



The heterogeneous effects of the first childbirth on women's income

Afrouz Azadikhah Jahromi¹ · Weige Huang²

Received: 19 December 2023 / Accepted: 23 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024,

Abstract

This study estimates the heterogeneous effects of the first childbirth on mothers' annual income, using data from several waves (1979-2018) of the National Longitudinal Survey of Youth. Women usually experience an immediate decrease in their income after childbirth, compared to what they would have earned if they had not become mothers. This gap closes somewhat over time, though mothers never fully catch up to their counterfactuals. Previous work tried to explain this "motherhood penalty" by estimating the average treatment effect of children on women's income; however, these effects can be quite heterogeneous across mothers with different observable characteristics. Instead, our analysis centers on the distribution of the *individual-level* effects of the first childbirth on mothers' income, using the Changes-in-Changes model and quantile regression. Identifying the features of this distribution is a challenging task as it requires knowledge of joint distribution. We find that around 73% of mothers have lower income after their first childbirth than they would have had if they had not had a child. These adverse effects are particularly pronounced among 10–20% of mothers. Our quantile regression analysis indicates that the first childbirth most negatively affects older, single/divorced, white, and more educated mothers.

Keywords Motherhood Penalty · Heterogeneous Treatment Effects · Changes in Changes · Joint Distribution of Potential Outcomes · Distribution of the Treatment Effect · Quantile Regression

These authors contributed equally: Afrouz Azadikhah Jahromi, Weige Huang

✉ Afrouz Azadikhah Jahromi
ajahromi@widener.edu

✉ Weige Huang
huangwg@smbu.edu.cn

¹ School of Business Administration, Widener University, Chester, PA, USA

² Faculty of Economics, Shenzhen MSU-BIT University, Shenzhen, China

JEL classification J16 · C21

1 Introduction

Recent research suggests that the wage disparity between mothers and childless women is more precisely characterized as a “family gap,” primarily attributable to the time and resources required to raise children. While previous studies have acknowledged the existence and persistent nature of this gap in different countries and time periods, they have primarily focused on *average effects*, leaving the distributional aspects unexplored (see Budig & England, 2001; Davies & Pierre, 2005; Budig & Hodges, 2010, Pal & Waldfogel, 2016, among others). By contrast, this paper investigates the role of children in explaining the motherhood penalty by estimating the *distribution* of the effect of the first childbirth on women’s income.¹ Specifically, we analyze the treatment effect heterogeneity of the first childbirth (see, among others, Manski, 1990; Angrist & Imbens, 1995, Abadie et al., 2002; Imbens, 2004) by estimating the individual-level quantile of the treatment effects on the treated (iQoTT). We use Changes-in-Changes analysis, detailed below, to compare the observed income of mothers after the birth of their first children to their income for the *same individual* if they had never given birth (which is not observable) while maintaining their rank in the income distribution as their rank before giving birth².

The knowledge of the distributional effect of the first childbirth on mothers’ income is important for several reasons. First, it will help us better understand the different effects of the first childbirth across the income distribution. The mean impact reported by prior research on motherhood penalty (Waldfogel, 1998, Kor-enman & Neumark, 1990, Moore & Wilson, 1982; Waldfogel, 1997) just represents the average of the positive and negative effects of having children on women’s income and cannot address the extent of distributional effects. Having children, which on average affects the women’s income negatively, may have no effects or even some beneficial effects on other mothers. More specifically, we want to distinguish between the case where all mothers experience the same negative effect after the birth of their first child and the case where the captured effect mixes together mothers who are experiencing large negative effects with those experiencing small negative effects (or even positive effects) after having their first child. Second, the distributional analysis has significant policy implications, as family policies based on the average effect may not be applicable in many cases. Given two maternal leave policies with the same mean effect, policymakers may be more interested in policies

¹ In our study, income refers to the total income from wages and salary from all jobs, before deductions for taxes or any other factors.

