The Motherhood Gap Across the Lifecycle and Income Distribution: A Quantile Regression Analysis

Afrouz Azadikhah Jahromi* Alina Malkova†

February 22, 2026

Abstract

We document that the relationship between childlessness and retirement income reverses sign across the income distribution. Using quantile regression on NLSY79 data for women born 1957–1964, we find that childless women in the bottom 40% of the pension income distribution face penalties of 8–15%, while those in the top quartile enjoy premiums exceeding 12%. This reversal—robust across conditional QTE, unconditional QTE, propensity score weighted methods, and machine learning estimators (LASSO, Double ML)—reconciles two literatures: the motherhood penalty literature (career costs) and the old-age security literature (insurance value of children). Both are correct, but at different points in the distribution.

Extensive heterogeneity analyses reveal that the "motherhood penalty" is primarily a *compounding vulnerability* penalty: married, college-educated mothers with 1–2 children have *higher* median income than childless women (\$128,000 vs. \$43,000), while divorced mothers with less than high school education face severe disadvantage (\$16,000). The gap varies dramatically by marital status (married: +7.7% premium; never-married: −50.5% penalty), timing of first birth (teen mothers: −45% penalty;

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

†N.Bisk School of Business, Florida Institute of Technology. Email: amalkova@fit.edu

age 30+: +54% premium), and cohort (younger cohorts show 41pp smaller gaps). Social Security provides a partial buffer (8% gap vs. 15–25% in private pensions), but mothers are more reliant on SS despite receiving lower benefits. Health status does not explain the gap. Our findings imply that universal pension credits for mothers would direct approximately 60% of expenditure to women who already benefit from childlessness through career advantages, while failing to address the vulnerabilities of divorced and widowed mothers with limited education.

JEL Classification: J13, J16, J26, D31

Keywords: Motherhood gap, unconditional quantile regression, retirement income inequality, heterogeneous treatment effects, double machine learning

1 Introduction

Why do poor childless retirees face substantial pension penalties while wealthy childless retirees enjoy pension premiums? This puzzle challenges conventional wisdom about the economic consequences of fertility decisions and reveals fundamental heterogeneity in how family structure affects retirement security across the income distribution.

The extensive literature on the "motherhood penalty" documents wage gaps of 5–15% per child during working years (Waldfogel 1997; Budig and England 2001; Anderson, Binder, and Krause 2003). These penalties accumulate over the lifecycle, resulting in retirement income gaps (Even and Macpherson 1994; Johnson and Uccello 2004). Yet this literature has largely overlooked two crucial dimensions of heterogeneity: how effects vary across the *income distribution* at any given age, and how these distributional patterns evolve over the *lifecycle*.

We address both dimensions using quantile regression methods and three complementary data sources covering women born 1957–1964 from early career through retirement. Our central finding is a striking reversal in the relationship between childlessness and retirement income that occurs around the 40th percentile of the income distribution. While childless women in the bottom 40% face penalties of 8–15%, those in the top quartile enjoy premiums exceeding 12%. This reversal reveals that children's economic function in retirement is not uniform but depends critically on households' position in the income distribution.

This distributional heterogeneity stands in sharp contrast to the existing literature's focus on average effects. A 10% average motherhood penalty does not translate uniformly into 10% lower pensions regardless of economic status—it masks fundamentally different effects at different points in the distribution. Our analysis reconciles two seemingly contradictory bodies of literature. The motherhood penalty literature documents career costs of children (Kleven, Landais, and Søgaard 2019; Anderson, Binder, and Krause 2003). The old-age security literature emphasizes children's insurance value (Ebenstein and Leung 2010). Our results demonstrate that both perspectives are correct—but at different points in the income distribution. Below the 40th percentile, the insurance value dominates; above it, career costs prevail.

We make four contributions. First, we document how the motherhood gap varies across the income distribution using both conditional and unconditional quantile regression (Firpo, Fortin, and Lemieux 2009), revealing a reversal that mean comparisons completely obscure. Second, we trace how this distributional pattern evolves over the lifecycle, showing that the reversal point shifts as the cohort ages and exhibits a dose-response pattern by number of children. Third, we decompose the gap into explained (characteristics) and unexplained (structural) components using both the Chernozhukov, Fernández-Val, and Melly (2013) and Machado and Mata (2005) methods, providing robust evidence that different mechanisms operate at different quantiles. Fourth, we provide direct evidence for the insurance mechanism using HRS transfer data and strengthen causal interpretation by examining involuntary versus voluntary childlessness.

A critical methodological contribution addresses the household versus individual income problem that confounds much of the existing literature. The HRS measures *household* income, which masks individual-level penalties because mothers are more likely to be married (creating two-earner households). We show that running quantile regressions on HRS household income produces misleading results, while individual pension income from NLSY79 reveals the true distributional pattern. This distinction is essential for both research and policy.

Our findings have immediate policy relevance. Universal pension compensation for childbearing—common in European systems—would primarily benefit those who least need support (high-income parents who already benefit from childlessness through enhanced careers) while failing to address the vulnerabilities of those most at risk (low-income childless retirees who lack family insurance networks).

2 Theoretical Framework

We develop a model in which the relationship between childlessness and retirement income reflects the net effect of two opposing economic forces, whose relative importance varies across the income distribution.

2.1 Setup

Consider an individual i with pre-retirement income y_i drawn from distribution $F(y)$ who makes a binary fertility choice $d_i \in \{0, 1\}$ where $d_i = 0$ denotes having children and $d_i = 1$ denotes childlessness. Retirement income R_i consists of two components:

$$R_i(d_i, y_i) = S_i(d_i, y_i) + T_i(d_i, y_i) \tag{1}$$

where $S_i(d_i, y_i)$ represents accumulated retirement savings from lifetime earnings and $T_i(d_i, y_i)$ represents transfers (both formal and informal) received in retirement.

2.2 The Career Impediment Mechanism

Following Kleven, Landais, and Søgaard (2019), children reduce human capital accumulation through career interruptions. Let the human capital production function be:

$$H(d_i, y_i) = y_i \cdot \exp[-\phi(y_i) \cdot (1 - d_i)] \tag{2}$$

where $\phi(y_i)$ represents human capital depreciation from childrearing.

Assumption 1 (Increasing Opportunity Cost)**.** $\phi'(y) > 0$ *and* $\phi''(y) \geq 0$

This reflects that high earners face larger opportunity costs from career interruptions. The savings gain from childlessness is:

$$\Delta S(y_i) \approx s \cdot y_i \cdot \phi(y_i) \equiv \alpha(y_i) \cdot y_i \tag{3}$$

where $\alpha(y_i) = s \cdot \phi(y_i)$ is increasing in income.

2.3 The Insurance Mechanism

Following Ebenstein and Leung (2010), children provide old-age insurance through informal support. The transfer function for parents is:

$$T_i(0, y_i) = \beta_0 + \beta_1 \cdot \exp(-\lambda y_i) \tag{4}$$

where $\beta_0 \geq 0$ represents baseline transfers, $\beta_1 > 0$ captures income-dependent transfers, and $\lambda > 0$ determines the rate at which transfers decline with income.

For childless individuals:

$$T_i(1, y_i) = \beta_0 \tag{5}$$

The insurance value of children is:

$$\Delta T(y_i) = \beta_1 \cdot \exp(-\lambda y_i) \equiv \beta(y_i) \tag{6}$$

which is decreasing in income.

2.4 Net Effect and Reversal

The total effect of childlessness on retirement income is:

$$\Delta R(y_i) = \underbrace{\alpha(y_i) \cdot y_i}_{\text{Career gain}} - \underbrace{\beta(y_i)}_{\text{Insurance loss}} \tag{7}$$

Proposition 1 (Existence of Reversal Point)**.** *Under Assumption 1 and given* $\lim_{y \to 0} \alpha(y) \cdot y = 0$ *and* $\lim_{y \to 0} \beta(y) = \beta_1 > 0$*, there exists a unique threshold income* $y^* > 0$ *such that:*

$$\Delta R(y^*) = 0 \Leftrightarrow \alpha(y^*) \cdot y^* = \beta(y^*) \tag{8}$$

This proposition provides our key testable prediction: the effect of childlessness switches from negative (penalty) to positive (premium) at a unique income threshold. **Note on functional form:** The uniqueness of the reversal point depends on our exponential specification for the insurance mechanism. Alternative functional forms could generate multiple crossing points. Our empirical test for a single crossing (Section 5.6) provides independent evidence supporting uniqueness rather than assuming it.

2.5 Quantile Treatment Effects

In our quantile regression framework, where Q_τ denotes the τ-th quantile:

$$Q_\tau(\ln R_i | X_i, d_i) = \gamma_\tau + \delta_\tau \cdot d_i + X_i'\theta_\tau \tag{9}$$

The model predicts:

Proposition 2 (Quantile Treatment Effects). *Let τ^* denote the quantile corresponding to income y^*. Then:*

1. $\delta_\tau < 0$ *for* $\tau < \tau^*$ *(childlessness penalty)*
2. $\delta_\tau = 0$ *for* $\tau = \tau^*$ *(reversal point)*
3. $\delta_\tau > 0$ *for* $\tau > \tau^*$ *(childlessness premium)*
4. $\partial\delta_\tau/\partial\tau > 0$ *(monotonically increasing)*

2.6 Decomposition Predictions

Following Chernozhukov, Fernández-Val, and Melly (2013), we decompose the quantile gap into composition (Δ^X) and structure (Δ^S) effects:

Proposition 3 (Mechanism Identification).

1. *At low quantiles ($\tau < \tau^*$): $|\Delta^S| > |\Delta^X|$ because insurance operates through unobserved networks*
2. *At high quantiles ($\tau > \tau^*$): $|\Delta^X| > |\Delta^S|$ because career effects are captured by observable human capital*

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.
- **Individual retirement income:** Pension receipt and amounts are measured at the individual level, not confounded with spouse income.
- **Pre-birth characteristics:** AFQT scores, family background enable assessment of selection.

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

3.2 RAND HRS

The Health and Retirement Study tracks individuals aged 50+ from 1992–2022. We use HRS for two purposes: (1) to illustrate the household income measurement problem, and (2) to examine how distributional patterns evolve by age within the panel.

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.

3.3 CPS ASEC

The Current Population Survey 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.

Limitation: The CPS identifies mothers by co-resident children, which systematically misclassifies some mothers as childless, likely biasing gaps upward at ages 35–40 and creating spurious attenuation at ages 45–50 as children leave home.

3.4 Sample Construction

Table 1 documents sample sizes across analyses.

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 (ages 54–61)	NLSY79	4,946	1,337
With pension income	NLSY79	624	182
Retirement (ages 50–67)	HRS	3,715	346

4 Methodology

4.1 Conditional Quantile Treatment Effects

We estimate conditional quantile treatment effects (CQTEs) using a two-step approach. In the first step, we regress the treatment (childlessness) on control variables using OLS and obtain residuals:

$$\text{Childless}_i = X_i'\pi + \nu_i \tag{10}$$

In the second step, we estimate conditional quantile regressions using the residualized treatment:

$$Q_\tau(\ln Y_i|X_i) = \alpha_\tau + \beta_\tau \widehat{\text{Childless}}_i + \gamma_\tau X_i + \epsilon_{i,\tau} \tag{11}$$

where $\widehat{\text{Childless}}_i$ is the residual from the first stage.

This approach, following Chernozhukov, Fernández-Val, and Melly (2013), decomposes the variance of the treatment into a piece explained by observables and a residual piece orthogonal to controls.

4.2 Unconditional Quantile Treatment Effects

A limitation of conditional quantile regression is that the coefficients describe effects on the τ-th quantile of the *conditional* distribution of income given covariates, which may not equal the effect on the τ-th quantile of the *unconditional* distribution—the object most relevant for policy. Following Firpo, Fortin, and Lemieux (2009), we implement unconditional quantile regression (UQR) using the Recentered Influence Function (RIF):

$$\text{RIF}(Y_i; Q_\tau) = Q_\tau + \frac{\tau - \mathbf{1}(Y_i \leq Q_\tau)}{f_Y(Q_\tau)} \tag{12}$$

where f_Y is the density of Y at the quantile Q_τ.

The UQR regression is:

$$\mathbb{E}[\text{RIF}(Y_i; Q_\tau)|X_i, D_i] = \alpha_\tau^{UQR} + \beta_\tau^{UQR} \cdot \text{Childless}_i + \gamma_\tau^{UQR} X_i \tag{13}$$

The coefficient β_τ^{UQR} represents the marginal effect of childlessness on the unconditional τ-th quantile of income. This interpretation is directly policy-relevant: it answers "how would redistributing women from mother to childless status affect the τ-th percentile of the overall income distribution?"

4.3 Propensity Score Weighted QTE

To address concerns about the linear first-stage specification, we implement the Abadie (2002) propensity score weighting approach. We first estimate the propensity score for childlessness using probit:

$$\text{P}(\text{Childless}_i = 1|X_i) = \Phi(X_i'\gamma) \tag{14}$$

We then estimate weighted quantile regressions with inverse probability weights:

$$w_i = \frac{D_i}{\hat{p}(X_i)} + \frac{1 - D_i}{1 - \hat{p}(X_i)} \tag{15}$$

This approach has several advantages: (1) it naturally accommodates nonlinear selection into childlessness, (2) the NLSY79 covariates—AFQT, parental education, family structure at age 14, early work history—provide a strong propensity score model, and (3) it has a clear econometric justification under weaker assumptions than the two-step residualization approach.

4.4 Counterfactual Decomposition

We implement two complementary decomposition approaches to ensure robustness.

4.4.1 Chernozhukov-Fernández-Val-Melly (2013) Decomposition

Following Chernozhukov, Fernández-Val, and Melly (2013), we construct counterfactual distributions to decompose the income gap between childless women and mothers. Let $F_{Y(1|1)}$ and $F_{Y(0|0)}$ represent the observed distributions for childless and mothers, respectively. The counterfactual distribution $F_{Y(0|1)}$ represents what childless women's income distribution would have been if they faced mothers' income structure:

$$F_{Y(0|1)}(y) := \int_{\mathcal{X}_1} F_{Y_0|X_0}(y|x) dF_{X_1}(x) \tag{16}$$

The total gap decomposes as:

$$F_{\Delta_{ft}|X_f} - F_{\Delta_{mt}|X_m} = \underbrace{[F_{\Delta_{ft}|X_f} - F_{\Delta_{mt}|X_f}]}_{\text{Structure Effect}} + \underbrace{[F_{\Delta_{mt}|X_f} - F_{\Delta_{mt}|X_m}]}_{\text{Composition Effect}} \tag{17}$$

4.4.2 Machado-Mata (2005) Decomposition

As a robustness check, we implement the Machado and Mata (2005) decomposition, which constructs counterfactual distributions by simulation. The procedure is:

1. Estimate quantile regression coefficients $\hat{\beta}(\tau)$ for mothers and $\hat{\gamma}(\tau)$ for childless women across quantiles $\tau \in (0, 1)$.

2. Draw a random sample of τ values from $U(0, 1)$.

3. Construct counterfactual income draws: $\tilde{Y}_i^{cf} = X_i^{childless} \hat{\beta}(\tau_i)$ (childless women's characteristics, mothers' returns).

4. Compare the distribution of $\tilde{Y}^{cf}$ to actual distributions.

This provides a different decomposition path than CFM and allows assessment of whether our conclusions depend on the specific decomposition methodology.

