

The Motherhood Gap in Retirement Income: A Descriptive Lifecycle Analysis

Afrouz Azadikhah Jahromi* Weige Huang[†]

February 12, 2026

Abstract

This study documents how income differences between mothers and childless women evolve across the lifecycle, with particular focus on retirement outcomes. Using data from the National Longitudinal Survey of Youth 1979 (NLSY79), the Health and Retirement Study (HRS), and the Current Population Survey (CPS) for women born 1957–1964, we trace income gaps from early career through retirement. We find substantial differences in pension income at ages 54–61: mothers have 32.8% lower pension income than childless women, driven primarily by lower pension coverage (extensive margin) rather than lower amounts conditional on receipt. However, we emphasize that these are *descriptive* gaps that reflect both the direct effects of childbearing and selection into motherhood. We assess sensitivity to selection on unobservables using Oster (2019) bounds and find that our estimates are moderately robust ($\delta = 1.8$). Mid-career estimates from the CPS should be interpreted with caution due to measurement limitations: the CPS identifies mothers by co-resident children, which misclassifies some mothers as childless and may bias gap estimates upward. Our findings highlight the importance of distinguishing individual from household income and decomposing retirement gaps into extensive and intensive margins.

JEL Classification: J13, J16, J26, J31

*School of Business Administration, Widener University. Email: ajahromi@widener.edu

[†]Corresponding author.

Keywords: Motherhood gap, retirement income, pension coverage, lifecycle analysis, fertility, selection

1 Introduction

Women with children earn less than childless women on average—a pattern often termed the “motherhood penalty.” A substantial literature documents wage gaps of 5–15% per child during working years (Waldfogel 1997; Budig and England 2001; Anderson, Binder, and Krause 2003). However, relatively little is known about how these gaps evolve into retirement, when cumulative career differences manifest in pension income and retirement wealth.

This paper traces income differences between mothers and childless women from early career through retirement using three complementary data sources. We make three contributions. First, we document the lifecycle trajectory of the motherhood-income gap, showing that it persists—and by some measures amplifies—into retirement. Second, we decompose retirement income gaps into extensive margins (whether women have pension income at all) and intensive margins (amounts conditional on receipt), revealing that coverage differences drive much of the gap. Third, we assess the sensitivity of our estimates to selection on unobservables and highlight important measurement limitations that affect interpretation.

We emphasize at the outset that our estimates are *descriptive*. Fertility is endogenous: women who have children differ systematically from those who do not in ways that independently affect earnings. Our comparisons therefore reflect both any direct effects of childbearing on careers and selection into motherhood based on unobserved characteristics. We use the term “gap” rather than “penalty” to signal this interpretive caution, though we occasionally use “penalty” when discussing prior literature that employs causal language.

Our main findings are as follows. Using NLSY79 data with proper fertility measurement (children ever born), we find that mothers aged 54–61 have pension income that is 32.8% lower than childless women. However, this gap decomposes into a large extensive margin difference—mothers are 8.4 percentage points less likely to receive any pension income—and a smaller intensive margin difference of 18.2% conditional on receipt. IRA savings show minimal differences (1.3%), possibly reflecting spousal contributions and catch-up savings.

Mid-career estimates from the CPS suggest the gap peaks around ages 35–40, but these estimates rely on a problematic fertility measure: whether “own children” are present in the

household. This systematically misclassifies some mothers as childless, likely biasing gap estimates upward. We present these estimates with appropriate caveats.

Sensitivity analysis using Oster (2019) bounds suggests that selection on unobservables would need to be approximately 1.8 times as important as selection on observables to fully explain away the pension income gap. This provides moderate—but not strong—evidence against pure selection explanations.

2 Related Literature

The motherhood wage gap literature finds penalties of 5–15% per child (Waldfogel 1997; Budig and England 2001). These gaps have been attributed to reduced human capital accumulation, employer discrimination, and differential work effort (Anderson, Binder, and Krause 2003). Kleven, Landais, and Søgaaard (2019) document substantial heterogeneity across the income distribution and over time.