² Kleven et al. (2019) examined two distinct impacts of childbirth on labor market outcomes. The first, termed the “post-child effect of realized fertility,” pertains to changes that mothers make in their working hours, industry, occupation, etc., following the birth of their first child. The second, known as the “pre-child effect of anticipated fertility,” involves career adjustments and reduced human capital investments made by mothers before the birth of their first child. The primary focus of our paper is on the post-child effects, given that the pre-child effects have diminished over time. This decline is attributed to the significant investments made by women in their careers and education during recent decades.

that are helping mothers at the lower tail of the income distribution than those at the top of the distribution. Similarly, if career interruptions due to the birth of children are more concentrated among specific groups of mothers, then targeted maternity leave policy or childcare assistance to those who experience the largest wage penalty may be more effective in diminishing the gap between mothers and childless women (Waldfogel, 1998; Carrasco, 2001). Lastly, for noisy outcomes like wages and income, the distributional analysis is more suitable as it accounts for outliers.

Learning about the treatment effect heterogeneity requires estimating the distributional effects of the first childbirth on mothers' income which is challenging for two reasons. First, we need to compare the income distribution of mothers with their income distribution had they not had any children (i.e., the counterfactual distribution), which is not observable. Following Athey and Imbens (2006) and Melly and Santangelo (2015), we use Changes-in-Changes (CIC) analysis to estimate the counterfactual distribution of annual income for mothers. The CIC model is an alternative to the Difference-in-Difference (DID) analysis, in which the goal is estimating the whole distribution of the counterfactual outcome for the treated group. The main assumption of CIC is the time invariance of the distribution of the unobservables within the treated and control groups.³ The heart of the DID setup is an additive structure for potential outcomes in the absence of the treatment, where the groups and time periods are treated symmetrically. In comparison, the CIC model treats groups and time periods asymmetrically. In other words, we use the entire 'before' and 'after' income distribution of childless women to nonparametrically estimate the changes over time, and assuming that mothers would have experienced the same changes over time, we can estimate the counterfactual distribution of income for mothers.

The second main challenge is that our parameter of interest for capturing the heterogeneous effects of the first childbirth depends on the joint distribution of annual income for mothers and their counterfactuals, which is not observable even under standard identification assumptions like selection on observables. Previous studies like Heckman et al. (1997) and Bitler et al. (2006) have discussed conditions under which the joint distribution can be identified. One condition is the common treatment effect assumption, which causes the treatment effect distribution to collapse and become equal to the mean impact. Another condition is the rank preservation assumption, where we assume that if an individual's rank falls within the q^{th} quantile of the counterfactual control distribution, that same individual will also have a rank in the q^{th} quantile within the counterfactual treated distribution. Neither condition is plausible in our case since the effects of having children are not the same among women and since women may keep or change their ranks in the income distribution after the birth of their children. We estimate a mother's counterfactual income by using her conditional rank in the income distribution before the birth of the first child and estimating the conditional quantile of counterfactual income (which is identified by CIC) at that specific rank. Our assumption is less restrictive than the rank preservation assumption, a concept further detailed in Section 3.

³ We use two-sample tests, including Kolmogorov-Smirnov, Cramer-von Mises, and Wasserstein Distance tests, to check the validity of the model assumptions.

Our research is somewhat aligned with prior studies by Budig and Hodges (2010), Cooke (2014), Killewald and Bearak (2014), England et al. (2016), and Glauber (2018), which have estimated the extent of wage and earnings disparities among mothers across the entire wage and earnings distribution using both conditional and unconditional quantile regression techniques. However, it is worth noting that no previous studies conducted an estimation of the whole income distribution of the counterfactual or compared the observed income of mothers to a counterfactual outcome for the same individual. Thus, our paper contributes to the existing literature on the motherhood penalty by estimating the entire distribution of the effects of a first childbirth on women's income. We achieve this by comparing the observed income of mothers with the incomes they would likely have earned had they not had children, while maintaining the same rank in the distribution of income as they had before they became mothers.