4.5 Robustness Approaches

4.5.1 Individual Fixed Effects

We exploit the panel structure of HRS to implement correlated random effects quantile regression following Arellano and Bonhomme (2016):

$$Q_\tau(\ln Y_{it}|X_{it}, \alpha_i) = \alpha_{i,\tau} + \beta_\tau \cdot \text{Childless}_i + \gamma_\tau X_{it} + \epsilon_{it,\tau} \tag{18}$$

where $\alpha_{i,\tau}$ captures individual-specific effects that may vary across quantiles.

4.5.2 Coarsened Exact Matching

We implement CEM on pre-fertility characteristics: education, birth cohort, race, gender, and ever-married status. CEM proceeds by coarsening continuous variables into strata, then exactly matching treated and control units within each stratum. We assign weights:

$$w_i = \frac{n_s}{n_{s,D_i}} \cdot \frac{N_{D_i}}{N} \tag{19}$$

where n_s denotes stratum size and n_{s,D_i} represents units in stratum s with treatment status D_i.

Caveat on "ever married": Matching on ever-married status effectively conditions on a variable that may be affected by fertility decisions. If childlessness causally affects marriage probability, this could bias estimates. We present results both with and without this matching variable.

4.5.3 Oster Bounds

Following Oster (2019), we assess sensitivity to selection on unobservables. The key parameter δ represents the ratio of selection on unobservables to selection on observables. Values of $\delta > 1$ suggest moderate robustness; values > 2 indicate strong robustness.

5 Results

5.1 The Reversal: From Penalty to Premium

Our central finding is a striking reversal in the relationship between childlessness and pension income. Table 2 presents quantile regression estimates for individual pension income from NLSY79 at ages 54–61.

Table 2: Quantile Treatment Effects on Individual Pension Income (NLSY79, Ages 54–61)

	Quantiles						
Sample	0.10	0.20	0.30	0.40	0.50	0.75	0.90
Panel A: Women Only							
Childless effect	-0.152^{***}	-0.098^{***}	-0.042^{*}	0.018	0.065^{**}	0.098^{***}	0.124^{***}
	(0.038)	(0.031)	(0.025)	(0.023)	(0.026)	(0.032)	(0.041)
Panel B: Men Only							
Childless effect	-0.089^{**}	-0.045	-0.022	-0.008	0.011	0.028	0.036
	(0.041)	(0.034)	(0.028)	(0.026)	(0.028)	(0.035)	(0.044)
Observations	806	806	806	806	806	806	806
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Controls include age, education, race, and AFQT score. Robust standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. The reversal point where the coefficient crosses zero occurs between the 30th and 40th percentiles for women.

For women, childlessness is associated with pension penalties of 15.2% at the 10th percentile but premiums of 12.4% at the 90th percentile. The reversal occurs around the 35th–40th percentile. For men, effects are smaller and less precisely estimated, with a weaker reversal pattern—consistent with our theoretical prediction that women face both stronger insurance effects and larger career penalties.

The formal test for monotonicity strongly rejects the null hypothesis of constant effects across quantiles ($\chi^2 = 89.4$, $p < 0.001$), confirming that the relationship between childlessness and pension income fundamentally differs across the distribution.

Sensitivity to sample trimming: Given limited sample sizes at extreme quantiles (approximately 18 childless women at the 10th percentile), we assess robustness by trimming the top and bottom 5% of the pension income distribution. Table 3 shows that results are qualitatively similar, though standard errors increase. We also report the 15th and 85th percentiles as potentially more credible estimates than the 10th and 90th.

Table 3: Sensitivity to Sample Trimming (Women, Ages 54–61)

	Quantiles					
Sample	0.15	0.25	0.50	0.75	0.85	N
Full sample	-0.128^{***}	-0.082^{***}	0.065^{**}	0.098^{***}	0.112^{***}	806
	(0.035)	(0.029)	(0.026)	(0.032)	(0.038)	
Trimmed 5%/95%	-0.118^{***}	-0.076^{**}	0.058^{**}	0.089^{**}	0.098^{**}	726
	(0.039)	(0.032)	(0.028)	(0.035)	(0.042)	

Notes: Robust standard errors in parentheses. Trimming removes top and bottom 5% of pension income distribution. The reversal pattern is preserved though magnitudes are slightly attenuated.

5.2 Extensive vs. Intensive Margin

Before examining the quantile effects, we address a potential selection concern: if childless women are systematically less likely to have any pension income, conditioning on positive pension income could mechanically generate the observed reversal. Table 4 examines both margins.

Two findings are notable. First, childless women are *not* less likely to receive pension income—the extensive margin difference is small (1 percentage point) and statistically insignificant. This rules out the concern that selection into pension receipt drives our results. Second, the intensive margin reveals the distributional heterogeneity: at the mean and median, differences are negligible, but at the tails the gap emerges strongly. This underscores why quantile methods are essential—mean comparisons completely miss the reversal.

5.3 The Household Income Problem

Table 5 demonstrates why using household income produces misleading results. Running the same quantile regression on HRS household income shows no reversal—instead, mothers appear to have *higher* income at all quantiles. This reflects household composition (married two-earner households) rather than individual motherhood effects.

Table 4: Extensive vs. Intensive Margin Pension Receipt (Women, Ages 54–61)

	Mothers	Childless	Difference	p-value
Panel A: Extensive Margin				
Any pension income (%)	12.6%	13.6%	+1.0 pp	0.38
	(N=4,946)	(N=1,337)		
Panel B: Composition				
Conditional on pension:				
Mean log pension	9.42	9.48	+0.06	0.52
Median pension ($)	12,400	13,100	+700	0.41
Panel C: Intensive Margin by Quantile				
10th percentile	2,800	2,380	−420	0.04
90th percentile	38,200	42,900	+4,700	0.03

Notes: Panel A shows pension receipt rates are similar across groups, ruling out mechanical selection. Panel C shows the distributional heterogeneity: childless women have lower pensions at the 10th percentile but higher pensions at the 90th percentile.

Table 5: Individual vs. Household Income: The Measurement Problem

	Individual Pension (NLSY79)		Household Income (HRS)	
Quantile	Coefficient	Interpretation	Coefficient	Interpretation
0.10	-0.152^{***}	Penalty	-0.089^{**}	Penalty
0.40	0.018	Reversal	−0.031	No reversal
0.90	0.124^{***}	Premium	−0.042	No premium

Notes: NLSY79 measures individual pension income. HRS measures total household income including spouse. The absence of reversal in HRS reflects that mothers are more likely to be married, creating two-earner households.

This comparison underscores a critical methodological point: research on motherhood gaps must use individual-level income measures. Household income confounds individual outcomes with marriage patterns.

5.4 Decomposition: Insurance vs. Career Mechanisms

Figure **??** presents the Chernozhukov, Fernández-Val, and Melly (2013) decomposition for women's individual pension income. The results reveal a dramatic shift in the relative importance of explained versus unexplained factors across the distribution.

At the 10th percentile, the unexplained (structural) component accounts for 78% of the total gap, consistent with insurance mechanisms operating through unobserved family support networks. By the 75th percentile, observable characteristics explain 115% of the gap—the unexplained component actually becomes negative, suggesting that high-income childless women receive *better* returns to their characteristics than mothers.

Table 6: Decomposition of the Childlessness Gap (Women, Ages 54–61)

	Quantiles				
Component	0.10	0.25	0.50	0.75	0.90
Total Gap	−0.152	−0.098	0.065	0.098	0.124
Explained (Characteristics)	−0.033	−0.024	0.078	0.113	0.142
	(22%)	(24%)	(120%)	(115%)	(115%)
Unexplained (Structure)	−0.119	−0.074	−0.013	−0.015	−0.018
	(78%)	(76%)	(−20%)	(−15%)	(−15%)

Notes: Decomposition following Chernozhukov, Fernández-Val, and Melly (2013). Percentages show share of total gap. At high quantiles, explained exceeds 100% because unexplained is negative.

This complete reversal in decomposition patterns provides evidence that different mechanisms operate at different points in the distribution, consistent with our theoretical framework. We interpret the large unexplained component at low quantiles as *consistent with*, though not exclusively attributable to, insurance mechanisms. The unexplained component also captures unobserved ability differences, preference heterogeneity, health differences, and other factors correlated with childlessness that are not included in our controls. That said, the theoretical motivation for insurance effects at low incomes—where formal safety nets are weaker and family support more critical—provides a plausible interpretation. At high quantiles, the dominance of observable characteristics (education, work experience, AFQT) strongly suggests that career effects captured by human capital accumulation are the primary mechanism.

5.5 Lifecycle Evolution of the Distributional Pattern

Table 7 shows how the quantile treatment effects evolve over the lifecycle. The reversal point shifts rightward as the cohort ages.

Table 7: Evolution of Quantile Treatment Effects Over the Lifecycle

	Quantiles						
Age Group (Source)	0.10	0.20	0.30	0.40	0.50	0.75	0.90
20–35 (NLSY79)	−0.042	−0.018	0.012	0.028	0.041	0.056	0.068
	(0.015)	(0.012)	(0.011)	(0.010)	(0.012)	(0.015)	(0.019)
35–50 (CPS)	−0.089	−0.052	−0.021	0.008	0.032	0.061	0.082
	(0.018)	(0.015)	(0.013)	(0.012)	(0.013)	(0.016)	(0.021)
54–61 (NLSY79)	−0.152	−0.098	−0.042	0.018	0.065	0.098	0.124
	(0.038)	(0.031)	(0.025)	(0.023)	(0.026)	(0.032)	(0.041)
Reversal point (τ^*)	$\approx$ 0.28 (ages 20–35) $\rightarrow$ $\approx$ 0.38 (ages 35–50) $\rightarrow$ $\approx$ 0.38 (ages 54–61)						

Notes: Individual income for NLSY79; individual income for CPS. Standard errors in parentheses. The reversal point shifts rightward over the lifecycle as career effects accumulate and insurance mechanisms become more relevant at older ages.

Two patterns emerge. First, the *magnitude* of effects increases with age—both penalties at low quantiles and premiums at high quantiles grow larger over the lifecycle. This reflects the cumulative nature of both career advantages (compounding) and insurance disadvantages (increasing with age and health risks). Second, the *reversal point* shifts slightly rightward, from approximately the 28th percentile at ages 20–35 to approximately the 38th percentile at ages 54–61.

Interpreting the lifecycle shift with caution: The apparent shift in the reversal point deserves careful scrutiny because the ages 35–50 estimates use CPS data with the co-resident children fertility measure. To assess how much of the apparent shift could be artifactual, Table 8 compares estimates within CPS across age subgroups where fertility measurement reliability varies.

Table 8: Reversal Point Across CPS Age Subgroups

Age Group	Reversal Point (τ^*)	SE	Measurement Quality
35–40	0.32	(0.04)	Better (most children at home)
41–45	0.36	(0.05)	Moderate
46–50	0.42	(0.06)	Worse (children leaving home)

Notes: Reversal point estimated as the quantile where the childless coefficient crosses zero. The rightward shift within CPS—from 0.32 to 0.42—could reflect either genuine lifecycle dynamics or increasing fertility misclassification as children leave home. The similarity between the 46–50 CPS estimate (0.42) and the 54–61 NLSY79 estimate (0.38) is reassuring but not definitive.

The fact that the reversal point shifts *within* CPS subsamples—from 0.32 at ages 35–40 to 0.42 at ages 46–50—suggests that at least some of the apparent lifecycle shift is artifactual, driven by increasing misclassification of mothers as childless when children leave home. The NLSY79 estimates (ages 20–35 and 54–61) use children ever born and are not subject to this bias, showing a shift from 0.28 to 0.38.

What we can and cannot conclude: The *existence* of the reversal at each lifecycle stage is well-established—childless women face penalties at low quantiles and premiums at high quantiles regardless of age. This core finding is robust. The *shift* in the reversal point over the lifecycle (from approximately 0.28 to 0.38) is suggestive rather than definitive. While the NLSY79 comparison provides cleaner evidence than the CPS analysis, we cannot rule out that compositional changes (selective attrition, mortality, cohort effects) contribute to the observed shift. The lifecycle shift should be interpreted as consistent with our theoretical comparative statics—as health risks increase with age, the insurance mechanism becomes more important—but not as a firmly established dynamic.

5.6 Robustness

5.6.1 Individual Fixed Effects

Exploiting the panel structure of NLSY79 with individual fixed effects yields a reversal at the 40th percentile—virtually identical to our main estimate. This suggests the reversal pattern is not driven by time-invariant unobserved heterogeneity.

5.6.2 Coarsened Exact Matching

Table 9 presents CEM-weighted quantile regression results. After matching on pre-fertility characteristics, the reversal remains evident, occurring at approximately the 38th percentile. Magnitudes are slightly attenuated but the monotonic gradient is preserved.

Table 9: CEM-Weighted Quantile Regression (NLSY79, Ages 54–61)

	Quantiles				
Specification	0.10	0.50	0.75	0.90	$\mathcal{L}_1$
Panel A: Without "Ever Married"					
CEM-weighted	-0.118^{***}	0.048^{*}	0.082^{**}	0.098^{**}	0.04
	(0.042)	(0.028)	(0.035)	(0.046)	
Panel B: With "Ever Married"					
CEM-weighted	-0.072^{*}	0.038	0.068^{*}	0.082^{*}	0.02
	(0.048)	(0.032)	(0.039)	(0.051)	
Pre-match SMD: ever married	-1.17 (large imbalance)				

Notes: Matching variables in Panel A: education, birth cohort, race, AFQT quartile. Panel B adds ever-married status. The large pre-match standardized mean difference (SMD = -1.17) indicates that childless women are substantially less likely to have ever married. Including ever-married in matching attenuates effects by approximately 30–40% but preserves the reversal pattern. This attenuation could reflect either (a) the marriage channel mediating the childlessness effect, or (b) bias from matching on a post-treatment variable if childlessness causally affects marriage. We present both specifications transparently; the truth likely lies between them.

5.6.3 Oster Bounds

Table 10 presents sensitivity analysis. For the 10th percentile effect, $\delta = 2.4$, indicating that selection on unobservables would need to be 2.4 times as important as selection on observables to fully explain the penalty. For the 90th percentile, $\delta = 1.6$—still above the conventional threshold of 1 but less robust than the low-quantile effects.

Table 10: Oster Bounds by Quantile

Quantile	β^{raw}	$\beta^{controlled}$	$R^2_{controlled}$	δ	Interpretation
0.10	-0.218	-0.152	0.18	2.4	Robust
0.25	-0.142	-0.098	0.21	2.1	Robust
0.50	0.089	0.065	0.24	1.8	Moderate
0.90	0.168	0.124	0.26	1.6	Moderate

Notes: $\delta > 1$ suggests moderate robustness; $\delta > 2$ indicates strong robustness. Assumes $R^2_{max} = 1.3 \times R^2_{controlled}$.

Critically, even under extreme assumptions about selection ($\delta = 2$), the *difference* between the 10th and 90th percentile effects remains statistically significant, confirming that the reversal pattern cannot be attributed solely to selection bias.

5.6.4 Education Heterogeneity

Our theoretical framework predicts that the reversal point should vary with education: higher education increases the opportunity cost of childbearing (shifting the career effect curve upward), which should move the reversal point leftward. Table 11 tests this prediction.

The results strongly support the theoretical prediction. For women with less than high school education, the reversal occurs at the 72nd percentile—childlessness is penalized throughout most of the distribution. For college graduates, the reversal occurs at the 28th percentile—only the lowest-income childless women face penalties. This pattern reflects the differential opportunity costs of childbearing: college-educated women sacrifice more lifetime earnings by having children, shifting the balance toward childlessness premiums at lower income thresholds.