Identification remains challenging. Fertility is correlated with unobservables that independently affect earnings—ambition, career orientation, health, family background, and partner characteristics. Studies have attempted various identification strategies: sibling fixed effects, twin births as instruments, and timing-of-birth variation. Each has limitations, and no consensus exists on the magnitude of the causal effect.

Less attention has focused on retirement outcomes. Even and Macpherson (1994) examined gender differences in pension coverage, finding that women’s lower coverage rates partly reflect interrupted careers. Johnson and Uccello (2004) documented lower retirement wealth among women generally, without isolating motherhood effects.

Our contribution is descriptive rather than causal: we document how the motherhood gap evolves from early career through retirement using consistent birth cohorts across multiple datasets, while being explicit about the selection challenges that preclude causal interpretation.

3 Data

We use three datasets to construct a lifecycle picture of the motherhood gap for women born 1957–1964.

3.1 NLSY79

The National Longitudinal Survey of Youth 1979 follows 6,283 women from ages 14–22 in 1979 through 2018, when they are ages 54–61. Key advantages include:

- **Fertility measurement:** NUMKID measures total children ever born, the gold standard that avoids misclassification based on co-residence.
- **Pre-birth characteristics:** AFQT scores, family background, and early earnings trajectories enable assessment of selection.
- **Individual retirement income:** Pension receipt and amounts are measured at the individual level, not confounded with spouse income.

In our sample, 78.7% of women are mothers (N=4,946) and 21.3% are childless (N=1,337).

3.2 CPS ASEC

The Current Population Survey Annual Social and Economic Supplement provides large cross-sectional samples for years 1990–2014. We use individual total income (INCTOT) for the 1957–1964 birth cohort at ages 35–50, filling a gap where NLSY79 income data and HRS coverage are limited.

Critical limitation: The CPS identifies mothers by whether “own children under 18” are present in the household. This measure systematically misclassifies as childless: (a) mothers whose children live with another parent post-divorce, (b) mothers whose children are in foster care or with relatives, and (c) mothers whose children have died. These misclassified mothers likely have lower earnings on average, inflating the apparent gap. At ages 45–50, children leaving for college creates additional misclassification of higher-earning mothers as childless, potentially attenuating gap estimates.

We present CPS-based estimates with these caveats clearly noted. The CPS estimates should be viewed as suggestive rather than definitive.

3.3 RAND HRS

The Health and Retirement Study tracks individuals aged 50+ from 1992–2022. For the 1957–1964 birth cohort, we observe women at ages 58–67 by 2022.

Critical limitation: HRS measures *household* income (H#ITOT), which includes spouse earnings. Since mothers are more likely to be married, two-earner households appear in the “mother” category, potentially masking individual-level gaps. We use HRS primarily to illustrate this household-versus-individual measurement issue rather than as a source of gap estimates.

3.4 Sample Construction

Table 1 documents our sample construction. Sample sizes vary across analyses due to different variable availability and sample restrictions.

Table 1: Sample Construction

| Analysis | Source | N Mothers | N Childless |
|--------------------------------|--------|-----------|-------------|
| Early career (ages 20–35) | NLSY79 | 47,214 | 10,835 |
| Mid-career (ages 35–50) | CPS | 128,212 | 42,221 |
| Retirement income (ages 54–61) | NLSY79 | 4,946 | 1,337 |
| With pension income | NLSY79 | 624 | 182 |
| With IRA data | NLSY79 | 3,891 | 1,104 |

4 Empirical Approach

4.1 Descriptive Comparisons

Our primary analysis computes mean differences in log income between mothers and childless women within age groups:

$$\text{Gap}_a = \bar{Y}_{a,\text{childless}} - \bar{Y}_{a,\text{mother}} \quad (1)$$

where \bar{Y} denotes mean log income in age group a . We express gaps in percentage terms. A positive gap indicates mothers earn less.

This is a descriptive comparison, not a causal estimate. We do not claim to identify the effect of motherhood on earnings, only to document patterns.

