How the Large-Scale Early Withdrawals from Private Pension Plans Were Used: Insights from Young Adults

Triin Bulõgina, Merike Kukk

Discussion by: Julius Ilciukas

2021 Estonian Pension Reform

- Policy goal: increase flexibility and personal responsibility in retirement saving
- Second pillar becomes voluntary
- Individuals can withdraw all accumulated savings
- 20% of contributors exited in the first wave (September 2021)

Data

- Account-level transactional data from the third largest commercial bank in Estonia
- June 2020 to October 2022
- Focus on young adults aged 18–35
- 8500 active bank customers
- Outcomes: spending, investment, loan, and deposit balances

Empirical Strategy

- Propensity score matching on detailed pre-reform covariates
  - Demographics: sex, age, education
  - Financial situation: student loan, unemployed, mortgage, bailiff payments, etc.
- Log-difference regressions over 1–12 month horizons

Key Results

- Having loans, receiving unemployment or disability benefits, or being subject to bailiff enforcement $\Leftrightarrow$ more likely to withdraw pension early
- Having savings in the third pillar or other financial investments $\Leftrightarrow$ less likely to withdraw pension early
- In the short run: withdrawers spend more and repay their loans
- In the long run: no differences observed

What I liked

- Amazing data from a Baltic country
- Increasingly important research question
- Super well written paper
- Two (or three?) sets of interesting results:
  - Who takes out their savings?
  - What do they do with it?
  - What are the long term consequences?

Comments/suggestions

"Who takes out their savings?" is a very interesting question in its own right. At times, the paper gives the impression that this is more of a challenge to be addressed

- i.e. the focus on matching

"What do they do with it?" is also a very interesting question, and the data are well suited to address it

- Is matching necessary or especially helpful for estimating short-run effects?
- Perhaps an event study around the time of withdrawal could provide useful insights?

Alternative methods for the long run?

- The second pillar was mandatory only for those born in 1983 or later
- Could an RDD or diff-in-diff around the cutoff be feasible?
- This group may have followed similar trends to the treated group prior to the reform

Minor:

- Does controlling for future income risk introducing bias?

Parenthood Timing and Gender Inequality

Julius Ilciukas

University of Amsterdam

Motivation

Gender inequality in Western labor markets emerges when individuals become parents

Quantifying the causal effect of parenthood is central to understanding and addressing gender inequity

Existing evidence is conflicting

At the core of the conflict are methodological challenges:

1. Parenthood (timing) may be selective: human capital, wealth, health, career prospects
2. Effects may depend on timing: age of children, career stage at childbirth

Existing methods cannot address both simultaneously

This Paper

How much can we say about the causal effect of parenthood?

- How would labor market outcomes of parents change if they did not have children

1. New methodological approach robust to selection and dynamic effects
   - Leverage quasi-experimental variation in a sequence of conception attempts

2. Empirical evidence using novel administrative Dutch data
   - Focus on couples undergoing **artificial insemination**

3. Framework to assess the extent of timing-dependent effects
   - Quantify bias in leading methods

Preview of Main Results

- Parenthood persistently reduces women's work hours and income
  - Yearly reductions between 9 and 24 percent
- Parenthood causes a large share, but far from all, of post-child gender inequality
  - Between 36 and 54 percent in work hours and up to 46 percent in income
- Both selection and dynamic effects are substantial
  - Bias large enough to conclude all or none of gender inequality is due to parenthood

Literature and Contribution

1. Existing evidence relies on restrictive assumptions about the role of timing
   - Hotz et al. (2005); Agüero & Marks (2008); Cristia (2008); Miller (2011); Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide the first estimates robust to timing-dependent effects

Literature and Contribution

1. Existing evidence relies on restrictive assumptions about the role of timing
   - Hotz et al. (2005); Agüero & Marks (2008); Cristia (2008); Miller (2011); Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide the first estimates robust to timing-dependent effects

2. The most credible evidence comes from unique samples
   - Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide evidence from a common and minimally invasive infertility treatment

Literature and Contribution

1. Existing evidence relies on restrictive assumptions about the role of timing
   - Hotz et al. (2005); Agüero & Marks (2008); Cristia (2008); Miller (2011); Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide the first estimates robust to timing-dependent effects

2. The most credible evidence comes from unique samples
   - Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide evidence from a common and minimally invasive infertility treatment

3. Nearly all gender inequality associated with parenthood can be explained by gaps in outcomes between mothers and childless women (Kleven et al., 2019)
   - Patterns are nearly identical among couples undergoing artificial insemination

I provide a framework to separate causal effects from selection

Literature and Contribution

1. Existing evidence relies on restrictive assumptions about the role of timing
   - Hotz et al. (2005); Agüero & Marks (2008); Cristia (2008); Miller (2011); Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide the first estimates robust to timing-dependent effects

2. The most credible evidence comes from unique samples
   - Lundborg et al. (2017); Bensnes et al. (2023); Gallen et al. (2023); Lundborg et al. (2024)

I provide evidence from a common and minimally invasive infertility treatment

3. Nearly all gender inequality associated with parenthood can be explained by gaps in outcomes between mothers and childless women (Kleven et al., 2019)
   - Patterns are nearly identical among couples undergoing artificial insemination

I provide a framework to separate causal effects from selection

4. Addressing selection and dynamic effects is a common challenge
   - Education programs with multiple admission cycles, assignment to judges, promotion tournaments

Method applicable to many other settings with sequential quasi-experiments

Baseline Instrumental Variable Setup

Addressing selection into parenthood requires a quasi-experiment: focus on women who undergo artificial insemination

- Outcome of the first procedure $Z_1 \in \{0, 1\}$
- Parenthood indicator $D \in \{0, 1\}$
- Potential labor market outcomes $Y_{z_1}(d)$
- Effect of interest $Y_1(1) - Y_0(0)$

Assuming first procedure success is random (unconfoundedness) and affects outcomes only via parenthood status (exclusion) enables identification

Baseline Instrumental Variable Setup

Addressing selection into parenthood requires a quasi-experiment: focus on women who undergo artificial insemination

- Outcome of the first procedure $Z_1 \in \{0, 1\}$
- Parenthood indicator $D \in \{0, 1\}$
- Potential labor market outcomes $Y_{z_1}(d)$
- Effect of interest $Y_1(1) - Y_0(0)$

Assuming first procedure success is random (unconfoundedness) and affects outcomes only via parenthood status (exclusion) enables identification

However:

- 75% of women have children after the first procedure fails
- Z_1 affects D, but also affects timing of parenthood
- Exclusion may be violated: $Y_1(1) \neq Y_0(1)$, leading to bias

Model

Women differ in two unobserved characteristics:

- "Willingness" to undergo ACPs, $W \in \{1, \dots, \overline{w}\}$
  - Would undergo W ACPs for the first child if all ACPs failed
- "Reliance" on ACPs, $R \in \{0, 1\}$
  - No child if all ACPs fail, $R = 1$

Model

Women differ in two unobserved characteristics:

- "Willingness" to undergo ACPs, $W \in \{1, \ldots, \overline{w}\}$
  - Would undergo W ACPs for the first child if all ACPs failed
- "Reliance" on ACPs, $R \in \{0, 1\}$
  - No child if all ACPs fail, $R = 1$

Observables:

- ACP j success indicator, Z_j, for procedures before having any children
- Number of realized ACPs:

$$A = \min\left(\{j : Z_j = 1\} \cup \{W\}\right)$$

- Parenthood indicator:

$$D = \max(Z_A, 1 - R)$$

Sequential Unconfoundedness

Assumption (Sequential Unconfoundedness)

$(Y_1(1), Y_0(0), R, W) \perp\!\!\!\perp Z_j | A \geq j.$

In words: once sperm/embryo at ACP j are implanted, whether this results in a conception is as-good-as-random

- **The decision to undergo the procedure can be endogenous**
- $Y_1(1), Y_0(0), R$ and W can be related
- Main method relaxes to covariate-conditional version: age at procedure, technology

Simple World: Max 2 ACPs, All Reliers

$W = 1$
(willing to try once)

$Z_1 = 1$

$Z_1 = 0$

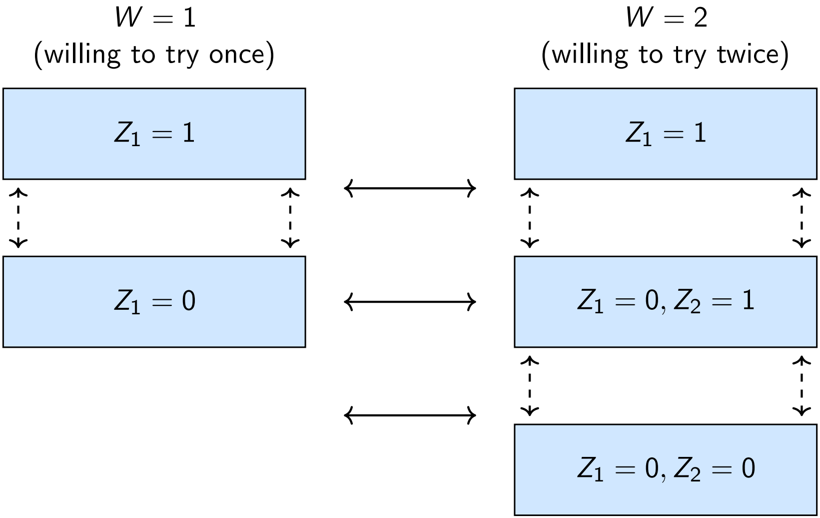
Simple World: Max 2 ACPs, All Reliers

$W = 1$ (willing to try once)	$W = 2$ (willing to try twice)
$Z_1 = 1$	$Z_1 = 1$
$Z_1 = 0$	$Z_1 = 0, Z_2 = 1$
	$Z_1 = 0, Z_2 = 0$