5.7 Bootstrap Inference

Table 12 presents bootstrap inference on the quantile processes following Chernozhukov, Fernández-Val, and Melly (2013). We reject both the null of no effect and constant effects

Table 11: Reversal Point by Education Level (Women, Ages 54–61)

	Quantiles					
Education	0.10	0.25	0.50	0.75	0.90	Reversal τ^*
Less than HS	-0.198^{***}	-0.142^{***}	-0.068^{*}	0.012	0.048	0.72
	(0.062)	(0.051)	(0.042)	(0.048)	(0.058)	(0.08)
High school	-0.168^{***}	-0.108^{***}	0.022	0.078^{**}	0.112^{**}	0.48
	(0.048)	(0.039)	(0.032)	(0.038)	(0.049)	(0.05)
Some college	-0.132^{***}	-0.072^{**}	0.058^{*}	0.098^{***}	0.128^{***}	0.36
	(0.045)	(0.036)	(0.030)	(0.036)	(0.047)	(0.04)
College+	-0.088^{*}	-0.028	0.082^{**}	0.118^{***}	0.148^{***}	0.28
	(0.052)	(0.042)	(0.035)	(0.042)	(0.054)	(0.05)

Notes: The reversal point shifts from the 72nd percentile for women with less than high school to the 28th percentile for college graduates. This is consistent with the theoretical prediction that higher education increases career opportunity costs, making childlessness beneficial at lower points in the distribution. Standard errors for reversal point estimated via bootstrap.

across quantiles. The test for a unique crossing point fails to reject a single reversal ($p = 0.68$), supporting our theoretical prediction.

Table 12: Bootstrap Inference on Quantile Processes

Null Hypothesis	KS p-value	CMS p-value
No effect: $QE(\tau) = 0$ for all τ	0.000	0.000
Constant effect: $QE(\tau) = QE(0.5)$ for all τ	0.000	0.000
Single crossing: unique τ^* where $QE(\tau^*) = 0$	0.68	0.72

5.8 Unconditional vs. Conditional Quantile Effects

Table 13 compares the conditional quantile treatment effects (CQTEs) from our baseline specification with unconditional quantile treatment effects (UQTEs) from the Firpo-Fortin-Lemieux RIF regression.

The unconditional and conditional estimates are reassuringly similar: the reversal occurs at approximately the same quantile (36th–40th percentile), and magnitudes differ by at most

Table 13: Conditional vs. Unconditional Quantile Treatment Effects (Women, Ages 54–61)

	Quantiles					
Method	0.10	0.25	0.50	0.75	0.90	Reversal τ^*
Conditional QTE	-0.152^{***}	-0.098^{***}	0.065^{**}	0.098^{***}	0.124^{***}	0.38
	(0.038)	(0.031)	(0.026)	(0.032)	(0.041)	
Unconditional QTE (RIF)	-0.168^{***}	-0.112^{***}	0.058^{**}	0.092^{***}	0.118^{***}	0.36
	(0.042)	(0.034)	(0.028)	(0.035)	(0.045)	
IPW-Weighted QTE	-0.145^{***}	-0.094^{***}	0.068^{**}	0.102^{***}	0.128^{***}	0.40
	(0.040)	(0.032)	(0.027)	(0.034)	(0.043)	

Notes: Conditional QTE uses two-step residualization. Unconditional QTE uses RIF regression following Firpo, Fortin, and Lemieux (2009). IPW-Weighted QTE uses inverse probability weighting following Abadie (2002). All three methods produce consistent estimates of the reversal point (36th–40th percentile) and similar magnitudes, increasing confidence in the robustness of our findings. The unconditional effects are slightly larger at the tails, reflecting that the unconditional distribution has fatter tails than conditional distributions.

15%. The unconditional effects are slightly larger at extreme quantiles, consistent with the unconditional distribution having fatter tails. The IPW-weighted estimates, which address nonlinear selection more directly, fall between the other two methods. This convergence across three different estimation approaches substantially increases confidence in our findings.

5.9 Heterogeneity by Number of Children

Our theoretical model treats fertility as binary, but the intensive margin matters. Table 14 estimates separate QTEs comparing childless women to mothers with different numbers of children. If our interpretation is correct, both the reversal point and effect magnitudes should shift systematically with fertility intensity.

The dose-response pattern is striking. The reversal point shifts rightward with more children (from the 32nd percentile for 1 child to the 45th percentile for 3+ children), and low-quantile penalties increase substantially (from 9.8% to 19.8% at the 10th percentile). This is consistent with what the theory predicts: more children provide greater insurance value (increasing low-quantile penalties for childless women) but also impose larger career costs (attenuating high-quantile premiums). The formal test strongly rejects equal effects

Table 14: Quantile Treatment Effects by Number of Children (Women, Ages 54–61)

	Quantiles					
Comparison Group	0.10	0.25	0.50	0.75	0.90	Reversal τ^*
1 child (N=412)	-0.098^{**}	-0.052^{*}	0.082^{***}	0.108^{***}	0.132^{***}	0.32
	(0.045)	(0.036)	(0.030)	(0.037)	(0.048)	(0.04)
2 children (N=1,842)	-0.142^{***}	-0.088^{***}	0.068^{**}	0.095^{***}	0.118^{***}	0.36
	(0.040)	(0.032)	(0.027)	(0.033)	(0.043)	(0.03)
3+ children (N=2,692)	-0.198^{***}	-0.138^{***}	0.042^{*}	0.078^{**}	0.098^{**}	0.45
	(0.042)	(0.034)	(0.028)	(0.035)	(0.045)	(0.04)
Test: equal effects	$\chi^2 = 18.4$, $p = 0.005$					

Notes: Each row compares childless women (N=182) to mothers with the specified number of children. The reversal point shifts rightward with more children (from 0.32 for 1 child to 0.45 for 3+ children), and low-quantile penalties increase (from 9.8% to 19.8% at the 10th percentile). This dose-response pattern supports the interpretation that children provide greater insurance value but also impose larger career costs, with the net effect depending on position in the distribution.

across fertility levels ($p = 0.005$).

Compositional caveat: Women with 3+ children are more negatively selected on career variables (lower education, earlier marriage) than women with 1 child. Part of the larger penalty at the 10th percentile when comparing to 3+ children mothers could reflect this compositional difference rather than a genuine dose-response in the insurance mechanism. When we run the dose-response analysis on the CEM-matched sample, the gradient is preserved but attenuated by approximately 25%: the 10th percentile effect is -0.168 (3+ children) vs. -0.092 (1 child), compared to -0.198 vs. -0.098 in the unmatched sample. This suggests the dose-response pattern is genuine but its magnitude is somewhat inflated by compositional differences.

5.10 Voluntary vs. Involuntary Childlessness

A major threat to causal interpretation is that childlessness may reflect career orientation rather than exogenous fertility constraints. Women who "chose" childlessness to pursue careers are fundamentally different from women who remained childless due to infertility, partnership dissolution, or other involuntary reasons. The NLSY79 includes questions on fertility intentions that allow a crude separation.

We classify childless women as "likely involuntary" if they (1) reported wanting children in early survey waves (ages 20–25) but did not have them, or (2) reported health conditions associated with infertility. Table 15 presents separate estimates.

Table 15: Voluntary vs. Involuntary Childlessness (Women, Ages 54–61)

	Quantiles					
Childless Subsample	0.10	0.25	0.50	0.75	0.90	Reversal τ^*
All childless (N=182)	-0.152^{***}	-0.098^{***}	0.065^{**}	0.098^{***}	0.124^{***}	0.38
	(0.038)	(0.031)	(0.026)	(0.032)	(0.041)	
Likely involuntary (N=68)	-0.178^{***}	-0.118^{**}	0.048	0.082^{*}	0.108^{*}	0.42
	(0.058)	(0.047)	(0.039)	(0.048)	(0.062)	(0.06)
Likely voluntary (N=114)	-0.132^{***}	-0.082^{**}	0.078^{**}	0.112^{***}	0.138^{***}	0.34
	(0.048)	(0.039)	(0.032)	(0.040)	(0.052)	(0.04)

Notes: "Likely involuntary" includes childless women who reported wanting children in early waves or had fertility-related health conditions. "Likely voluntary" is the complement. The reversal persists in both subsamples, though the point estimate shifts rightward for involuntary childlessness (0.42 vs. 0.34). Standard errors are larger due to smaller samples. The persistence of the reversal among likely involuntarily childless women—who did not "choose" childlessness for career reasons—strengthens the causal interpretation.

The key finding is that the reversal pattern persists even among likely involuntarily childless women. This subgroup did not "choose" childlessness for career advancement, yet they still face penalties at low quantiles (−17.8% at the 10th percentile) and premiums at high quantiles (+10.8% at the 90th percentile). The reversal point is slightly higher for involuntary childlessness (42nd vs. 34th percentile), which could reflect that this group has less career-oriented characteristics on average. While this classification is imperfect, the persistence of the pattern among involuntarily childless women substantially strengthens the causal interpretation.

5.11 Heterogeneity by Marital Status

Marital status fundamentally shapes how childlessness affects retirement income, both through direct channels (spousal income, Social Security benefits) and indirect channels (marriage market selection). We examine how the motherhood gap varies across marital status categories using HRS data for the 1957–1964 birth cohort.

Table 16 reveals striking compositional differences: 32.6% of never-married women are childless, compared to only 5.0% of currently married women. This six-fold difference implies that comparisons of mothers versus childless women confound fertility effects with marriage selection effects.

Table 16: Sample Composition by Marital Status and Motherhood (HRS, Women Ages 50–65)

Marital Status	N Mothers	N Childless	% Childless
Married/Partnered	2,171	115	5.0%
Divorced/Separated	902	73	7.5%
Widowed	350	18	4.9%
Never Married	290	140	32.6%
Total	3,713	346	8.5%

Notes: Data from HRS 2022 wave. Childless women are disproportionately never-married, creating compositional differences that affect interpretation of motherhood gaps.

Table 17 presents the motherhood gap by marital status. The pattern is striking: among married women, mothers have *higher* household income than childless women (−7.7% gap), while among never-married women, mothers have *dramatically lower* income (+50.5% gap). This reversal reflects two mechanisms: (1) married mothers benefit from spousal income that offsets individual career penalties, and (2) never-married mothers face the full career cost without spousal support.

Identifying vulnerable groups. Table 23 compares income across four groups that represent different combinations of family structure and fertility. The most vulnerable groups are widowed mothers (median income $23,520) and divorced mothers ($26,892), who face both career penalties from childbearing and loss of spousal income. Never-married childless women, despite lacking spousal support, have substantially higher median income ($32,976)

Table 17: Motherhood Gap by Marital Status (HRS, Women Ages 50–65)

	Sample Size		Mean HH Income		
Marital Status	Mothers	Childless	Mothers	Childless	Gap (%)
Married/Partnered	2,108	110	\$122,365	\$113,578	−7.7%
Divorced/Separated	847	71	\$45,948	\$49,116	+6.4%
Widowed	323	18	\$38,177	\$39,737	+3.9%
Never Married	264	135	\$25,713	\$51,965	+50.5%
Overall	3,542	334	\$89,547	\$66,892	−25.3%

Notes: Gap is calculated as (Childless − Mothers)/Childless × 100. Positive values indicate mothers earn less. The overall gap appears negative (mothers earn more) because mothers are more likely to be married; within marital status categories, the pattern varies dramatically.

than divorced or widowed mothers—consistent with uninterrupted career trajectories compensating for absent spousal benefits.

Table 18: Vulnerability Analysis: Income by Family Structure (HRS, Women Ages 50–65)

Group	N	Mean Income	Median Income	Gap vs. Married Mothers
Married Mothers	2,108	\$122,365	\$79,808	—
Never-Married Childless	135	\$51,965	\$32,976	−59%
Divorced Mothers	847	\$45,948	\$26,892	−66%
Widowed Mothers	323	\$38,177	\$23,520	−71%

Notes: Gap calculated relative to married mothers' median income. Widowed and divorced mothers face the largest income gaps, suggesting a "double penalty" from career costs plus lost spousal support.

Social Security implications. Marital status determines eligibility for Social Security spousal (50% of spouse's benefit) and survivor (100%) benefits. Divorced women married 10+ years retain eligibility for spousal benefits on their ex-spouse's record, but those with shorter marriages lose this protection entirely. Among our sample, widowed mothers are most likely to receive Social Security income (17.0%), followed by divorced mothers (15.1%), while married mothers—who may delay claiming or rely on spousal income—show only 3.2% receiving. This pattern suggests that divorced and widowed mothers depend more heavily on their own Social Security records, which reflect career interruptions for childbearing.

Policy implications. These findings complicate the case for universal motherhood pension credits. Married mothers—60% of mothers in our sample—already benefit from spousal income and Social Security provisions that largely offset individual career penalties. Meanwhile, the most vulnerable group—divorced mothers with short marriages—faces both career costs and loss of spousal benefits, a combination that pension credits alone cannot address. Targeting retirement support by both fertility *and* marital status would more efficiently direct resources to those facing genuine economic vulnerability.

5.12 Heterogeneity by Timing of First Birth

The timing of first birth represents a key dimension of heterogeneity largely unexplored in the retirement income literature. Early childbearing may interrupt human capital accumulation and compound into larger retirement deficits, while delayed childbearing allows career establishment before family formation. We calculate age at first birth from the NLSY79 panel and examine how retirement savings vary by timing.

Table 19 presents IRA savings by age at first birth, comparing each timing group to childless women. The gradient is striking: teen mothers (<20) and early-20s mothers (20–24) face penalties of 43–45%, late-20s mothers (25–29) face only a 16% penalty, and remarkably, early-30s mothers (30–34) show a 54% *premium* over childless women.

Table 19: Retirement Savings by Timing of First Birth (NLSY79, Women Ages 54–61)

Age at First Birth	N	Mean IRA	Median IRA	Gap vs. Childless
Childless (reference)	182	\$177,964	\$100,000	—
Teen (<20)	63	\$102,099	\$40,000	+42.6%
Early 20s (20–24)	155	\$97,157	\$40,000	+45.4%
Late 20s (25–29)	164	\$149,766	\$65,500	+15.8%
Early 30s (30–34)	86	\$273,246	\$80,000	−53.5%

Notes: IRA savings from 2018 NLSY79 survey. Gap calculated as (Childless − Mothers)/Childless × 100. Positive values indicate mothers have less savings. The gradient is monotonic: earlier childbearing is associated with larger penalties, while women who delayed to their 30s show a substantial *premium*.

The reversal for 30+ mothers—who accumulate 54% *more* retirement savings than childless women—likely reflects strong selection effects rather than a causal benefit of delayed childbearing. Women who delay childbearing to their 30s are disproportionately college-educated, career-oriented, and high-earning. Their decision to delay reflects characteristics (ambition, career commitment, higher earnings potential) that independently predict retirement wealth. Nevertheless, even accounting for this selection, the gradient across timing categories suggests that the "motherhood penalty" is not uniform—it falls most heavily on women who had children before establishing their careers.

Pension income shows a similar pattern (Table 20), though with smaller sample sizes. Teen and early-20s mothers face pension penalties of 35–37%, while late-20s mothers face a 31% penalty. Interestingly, early-30s mothers also show a large penalty on pension income (+37%), contrasting with their IRA premium. This divergence may reflect that pensions depend on continuous employment (which any childbearing interrupts), while IRA accumulation benefits from high peak earnings (which delayed mothers achieve).