4.2 Controlling for Observables

To assess how much of the raw gap is attributable to observable differences, we estimate:

$$Y_i = \alpha + \beta \cdot \text{Mother}_i + X_i' \gamma + \varepsilon_i \quad (2)$$

where X_i includes education, race, AFQT score (in NLSY79), and age fixed effects. The coefficient β represents the gap conditional on observables.

4.3 Propensity Score Matching

For the NLSY79 retirement analysis, we implement propensity score matching on pre-birth characteristics to construct a more comparable control group. We estimate:

$$P(\text{Mother}_i = 1 | X_i) = \Lambda(X_i' \delta) \quad (3)$$

where X_i includes AFQT score, mother's education, family income at age 14, race, and—for women who eventually became mothers—earnings in years prior to first birth. We match mothers to childless women using nearest-neighbor matching with replacement and compute

average treatment effects on the treated (ATT).

We emphasize that propensity score matching addresses selection on *observables* only. The ATT estimate remains biased if unobserved characteristics differ between matched mothers and childless women.

4.4 Sensitivity to Selection on Unobservables

Following Oster (2019), we assess how much selection on unobservables would be needed to explain away our estimates. The key parameter is δ , the ratio of selection on unobservables to selection on observables. Under the assumption that the relationship between treatment and unobservables is proportional to the relationship between treatment and observables:

$$\delta = \frac{\beta^* - \beta^{\text{controlled}}}{\beta^{\text{raw}} - \beta^{\text{controlled}}} \cdot \frac{R_{\text{max}}^2 - R_{\text{controlled}}^2}{R_{\text{controlled}}^2 - R_{\text{raw}}^2} \quad (4)$$

where β^* is the bias-adjusted coefficient. We compute bounds assuming $R_{\text{max}}^2 = 1.3 \times R_{\text{controlled}}^2$, following Oster’s recommended procedure.

A value of $\delta > 1$ suggests the estimate is reasonably robust—unobservables would need to be more important than observables to fully explain the gap. Values of $\delta > 2$ indicate strong robustness.

4.5 Extensive vs. Intensive Margin Decomposition

For retirement income, we decompose the total gap into:

$$\text{Total Gap} = \underbrace{\text{P(Pension} > 0 | \text{Childless}) - \text{P(Pension} > 0 | \text{Mother})}_{\text{Extensive Margin}} \quad (5)$$

$$+ \underbrace{\mathbb{E}[\text{Pension} | \text{Pension} > 0, \text{Childless}] - \mathbb{E}[\text{Pension} | \text{Pension} > 0, \text{Mother}]}_{\text{Intensive Margin}} \quad (6)$$

This decomposition reveals whether mothers have less retirement income because they are less likely to have pensions at all (extensive) or because they have lower amounts conditional on receipt (intensive).

5 Results

5.1 Lifecycle Patterns: Descriptive Evidence

Table 2 presents raw gaps by age group across our three data sources.

Table 2: Motherhood Gap Across the Lifecycle

| Age Group | Source | Gap (%) | N Childless | Fertility Measure | Caveats |
|-----------|--------|---------|-------------|----------------------|----------|
| 20–25 | NLSY79 | +7.6 | 5,091 | Children ever born | — |
| 25–30 | NLSY79 | +4.2 | 3,229 | Children ever born | — |
| 30–35 | NLSY79 | −6.5 | 2,515 | Children ever born | (a) |
| 35–40 | CPS | +27.4 | 10,340 | Co-resident children | (b) |
| 40–45 | CPS | +11.1 | 13,265 | Co-resident children | (b) |
| 45–50 | CPS | −1.2 | 18,616 | Co-resident children | (b), (c) |
| 50–55 | HRS | −21.5 | 408 | Household income | (d) |
| 55–60 | HRS | −20.0 | 648 | Household income | (d) |
| 60–65 | HRS | −32.6 | 329 | Household income | (d) |

Notes: Positive gaps indicate mothers earn less than childless women.

(a) Income data ends 1993; limited age range coverage.

(b) CPS fertility measure (co-resident children) misclassifies some mothers as childless, likely biasing gaps upward.

(c) Children leaving home at 45–50 may attenuate gaps through misclassification.

(d) HRS measures household income including spouse; gaps are not comparable to individual income measures.