Simple World: Max 2 ACPs, All Reliers

$W = 1$ (willing to try once)		$W = 2$ (willing to try twice)
$Z_1 = 1$	$\longleftrightarrow$	$Z_1 = 1$
$Z_1 = 0$	$\longleftrightarrow$	$Z_1 = 0, Z_2 = 1$
	$\longleftrightarrow$	$Z_1 = 0, Z_2 = 0$

Simple World: Max 2 ACPs, All Reliers



Simple World: Max 2 ACPs, All Reliers

$W = 1$ (willing to try once)		$W = 2$ (willing to try twice)
$Z_1 = 1$	$\longleftrightarrow$	$Z_1 = 1$
$\mathbb{E}[Y_1(1) \vert W = 1]$		$\mathbb{E}[Y_1(1) \vert W = 2]$
$Z_1 = 0$	$\longleftrightarrow$	$Z_1 = 0, Z_2 = 1$
$\mathbb{E}[Y_0(0) \vert W = 1]$		$\mathbb{E}[Y_0(1) \vert W = 2]$
	$\longleftrightarrow$	$Z_1 = 0, Z_2 = 0$
		$\mathbb{E}[Y_0(0) \vert W = 2]$

Simple World (Observed): Max 2 ACPs, All Reliers

$W = 1$ (willing to try once)	$W = 2$ (willing to try twice)
$Z_1 = 1$	$Z_1 = 1$
$Z_1 = 0$	$Z_1 = 0, Z_2 = 1$
	$Z_1 = 0, Z_2 = 0$

Simple World (Observed): Max 2 ACPs, All Reliers

$W = 1$ (willing to try once)	$W = 2$ (willing to try twice)
$Z_1 = 1$	
$\mathbb{E}[Y_1(1)]$	
$Z_1 = 0$	$Z_1 = 0, Z_2 = 1$
$\mathbb{E}[Y_0(0) \mid W = 1]$	$\mathbb{E}[Y_0(1) \mid W = 2]$
	$Z_1 = 0, Z_2 = 0$
	$\mathbb{E}[Y_0(0) \mid W = 2]$

Simple World (Observed): Max 2 ACPs, All Reliers

$W = 1$ (willing to try once)	$W = 2$ (willing to try twice)

$Z_1 = 1$

$\mathbb{E}[Y_1(1)]$

$W = 1$	$W = 2$
$Z_1 = 0$	$Z_1 = 0, Z_2 = 1$
$\mathbb{E}[Y_0(0) \mid W = 1]$	$\mathbb{E}[Y_0(1) \mid W = 2]$
	$Z_1 = 0, Z_2 = 0$
	$\mathbb{E}[Y_0(0) \mid W = 2]$

Simple World (Observed): Max 2 ACPs, All Reliers

$W = 1$
(willing to try once)

$W = 2$
(willing to try twice)

$Z_1 = 1$

$\mathbb{E}[Y_1(1)]$

$Z_1 = 0$

$\mathbb{E}[Y_0(0)|W = 1]$

$Z_1 = 0, Z_2 = 1$

$\mathbb{E}[Y_0(1)|W = 2]$

$Z_1 = 0, Z_2 = 0$

$\mathbb{E}[Y_0(0)|W = 2]$

$Pr(W = 1) =$

Simple World: Max 1 ACP with Non-reliers

$R = 1$ (no child if fail)	$R = 0$ (child if fail)

Simple World: Max 1 ACP with Non-reliers

$R = 1$
(no child if fail)

$R = 0$
(child if fail)

$$Z_1 = 1$$

Simple World: Max 1 ACP with Non-reliers

$R = 1$
(no child if fail)

$R = 0$
(child if fail)

$Z_1 = 1$

$Z_1 = 0, D = 0$

$Z_1 = 0, D = 1$

Simple World: Max 1 ACP with Non-reliers

$R = 1$
(no child if fail)

$R = 0$
(child if fail)

$Z_1 = 1$

Distribution of $Y_1(1)$

$Z_1 = 0, D = 0$

$\mathbb{E}[Y_0(0)|R = 1]$

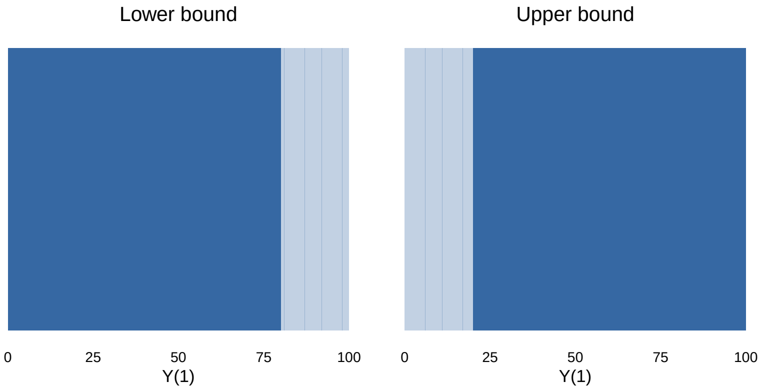
$Z_1 = 0, D = 1$

$\mathbb{E}[Y_0(1)|R = 0]$

$Pr(R = 1) =$

Intuition: Motherhood Outcome $Y_1(1)$

1. Treated group is a representative sample but their types are unobserved
2. Identify $\Pr(R = 1) = 0.8$ on control group
3. Assume most extreme distributions of types in treated group
4. Bound $\mathbb{E}[Y_1(1)|R = 1]$



Background and Data

Assisted conception procedures →

- In-vitro fertilization: invasive medical procedure, first 3 free
- **Intrauterine insemination (new)**: direct sperm injection, minimally invasive, free

Dutch family policies and labor market similar to OECD average →

- 16 weeks maternity + pregnancy leave, 1 week paternity leave

Data combining ACP medical records with tax records (new) →

- Work hours and income include leave; results for hours corrected for uncertainty
- 15,523 cohabiting opposite-sex couples
- Balance: ACP success at each attempt uncorr. with past outcomes cond. on age

Details | Balance in 1st ACP first | Balance in later ACPs | Success and willingness | Rep. samp.