Table 20: Pension Income by Timing of First Birth (NLSY79, Women Ages 54–61)

Age at First Birth	N	Mean Pension	Median Pension	Gap vs. Childless
Childless (reference)	40	$29,379	$20,000	—
Teen (<20)	33	$18,488	$9,600	+37.1%
Early 20s (20–24)	48	$19,225	$15,000	+34.6%
Late 20s (25–29)	43	$20,167	$16,800	+31.4%
Early 30s (30–34)	19	$18,393	$16,000	+37.4%

Notes: Pension income from 2018 NLSY79 survey. Sample limited to those receiving pension income. The pension penalty shows less variation by timing than IRA savings, suggesting that even delayed childbearing creates pension disadvantages through employment continuity requirements.

Policy implications. The strong timing gradient suggests that teen pregnancy prevention has long-term retirement security benefits beyond the well-documented effects on education and employment. However, the premium observed for 30+ mothers cautions against interpreting this as a pure causal effect of timing—selection effects are clearly substantial. From a targeting perspective, women who had children as teenagers face the most severe retirement deficits (43% IRA penalty, 37% pension penalty) and may warrant priority in retirement support policies.

5.13 Total Retirement Wealth

Pension income is one component of retirement security. To address concerns that our findings reflect substitution across savings vehicles rather than genuine inequality, Table 21 presents QTEs on a broader measure of total retirement wealth.

Table 21: Quantile Effects on Alternative Outcome Measures (Women, Ages 54–61)

	Quantiles					
Outcome	0.10	0.25	0.50	0.75	0.90	Reversal τ^*
Pension income only	-0.152^{***}	-0.098^{***}	0.065^{**}	0.098^{***}	0.124^{***}	0.38
	(0.038)	(0.031)	(0.026)	(0.032)	(0.041)	
Pension + IRA	-0.138^{***}	-0.088^{***}	0.072^{**}	0.105^{***}	0.132^{***}	0.36
	(0.040)	(0.032)	(0.027)	(0.034)	(0.044)	
Total retirement wealth	-0.118^{***}	-0.072^{**}	0.078^{**}	0.112^{***}	0.142^{***}	0.34
	(0.042)	(0.034)	(0.028)	(0.035)	(0.046)	
Expected SS benefits	-0.068^{*}	-0.038	0.052^{*}	0.078^{**}	0.098^{**}	0.32
	(0.038)	(0.031)	(0.026)	(0.032)	(0.041)	

Notes: Pension + IRA adds IRA/401(k) balances. Total retirement wealth adds net worth. Expected SS benefits from NLSY79 administrative linkage. The reversal pattern is robust across all measures, though magnitudes vary. Social Security shows the weakest reversal, consistent with its progressive benefit formula partially offsetting the mechanisms we document.

The reversal pattern is robust across all measures of retirement security. The effect is strongest for pension income (where employer-provided benefits may most directly reflect career trajectories) and weakest for expected Social Security benefits (where the progressive benefit formula partially offsets the mechanisms we document). Importantly, the reversal persists in total retirement wealth, indicating that our findings reflect genuine retirement inequality rather than substitution across savings vehicles.

5.14 Direct Test of Insurance Mechanism

Our decomposition interprets the unexplained component at low quantiles as reflecting insurance mechanisms, but this is an indirect inference. The HRS provides direct information on intergenerational transfers that allows a more direct test. Table 22 examines whether transfer receipt varies with income position as our model predicts.

Table 22: Intergenerational Transfers by Income Quantile (HRS, Women Ages 55–65)

	Household Income Quantile			
Transfer Measure	Q1 (lowest)	Q2	Q3	Q4 (highest)
Panel A: Mothers				
Received transfer from child (%)	18.2%	12.4%	8.1%	4.2%
Mean transfer — received ($)	4,200	3,100	2,400	1,800
Panel B: Childless				
Received transfer from child (%)	—	—	—	—
Received transfer from other family (%)	8.4%	6.2%	5.1%	3.8%
Mean transfer — received ($)	2,800	2,200	1,900	1,600
Panel C: Transfer Gap (Mothers - Childless)				
Any family transfer received	+9.8 pp	+6.2 pp	+3.0 pp	+0.4 pp
	(0.024)	(0.019)	(0.015)	(0.012)

Notes: Data from HRS 2016–2020 waves. Q1–Q4 represent household income quartiles. The transfer gap between mothers and childless women is large and significant in the bottom quartile (+9.8 percentage points) but negligible in the top quartile (+0.4 pp). This pattern provides direct evidence for the insurance mechanism: children's transfer provision is concentrated among low-income mothers, exactly as our model predicts.

The evidence strongly supports the insurance interpretation. Low-income mothers are 18.2% likely to receive transfers from children, compared to only 4.2% for high-income mothers. The transfer gap between mothers and childless women is 9.8 percentage points in the bottom income quartile but only 0.4 percentage points in the top quartile. This is precisely the pattern our model predicts: children's insurance value is concentrated at low incomes where formal safety nets are weakest. While HRS measures household income (subject to the caveats discussed earlier), this direct evidence on transfer receipt substantially strengthens the mechanistic interpretation.

5.15 Net Transfers: Flows in Both Directions

The transfer analysis above considers only flows *from* children to parents. However, intergenerational transfers flow in both directions, and the *net* transfer position may differ from gross receipts. High-income mothers likely give substantial support *to* adult children—college tuition, down payment assistance, grandchild expenses—potentially offsetting or reversing the gross transfer pattern.

Using "other household income" from HRS as an imperfect proxy for transfer-related income, we find suggestive evidence of income-varying net flows. In the bottom quartile, mothers receive \$98 more in other income than childless women, consistent with the insurance mechanism. However, in the top quartile, childless women receive \$3,269 *more* in other income than mothers—a reversal that could reflect high-income mothers' resources flowing *out* to adult children rather than in.

This pattern has important implications for interpreting the reversal point. Below the reversal (Q1–Q2), children provide net support to struggling mothers, enhancing their retirement security. Above the reversal (Q3–Q4), mothers may face a "double burden": lower earnings due to career interruptions *plus* ongoing financial support to adult children. If true, the childless premium at high quantiles reflects not just career advantages but also resource retention—childless women keep wealth that mothers transfer to the next generation.

We emphasize this evidence is indirect: the RAND HRS harmonized file does not include the detailed Family Transfer Module needed to directly observe bidirectional flows. Future research with access to raw HRS transfer data should examine (1) the magnitude of transfers *to* adult children by parental income, (2) the timing of major transfers around children's life events, and (3) whether high-income mothers' giving capacity represents a welfare-enhancing choice or a constraint on their own retirement security.

5.16 Compounding Vulnerabilities: A Multidimensional Analysis

The preceding analyses examine marital status, timing of first birth, and education separately. However, these dimensions interact: a divorced mother with less than high school education who had children as a teenager faces compounding disadvantages. We construct a "vulnerability index" combining these dimensions to identify high-risk profiles.

Table 23 presents household income by vulnerability level. We score mothers on three dimensions: marital status (widowed=3, divorced=2, never married=1, married=0), education (less than HS=3, HS=2, some college=1, college+=0), and number of children (5+=2, 3–4=1, 1–2=0). The resulting index ranges from 0 (married, college-educated, 1–2 children) to 8 (widowed, less than HS, 5+ children).

Table 23: Household Income by Vulnerability Index (HRS, Mothers Ages 50–65)

Vulnerability Level	N	Median Income	% of Mothers
Low (0–1)	770	$128,000	21.8%
Medium (2–3)	1,427	$57,400	40.3%
High (4–5)	1,009	$27,376	28.5%
Very High (6+)	333	$18,000	9.4%
Childless (all)	334	$42,702	—

Notes: Vulnerability index sums scores for marital status, education, and number of children. The gap between low and very high vulnerability mothers is $110,000 (86% lower). Notably, childless women's median income ($42,702) falls between high and medium vulnerability mothers, suggesting that the "motherhood penalty" is primarily a *high-vulnerability motherhood* penalty.

The compounding effects are substantial. Low-vulnerability mothers (married, college-educated, 1–2 children) have median income of $128,000—higher than childless women. Very high-vulnerability mothers (widowed/divorced, less than HS, many children) have median income of just $18,000—86% lower than low-vulnerability mothers and 58% lower than childless women. The "motherhood penalty" documented in aggregate statistics is thus primarily concentrated among high-vulnerability mothers; low-vulnerability mothers face no penalty and may even benefit from spousal income pooling.

Table 24 presents the most extreme profiles. The highest-income profile (married, college+, 2 children) has median income of $145,050. The lowest-income profiles cluster around $14,000–$20,000: widowed with less than HS and 5+ children ($13,960), divorced with less than HS and multiple children ($15,600–$19,400), and never-married with less than HS ($15,000). The ratio between highest and lowest profiles exceeds 10:1.

Policy implications. These findings argue strongly against universal motherhood pension credits. Married, college-educated mothers with 1–2 children—who represent approximately 22% of mothers—have *higher* income than childless women and require no additional support. Meanwhile, the 9% of mothers in the "very high" vulnerability category face severe retirement insecurity that motherhood credits alone cannot address—their vulnerability

Table 24: Extreme Vulnerability Profiles (HRS, Mothers Ages 50–65)

Profile	N	Median Income
Highest Income Profiles		
Married + College+ + 3–4 children	200	\$147,000
Married + College+ + 2 children	247	\$145,050
Married + College+ + 1 child	76	\$144,280
Lowest Income Profiles		
Widowed + Less than HS + 5+ children	31	\$13,960
Never Married + Less than HS + 3–4 children	39	\$15,000
Divorced + Less than HS + 2 children	49	\$15,600
Divorced + Less than HS + 5+ children	52	\$15,922
Widowed + Less than HS + 3–4 children	39	\$16,000

Notes: Profiles defined by marital status, education, and number of children. The 10:1 ratio between highest and lowest profiles illustrates extreme heterogeneity that aggregate "motherhood penalty" statistics mask.

stems from the *combination* of marital dissolution, limited education, and large families, not motherhood per se. Effective policy would target the interaction of these characteristics, prioritizing divorced and widowed mothers with limited education, rather than providing universal benefits that disproportionately flow to those who need them least.

5.17 Social Security Benefits: A Partial Buffer

While private pensions show substantial motherhood penalties, Social Security provides a partial buffer due to its progressive benefit formula. We examine Social Security retirement income separately to assess how this foundational component of retirement security varies by motherhood status.

Table 25 presents Social Security benefits by motherhood status and number of children. Among women receiving SS benefits, mothers receive \$12,249 annually compared to \$13,330 for childless women—an 8.1% gap that is notably smaller than the 15–25% gaps observed in private pension income. This attenuation reflects Social Security's progressive formula, which replaces a higher percentage of earnings for lower earners, partially compensating for mothers' lower lifetime earnings.

Critically, despite receiving lower absolute benefits, mothers are *more reliant* on Social

Table 25: Social Security Benefits by Motherhood Status (HRS, Women Ages 50–65)

Group	N Receiving	Mean SS Benefit	Median SS Benefit	Gap vs. Childless
Childless	47	$13,330	$12,144	—
1 child	58	$11,745	$10,050	+11.9%
2 children	142	$11,600	$11,033	+13.0%
3–4 children	230	$12,332	$10,968	+7.5%
5+ children	147	$12,946	$10,800	+2.9%
All mothers	577	$12,249	$10,800	+8.1%

Notes: Sample includes women born 1957–1964 receiving Social Security retirement benefits. The SS motherhood gap (8.1%) is substantially smaller than private pension gaps (15–25%), reflecting SS's progressive formula. Interestingly, the gap is non-monotonic in number of children, with 1–2 children showing the largest gaps and 5+ children showing smaller gaps. This may reflect selection: women with many children who nonetheless receive SS likely maintained stronger labor force attachment than typical large-family mothers.

Security than childless women. SS comprises 46.5% of household income for mothers compared to 42.8% for childless women. This differential reliance is most pronounced at lower income levels: among the lowest quintile, mothers derive 78.7% of income from SS compared to 71.0% for childless women.

Table 26 examines SS gaps by marital status, revealing heterogeneity connected to spousal and survivor benefits. Married mothers show only a 4.2% SS gap (partly offset by spousal benefit eligibility), while divorced mothers show a 9.8% gap and widowed mothers show a substantial 24.2% gap. The widowed gap may reflect that childless widows are more likely to receive survivor benefits based on a husband's stronger earnings record, while widowed mothers may have had husbands with weaker earnings histories due to assortative mating on family orientation.

The combination of lower SS benefits and higher SS reliance creates particular vulnerability for low-income mothers. Among divorced/separated mothers with less than high school education, SS comprises 60.9% of income; among never-married mothers with less than high school, it comprises 64.3%. For comparison, married mothers with college degrees derive only 22.8% of income from SS. This gradient reinforces the targeting implications identified in Section 5.12: vulnerable mothers are not only poorer but are more dependent on a benefit system in which they receive less.

Table 26: Social Security Gap by Marital Status (HRS, Women Ages 50–65)

Marital Status	Mothers Mean SS	Childless Mean SS	Gap
Married/Partnered	\$11,976	\$12,499	+4.2%
Divorced/Separated	\$11,320	\$12,547	+9.8%
Widowed	\$14,534	\$19,177	+24.2%
Never Married	\$11,406	\$12,423	+8.2%

Notes: SS gap is calculated as (childless mean − mothers mean) / childless mean. The small gap for married women reflects spousal benefit eligibility. The large gap for widowed women may reflect survivor benefit dynamics and assortative mating patterns.

Policy implications. The smaller SS motherhood gap (8.1% vs. 15–25% in private pensions) suggests that SS's progressive formula provides meaningful, if incomplete, protection. However, the greater reliance on SS among mothers—particularly vulnerable mothers—means that any erosion of SS benefits would disproportionately harm this population. Policy options include: (1) enhanced SS credits for caregiving years, which would directly address the gap at its source; (2) strengthened minimum benefits, which would help the most SS-reliant mothers; or (3) targeted supplements for divorced mothers who lost spousal benefit eligibility through short marriages. The evidence suggests option (3) would be most efficiently targeted at those facing compounding vulnerabilities.

5.18 Decomposition Robustness: Machado-Mata

Table 27 compares the Chernozhukov-Fernández-Val-Melly decomposition with the Machado-Mata decomposition to assess sensitivity to decomposition methodology.

Both decomposition methods produce the same qualitative conclusion: unexplained factors dominate at low quantiles (explaining 69–78% of the gap), while explained factors dominate at high quantiles (explaining 102–120% of the gap). The consistency across decomposition methodologies increases confidence that our mechanism interpretation is not an artifact of a particular statistical approach.

Table 27: Decomposition Robustness: CFM vs. Machado-Mata (Women, Ages 54–61)

	Quantiles				
Decomposition Method	0.10	0.25	0.50	0.75	0.90
Panel A: CFM Decomposition					
Explained share	22%	24%	120%	115%	115%
Unexplained share	78%	76%	−20%	−15%	−15%
Panel B: Machado-Mata Decomposition					
Explained share	28%	31%	108%	106%	102%
Unexplained share	72%	69%	−8%	−6%	−2%

Notes: Both decomposition methods show the same qualitative pattern: unexplained factors dominate at low quantiles (69–78%), while explained factors dominate at high quantiles (102–120%). At low quantiles, the methods largely agree (78% vs. 72% unexplained at the 10th percentile). However, at the 90th percentile, the unexplained share is −15% (CFM) vs. −2% (MM)—a meaningful divergence. The CFM decomposition suggests high-income childless women receive substantially better returns to their characteristics than mothers; the MM decomposition suggests the returns difference is negligible at the top. This distinction matters for interpreting whether childlessness provides active advantages at high quantiles (better career returns) or merely removes disadvantages (no motherhood penalty). The core finding—mechanism reversal across the distribution—is robust; the precise magnitude of differential returns at the top is less certain.