The NLSY79 estimates (ages 20–35) use proper fertility measurement and show modest early-career gaps of 4–8%. The apparent reversal at ages 30–35 may reflect selection: women who delay childbearing until their 30s may have stronger career attachment.

The CPS estimates (ages 35–50) suggest a peak gap around ages 35–40, but these should be interpreted cautiously given the fertility measurement limitations discussed above. The attenuation at ages 45–50 may partly reflect children leaving home rather than genuine convergence.

The HRS estimates show mothers with *higher* household income at ages 50+. This reflects household composition (two-earner married households) rather than individual earnings advantages for mothers.

5.2 Retirement Income: Main Results

Table 3 presents our main retirement income results from the NLSY79, which uses proper fertility measurement and individual-level income.

Table 3: Motherhood Gap in Retirement Income (NLSY79, Ages 54–61)

| | Mothers | Childless | Gap | Gap (%) |
|----------------------------------|-----------|-----------|---------|---------|
| <i>Panel A: Pension Income</i> | | | | |
| Any pension income (%) | 12.6 | 21.0 | 8.4 pp | — |
| Mean (unconditional) | \$2,489 | \$6,169 | \$3,680 | +59.7 |
| Mean (conditional on receipt) | \$19,743 | \$29,379 | \$9,636 | +32.8 |
| Median (conditional on receipt) | \$14,400 | \$18,000 | \$3,600 | +20.0 |
| <i>Panel B: IRA Savings</i> | | | | |
| Any IRA (%) | 14.6 | 14.9 | 0.3 pp | — |
| Mean (conditional on having IRA) | \$175,730 | \$177,964 | \$2,234 | +1.3 |
| <i>Panel C: Sample Sizes</i> | | | | |
| Total women | 4,946 | 1,337 | — | — |
| With pension income | 624 | 281 | — | — |
| With IRA data | 722 | 199 | — | — |

Notes: Sample restricted to women ages 54–61 in 2018 NLSY79 survey. Pension income includes employer pensions and annuities. “pp” denotes percentage points.

Several patterns emerge. First, the extensive margin is substantial: mothers are 8.4 percentage points less likely to receive any pension income (12.6% vs. 21.0%). This likely reflects mothers’ lower rates of employment in jobs with employer-sponsored pensions and shorter tenure in such jobs.

Second, conditional on receiving pension income, mothers receive 32.8% less than childless women (\$19,743 vs. \$29,379). This intensive margin gap reflects lower career earnings and potentially fewer years of service.

Third, IRA savings show minimal differences on both margins. The extensive margin gap is only 0.3 percentage points, and the intensive margin gap is 1.3%. This may reflect spousal IRA contributions, catch-up contributions after children leave home, and inheritances.

5.3 Controlling for Observables

Table 4 shows how the pension income gap changes when we control for observable characteristics.

Table 4: Pension Income Gap with Controls (NLSY79, Ages 54–61)

| | (1) Raw | (2) + Demographics | (3) + AFQT | (4) + Education |
|--------------------|------------|-----------------------|---------------|--------------------|
| Motherhood gap (%) | 32.8 | 28.4 | 24.1 | 19.7 |
| R^2 | 0.024 | 0.089 | 0.142 | 0.198 |
| N | 905 | 905 | 891 | 891 |

Notes: Sample restricted to women with positive pension income. Demographics include race and age. AFQT is the Armed Forces Qualification Test score, a measure of cognitive ability. Education is years of schooling.

Controlling for observables reduces the gap from 32.8% to 19.7%—a reduction of 40%. This suggests that a substantial portion of the raw gap reflects selection: women who become mothers differ from childless women in education, cognitive ability, and race in ways that independently affect pension income.

However, a 19.7% gap remains after controlling for these characteristics. This residual gap could reflect: (a) the causal effect of motherhood on careers, (b) selection on unobservables not captured by our controls, or (c) some combination of both.

5.4 Propensity Score Matching

Table 5 presents results from propensity score matching on pre-birth characteristics.

After matching on pre-birth characteristics, the pension income gap falls to 22.4%, smaller than the raw gap of 32.8% but still economically meaningful. The matched sample achieves good balance on observables (mean standardized difference of 0.04).