Bounds

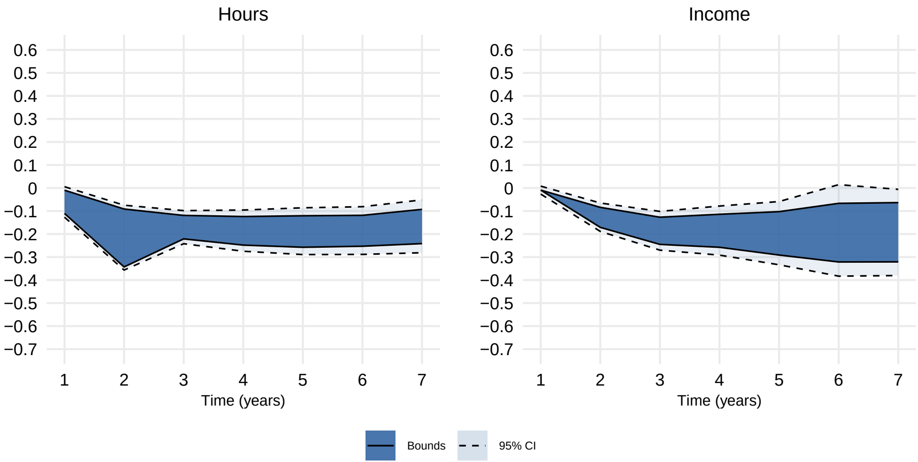


Figure 1: Bounds for Women Under Monotonicity

Bounds for Men

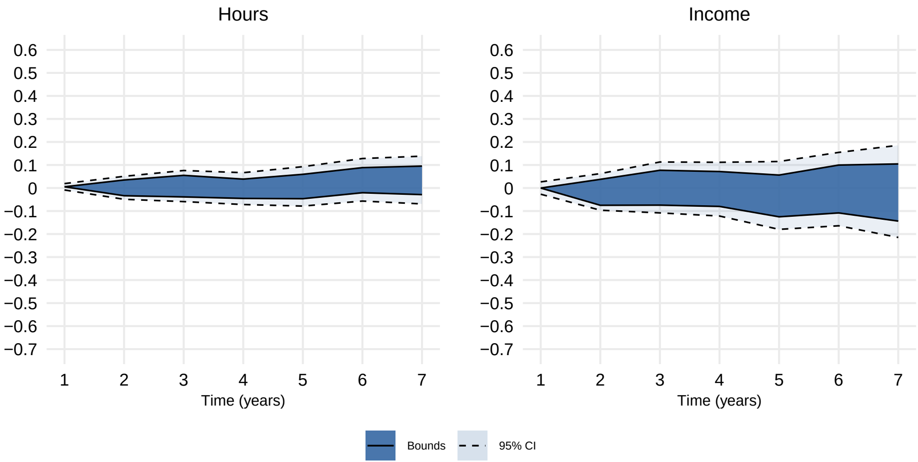


Figure 2: Bounds fo Men

Gender Inequality

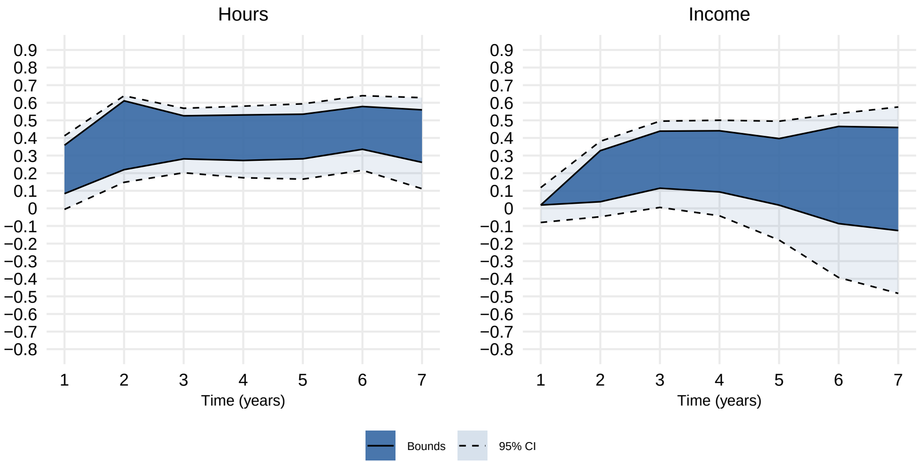


Figure 3: Share of Gender Inequality Caused by Parenthood (Effect on Gender Gap relative to Gap Under Parenthood)

Extensions

- Comparing methods Naive comp. Reasons for difference Less naive comp.
- **Bias in event study** Pop. ES vs bounds for ineq. Formal procedure ES bias estimates
- Bias in IV Formal procedure IV bias estimates
- Mental health side effects Discussion Bounds for non-depresses
- Relation to methodological literature Theoretical comparison Results
- Confidence inveral comparison Confidence intervals
- Inequality correcting for age De-aging partners
- Stable complier group Childless final period
- Estimator without DML Estimates
- Monotonicity Discussion Direction Partnered only Partnership and depression Test
- Testing Bensnes et al. (2023); Gallen et al. (2023) Estimates
- Heterogeneity Willingness to try
- Population imputation ES pop. Mother. imp. Childless imp. Effect imp.

Conclusion

Method for evaluating treatment effects under dynamic non-compliance:

- Assignment to job training and educational programs, legal settings with assignment to varying leniency "judges", promotion tournaments, clinical trials in extension phase
- Also works with two-sided non-compliance and without sequential randomization

Application to estimate the career cost of parenthood in the Netherlands:

- First evidence robust to selective fertility and timing-dependent effects
- Motherhood reduces work hours and income by 9% to 24%
- Parenthood causes up to 50% of post-child gender inequality

I show that accounting for bias is key to reconciling key conflicting findings in the literature:

- IV might understate the role of parenthood in gender ineq., ES overstates it
- Factors other than bias play a minimal role in the difference

External relevance:

- Alternative methods give very similar results on ACP sample and representative sample
- My method can be used to validate alternatives (e.g. imputation or structural)

Policy: large share of gender inequality may not be due to parenthood per se

Why IV and ES Estimates May Differ Within the Same Sample

1. Difference in weights
   - IV: local average treatment effect
   - ES: closer to the average treatment effect on the treated
2. Different treated outcomes
   - IV: motherhood at first attempt
   - ES: motherhood at any point
3. Difference in control outcomes
   - IV: trying and failing (mental health and relationship side effects)
   - ES: potentially not trying yet
4. Bias due to dynamic effects or selective timing
   - IV: biased under dynamic effects
   - ES: biased under selective timing

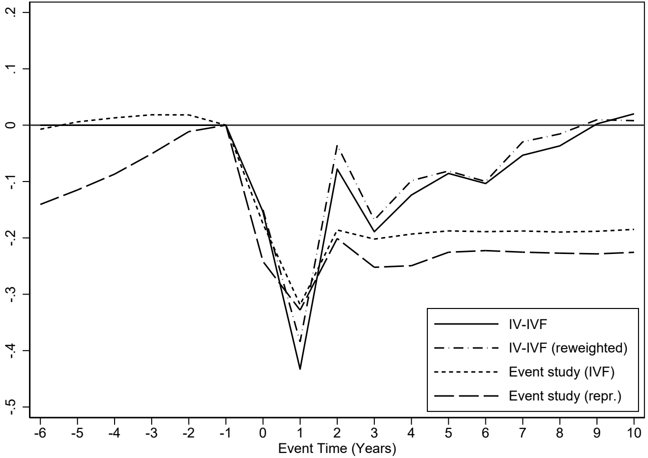
I demonstrate that:

1. No explanatory power
2. Very limited explanatory power
3. Sufficient to explain the difference

Back (extensions) Conclusion

ToC

Instrumental Variable vs Event Study: Percent Reduction in Earnings



Source: (Lundborg et al., 2024)

- “Naive” comparison with differing sub-populations and treatment definitions

Back (literature) Back (extensions)

ToC

Theoretical Bias

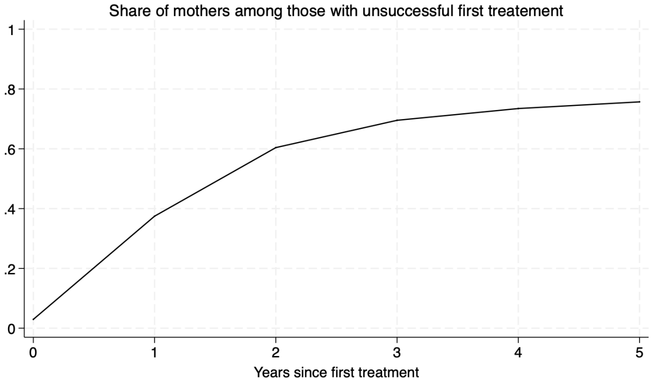


Figure 4: Motherhood among unsuccessfully treated

$$\tau_{RF} = 0.25\tau_{Parenthood} - 0.75\tau_{Delay}$$

$$\tau_{IV} = \tau_{Parenthood} - 3\tau_{Delay}$$

ToC Back (for illustration restricting heterogeneity between individuals)

Bounding τ_{ATR}

Construct the moment:

$$m^L(G,\eta^0) = Y1_{\{Y<q(r(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D)\prod_{j=1}^{A}\frac{(1-Z_j)}{(1-e_j(X_j))}$$

- G is the observed data vector
- η^0 contains the following:
  - $e_j(X_j) = \Pr(Z_j = 1|X_j)$
  - $q(r(X_1), X_1)$ is the $r(X_1)$-th quantile of Y given X_1 and $Z_1 = 1$
  - $r(X_1)$ identifies the covariate-conditional relier share

Assumption (Conditional Sequential Unconfoundedness)

$(Y(k), R, W) \perp\!\!\!\perp Z_j \mid X_j$ *for all* j, k*, and* $X_j, A \geq j$.

Theorem (Lower Bound)

Under conditional sequential unconfoundedness and regularity, the sharp lower bound on τ_{ATR} *is* $\mathbb{E}[m^L(G,\eta^0)]/\mathbb{E}[r(X_1)]$.

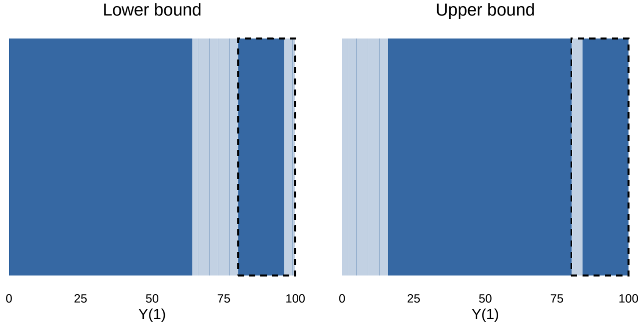
Back

ToC

Intuition: Motherhood Outcome $Y_1(1)$—Covariates

Pre-ACP covariates can help narrow the bounds:

- Can identify relier share at each covariate value
- Baseline bounds assume extreme scenarios where reliers have highest or lowest treated outcomes
- These distributions of treated outcomes might be inconsistent with conditional relier shares



Estimating the Bounds

Distribution of $m^L(G, \eta^0)$ is complicated by $q(r(X_1), X_1)$

- Semenova (2023) addresses a closely related inference challenge
- Double/debiased machine learning approach
  1. Adjust $m^L(G, \eta^0)$ to make it insensitive to small error in $q(r(X_1), X_1)$
  2. Sample splitting
- Asymptotic inference as if $q(r(X_1), X_1)$ was known

New moment:

$$\psi^L(G, \xi^0) = m^L(G, \eta^0) + corr(G, \xi^0)$$

Identifies same parameter:

$$\mathbb{E}[\psi^L(G, \xi^0)] = \mathbb{E}[m^L(G, \eta^0)]$$

Insensitive to estimation error in $q(r(X_1), X_1)$:

$$\partial_{q(.)} \mathbb{E}[\psi^{L+}(G, \xi_r)|X_1]|_{\xi_r = \xi_r^0} = 0 \ a.s.$$