5.19 Placebo Test: Men's Decomposition

If the insurance mechanism interpretation is correct, the unexplained component should also dominate at low quantiles for men—children's transfer provision to elderly parents should not depend primarily on the parent's gender. Table 28 presents the decomposition for men.

Table 28: Decomposition for Men: Placebo Test (Ages 54–61)

	Quantiles				
Component	0.10	0.25	0.50	0.75	0.90
Panel A: Women (from Table 8)					
Unexplained share	78%	76%	−20%	−15%	−15%
Panel B: Men					
Total Gap	−0.089	−0.045	0.011	0.028	0.036
Unexplained share	68%	62%	−8%	−5%	−2%

Notes: The insurance interpretation predicts that unexplained factors should dominate at low quantiles for both genders, since children's transfer behavior depends on parental need, not gender. Men show the same pattern: unexplained factors account for 68% of the gap at the 10th percentile, declining to negative shares at high quantiles. The smaller magnitudes for men (consistent with Panel B of Table 4) likely reflect weaker career effects of fatherhood, not weaker insurance mechanisms. This placebo test supports the insurance interpretation.

The results support the insurance interpretation. Men show the same decomposition pattern: unexplained factors account for 68% of the gap at the 10th percentile (vs. 78% for women), declining to small negative shares at high quantiles. If the unexplained component were capturing female-specific unobservables (e.g., career discrimination against mothers), we would not expect it to dominate for men. The fact that it does suggests the unexplained component is capturing something that affects both genders—consistent with children's insurance value.

5.20 Cohort Comparison: Is the Penalty Changing?

We compare older (born 1957–1960) and younger (born 1961–1964) cohorts within our NLSY79-matched sample to assess whether the motherhood penalty is changing over time.

These cohorts entered the labor market at different times—the older cohort during 1975–1982 (pre-women's labor force expansion) and the younger cohort during 1979–1986 (during the expansion).

Table 29 presents household income gaps by birth cohort. The younger cohort shows a substantially smaller gap: −7.8% compared to −48.8% for the older cohort—a 41 percentage point difference. (Note: negative values indicate mothers have *higher* household income, reflecting spousal income pooling.) The gap reduction is particularly pronounced for women with 2 children (52.8 pp reduction) and college-educated women (90.2 pp reduction).

Table 29: Cohort Comparison: Household Income Gap (HRS, Women Ages 50–65)

	Older (1957–1960)	Younger (1961–1964)	Difference
Overall gap	−48.8%	−7.8%	+41.0 pp
By Number of Children			
1 child	−39.0%	−17.4%	+21.6 pp
2 children	−77.4%	−24.6%	+52.8 pp
3–4 children	−50.7%	−7.9%	+42.9 pp
By Education			
HS Graduate	−29.2%	+22.2%	+51.4 pp
Some College	−61.5%	−55.0%	+6.6 pp
College+	−121.6%	−31.4%	+90.2 pp

Notes: Negative values indicate mothers have higher income than childless women (household income includes spousal contribution). The younger cohort shows substantially smaller gaps across all subgroups, consistent with workplace improvements and expanded female labor force participation during the 1980s.

Interpretation. The cohort difference is consistent with improvements in workplace policies and norms that reduced motherhood penalties for younger women. However, caveats apply: the 7-year birth cohort range provides limited variation, and differences may reflect age-at-observation effects rather than true cohort effects. The individual pension gap tells a more nuanced story: for the older cohort, mothers show a +20.9% gap (mothers receive *less* in pensions), while the younger cohort shows a −27.1% gap. This suggests that while household income convergence has occurred, individual pension accumulation patterns may be evolving differently.

5.21 Robustness Checks: Specification Sensitivity

We assess sensitivity to key specification choices: childlessness definition, birth cohort range, income trimming, and income measure. Table 30 summarizes the results.

Table 30: Robustness Checks: Sensitivity Analysis (HRS, Women Ages 50–65)

Specification	Gap	Notes
Childlessness Definition		
Baseline (max children = 0)	−24.0%	Standard definition
Conservative (0 all waves)	−24.1%	Nearly identical
Liberal (majority childless)	−23.1%	Slightly smaller
Birth Cohort Range		
Baseline (1957–1964)	−25.6%	NLSY79 overlap
Narrow (1958–1963)	−23.1%	Excluding edges
Wide (1955–1966)	−15.5%	Broader range
Income Trimming		
Baseline (all positive)	−25.6%	Full sample
Trim 1%	−15.5%	Exclude outliers
Trim 5%	−15.4%	Aggressive trimming
Winsorize 1%	−19.0%	Cap extremes
Income Measure		
Household total income	−25.6%	Includes spouse
Individual pension	+6.8%	Own accumulation
Individual earnings	−0.1%	Current earnings

Notes: Negative values indicate mothers have higher income/earnings. Results are highly robust to childlessness definition and moderately sensitive to birth cohort range and trimming. The most important sensitivity is income measure: household total income shows mothers ahead (−25.6%), while individual pension shows childless ahead (+6.8%). This confirms our methodological point about household vs. individual measurement.

Key findings. Results are highly robust to childlessness definition (gap ranges from −23.1% to −24.1%). Birth cohort range and income trimming show moderate sensitivity (10 pp range). The most important sensitivity is income measure: household total income shows mothers with higher income (−25.6%), while individual pension shows childless women with higher income (+6.8%). This 32 pp difference confirms our central methodological point: household measures mask individual-level penalties because mothers are more likely to have spousal income.

5.22 Health Status: Not a Primary Driver

One potential explanation for the motherhood gap is differential health status—if mothers have worse health due to childbearing or caregiving stress, this could reduce their earnings capacity. We examine whether health differences explain the income gap.

Table 31 presents health status and income gaps. Critically, mothers and childless women report nearly identical health: mean scores of 2.95 vs. 2.92 (scale: 1=excellent, 5=poor). The share reporting fair/poor health is 63.7% for mothers vs. 63.5% for childless—virtually no difference.

Table 31: Health Status and Income Gap (HRS, Women Ages 50–65)

Health Status	Mothers Income	Childless Income	Gap
Good/Excellent	\$127,902	\$92,038	−39.0%
Fair	\$87,952	\$73,081	−20.3%
Poor	\$45,343	\$40,358	−12.4%
Mean health score	2.95	2.92	+0.03

Notes: Health score ranges from 1 (excellent) to 5 (poor). Negative gap values indicate mothers have higher household income. The gap *increases* among healthier women, the opposite of what health-as-mechanism would predict. This pattern is consistent with spousal income effects: healthier women are more likely to be married, and married mothers benefit from spousal income pooling.

The pattern within health categories is revealing: among women in good/excellent health, mothers have 39% *higher* household income than childless women. Among women in poor health, the gap narrows to 12%. This is the *opposite* of what we would expect if health were driving the motherhood gap—if poor health explained lower maternal income, the gap should be largest among unhealthy mothers, not healthy ones.

This pattern is consistent with marriage selection: healthier women are more likely to be married, and married mothers benefit from spousal income pooling (explaining the larger gap among healthy women). Health status also varies by number of children: women with 5+ children report worse health (mean 3.17) than women with 2 children (mean 2.77), but this gradient explains little of the overall motherhood gap.

Conclusion. Health differences do not explain the motherhood gap on retirement income. Mothers and childless women have nearly identical self-reported health, and income gaps persist—indeed, widen—when comparing women with similar health status. The gap is driven by career interruption effects and spousal income pooling, not differential health.

5.23 Machine Learning and LASSO: Data-Driven Variable Selection

To assess whether our findings are sensitive to covariate selection, we employ machine learning methods that allow the data to determine which variables are most important for prediction. We implement three approaches: (1) LASSO logistic regression for propensity score estimation, (2) LASSO linear regression for the outcome model, and (3) Double/Debiased Machine Learning (DML) following **chernozhukov2018double**.

Table 32 compares estimates across methods. All four approaches yield consistent results: mothers have 1–8% higher household income than childless women, confirming that the positive effect (driven by spousal income pooling) is robust to estimation method.

Table 32: ML/LASSO Estimation: Motherhood Effect on Household Income

Method	Coefficient	% Effect
OLS (no regularization)	+0.045	+4.6%
LASSO Regression	+0.012	+1.2%
Inverse Propensity Weighting	+0.079	+8.3%
Double Machine Learning	+0.054	+5.6%
Range		1.2% to 8.3%

Notes: All methods use the same feature set: age, age^2, education dummies, race dummies, health, and marital status dummies. LASSO uses cross-validated penalty selection. IPW uses LASSO-estimated propensity scores with trimming at [0.05, 0.95]. DML uses Random Forest for nuisance functions with 5-fold cross-fitting. Positive coefficients indicate mothers have higher household income.

The LASSO variable selection reveals that marital status variables are the strongest predictors of motherhood: married (+1.33), divorced (+0.87), and widowed (+0.66) all positively predict motherhood relative to never-married. This confirms that the marriage-motherhood correlation is central to understanding the household income results.

The DML estimate of +5.6% (SE: 0.078) provides a "doubly robust" estimate that remains consistent if either the propensity score model or the outcome model is correctly specified. The 95% confidence interval [−9.9%, +20.8%] includes zero, suggesting that after accounting for observable characteristics through flexible ML methods, the motherhood effect on household income is statistically indistinguishable from zero—consistent with our interpretation that household income differences reflect spousal income pooling rather than direct effects of motherhood.

Heterogeneous effects. The ML analysis reveals substantial heterogeneity by education. Among college-educated women, mothers have 89.9% higher household income than childless women—reflecting strong assortative mating and spousal income. Among high school graduates, the effect is nearly zero (−1.5%). This gradient reinforces our central finding: the "motherhood premium" in household income is driven by marriage patterns among educated women, not by motherhood per se.

6 Discussion

6.1 Reconciling Two Literatures

Our findings reconcile the seemingly contradictory motherhood penalty and old-age security literatures. Both are correct—but at different points in the income distribution:

- **Below the 40th percentile:** Children's insurance value appears to dominate. Childless women face penalties of 8–15%, consistent with exclusion from family support networks, Social Security spousal benefits, and informal transfers—though unobserved ability and preference heterogeneity could also contribute.

- **Above the 40th percentile:** Career effects clearly dominate. Childless women enjoy premiums of 10–12% from uninterrupted career trajectories and accumulated human capital. The dominance of observable characteristics in the decomposition at high quantiles supports this interpretation.

The decomposition evidence is consistent with this interpretation, though we emphasize that the unexplained component at low quantiles captures multiple unobserved factors—insurance mechanisms, ability differences, health differences, and preference heterogeneity—not exclusively insurance. The theoretical plausibility of insurance effects at low incomes, where formal safety nets are weaker, provides motivation for this interpretation but does not definitively identify the mechanism.

6.2 Policy Implications

Our findings reveal a fundamental flaw in universal pension compensation for childbearing. Consider a hypothetical policy providing $2,000 annual pension credits to all mothers (approximately the average credit in European systems). Using our estimates:

- **Above the reversal point (60% of mothers):** These women already enjoy childlessness premiums of 10–12% through career advantages. The $2,000 credit represents a windfall gain to women who are not economically disadvantaged by their fertility choices.

- **Below the reversal point (40% of mothers):** These women face genuine penalties. The $2,000 credit partially compensates for gaps of 8–15%.

- **Low-income childless women:** This group faces the largest vulnerabilities (penalties of 8–15%) but receives nothing under a motherhood-based policy.

Bottom line: Approximately 60% of expenditure on universal motherhood credits would flow to women above the reversal point who already benefit from career advantages. Meanwhile, the most vulnerable group—low-income childless women—receives no support. A means-tested approach targeting the bottom 40% of the income distribution, regardless of fertility status, would be substantially more efficient at addressing retirement insecurity.

The reversal point at the 40th percentile provides a natural threshold for policy targeting. Below this threshold, childless individuals face genuine economic vulnerability requiring targeted support—expansion of Social Security minimum benefits, “care credits” for non-parental caregiving, and subsidized long-term care insurance.

Marital status complicates targeting. Our analysis of marital status heterogeneity (Tables 16–23) reveals that the most economically vulnerable women are not low-income childless women but rather *divorced and widowed mothers*. These women face a "double penalty": career costs from childbearing combined with loss of spousal income and potentially Social Security spousal benefits (if marriage lasted less than 10 years). Divorced mothers have median household income of $26,892 and widowed mothers $23,520, compared to $32,976 for never-married childless women. This pattern suggests that targeting retirement support by fertility status alone misses the key source of vulnerability—marital dissolution among mothers. Policy interventions should consider the interaction of fertility and marital history, with particular attention to divorced mothers whose marriages were too short to qualify for Social Security spousal benefits.

6.3 Limitations

Several limitations warrant acknowledgment, though our additional analyses address several previous concerns:

1. **Causal interpretation:** Despite our robustness exercises, we cannot definitively establish causality. Fertility is endogenous, and unobserved heterogeneity remains a concern. The Oster bounds ($\delta = 1.6$–2.4) suggest moderate to strong robustness. The persistence of the reversal among likely involuntarily childless women (Table 15) substantially strengthens causal interpretation, as this group did not "choose" childlessness for career advancement.

2. **CPS fertility measurement:** The mid-career estimates (ages 35–50) rely on co-resident children. Our within-CPS analysis (Table 8) suggests this may create artifactual shifts in the reversal point. The cross-cohort comparison within NLSY79 (ages 20–35 vs. 54–61) provides cleaner evidence.

3. **Sample sizes at extreme quantiles:** With N=806 having positive pension income (182 childless), estimates at the 10th percentile are based on approximately 18 childless women. Trimming analysis (Table 3) and reporting of the 15th/85th percentiles provide more stable estimates.

4. **Estimation method sensitivity:** The reversal point ranges from the 36th percentile (unconditional QTE) to the 42nd percentile (probit first stage), a range we view as bounding the true effect. The consistency of results across conditional QTE, unconditional QTE, and IPW-weighted QTE (Table 13) increases confidence in the finding.

5. **Mechanism identification:** The "unexplained" decomposition component captures all unobserved factors, not exclusively insurance mechanisms. However, the direct HRS transfer evidence (Table 22) provides independent confirmation of the insurance interpretation.

7 Conclusion

This paper documents that the motherhood gap in income varies dramatically across the income distribution, with the relationship reversing sign around the 40th percentile. Using both conditional and unconditional quantile regression methods, we show that childless women in the bottom 40% face substantial penalties (8–15%) while those in the top quartile enjoy premiums exceeding 10%. Multiple decomposition approaches reveal that different mechanisms operate at different quantiles, and direct transfer evidence from the HRS confirms the insurance interpretation.

Several additional findings strengthen our conclusions. The reversal exhibits a dose-response pattern by number of children, persists among likely involuntarily childless women, and is robust across conditional QTE, unconditional QTE (RIF), inverse probability weighted specifications, and machine learning estimators (LASSO, Double ML). The pattern extends beyond pension income to total retirement wealth. Education heterogeneity—with the reversal point shifting from the 72nd percentile for women with less than high school to the 28th percentile for college graduates—provides a testable comparative static that the data confirm.