For our analysis, we pooled data from the 1979 to 2018 waves of the National Longitudinal Survey of Youth (NLSY), a nationally representative sample comprising 12,686 U.S. individuals aged between 14 to 22 years old in 1979, with an equal gender distribution (50% women). Our examination of the distributional impact of the first childbirth on women's income reveals that approximately 73% of mothers experience a decrease in income after the birth of their first child, compared to what their income would have been without any children, while maintaining the same rank in the income distribution as before childbirth. These effects exhibit significant heterogeneity among different groups of mothers. For example, at the 5th quantile, the estimated annual income of mothers is \$21,005 lower than what it would have been in the absence of children, while at the 95th quantile, their annual income is estimated to be \$9887 higher than their counterfactual income.

Our primary findings rely on the assumption that mothers, in the absence of any childbirth, would maintain the same ranks in the *conditional* income distribution as they held before giving birth. However, it is acknowledged that mothers may experience rank changes in the income distribution irrespective of childbirth (i.e., promotions, changing jobs, etc.). Conditional on the identification assumptions of the CIC model, we compare the actual income for each mother to a transformed version of their income in the previous period. More specifically, for each mother we match the observed income after childbirth with an estimate of income in the absence of any birth. However, it is possible that some of our results are driven by individuals changing their ranks in the income distribution over time, rather than being due to heterogeneous effects of the first childbirth. To explore potential heterogeneity beyond the assumed condition, we conduct several robustness checks. We follow the suggested tests in Azadikhah et al. (2022) to study how much childless women change their ranks in the unconditional income distribution⁴ over time, relative to mothers, and additionally compare the standard deviation of the imputed untreated potential outcome for childless women under rank invariance over time, with analogous standard deviation for mothers. The results of our robustness checks suggest that, while some observed heterogeneity may stem from the violation of rank invariance over time, meaningful heterogeneity still persists in our findings. We also

⁴ These tests assess unconditional rank invariance while our main assumption is weaker, focusing on conditional rank invariance over time in the absence of the treatment.

discussed several other robustness checks, including an alternative methodology by Callaway et al. (2018), to estimate the counterfactual income distribution for mothers in the first step and placebo tests related to regression toward mean over time.

The rest of this paper is organized as follows. In Section 2, we provide a review of the existing work on motherhood and the labor market. In Section 3, we propose the model and the estimators of marginal distributions, as well as the parameters of interest for analyzing the heterogeneous effects of the first childbirth on women's income. Section 4 contains the data description and key variables, then we present the main results in Section 5. Several robustness check results are shown in Section 6. Section 7 concludes the main paper, while additional analyses are collected in the Appendix.

2 Motherhood and the labor market

The motherhood penalty represents the immediate income decline experienced by women after childbirth, and it persists throughout their careers (Anderson et al., 2003; Killewald & Bearak, 2014). The existing studies on the motherhood penalty either focus on the “parenthood gap”, representing the pay disparity between men and women after having children, or the “family gap”, representing the wage gap between women with and without children. Despite advancements in narrowing gender and family wage gaps, mothers still face significant income disparities compared to childless women or fathers (Pal & Waldfogel, 2016). Waldfogel (1997) estimated a wage penalty of 5% to 15% between mothers and non-mothers, while Zhang (2009) estimated an average hourly earnings gap of 30%, as of age 40, between women without children and mothers with more than three years of career interruption. Several mechanisms can account for the drop in mothers' income even after controlling for observable characteristics; employer discrimination (Correll et al., 2007; Bertrand & Mullainathan, 2004), adjustments in human capital investment (e.g., Hill, 1979; Budig & England, 2001; Budig & Hodges, 2010, Becker, 1985), family structure and resources Budig and Hodges (2010), and even shifts in occupations (Goldin, 2014, Adda et al. 2017; Cortes & Pan, 2018).