Assisted Conception Procedures

- IUI (main procedure): sperm injected into uterus
  - Minimally invasive, primary ACP in most countries
  - "Free" in NL
- IVF (secondary procedure): embryo inserted into uterus
  - Invasive treatment, performed under sedation/anesthesia
  - Eggs retrieved through the vaginal wall using a specialized needle
  - In NL, first 3 free; each subsequent costs between 1000 and 4000 EUR

Back

Institutions

- Dutch family friendly polices similar to OECD average
  - 16 weeks of fully paid pregnancy+maternity leave
  - 1 week of paternity leave
  - Average time in child care similar to OECD average
  - Net child care cost 10% median household income
- Dutch employment intensity similar to OECD average
  - Employment among parents and non-parents relatively high
  - Part time work much more common
  - Approximately 15% two-parent families have both partners working part-time

Back

Data

Administrative data from Statistics Netherlands

- Comprehensive hospital records cover fertility treatments from 2012 to 2017: procedure date and type
  - Success imputed as having child born within 10 months
- Tax records cover work hours and income from 2011 to 2023
  - Include maternity leave and pay
  - Main bounds account for uncertainty around actual work hours
- Birth dates, legal family connections, cohabitation
- Dispensed medication registry

Main sample: cohabiting opposite-sex couples undergoing IUI for their first child between 2013 and 2016: 15,523

Back

Overview of Descriptives

- First and subsequent ACP success uncorrelated with past labor market outcomes condiditional on age Table first Table later
  - Support for independence of Z_j and $(Y_1(1), Y_0(0))$
- Success probability stable across ACPs conditional on age Figure
  - Support for independence of Z_j and W
- Representative sample worked less and had lower income before parenthood, but differences relatively small Table
  - ACP sample older before parenthood

Back (summary)

Balance in 1st ACP

Table 1: First ACP Outcomes and Descriptives

	Success (1)	Fail (2)	Difference (1)-(2)	Dif. cond. age & educ. (1)-(2) cond.
Work (W)	0.882	0.863	0.019	0.008
	[0.323]	[0.344]	(0.009)	(0.009)
Work (P)	0.884	0.865	0.019	0.013
	[0.320]	[0.342]	(0.009)	(0.009)
Hours (W)	1240.315	1207.860	32.455	18.702
	[604.666]	[635.194]	(16.183)	(16.560)
Hours (P)	1474.530	1438.590	35.940	18.579
	[658.231]	[695.692]	(17.713)	(17.870)
Income 1000s € (W)	28.065	27.418	0.647	0.745
	[19.559]	[20.219]	(0.516)	(0.546)
Income 1000s € (P)	37.205	36.952	0.252	0.364
	[26.482]	[29.452]	(0.746)	(0.730)
Bachelor deg. (W)	0.480	0.451	0.029	
	[0.500]	[0.498]	(0.013)	
Bachelor deg. (P)	0.394	0.381	0.013	
	[0.489]	[0.486]	(0.012)	
Age (W)	31.638	32.388	-0.750	
	[4.015]	[4.383]	(0.111)	
Age (P)	34.675	35.461	-0.786	
	[5.513]	[5.996]	(0.152)	
Observations	1,714	13,809		
Joint *p*-val.			0.000	0.928

Note: Labor market outcomes measured year before first ACP. (W) - woman, (P) - partner. Last column uses inverse prbability weights for the first ACP that follow the main specificaition. Standard deviations in brackets. Standard errors in parentheses.

Balance in Subsequent ACPs

Table 2: Balance in Later ACPs

	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
Work (W)	0.009	-0.004	0.022	0.014	0.039	-0.003	-0.011	0.022	0.030
	(0.010)	(0.011)	(0.011)	(0.012)	(0.012)	(0.017)	(0.018)	(0.019)	(0.024)
Work (P)	0.006	0.016	0.012	0.020	-0.004	-0.004	-0.019	0.017	0.030
	(0.010)	(0.010)	(0.012)	(0.012)	(0.015)	(0.015)	(0.019)	(0.020)	(0.027)
Hours (W)	32.885	-4.482	52.999	41.332	81.957	11.894	-18.836	72.659	24.819
	(18.721)	(20.032)	(21.045)	(22.686)	(25.131)	(31.187)	(32.937)	(38.210)	(48.490)
Hours (P)	21.655	24.730	23.756	38.965	9.666	-6.580	-28.458	30.525	43.722
	(21.018)	(21.089)	(23.574)	(25.255)	(30.585)	(31.513)	(37.976)	(44.856)	(52.821)
Income 1000s € (W)	1.481	-0.015	1.685	1.802	2.086	0.150	-0.043	0.866	-0.444
	(0.615)	(0.624)	(0.767)	(0.830)	(0.913)	(1.000)	(1.092)	(1.234)	(1.629)
Income 1000s € (P)	-0.749	1.002	2.040	0.800	0.774	0.025	0.259	-0.324	0.149
	(0.835)	(0.912)	(1.066)	(1.115)	(1.424)	(1.424)	(1.563)	(1.737)	(2.203)
Observations	12,974	10,774	8,726	6,977	5,411	3,944	2,723	1,850	1,174
Joint *p*-val.	0.175	0.976	0.234	0.303	0.140	1.000	0.956	0.704	0.917

Note: Each column describes the difference in average characteristics between women for whom the respective ACP succeeds and those for whom it fails, among those who undergo the procedure, using inverse probability weights for each ACP following the main specification. Labor market outcomes and age measured year before first treatment. (W) - woman, (P) - partner. Standard errors in parentheses.

ACP histories Back (summary) Back (detailed descriptives)

ToC

Estimated Success Probabilities

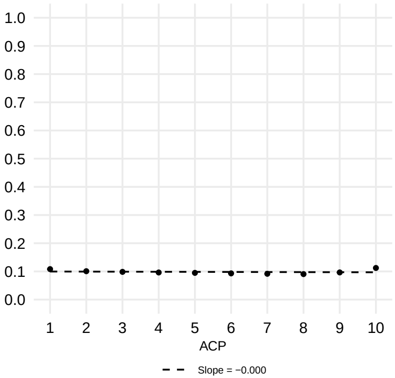


Figure 5: Estimated Success Probabilities

ACP histories Back (summary) Back (detailed descriptives)

ToC

Comparison to Representative Sample

Table 3: Full Sample, Reliers, and Representative Sample

	Success (1)	Fail (2)	Reliers (3)	Rep. (4)	Success vs rep. (1)-(4)	Rel. vs rep. (3)-(4)
Work (W)	0.882	0.863	0.820	0.801	0.080	0.019
	[0.323]	[0.344]	[0.333]	[0.399]	(0.010)	(0.005)
Work (P)	0.884	0.865	0.849	0.783	0.101	0.066
	[0.320]	[0.342]	[0.344]	[0.412]	(0.010)	(0.005)
Hours (W)	1240.315	1207.860	1117.711	1076.204	164.111	41.508
	[604.666]	[635.194]	[582.334]	[696.245]	(16.856)	(8.412)
Hours (P)	1474.530	1438.590	1390.699	1250.948	223.582	139.752
	[658.231]	[695.692]	[662.920]	[793.536]	(19.211)	(9.576)
Income 1000s € (W)	28.065	27.418	24.976	21.362	6.703	3.615
	[19.559]	[20.219]	[15.359]	[18.330]	(0.444)	(0.222)
Income 1000s € (P)	37.205	36.952	35.299	28.107	9.098	7.193
	[26.482]	[29.452]	[24.304]	[29.076]	(0.704)	(0.351)
Bachelor deg. (W)	0.480	0.451	0.398	0.411	0.069	-0.012
	[0.500]	[0.498]	[0.411]	[0.492]	(0.012)	(0.006)
Bachelor deg. (P)	0.394	0.381	0.329	0.345	0.049	-0.015
	[0.489]	[0.486]	[0.397]	[0.475]	(0.012)	(0.006)
Age (W)	31.638	32.388	33.480	28.713	2.926	4.767
	[4.015]	[4.383]	[3.897]	[4.658]	(0.113)	(0.056)
Age (P)	34.675	35.461	36.580	28.713	5.962	7.868
	[5.513]	[5.996]	[3.928]	[4.665]	(0.113)	(0.057)
Observations	1,714	13,809	4,882	376,152		

Note: Labor market outcomes measured year before first ACP for main sample and year and 9 months before birth of first child for the represenstative sample. Representative sample is selected to match the main sample by year of conception. Average relier outcomes are based on sample of women who remain childless 7 years after their first ACP with weights described under implementation. (W) - woman, (P) - partner. Standard deviations in brackets. Standard errors in parentheses.