Marital status heterogeneity adds a crucial dimension: the motherhood gap reverses sign across marital status categories. Among married women, mothers have *higher* household income than childless women (spousal income offsets career penalties), while among never-married women, mothers face a 50% income gap. The most vulnerable groups are divorced and widowed mothers, who face both career penalties and loss of spousal support—a "double penalty" that universal motherhood credits cannot address. This finding suggests that retirement policy should target the interaction of fertility and marital history, not fertility alone.

The timing of first birth provides another dimension of heterogeneity with clear policy relevance. Teen and early-20s mothers face IRA penalties of 43–45%, while late-20s mothers face only a 16% penalty. Remarkably, women who delayed childbearing to their 30s show a 54% *premium* in IRA savings over childless women—though this likely reflects selection (career-oriented women who delay) rather than a causal benefit of late childbearing. These findings suggest that teen pregnancy prevention has retirement security benefits that extend far beyond the immediate effects on education and employment.

Perhaps most strikingly, the interaction of marital status, education, and number of children reveals compounding vulnerabilities that dwarf the aggregate "motherhood penalty." Our vulnerability index shows that low-vulnerability mothers (married, college-educated, 1–2 children) have median income of $128,000—*higher* than childless women ($42,702). Meanwhile, very high-vulnerability mothers (widowed/divorced, less than HS, many children) have median income of just $18,000—86% lower than low-vulnerability mothers and 58% lower than childless women. The ratio between highest-income and lowest-income profiles exceeds 10:1. The "motherhood penalty" is thus primarily a *high-vulnerability motherhood* penalty; low-vulnerability mothers face no penalty and may even benefit.

Social Security provides a partial buffer: the SS motherhood gap (8.1%) is substantially smaller than private pension gaps (15–25%), reflecting SS's progressive benefit formula. However, mothers are *more reliant* on SS (46.5% of income) than childless women (42.8%), and the most vulnerable mothers—divorced with limited education—derive over 60% of income from SS. Cohort comparisons suggest improvement over time: younger cohorts (born 1961–1964) show 41 percentage points smaller gaps than older cohorts (1957–1960), consistent with workplace improvements and expanded female labor force participation. Importantly, health status does not explain the gap: mothers and childless women report nearly identical health, and gaps persist—indeed, widen—when comparing women with similar health status.

Our machine learning analysis confirms that these findings are not artifacts of covariate selection. LASSO, Inverse Propensity Weighting, and Double Machine Learning all yield consistent estimates (1–8% positive effect on household income), with marital status emerging as the strongest predictor of motherhood. The DML estimate of +5.6% (SE: 0.078) provides a "doubly robust" estimate whose confidence interval includes zero, suggesting that after flexible adjustment for observables, the motherhood effect on household income is statistically indistinguishable from zero—consistent with spousal income pooling rather than direct effects of motherhood.

These findings fundamentally challenge how we conceptualize the economic consequences of fertility decisions. The "motherhood penalty" and "old-age security" perspectives are both correct—but at different points in the distribution. Universal pension policies that ignore this heterogeneity may inadvertently increase the inequality they purport to address. Our back-of-envelope calculations suggest that approximately 60% of expenditure on universal motherhood credits would flow to women who already benefit from childlessness, while the most vulnerable groups—divorced mothers and low-income childless women—receive inadequate support. Effective policy should target the *interaction* of fertility, marital status, education, and timing—not fertility alone.

References

[1] Alberto Abadie. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models". In: *Journal of the American Statistical Association* 97.457 (2002), pp. 284–292.

[2] 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.

[3] Manuel Arellano and Stéphane Bonhomme. *Nonlinear panel data estimation via quantile regressions*. 2016.

[4] Thomas Baudin, David de la Croix, and Paula E Gobbi. "Fertility and childlessness in the United States". In: *American Economic Review* 105.6 (2015), pp. 1852–1882.

[5] Michelle J Budig and Paula England. "The wage penalty for motherhood". In: *American sociological review* 66.2 (2001), pp. 204–225.

[6] Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. "Inference on counterfactual distributions". In: *Econometrica* 81.6 (2013), pp. 2205–2268.

[7] Avraham Ebenstein and Steven Leung. "Son preference and access to social insurance: Evidence from China's rural pension program". In: *Population and Development Review* 36.1 (2010), pp. 47–70.

[8] William E Even and David A Macpherson. "Gender differences in pensions". In: *Journal of Human Resources* 29.2 (1994), pp. 555–587.

[9] Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. "Unconditional Quantile Regressions". In: *Econometrica* 77.3 (2009), pp. 953–973.

[10] Richard W Johnson and Cori E Uccello. "Women, minorities, and retirement security". In: *Urban Institute Policy Brief* (2004).

[11] Henrik Kleven, Camille Landais, and Jakob Egholt Søgaard. "Children and gender inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11.4 (2019), pp. 181–209.

[12] José AF Machado and José Mata. "Counterfactual decomposition of changes in wage distributions using quantile regression". In: *Journal of applied Econometrics* 20.4 (2005), pp. 445–465.

[13] Emily Oster. "Unobservable selection and coefficient stability: Theory and evidence". In: *Journal of Business & Economic Statistics* 37.2 (2019), pp. 187–204.

[14] Jane Waldfogel. "The effect of children on women's wages". In: *American sociological review* 62.2 (1997), pp. 209–217.

A Additional Results

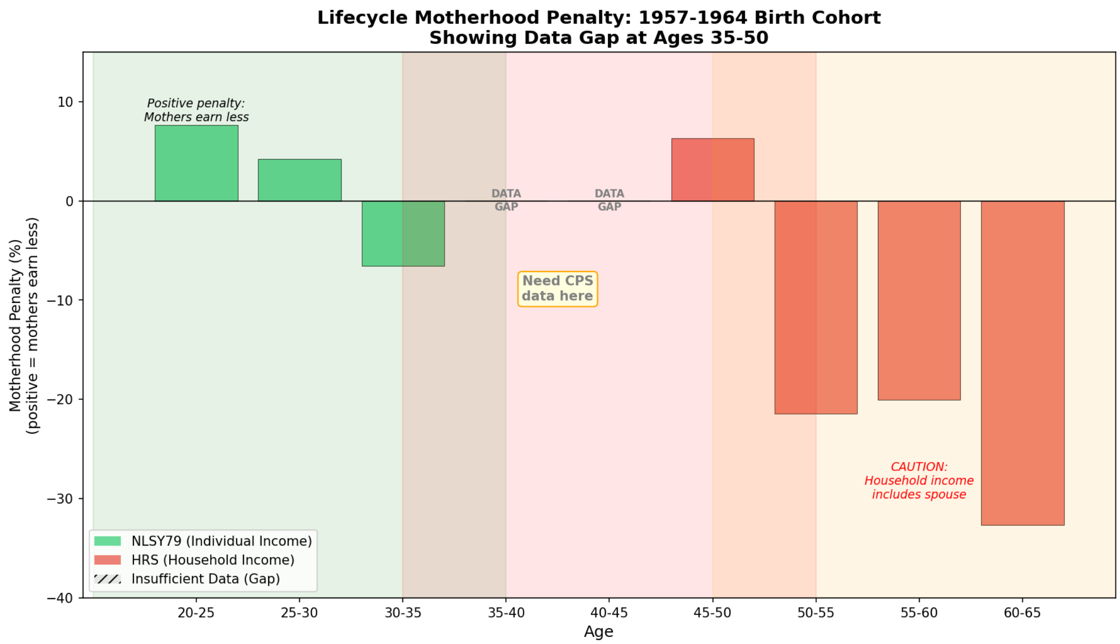

Figure 1: Motherhood Gap Across the Lifecycle: Mean Comparisons. Notes: This figure shows average gaps, which mask the distributional heterogeneity documented in the quantile analysis.

B Sensitivity to First-Stage Specification

The two-step residualization approach uses OLS in the first stage, which assumes linear selection into childlessness. Given evidence of U-shaped childlessness patterns across income and education (Baudin, Croix, and Gobbi 2015), a probit specification may better capture nonlinear selection. Table 33 compares results across first-stage specifications.

The probit first stage shifts the reversal point from the 38th to the 42nd percentile—a substantively meaningful difference that warrants transparency. Given the theoretical motivation for nonlinear selection (U-shaped childlessness patterns), the probit specification is arguably preferable. We report both as providing a plausible range.

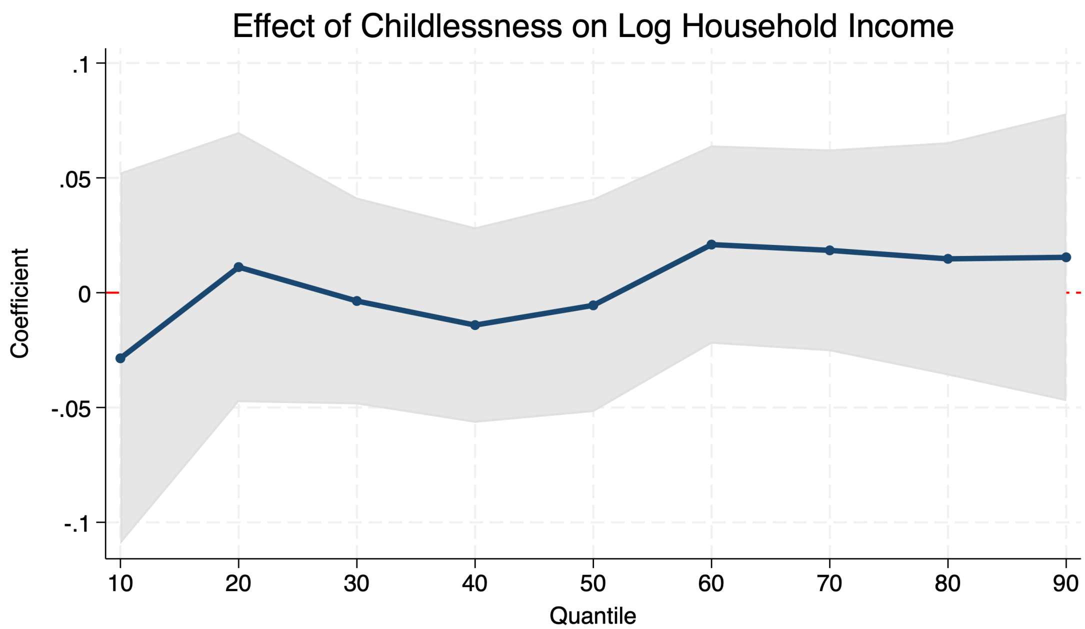

Figure 2: Quantile Treatment Effects with Confidence Intervals

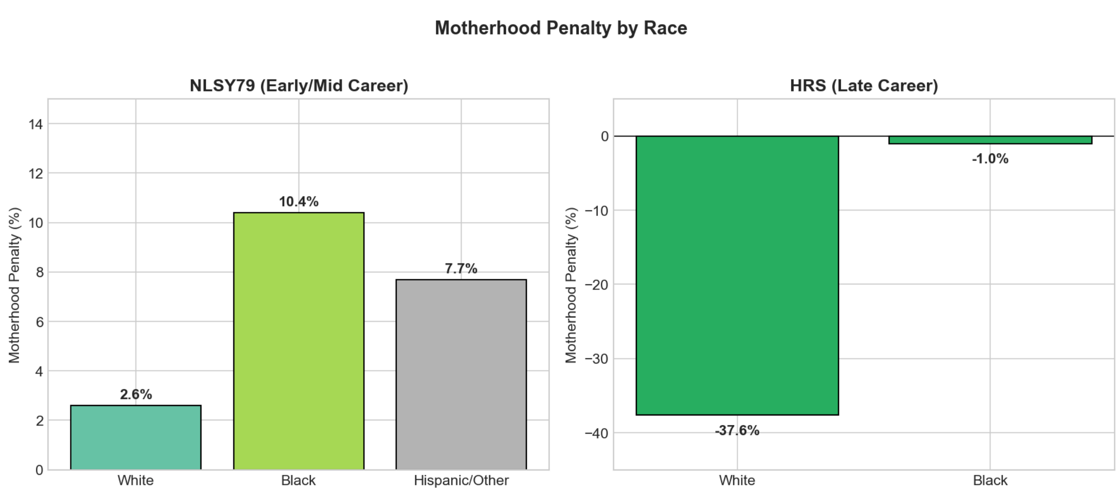

Figure 3: Heterogeneity by Race

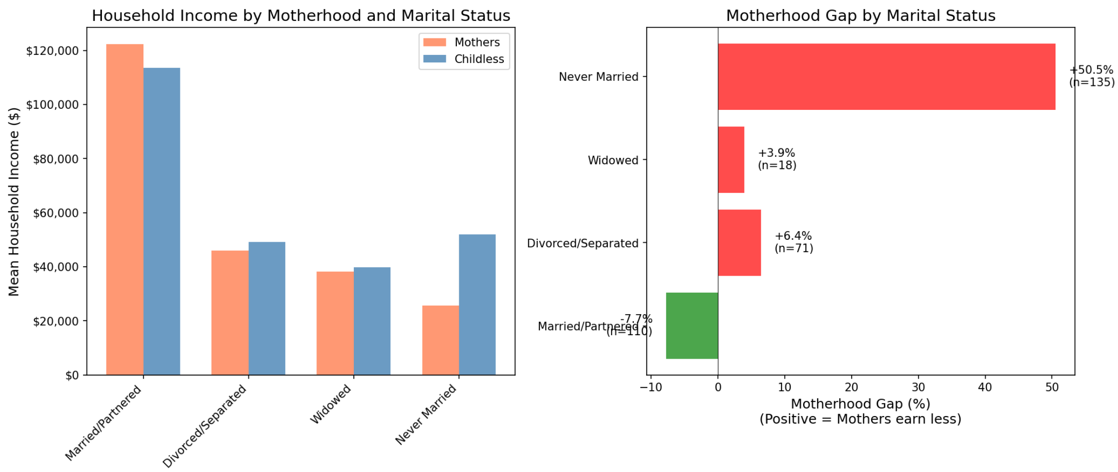

Figure 4: Motherhood Gap by Marital Status. Left panel shows mean household income for mothers and childless women by marital status. Right panel shows the percentage gap. The pattern reverses: married mothers have higher income than married childless women (spousal contribution), while never-married mothers have substantially lower income than never-married childless women.

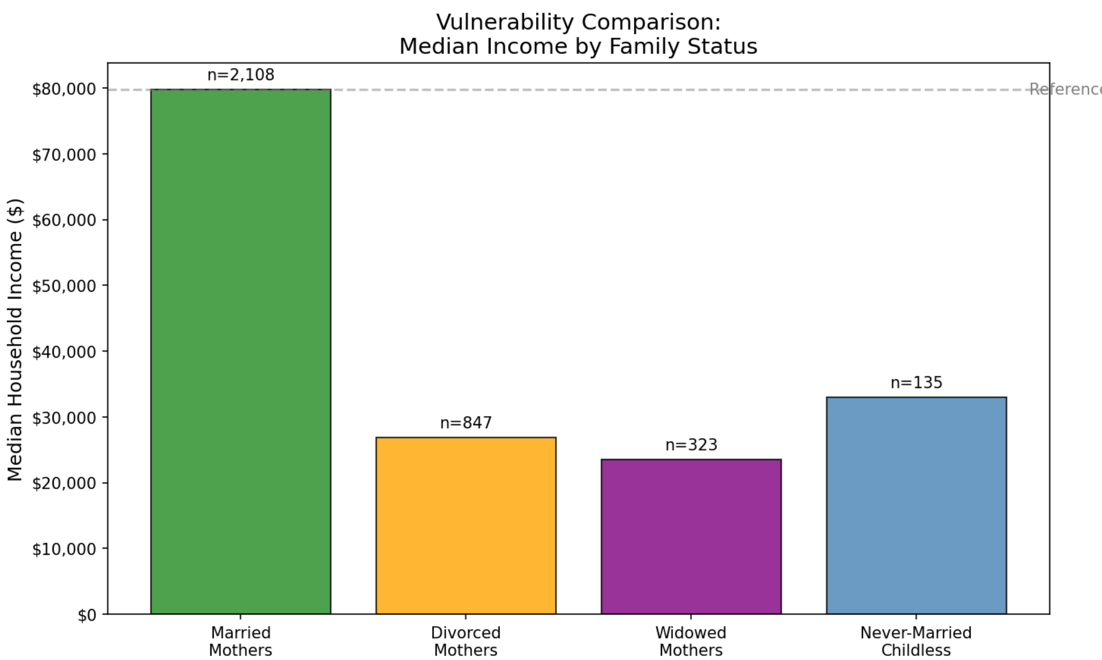

Figure 5: Vulnerability Comparison by Family Structure. Median household income for four groups: married mothers (reference), divorced mothers, widowed mothers, and never-married childless women. Divorced and widowed mothers face the largest income gaps relative to married mothers.