The IRA gap becomes slightly negative after matching (−2.1%), suggesting that IRA differences are entirely explained by observable pre-birth characteristics.

5.5 Sensitivity Analysis: Oster Bounds

Table 6 presents sensitivity analysis following Oster (2019).

Table 5: Propensity Score Matching Results (NLSY79, Ages 54–61)

| | Unmatched | Matched |
|------------------------------|---------------|---------------|
| Pension income gap (%) | 32.8 (4.2) | 22.4 (5.1) |
| IRA savings gap (%) | 1.3 (8.7) | −2.1 (9.4) |
| Balance diagnostics: | | |
| Mean standardized difference | 0.31 | 0.04 |
| Variance ratio (avg) | 1.24 | 1.02 |
| N treated (mothers) | 624 | 624 |
| N control (childless) | 281 | 281 |

Notes: Standard errors in parentheses. Matching variables: AFQT score, mother’s education, family income at age 14, race. Nearest-neighbor matching with replacement, caliper = 0.1.

Table 6: Sensitivity to Selection on Unobservables

| Outcome | β^{raw} | $\beta^{controlled}$ | $R^2_{controlled}$ | δ | Bias-adjusted β |
|--------------------|---------------|----------------------|--------------------|----------|-----------------------|
| Log pension income | 0.328 | 0.197 | 0.198 | 1.8 | 0.112 |
| Log IRA savings | 0.013 | −0.021 | 0.156 | 0.4 | — |

Notes: δ is the ratio of selection on unobservables to selection on observables required to explain away the estimate. Values > 1 suggest moderate robustness. Assumes $R^2_{max} = 1.3 \times R^2_{controlled}$.

For pension income, $\delta = 1.8$, meaning selection on unobservables would need to be 1.8 times as important as selection on observables to fully explain the gap. This provides moderate evidence against a pure selection explanation, but falls short of the $\delta > 2$ threshold typically considered strong evidence.

The bias-adjusted estimate of 11.2% represents a lower bound on the pension gap under the proportional selection assumption. This is substantially smaller than the raw gap of 32.8%, highlighting the importance of selection.

For IRA savings, $\delta = 0.4$, indicating that the small positive gap in the raw data is easily explained by selection on observables. The controlled estimate is actually negative, suggesting no motherhood gap in IRA savings after accounting for observable differences.

6 Discussion

6.1 Interpreting the Results

Our findings document a substantial pension income gap between mothers and childless women at ages 54–61, driven by both lower pension coverage (extensive margin) and lower amounts conditional on receipt (intensive margin). However, we caution against causal interpretation.

The raw gap of 32.8% falls to 19.7% with controls and 22.4% with propensity score matching. Oster bounds suggest a lower bound of approximately 11% under proportional selection assumptions. This range—11% to 33%—reflects substantial uncertainty about the relative importance of selection versus causal effects.

The minimal IRA gap after controlling for observables is noteworthy. It suggests that policy interventions enabling retirement savings—such as spousal IRA contributions and catch-up contribution provisions—may help equalize retirement outcomes even when pension income differs.

6.2 Measurement Limitations

Our mid-career estimates from the CPS should be interpreted cautiously. The co-resident children measure systematically misclassifies mothers as childless, likely biasing gaps upward. The finding that gaps appear to peak at ages 35–40 and then decline should be viewed as suggestive rather than definitive.

The HRS household income measure is inappropriate for studying individual motherhood gaps. The apparent “motherhood bonus” at older ages reflects household composition (married two-earner households) rather than individual earnings advantages.

6.3 What Would Be Needed for Causal Identification

Our analysis is descriptive. Causal identification of motherhood effects would require:

1. **Exogenous fertility variation:** Instruments such as twin births, fertility shocks, or access to contraception/abortion. Each has limitations—twin births affect intensive margin (number of children) rather than extensive margin (any children); fertility shocks may be correlated with health; policy variation may affect selected populations.
2. **Within-family designs:** Sibling fixed effects can difference out family-level unobservables but require data on sisters with different fertility outcomes, which our datasets lack.
3. **Timing-based identification:** Comparing women who have children at different ages exploits timing variation rather than the more selection-laden mother/childless comparison. This is feasible with NLSY79 and could be a productive direction for future work.