ACP Histories

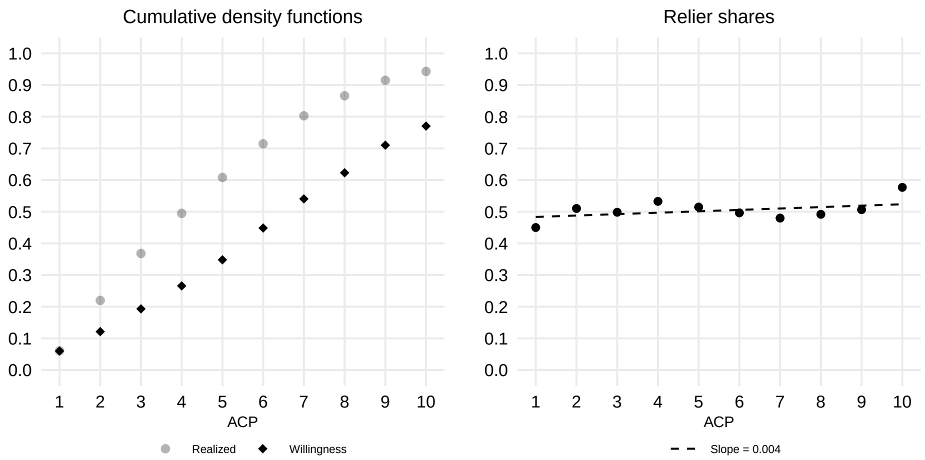


Figure 6: ACP Histories and Reliance

What are the counterfactuals?

Broadly:

- Do not want/plan children
- **Want/plan children**

Motherhood outcome:

- Get immediately
- Get naturally after few attempts
- **Get with medical assistance**

Childless outcome:

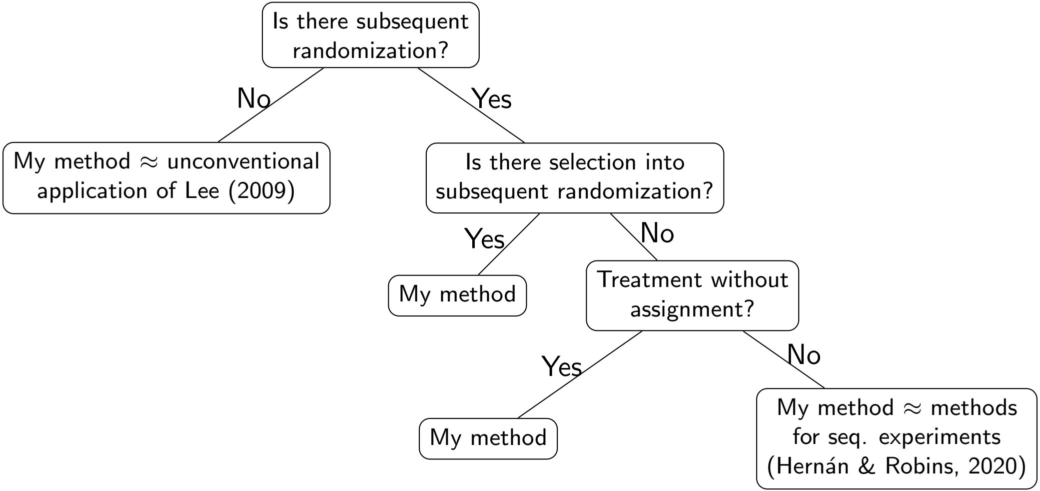
- Do not try
- Try and fail naturally
- **Try and fail with medical assistance (+ naturally?)**

Extrapolation requires carefully addressing mental health consequences of failure (and medical procedures)

Back (model) Back (extensions) Depr. effect Bounds non-depr. Arguments

ToC

Relation to Methodological Literature



Comparison with Lee (2009)

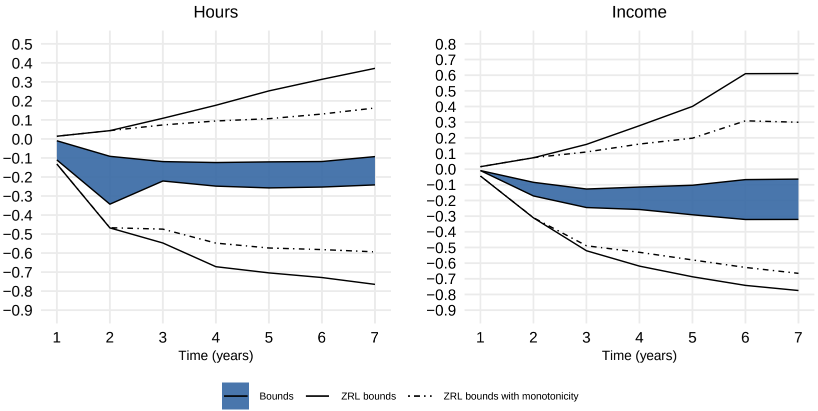


Figure 7: Comparison with Lee (2009) Bounds for Effects on Women

Back (extensions) Conclusion

ToC

Less Naive Comparison to Existing Methods

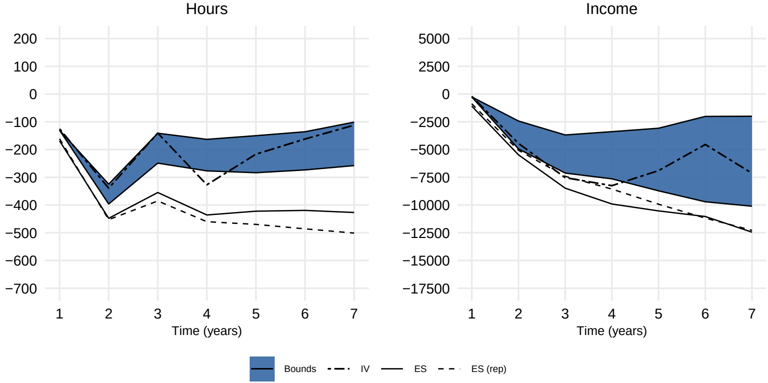


Figure 8: Estimates Using on Different Methods (Absolute Effects)

- Using women whose ACP succeeds for ES makes treatment definition consistent
- The three methods still target different sub-populations

Baseline Gender Inequality: Representative ES vs Bounds

- ES imputes childless trajectories using pre-parenthood outcomes of later mothers
- Comparing ES estimates from the representative sample to bounds for ACP sample:

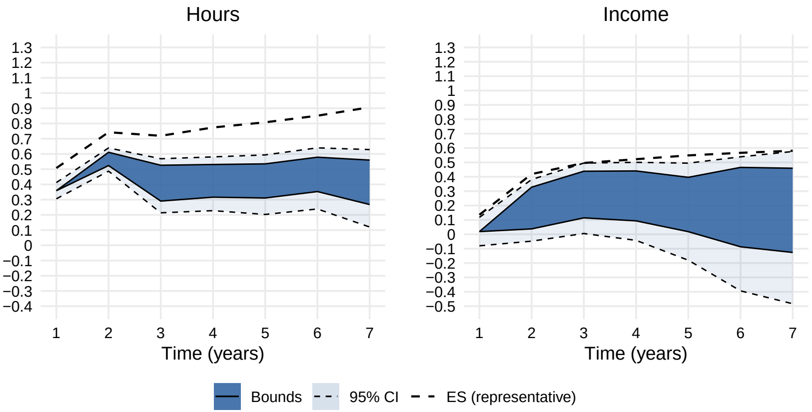


Figure 9: Share of Gender Inequality Attributed to Parenthood

Quantifying Bias in Event Study

- ES estimates may be biased due to selective timing
- Difference from bounds need not imply bias, even when using the same sample:
  - Different (sub)populations, average age at first birth, and control definitions
- Can the difference be explained by selective timing alone?
  - I aim to quantify selection specifically for reliers and incorporate it into the bounds

1. I proxy when women chose to have children using the timing of their first ACP
2. I compare relier childless trajectories identified using my method to those imputed from pre-ACP outcomes of women initiating ACP at older ages, à la event study

- Interpretation: average bias in event study estimates for reliers due to selective fertility timing at each moment of parenthood
- **Same population allows for comparison with bounds**

Back (extensions) Conclusion

ToC

Placebo Event Study

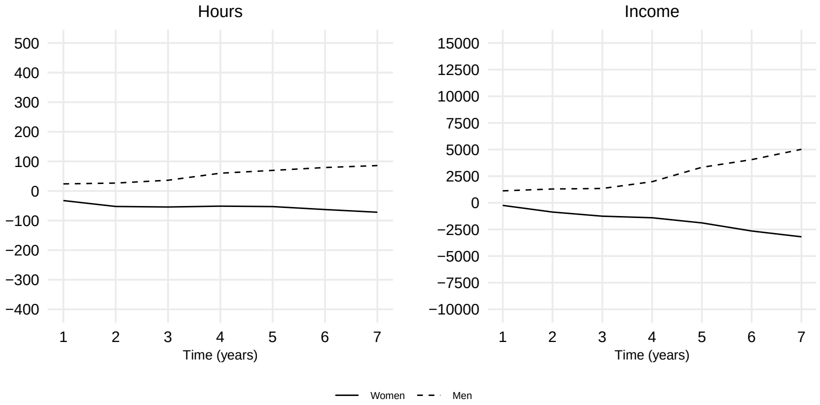


Figure 10: Placebo Event Study (Absolute)