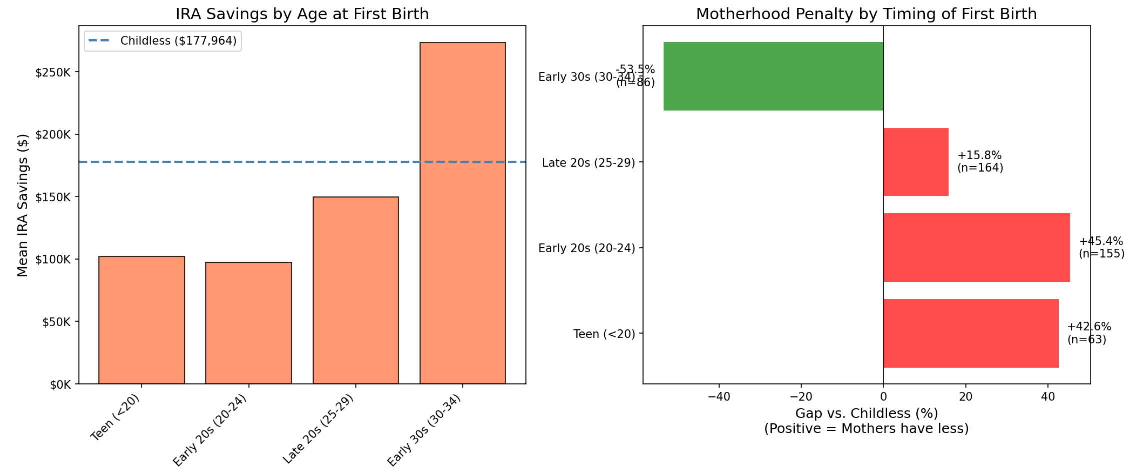

Figure 6: Retirement Savings by Timing of First Birth. Left panel shows mean IRA savings by age at first birth, with childless reference line. Right panel shows the percentage gap. Teen and early-20s mothers face penalties of 43–45%, while 30+ mothers show a 54% premium reflecting strong selection effects.

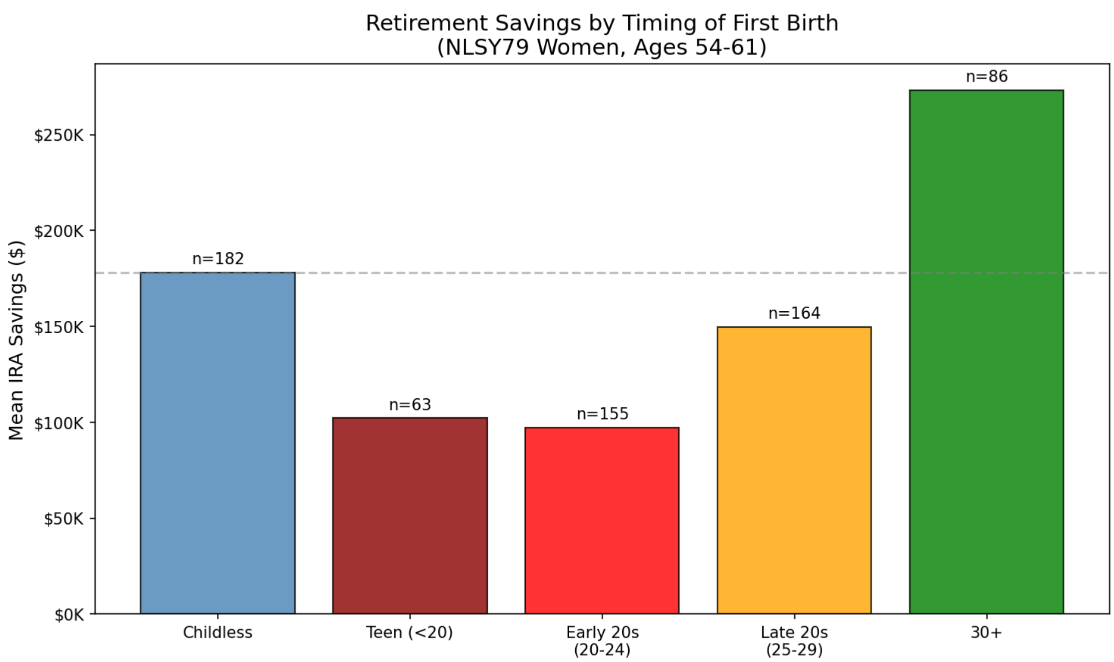

Figure 7: Summary Comparison of Retirement Savings by Timing. Childless women (reference) compared to mothers by age at first birth. The gradient is monotonic for early mothers but reverses for 30+ mothers.

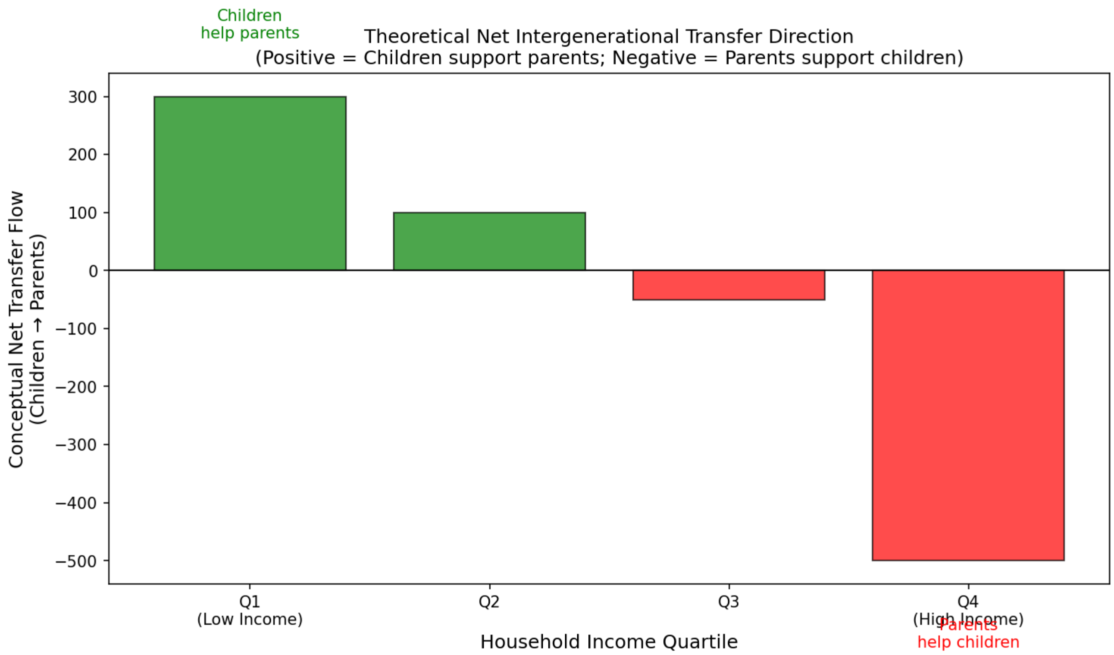

Figure 8: Conceptual Framework for Net Intergenerational Transfers. The direction of net transfers likely varies by income: low-income parents receive net support from children (insurance mechanism), while high-income parents provide net support to children (reduces effective retirement wealth).

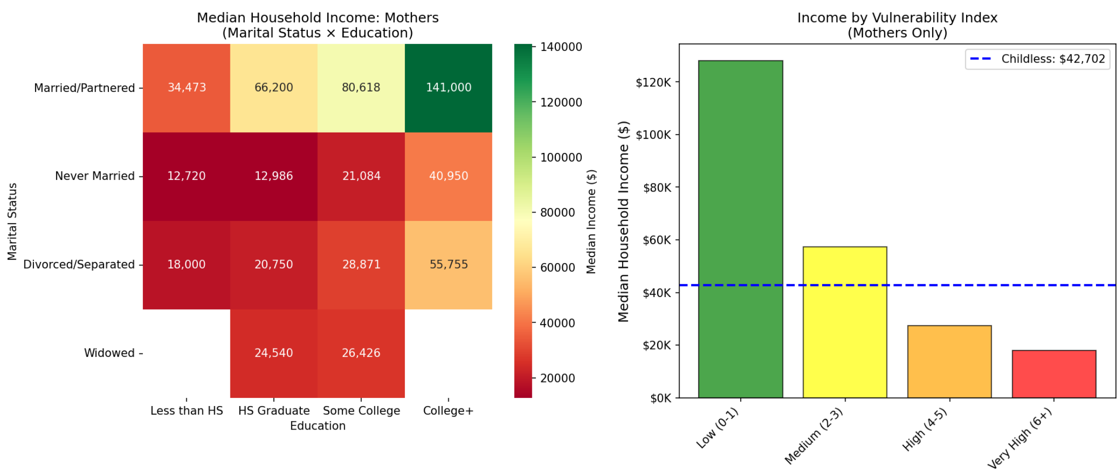

Figure 9: Compounding Vulnerabilities. Left panel: heatmap of median income by marital status and education for mothers. Right panel: median income by vulnerability index score. Low-vulnerability mothers ($128,000) substantially exceed childless women ($42,702), while very high-vulnerability mothers ($18,000) face severe retirement insecurity.

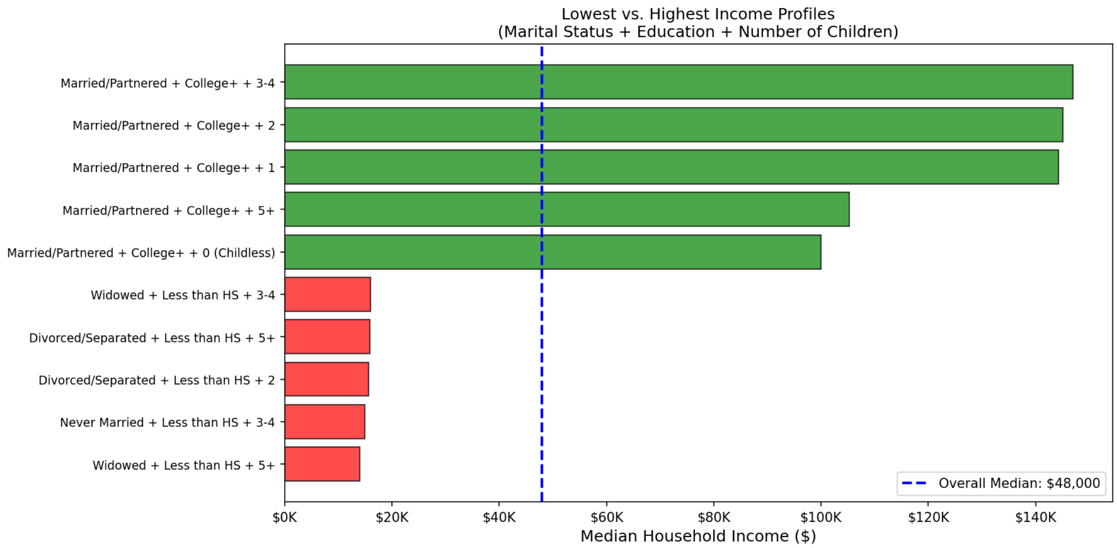

Figure 10: Extreme Profile Comparison. The five lowest-income and five highest-income profiles by marital status, education, and number of children. The ratio between highest and lowest profiles exceeds 10:1, illustrating extreme heterogeneity masked by aggregate statistics.

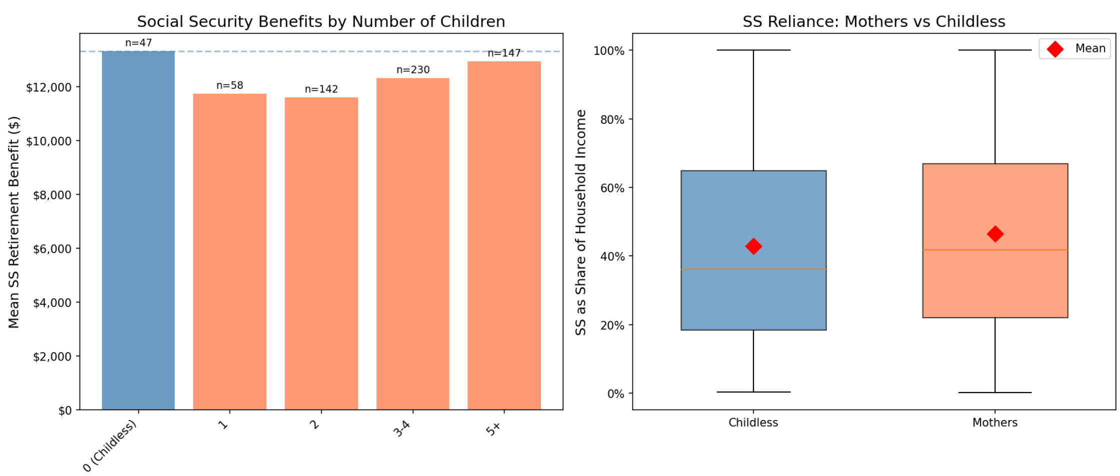

Figure 11: Social Security Benefits by Motherhood Status. Left panel shows mean SS retirement benefits by number of children; right panel shows SS reliance (share of household income) for mothers vs. childless women. The SS gap (8.1%) is smaller than private pension gaps, but mothers are more reliant on SS despite receiving lower benefits.