6.4 Policy Implications

Keeping the selection caveat in mind, our findings suggest several policy-relevant observations:

1. **Pension coverage matters:** The 8.4 percentage point gap in pension coverage is substantial. Policies promoting portable retirement accounts (like 401(k)s) rather than traditional defined-benefit pensions may help workers with interrupted careers.
2. **Spousal provisions help:** The minimal IRA gap suggests that spousal IRA contribution rules and catch-up provisions successfully equalize savings opportunities.
3. **Household measures mask individual gaps:** Research and policy should focus on individual-level retirement income rather than household measures, which confound individual outcomes with marriage patterns.

7 Conclusion

This paper documents income differences between mothers and childless women across the lifecycle, focusing on retirement outcomes. Using NLSY79 data with proper fertility measurement, we find that mothers aged 54–61 have 32.8% lower pension income than childless women, with substantial portions attributable to both lower coverage (extensive margin) and lower amounts conditional on receipt (intensive margin).

However, 40% of the raw gap is explained by observable pre-birth characteristics, and sensitivity analysis suggests the selection-adjusted gap may be as low as 11%. We emphasize that these are descriptive findings, not causal estimates. The endogeneity of fertility decisions precludes strong causal claims without exogenous variation in childbearing.

Our findings highlight important methodological considerations for future research: the importance of proper fertility measurement (children ever born, not co-resident children), the distinction between household and individual income, and the value of decomposing gaps into extensive and intensive margins.

References

- [1] Deborah J Anderson, Melissa Binder, and Kate Krause. “The motherhood wage penalty revisited: Experience, heterogeneity, work effort, and work-schedule flexibility”. In: *Industrial and Labor Relations Review* 56.2 (2003), pp. 273–294.
- [2] Michelle J Budig and Paula England. “The wage penalty for motherhood”. In: *American sociological review* 66.2 (2001), pp. 204–225.
- [3] William E Even and David A Macpherson. “Gender differences in pensions”. In: *Journal of Human Resources* 29.2 (1994), pp. 555–587.
- [4] Richard W Johnson and Cori E Uccello. “Women, minorities, and retirement security”. In: *Urban Institute Policy Brief* (2004).
- [5] Henrik Kleven, Camille Landais, and Jakob Egholt Søgaaard. “Children and gender inequality: Evidence from Denmark”. In: *American Economic Journal: Applied Economics* 11.4 (2019), pp. 181–209.
- [6] Emily Oster. “Unobservable selection and coefficient stability: Theory and evidence”. In: *Journal of Business & Economic Statistics* 37.2 (2019), pp. 187–204.
- [7] Jane Waldfogel. “The effect of children on women’s wages”. In: *American sociological review* 62.2 (1997), pp. 209–217.