- Negative selection of early mothers and positive selection of fathers

Back (extensions) Conclusion

ToC

Gender Inequality: Population ES vs Bounds with Selection

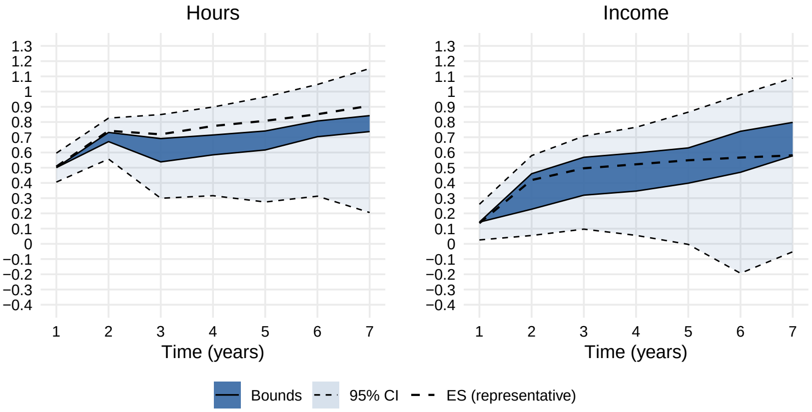


Figure 11: Share of Gender Inequality Explained by Selection and Parenthood

- Consistent with ES estimates attributing almost all gender inequality to parenthood

Back (extensions) Conclusion

ToC

Quantifying Bias in IV

Instrumental Variable (Lundborg et al., 2017):

- I have bounded τ_{ATR} allowing for dynamic effects
- I can point-identify τ_{ATR} assuming static effects à la instrumental variable
- Subtracting it from bounds on τ_{ATR} gives bounds on bias

Back (extensions) Conclusion

Effect of Delaying Motherhood

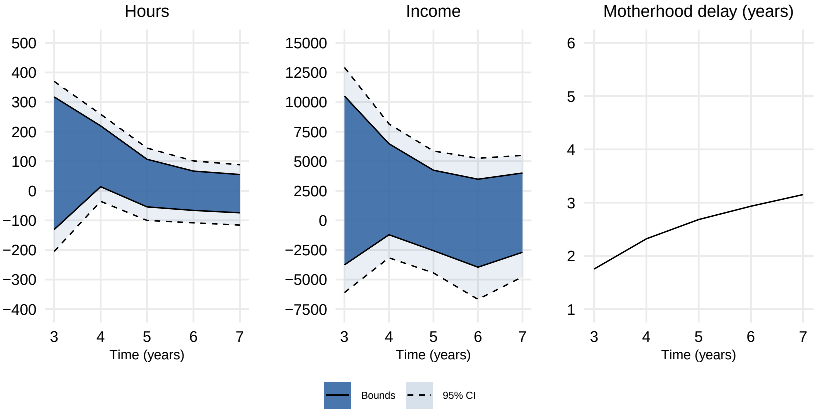


Figure 12: Effect of Delaying Relative to Motherhood at First Attempt (Absolute)

$$\tau_{IV} = \tau_{Parenthood} - 3\tau_{Delay}$$

ToC Back (extensions) Conclusion

Mental Health and ACPs

Mental health consequences associated with failure to conceive are a part of the story:

- Unmet fertility goals may negatively impact mental health, and in turn, labor market outcomes

There are, however, additional concerns:

- Mental health issues caused specifically by failed conception or ACPs (external)
  - Focusing on artificial insemination helps mitigate this
- Large impacts unique to ACP families (external)
- Worsened mental health by threatening monotonicity (internal)

In practice, these impacts are likely small (Lundborg et al., 2024)

Antidepressant uptake | Back (extensions) | Conclusion

ToC

Monotone Bounds for Non-depressed Childless Women

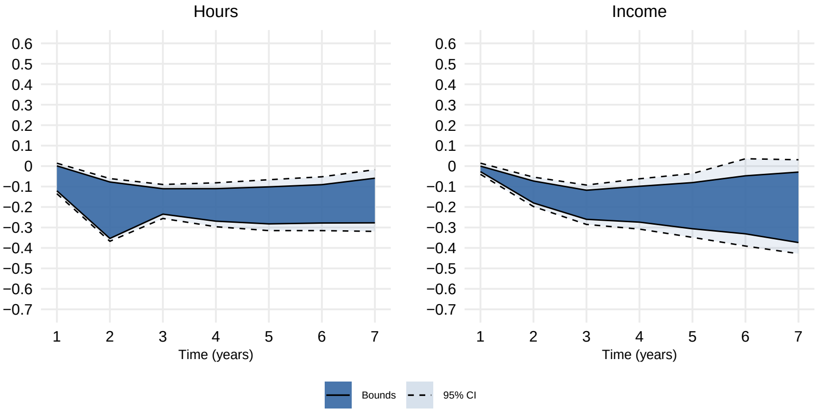


Figure 13: Monotone Bounds for Women Who Would Not Uptake Antidepressants if They Were to Remain Childless

Back (extensions) Conclusion

ToC

Confidence Intervals

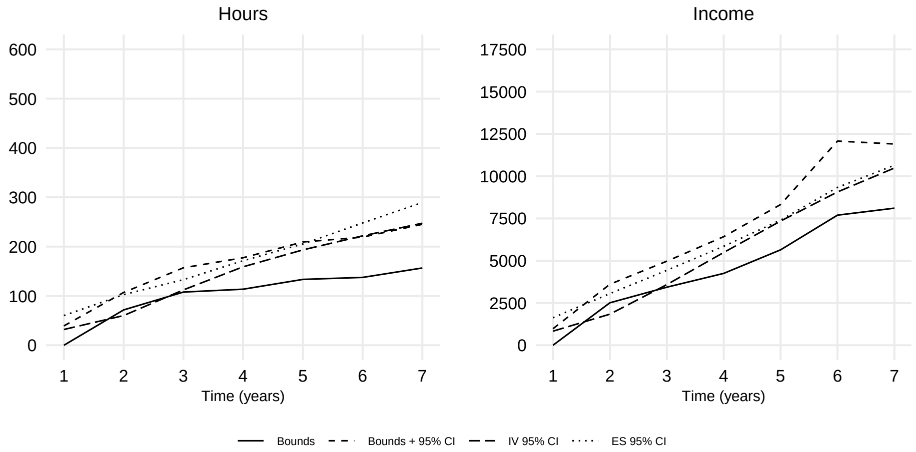


Figure 14: 95% CI for Different Methods (Absolute)

Monotonicity

Is monotonicity realistic?

- Yes, if families are determined to have at least one child
  - Decreasing marginal returns to children.
  - Stronger sufficient assumption: success cannot increase natural births
- No, if first treatment success increases the likelihood of attempting to conceive naturally
  - Couples may realize they are fertile and try more
  - First child may "save the relationship" resulting in more attempts to conceive

Back

Monotone Bounds: Women who Remain Childless

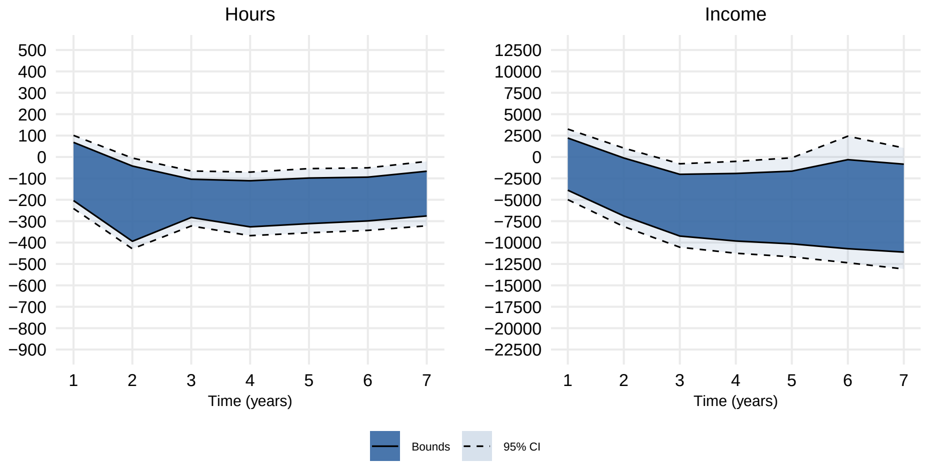
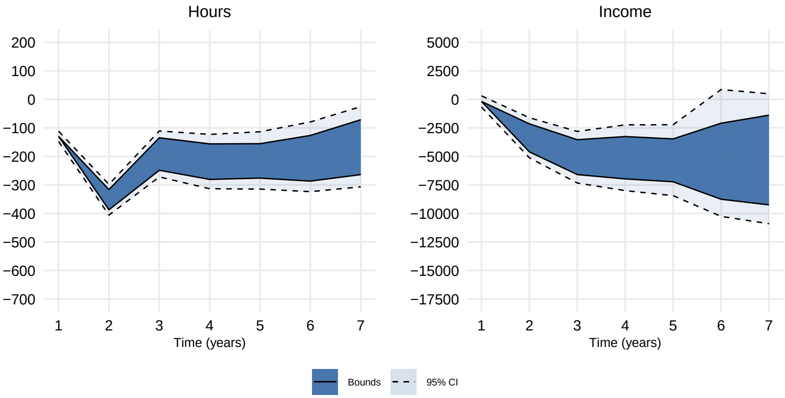


Figure 15: Monotone Bounds Using Completed Fertility (Absolute)