Among the existing studies, only a few existing studies have explored the impact of motherhood on the income distribution (e.g., Budig & Hodges, 2010; Cooke, 2014; Killewald & Bearak, 2014; England et al., 2016, and Glauber, 2018). For instance, Budig and Hodges (2010) conducted research using conditional quantile regression with fixed effects. This approach aimed to clarify how factors like family resources, time dedicated to work, and transitioning to more family-friendly jobs impact the motherhood penalty across different income levels. Their findings indicate that women with higher earnings tend to face relatively smaller income reductions compared to those with lower incomes. The study emphasized the significant roles played by factors such as family resources (including husbands' income) and work effort (such as the number of weeks worked per year and weekly work hours) in explaining the income losses experienced by mothers. Similarly, Glauber (2018) employed unconditional quantile regression with fixed effects to examine changes in the motherhood penalty at different points across the income distribution over time, and observed a more notable decrease in the motherhood penalty among women with

higher earnings as time progressed. Surprisingly, no prior studies on the distributional treatment effect of having children have estimated the complete income distribution for mothers who have never had children or explored the specific “individual” treatment effect, two key goals of this paper.

The potential endogeneity of motherhood decisions has been discussed in several studies (e.g., Shapiro & Mott, 1979; Korenman & Neumark, 1990, Browning, 1992; Bernhardt 1993; Kalwij, 2000; Carrasco, 2001, Ahn & Mira, 2002; Stanca, 2012). To address the possibility that the decision to have children might be correlated with unobservable characteristics, particularly those linked with wage determination (e.g., work effort), researchers have employed various strategies. Fixed effects models, for instance, aim to account for time-invariant unobservable factors within the linear labor supply model (Jee et al., 2019). Some prior studies have used instrumental variables (IVs) to mitigate the endogeneity of fertility (Korenman & Neumark, 1990; Winder, 2008); however, previously used IVs (e.g., the sex mix of the first two children) can mainly provide insights on the “local” effect of having more children rather than the total effects of children or the specific effects of the first childbirth, which limits their usefulness for our analysis. Another approach to counter endogeneity involves utilizing natural experiments. For example, Rosenzweig and Wolpin (1980) examined the labor supply decisions of mothers by studying multiple birth events during their first pregnancies. Nonetheless, this method often suffers from limited sample sizes, potentially restricting its applicability.

Browning (1992) provided a thorough overview of various modeling approaches utilized in prior research to account for the endogeneity of fertility, but the overall investigation has yielded inconclusive results. Several studies document that fertility either has no effect (for example, Cramer (1980)) or small but significant positive effects (for example, Cain and Dooley (1976)) on the female labor supply. Moreover, in some studies the consideration of endogeneity has led to alterations in the observed impact of having more children on women’s labor market participation. For instance, assuming exogenous fertility, Iacovou (2001) reported a reduction in women’s labor supply with the birth of a third child. Yet when considering fertility endogeneity, the author observed either no effect or a potentially positive influence on women’s labor market participation following the arrival of a third child. This discrepancy might be attributed to the income effect outweighing the substitution effect. Post-childbirth, working mothers might increase their labor supply to counterbalance the decline in family income due to childcare expenses (an income effect). Conversely, they might opt to work less or discontinue work to avoid these childcare costs (a substitution effect).

Some existing work analyzes the motherhood penalty across countries. Gangl and Ziefle (2009) conducted a comparative analysis in Germany, Britain, and the United States, estimating a wage penalty per child ranging from 10 to 18%. Notably, this effect was observed to be more pronounced among German mothers compared to their American and British counterparts. Molina and Montuenga (2009) studied the motherhood wage penalty for Spanish women between 1994 and 2001, and measured wage losses of 6%, 14%, and 15% for women with one, two, and three or more children, respectively. Kuziemko et al. (2018) used the British Household Panel Survey (BHPS) and three different U.S. datasets to show the increasing cost of having children over time. Similarly, Kleven et al. (2019) used Danish administrative

data from 1980 to 2013 to examine the maternal wage penalty's persistence over time. They concluded that the motherhood penalty is about 20% in the long run and that the wage gap between women with children and without children increases over time substantially.

