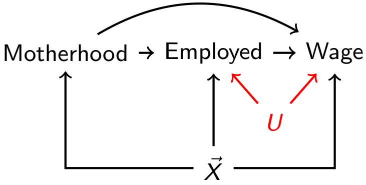
Alternative: Heckman selection

Premise: Everyone has a wage offer. Non-employed don't take it.

Goal: Infer about everyone, despite sample selection.

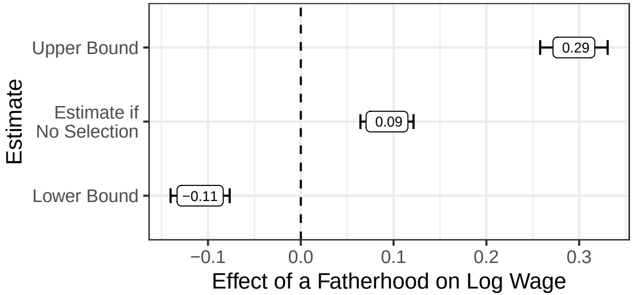


Heckman DAG

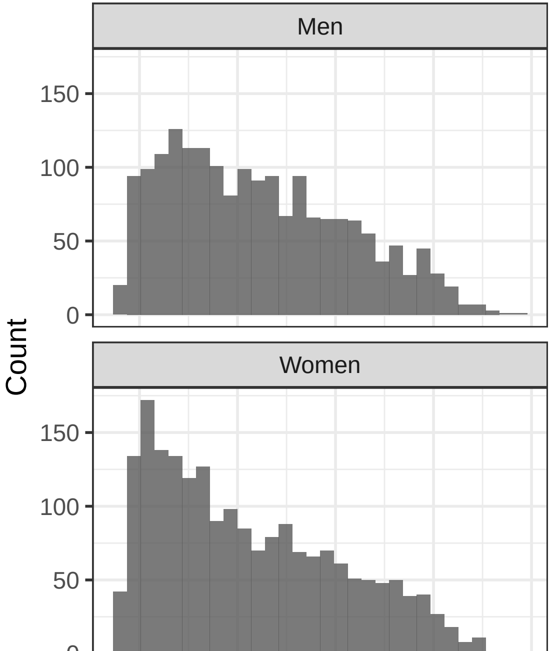


Our DAG

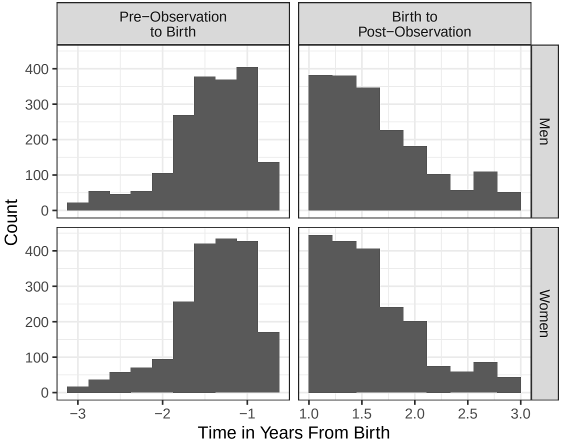
Fatherhood effects



Histogram: Age at birth



Histogram: Observation timing around birth



Histogram: Age of wage measurement

Histogram: Year of wage measurement