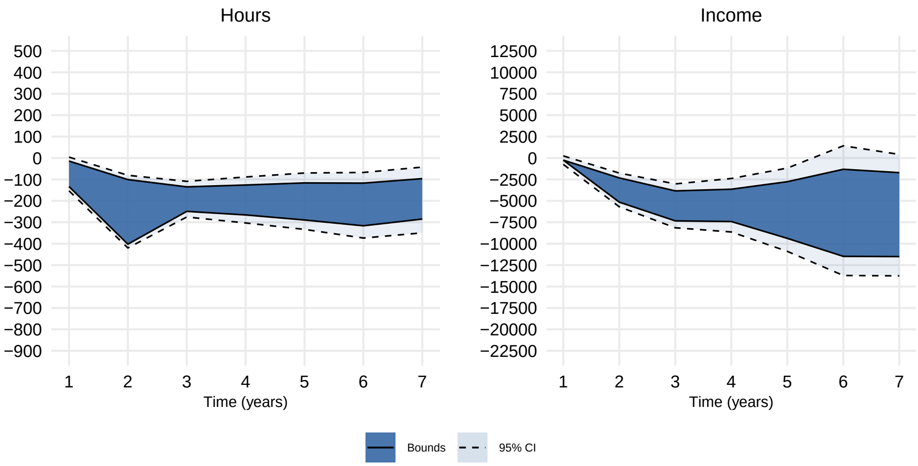
ToC Back

Simple estimator



- $\mathbb{E}[Y_1(1)|R=1] = \mathbb{E}[g(X_1)+\varepsilon|R=1]$
- $\mathbb{E}[g(X_1)|R=1]$ identified on chillness reliers using baseline method
- Only need to bound $\mathbb{E}[\varepsilon|R=1]$

Relaxing Monotonicity Direction



Back
ToC

Relaxing Monotonicity to Partnered Women

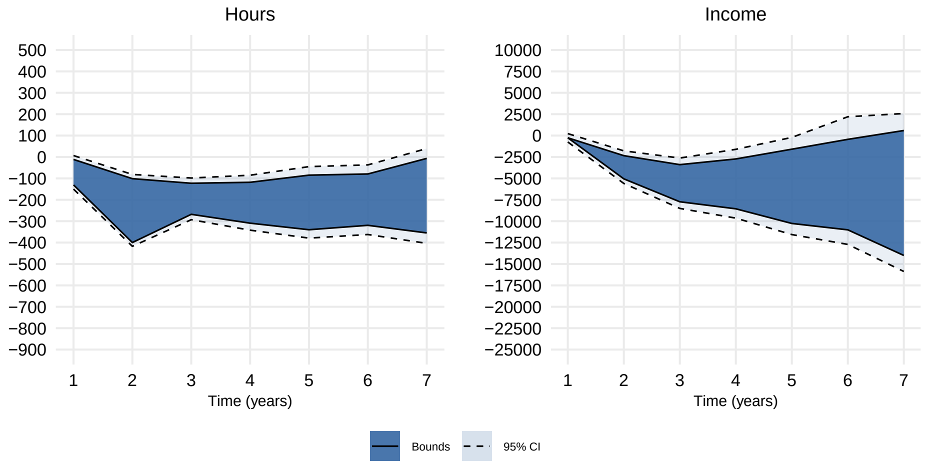


Figure 16: Monotone Bounds Using Women Who Stay Partnered

ToC Back (extensions)

Relaxing Monotonicity for Depression and Partnership

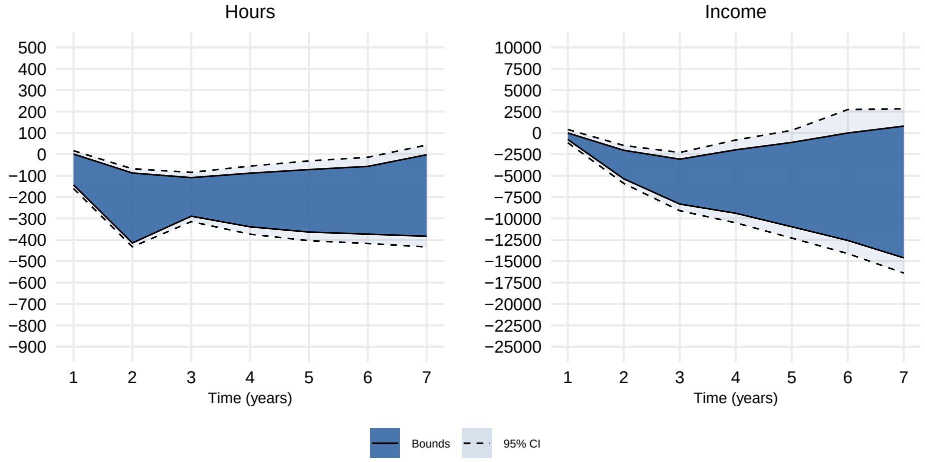
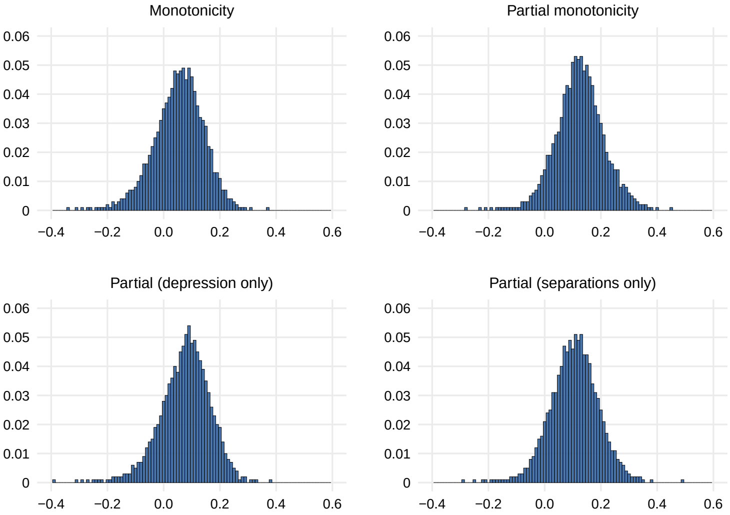


Figure 17: Monotone Bounds For Women Who Stay Partnered and Do Not Uptake Antidepress.

Testing Monotonicity



Heterogeneity by Willingness to Undergo Procedures

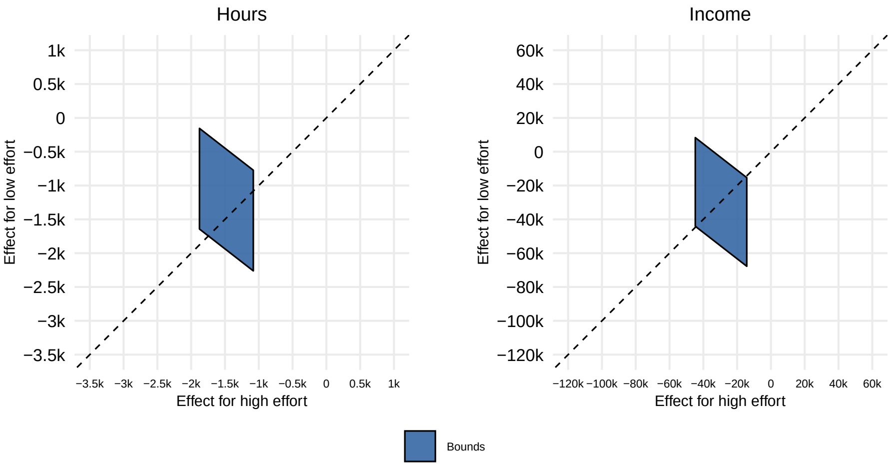


Figure 18: Cumulative Outcomes 6 Years After, G Above or Below 6

Monotone Bounds: Excluding Depression

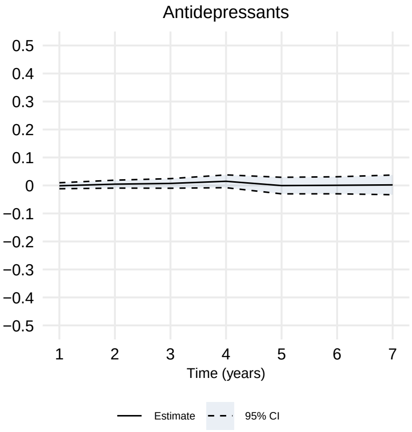


Figure 19: Effect on Antidepressant Take-Up

ToC
Back (extensions) Back (mental health)