3 Model and estimation

We are interested in the heterogeneous effects of a binary treatment D , whether a woman has any children or not, on the outcome Y , which is the annual income from wages and salary. The parameter of interest to estimate the distribution of the treatment effect is the *individual-level* Quantile of the Treatment Effect on the Treated (*iQoTT*), which is obtained in two steps. First, we need to estimate the outcome distribution for the treated group if they had not been treated (i.e., the counterfactual distribution). Second, we need to associate each treated potential outcome with untreated potential outcomes for the exact same unit in the treated group to identify the distributional treatment effect. In this paper, we name the distributional treatment effect as *iQoTT* because we want to distinguish it from *QoTT* which captures the gap between the distribution of treated potential outcomes and the distribution of untreated potential outcomes at each quantile. In other words, the *QoTT* compares the distribution of income that mothers actually experienced to the distribution of income that they would have experienced in the absence of the first childbirth. Unlike our *iQoTT*, comparing these two distributions does not require identifying the joint distribution of treated and untreated potential outcomes, and *iQoTT* captures the *quantile of the difference* rather than the *difference between quantiles*.

Following Melly and Santangelo (2015), we consider the case of two groups and two periods. Each individual belongs to group $G \in \{0, 1\}$ and is observed in period $T \in \{-1, 1\}$. We assume women who have children are treated and belong to $G = 1$ and women without children belong to $G = 0$. The time of the treatment is defined based on the birth of the first child, and all other times are indexed relative to that base year. We consider a year before having the first child, $T = -1$, as the untreated period, and a year after the first childbirth, $T = 1$, as the treated period.⁵ To establish the time frame for childless women, we employ a random assignment technique by allocating placebo births to childless women sampled from the estimated conditional distribution of mothers' age at first birth. A similar methodology has been utilized in prior studies such as Kleven et al. (2019) and Kuziemko et al. (2018). Our approach assumes that the age at a woman's first birth follows a log-normal distribution within specific cells determined by birth cohort and education level. These cells are based on the actual mean and variance of the age at first birth among mothers. We categorize birth cohorts into four distinct groups (1957–1958, 1959–1960, 1961–1962, 1963–1964) and define education based on the highest level attained within four cells (less than high school, high school, some college, college).

Let $Y(1)$ correspond to the income that a particular woman would have earned if she had children at a particular point in time and $Y(0)$ indicates the income she would

⁵ Other time periods were examined and the findings were consistent.

have had if she hadn't had any children.

$$Y = Y(0) \cdot (1 - D) + Y(1) \cdot D$$

Most of the previous literature has focused on identifying the Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT), which are defined as:

$$ATE = E[Y(1) - Y(0)]$$

$$ATT = E[Y(1) - Y(0)|D = 1]$$

In this paper, we focus on the distributional treatment effects rather than the average treatment effect which is given by

$$F_{Y(1)-Y(0)|D=1}(\delta) = P(Y(1) - Y(0) \leq \delta|D = 1) = E[1\{Y(1) - Y(0) \leq \delta|D = 1\}] \quad (1)$$

Where $0 < \delta < 1$. Identifying the distribution of the treatment effect is challenging as it depends the joint distribution of $Y(1)$ and $Y(0)$ for the treated group, which is not observable even under standard identification assumptions. In the next section, we explain how to obtain the distribution of the effects of first childbirth on women's income.

3.1 Model

We consider a setting with staggered treatment adoption, which is described in the following assumption.

Assumption 1 (Staggered treatment adoption). For all units and for all time periods $t = 2, \dots, T$, $D_{it-1} = 1 \Rightarrow D_{it} = 1$.

Staggered treatment adoption holds in our setting in the sense that motherhood is “scarring”; i.e., once a woman gives birth, she is considered “treated” in all subsequent periods. See Sun and Abraham (2021) for more discussion of staggered treatment adoption. In the first step, we estimate the counterfactual distribution of income for mothers using the Changes-in-Changes (CIC) method that was initially introduced by Athey and Imbens (2006).⁶

The CIC approach is designed for the case with two groups and two periods in which the counterfactual distribution of potential outcomes is obtained from three known and observed outcome distributions; the outcome distribution of the treated group at the pre-treatment period and two outcome distributions of the untreated group at post- and pre-treatment period. We use quantile regression to estimate these three marginal distributions (Koenker & Bassett Jr, 1978; Koenker, 2005). Following Athey and Imbens (2006), we introduce a set of assumptions to identify the CIC model. For simplicity, the subscript i was dropped from the notations. The