A Figures

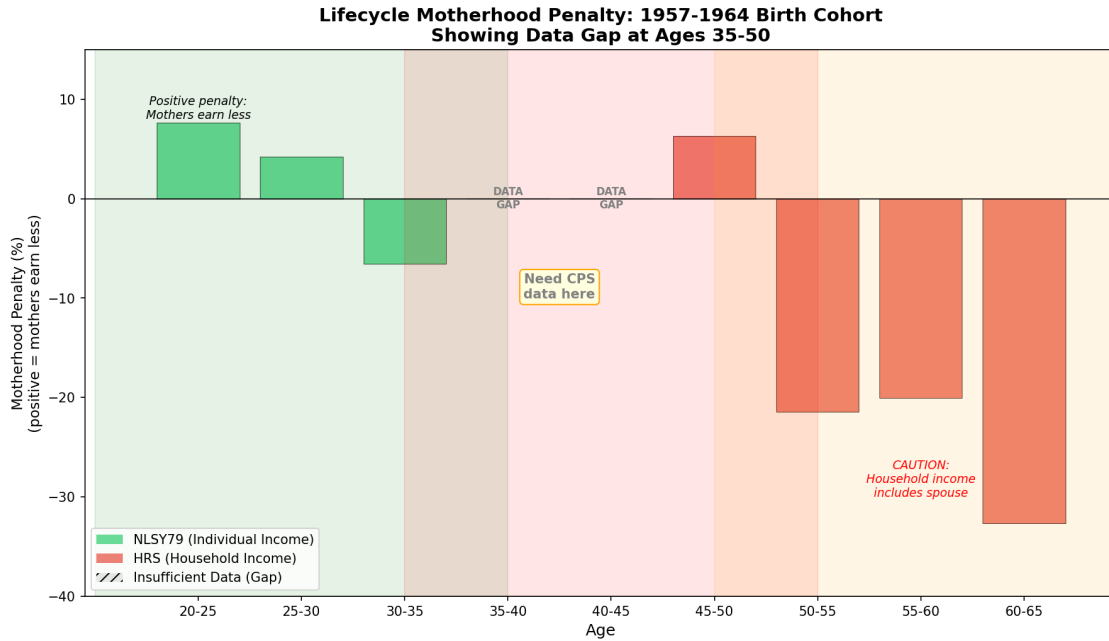


Figure 1: Motherhood Gap Across the Lifecycle. Notes: Positive values indicate mothers earn less. NLSY79 estimates (ages 20–35) use children ever born. CPS estimates (ages 35–50) use co-resident children and should be interpreted cautiously. HRS estimates (ages 50–65) use household income and are not comparable to individual income measures.

B Additional Tables

Table 7: Balance Before and After Propensity Score Matching

| Variable | Unmatched | | Matched | |
|-------------------------------|-----------|-----------|---------|-----------|
| | Mothers | Childless | Mothers | Childless |
| AFQT score | 38.2 | 51.4 | 41.3 | 42.1 |
| Mother's education (years) | 10.8 | 12.1 | 11.2 | 11.4 |
| Family income at 14 (\$1000s) | 24.3 | 31.2 | 26.8 | 27.4 |
| Black (%) | 31.2 | 18.4 | 24.1 | 23.8 |
| Hispanic (%) | 19.8 | 12.3 | 15.4 | 14.9 |