Monotone Bounds: Correcting for Partner's age

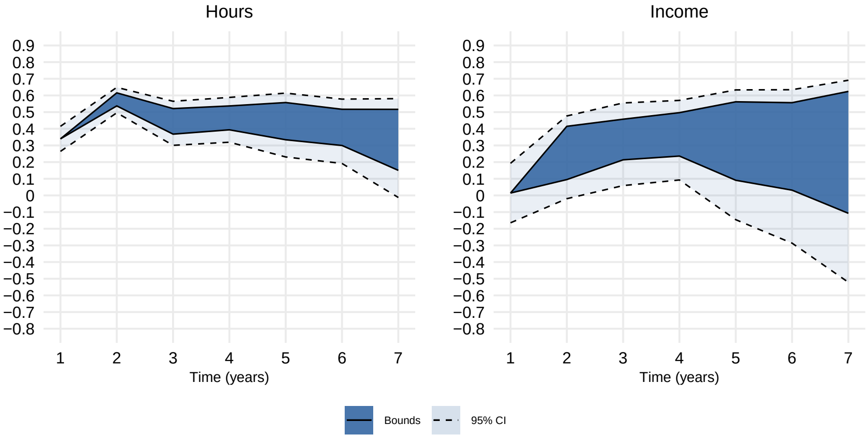


Figure 20: Monotone Bounds Using Male Income at Same Age as Female

ToC

Back

Testing the Plug-in Approach

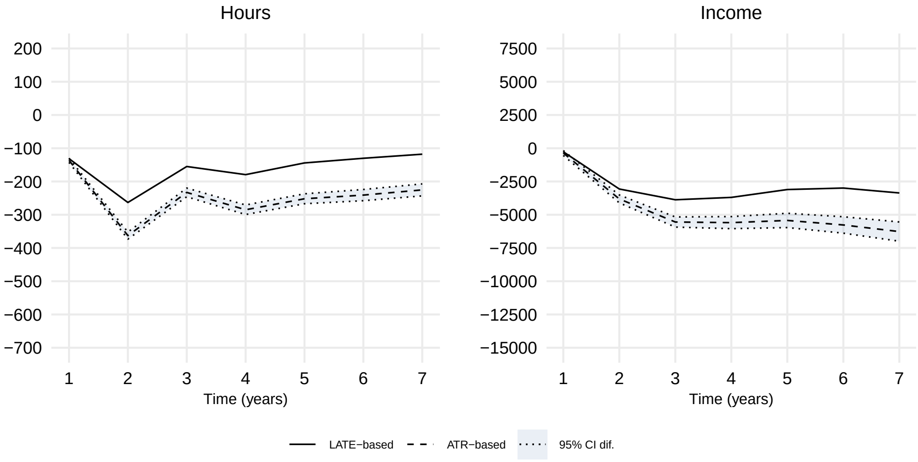


Figure 21: Plug-In Estimators Exploiting Different Numbers of Treatments

Event Study: Population vs IUI Sample

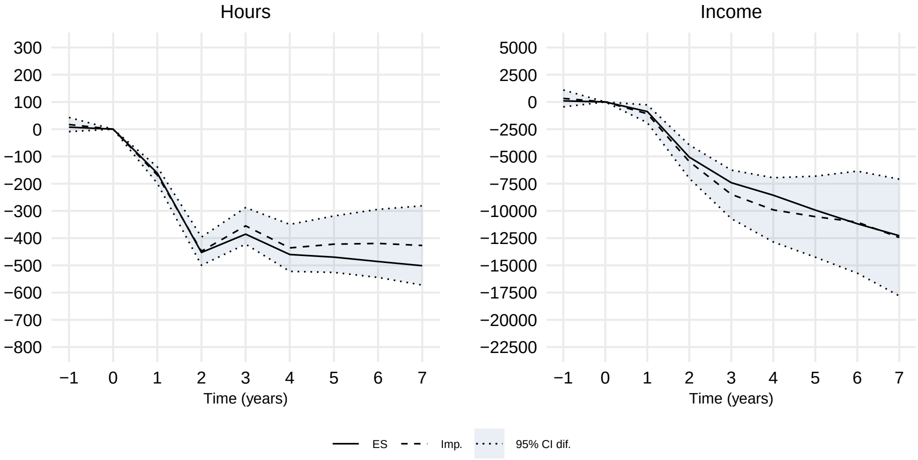


Figure 22: ES for Population and Women with First IUI Success

Imputing Population Motherhood Outcomes Using IUI Sample

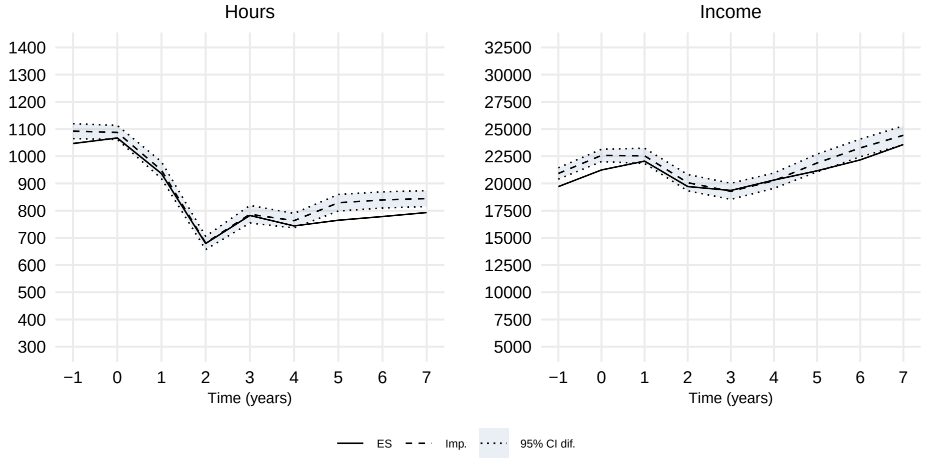


Figure 23: Population Outcomes vs IUI-Imputation (Age & Education)

Imputing Population Childless Outcomes Using IUI Sample

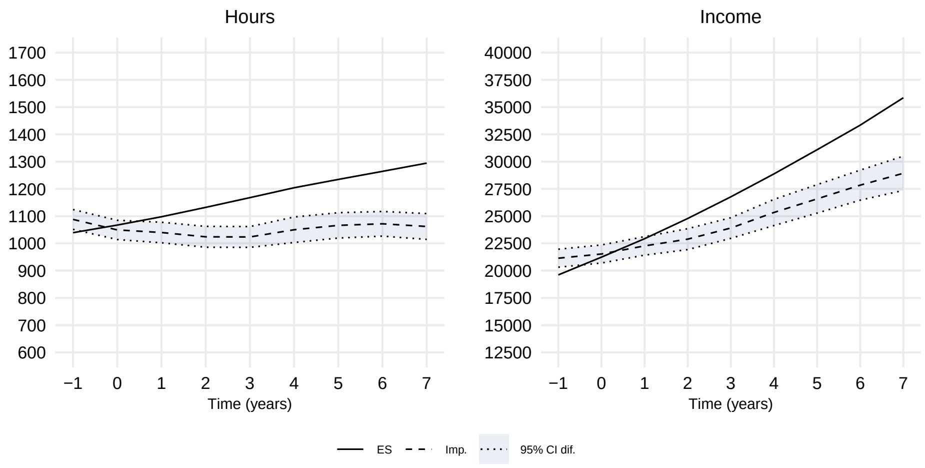


Figure 24: Population Outcomes vs IUI-Imputation (Age & Education)

Event Study vs IUI-imputation for Population

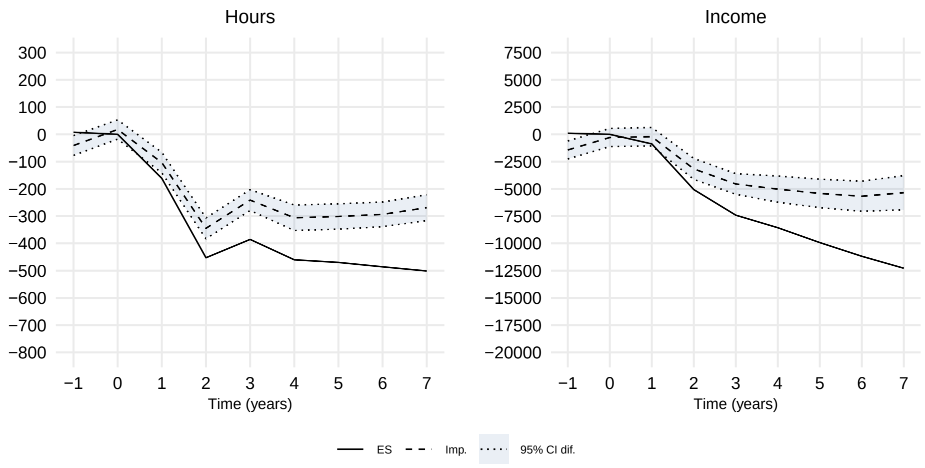


Figure 25: Event Study vs IUI-Imputation for Population (Age & Education)

References I

Agüero, J. M., & Marks, M. S. (2008). Motherhood and female labor force participation: evidence from infertility shocks. *American Economic Review*, *98*(2), 500–504.

Bensnes, S., Huitfeldt, I., & Leuven, E. (2023). Reconciling estimates of the long-term earnings effect of fertility. *IZA Discussion Paper*.

Cristia, J. P. (2008). The effect of a first child on female labor supply: Evidence from women seeking fertility services. *Journal of Human Resources*, *43*(3), 487–510.

Gallen, Y., Joensen, J. S., Johansen, E. R., & Veramendi, G. F. (2023). The labor market returns to delaying pregnancy. *Available at SSRN 4554407*.

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Boca Raton: Chapman & Hall/CRC.

Hotz, V. J., McElroy, S. W., & Sanders, S. G. (2005). Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *Journal of Human Resources*, *40*(3), 683–715.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209.

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

Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review*, *107*(6), 1611–37.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2024). Is there really a child penalty in the long run? new evidence from ivf treatments. *IZA Discussion Paper*.

Miller, A. R. (2011). The effects of motherhood timing on career path. *Journal of population economics*, *24*, 1071–1100.

Semenova, V. (2023). Generalized lee bounds. *arXiv preprint arXiv:2008.12720v3*.