⁶ Melly and Santangelo (2015) add covariates to the model. Other related work includes Melly (2006), Albrecht, et al., 2009; Chernozhukov et al. (2013), Athey and Imbens (2006), Callaway and Li (2017), Callaway et al. (2018).

shorthanded notations in this paper are as follows:

$$Y(0)_{gtx} \stackrel{d}{\sim} Y(0)|G = g, T = t, X = x, \quad Y(1)_{gtx} \stackrel{d}{\sim} Y(1)|G = g, T = t, X = x$$

$$Y_{gtx} \stackrel{d}{\sim} Y|G = g, T = t, X = x, \quad U_{gt} \stackrel{d}{\sim} U|G = g, T = t$$

In which $\stackrel{d}{\sim}$ stands for “is distributed as.” The random variable U represents the unobservable characteristics. The corresponding conditional distribution functions are $F_{Y(0)|gtx}$, $F_{Y(1)|gtx}$, $F_{Y|gtx}$, $F_{U|gtx}$ and the three main assumptions are as below.

Assumption 2 (Model for untreated potential outcomes). The outcome for an individual in the absence of the treatment is defined as $Y_t(0) = h_t(X, U_t)$

Assumption 3 (Strict monotonicity). The production function $h_t(x, u_t)$ is strictly increasing in u_t for all $t = 1, \dots, T$ almost surely

Assumption 4 (Time invariance). The distribution of $U_t|G, X$ is constant over time

Assumption 5 (Support). $\cup_{1x} \subset \cup_{0x}$ for $\forall x \in \mathbb{X}$.

Assumption 2 implies that all the unobservable characteristics are captured by U and for an individual with $U = u$ the random variable $Y(0)$ is the same in a given time and does not depend on the group indicator, so $Y_t(0)$ can be expressed as $h_t(X, U_t)$ where $h(\cdot)$ denotes an unrestricted function. Assumption 3 requires that higher U is associated with higher outcomes. Assumption 4 is the main assumption that implies the conditional distributions of unobservables are the same over time within each group (treated and control). It corresponds to the common trend assumption in the DID model within the context of CIC. This assumption is called rank similarity by Chernozhukov and Hansen (2005) and is less restrictive than the rank preservation assumption in which the ranks are assumed to be identical in all treatment states.⁷ We use two-sample tests, including Kolmogorov-Smirnov, Cramer-von Mises, and Wasserstein Distance tests, to check the validity of CIC model assumptions in particular Assumption 4.

Suppose that Assumptions 1–3 hold and $0 < \tau < 1$. Then the distribution of $Y(0)$ for the treated group, $F_{Y(0)|1x}$, is identified for $\forall x \in \mathbb{X}$ with;

$$F_{Y(0)|1x}(y) = F_{Y|1-x}(F_{Y|0-1x}^{-1}(F_{Y|01x}(y))) \tag{2}$$

The proof of Eq. (2) is presented in Melly and Santangelo (2015). Using Eq. (2), we can identify the unobserved distribution of untreated potential outcomes for the treated group, $F_{Y(0)|1x}$, with the knowledge of three observed distributions: the distribution of income for mothers at $T = -1$, $F_{Y|1-x}$, the quantile function of income for childless women at $T = -1$, $F_{Y|0-1x}^{-1}$, and the distribution of income for childless women at $T = 1$, $F_{Y|01x}$. Following Koenker and Bassett Jr (1978)⁸, we use linear quantile regression estimators to obtain the two conditional distributions, $F_{Y|1-x}$ and $F_{Y|01x}$, and the quantile function, $F_{Y|0-1x}^{-1}$, such that for

⁷ The rank preservation assumption is discussed by Chernozhukov and Hansen (2005) and Bitler et al. (2006) in depth.