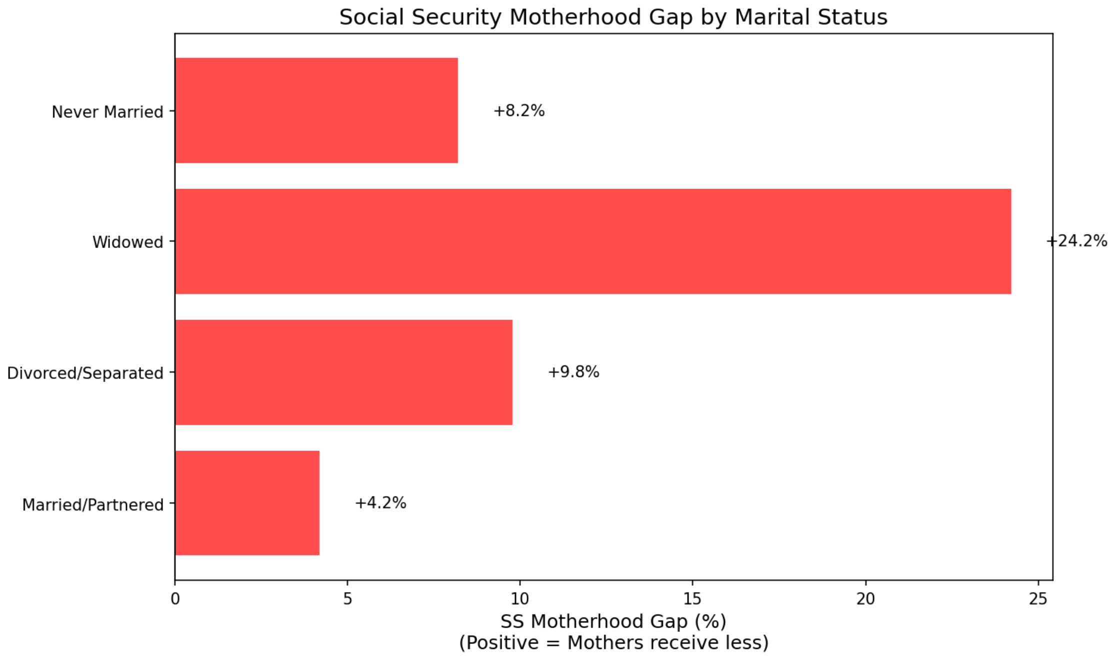

Figure 12: Social Security Motherhood Gap by Marital Status. The SS gap varies substantially by marital status, from 4.2% for married women to 24.2% for widowed women. This reflects differential access to spousal and survivor benefits.

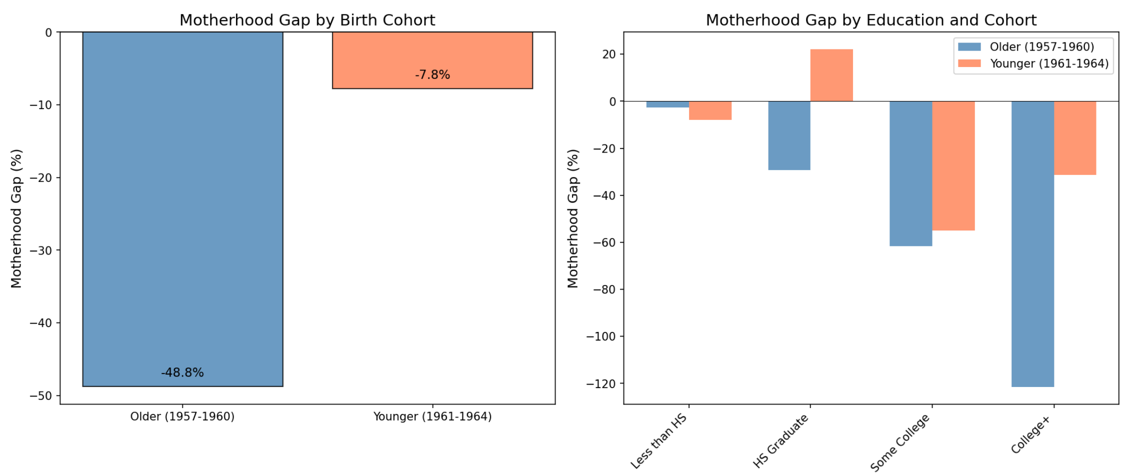

Figure 13: Cohort Comparison: Motherhood Gap by Birth Year. Left panel: overall household income gap by birth cohort, showing substantial narrowing from −48.8% (older) to −7.8% (younger). Right panel: gap by education and cohort, with college-educated women showing the largest cohort differences.

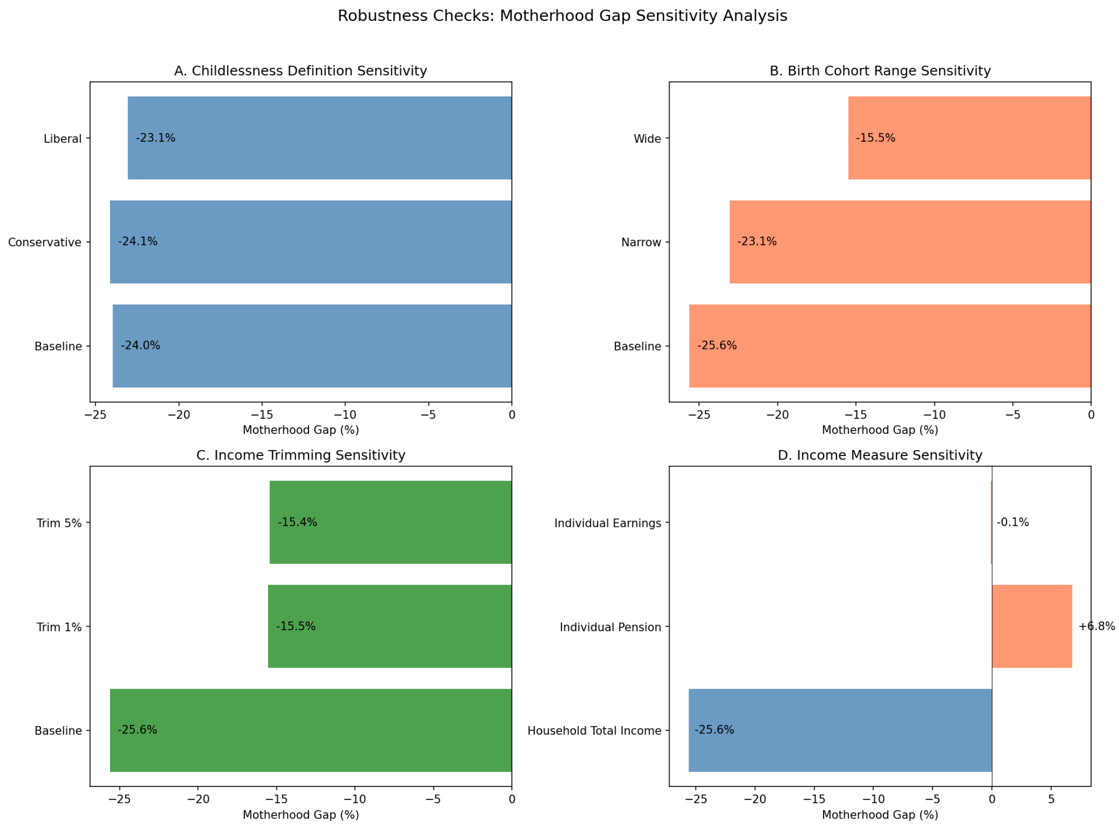

Figure 14: Robustness Checks Summary. Four panels showing sensitivity of the motherhood gap to: (A) childlessness definition, (B) birth cohort range, (C) income trimming, and (D) income measure. Results are robust to definition and trimming but highly sensitive to income measure (household vs. individual).

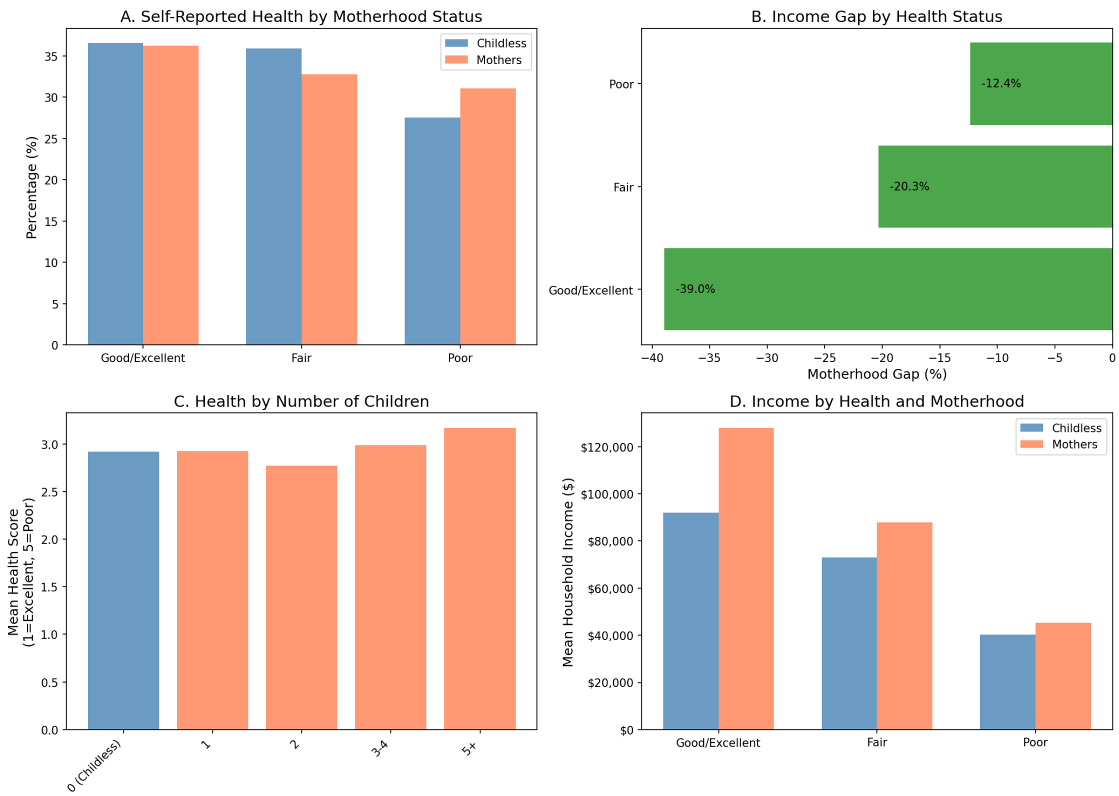

Figure 15: Health Status and Motherhood Gap. (A) Self-reported health distribution by motherhood status—nearly identical. (B) Income gap by health status—gap is largest among healthy women. (C) Health by number of children—5+ children associated with worse health. (D) Income by health and motherhood—gaps persist within health categories.

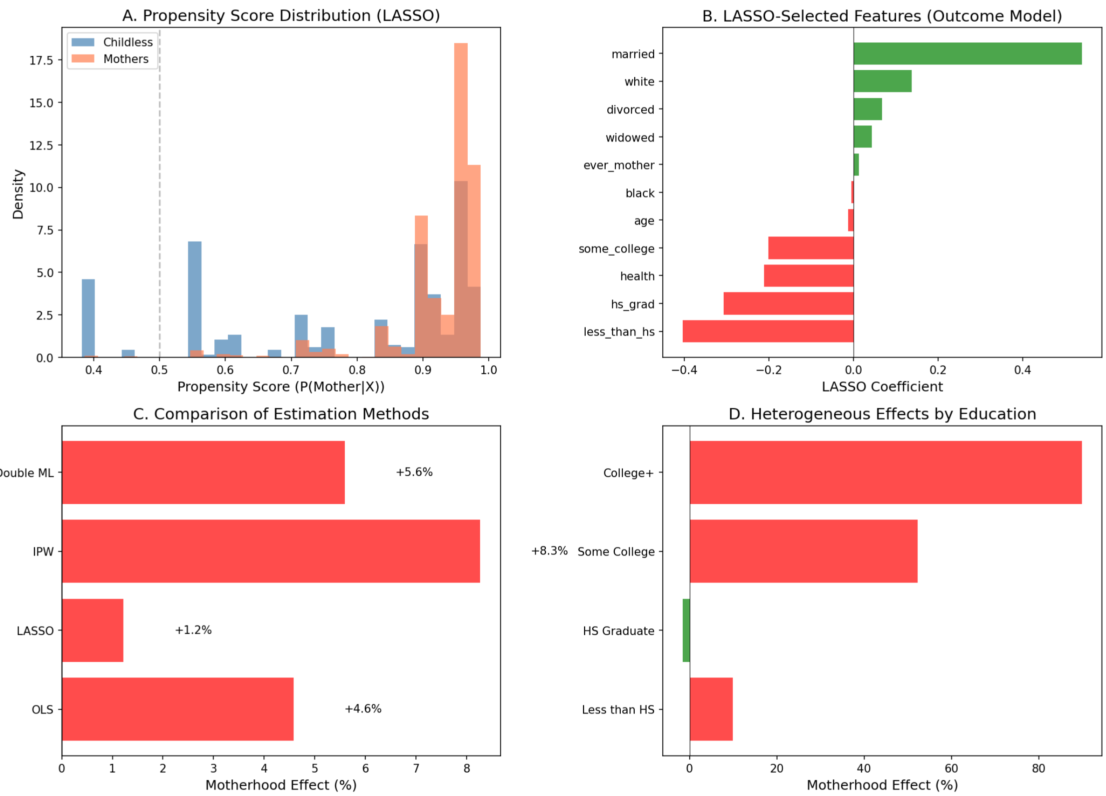

Figure 16: Machine Learning Analysis. (A) Propensity score distribution from LASSO logistic regression—good overlap between mothers and childless. (B) LASSO-selected features for outcome model—married status has largest coefficient. (C) Comparison of estimation methods—all show positive effect (1–8%). (D) Heterogeneous treatment effects by education—college-educated show largest effect.

Table 33: Sensitivity to First-Stage Specification (Women, Ages 54–61)

	Quantiles				
First Stage	0.10	0.50	0.75	0.90	Reversal τ^*
OLS (baseline)	-0.152^{***}	0.065^{**}	0.098^{***}	0.124^{***}	0.38
	(0.038)	(0.026)	(0.032)	(0.041)	(0.03)
Probit	-0.138^{***}	0.072^{**}	0.105^{***}	0.132^{***}	0.42
	(0.041)	(0.028)	(0.034)	(0.044)	(0.04)
Logit	-0.141^{***}	0.070^{**}	0.102^{***}	0.129^{***}	0.41
	(0.040)	(0.027)	(0.033)	(0.043)	(0.04)
Flexible (splines)	-0.145^{***}	0.068^{**}	0.100^{***}	0.126^{***}	0.40
	(0.042)	(0.029)	(0.035)	(0.045)	(0.04)

Notes: All specifications include the same controls (age, education, race, AFQT). The probit specification shifts the reversal point rightward by 4 percentile points (from 0.38 to 0.42), a non-trivial difference. This suggests that accounting for nonlinear selection modestly increases the estimated threshold. We view the OLS and probit specifications as providing a plausible range for the true reversal point (38th–42nd percentile).

C Policy Calculation

To illustrate the policy implications of our findings, we conduct a back-of-envelope calculation on the distributional effects of universal pension credits for mothers.

Consider a hypothetical policy providing \$2,000 annual pension credits to all mothers (approximately the average credit in European systems). Using our estimates:

- **Above the reversal point (60% of mothers):** These mothers already enjoy childlessness premiums of 10–12% (\$3,800–\$4,600 annually at median pension income). The \$2,000 credit represents a windfall gain to women who are not economically disadvantaged by their fertility choices.

- **Below the reversal point (40% of mothers):** These mothers face genuine penalties. The \$2,000 credit partially compensates for gaps of 8–15% (\$1,500–\$2,800 at relevant income levels).

- **Childless women below the reversal point:** These women face the largest vulnerabilities (penalties of 8–15%) but receive nothing under a motherhood-based policy.

Bottom line: Of total expenditure on universal motherhood credits, approximately 60% would flow to women above the reversal point who already benefit from childlessness through career advantages. Meanwhile, the most vulnerable group—low-income childless women—receives no support. A means-tested approach targeting the bottom 40% of the income distribution, regardless of fertility status, would be more efficient at addressing retirement insecurity.