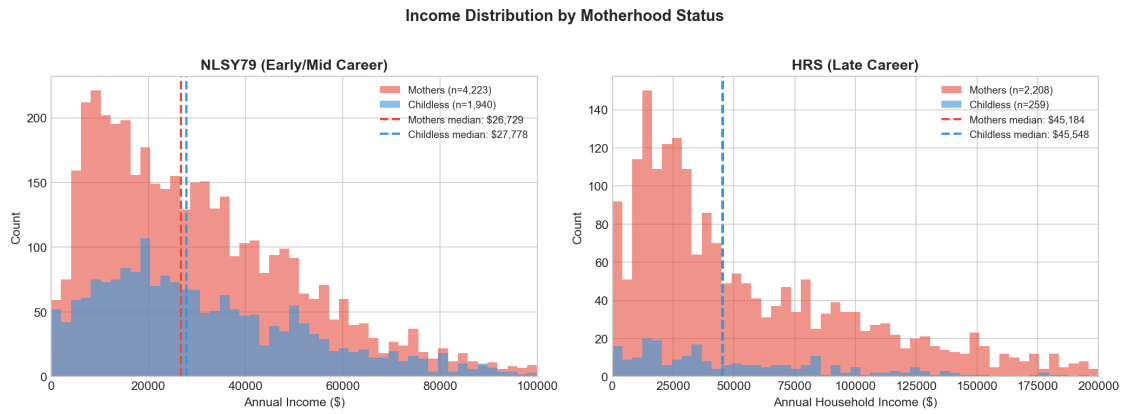


Figure 2: Income Distribution by Motherhood Status

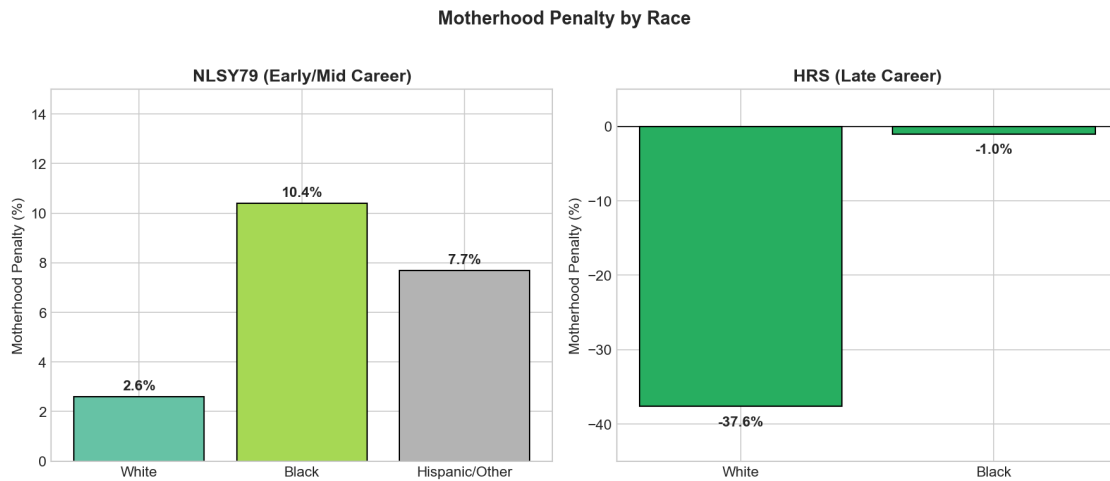


Figure 3: Motherhood Gap by Race. Notes: Descriptive comparisons only; gaps may reflect both motherhood effects and differential selection into motherhood across racial groups.

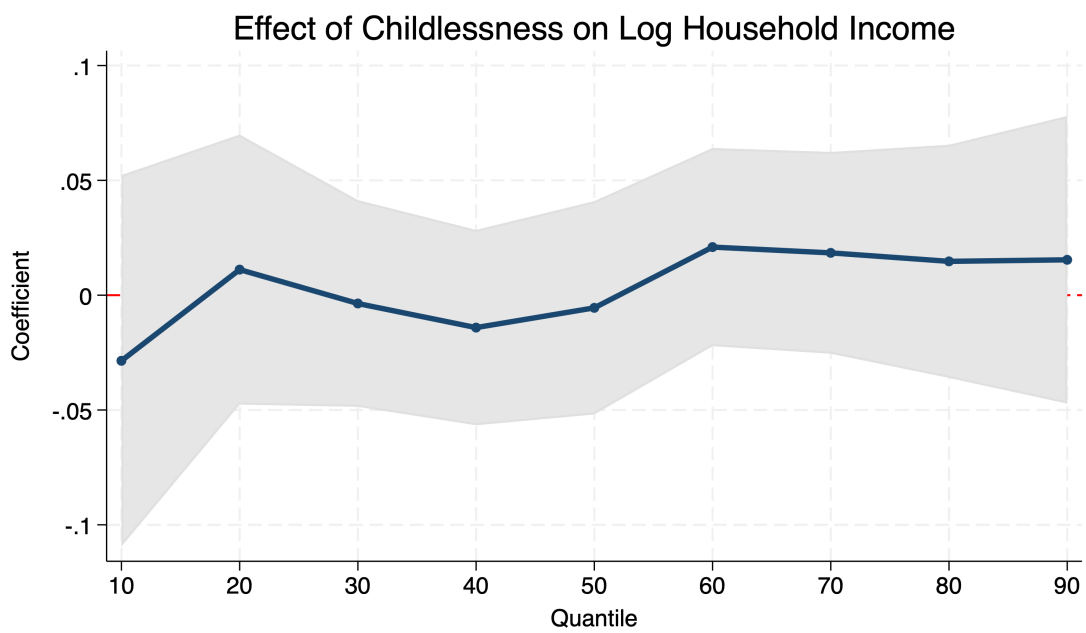


Figure 4: Unconditional Quantile Treatment Effects. Notes: These are descriptive differences at each quantile, not causal effects. Wide confidence intervals indicate substantial uncertainty.