⁸ Other methods like distribution regression by Chernozhukov et al. (2013) can be used to estimate two conditional marginal distributions of $F_{Y|1-x}$ and $F_{Y|01x}$.

all $(T, D) \in \{0, 1\} \times \{0, 1\}$,

$$Q_{Y|gtx}(\tau) = F_{Y|gtx}^{-1}(Y) = P_{gt}(X)' \beta_{gt}(\tau), \tag{3}$$

In which $\tau|X, D \sim \tau[0, 1]$ and $P_{gt}(X)$ are some transformations of the vector of covariates X .

The knowledge of marginal distributions of income and the counterfactual distribution of untreated potential outcomes for the treated group is not enough for our distributional analysis. Estimating the distribution of the treatment effects is more challenging than the Average Treatment Effect as it depends on the joint distribution of treated and untreated potential outcomes. Women may keep or change their ranks in the income distribution after the birth of their first child, which is not observable even under standard identification assumptions like the selection on observables. Therefore, additional information is required to associate mothers' income with their counterfactuals and eventually compute the gap. We utilize Eq. (4) to estimate the untreated potential outcome for each individual in the treated group. Ultimately, this approach helps us pinpoint and identify the distribution of the effects of the first childbirth on women's income,

$$\Rightarrow Y(0)_{11x} = F_{Y(0)|11x}^{-1}(F_{Y|1-1x}(Y_{1-1x})) \tag{4}$$

$F_{Y|1-1x}(Y_{1-1x})$ represents the individual's rank at $T = -1$. Intuitively, Eq. (4) means that if treated individuals had not been treated, then they would have had the same ranks in the potential outcome distribution as their ranks at the pre-treatment period ($T = -1$). In other words, after controlling for a set of covariates, if mothers have not given any birth, they would possess the same ranks in the income distribution as their ranks before giving birth.

By employing Eq. (4), we can associate the income of each mother, $Y(1)$, with her counterfactual, $Y(0)$, and obtain the pair of $(Y(1), Y(0))$, and eventually estimate the joint distribution of treated and untreated potential outcomes for mothers. To assess the distributional impact of the first child's birth on women's annual income, the parameter of interest is the quantile of the difference between $Y(0)$ and $Y(1)$ for each mother (i.e., $iQoTT$), which is useful to understand the heterogeneity of the treatment effect among *individuals*. Define,

$$iQoTT(\tau) = F_{Y(1)-Y(0)|11x}^{-1}(\tau) = Q_{Y(1)-Y(0)|11x}(\tau) \tag{5}$$

It's important to highlight that the Quantile of the Treatment Effect on the Treated ($iQoTT$) differs from the Quantile Treatment Effect on Treated ($QoTT$) which is simply defined as the gap between two marginal distributions, as below:

$$QoTT(\tau) = F_{Y(1)|11x}^{-1}(\tau) - F_{Y(0)|11x}^{-1}(\tau)$$

Under the rank preservation assumption, both $iQoTT$ and $QoTT$ measures would be equivalent. This assumption implies that an individual maintains their rank within the outcome distribution regardless of their treatment status. In other words, there is a perfect positive dependency between the distribution of $Y(0)$ and $Y(1)$ and the knowledge of the joint distribution is not required to estimate the $QoTT$. However, within our study, we make a distinction between $iQoTT$ and $QoTT$, where the $iQoTT$ relies on the knowledge of the joint distribution. We compare the observed outcome

for the mothers to the counterfactual outcome for the exact same individual under rank invariance over time assumption which is less restrictive than the rank preservation.

3.2 Estimation

In this section, we outline the steps for estimating the conditional distributions of annual income for mothers and their counterfactuals, along with determining the distributional effects of the first childbirth on women's income. To compute the counterfactual income distribution for mothers, it's essential to estimate these three observed distributions: $F_{Y|01x}$, $F^{-1}Y|0 - 1x$, and $FY|1 - 1x$. As a first step, we utilize the quantile regression method to estimate the conditional distribution of observed income for $G = 0$ at $T = 1$, thereby initiating the process.

$$\hat{F}_{Y|01x}(y) = X'\hat{\beta}_{01}(y). \tag{6}$$

After computing $\hat{F}_{Y|01x}(y)$, in the next step we estimate the quantiles of income for childless women, $G = 0$, at time $T = - 1$ for particular values of $u = \hat{F}_{Y|01x}(y)$ such that,

$$F_{Y|0-1x}^{-1}(u) = \hat{Q}_{Y|0-1x}(u) = X'\hat{\beta}_{0-1}(u). \tag{7}$$

In the final step, we use the quantile regression to estimate the conditional distribution of annual income for mothers, $G = 1$, at time $T = - 1$ for particular values of $\hat{F}_{Y|0-1x}^{-1}(u)$. So, the estimator for counterfactual distribution is defined by,

$$\hat{F}_{Y(0)|11x}(y) = \hat{F}_{Y|1-1x}(F^{-1}_{Y|0-1x}(\hat{F}_{Y|01x}(y))).$$

With the knowledge of $\hat{F}_{Y(0)|11x}(y)$ and Eq. (4), we can estimate the untreated potential outcomes for mothers, $\hat{Y}(0)_{11x}$, as follows:

$$\hat{Y}(0)_{11x} = \hat{F}_{Y(0)|11x}^{-1}(\hat{F}_{Y|1-1x}(Y_{1-1x})).$$

Eventually, we create the pair of $(\hat{Y}(1), \hat{Y}(0))$ and measure the gap to get the *iQoTT* by,

$$i\widehat{QoTT}(\tau) = F^{-1}_{Y(1)-Y(0)|11x}(\tau) = \hat{q}_{Y(1)-Y(0)|11x}(\tau).$$

As discussed above, our main goal is to compare the observed income for a mother following childbirth to a counterfactual outcome that she would have experienced if she had not given birth and maintained her pre-childbirth rank in the income distribution.

4 Data and key variables

Studying the impacts of the first childbirth on mothers' annual income requires panel data with information on labor market outcomes and children. We pooled the 1979 to 2018 waves of the National Longitudinal Survey of Youth (NLSY), a comprehensive national sample consisting of 12,686 individuals aged between 14 and 22 in 1979, with women comprising 50% of the sample. Participants were interviewed annually until 1998 and biannually thereafter. The NLSY79 dataset is ideal for our study on first childbirth timing, as it follows women throughout their entire reproductive years,

Table 1 Summary statistics—mothers and childless women

	Mothers	Childless women	Difference	P val on Difference
Income at T=-1	15.74	16.23	-0.497	0
Income at T=1	11.78	17.51	-5.737	0
Change in Income	-3.96	1.3	-5.257	0
White	0.65	0.63	0.026	0.27
Married	0.74	0.22	0.523	0
Education beyond High School	0.49	0.59	-0.096	0
Age	25.41	25.73	-0.32	0.06
Age at 1st Birth	25.61	26.4	-0.79	0
N	1729	549		

Source: National Longitudinal Survey of Youth 1979

up to their mid-50s. Notably, this dataset provides information about the number of children, year of the first childbirth, age of the respondent at the time of the first birth, annual income from wages and salary, and a comprehensive set of key variables that were consistently collected throughout the survey period.

The dependent variable is the respondent's real income from wages and salary in her current job (in 2000 US dollars). The vector of covariates, X , includes a dummy variable for age (=1 if older than 26 years old), a dummy variable for marital status (=1 if married), and race (=1 if White, defined in the data as non-Black and non-Hispanic), a dummy variable for education level (=1 beyond high school degree), and a dummy for age at the first childbirth (=1 if older than 26 years old). We include age to control life-cycle trends, and add education and race to capture other factors affecting human capital accumulation. Two points are worth noting here. First, while our analysis is based on the impact of the first child, long-term studies will include the impact of other children as well. Second, we did not make any difference between biological, step, or adopted children in the data set, as we do not have that information in some years.⁹

To investigate the family gap, we consider childless women as the control group.¹⁰ As our sample comprises women who have completed their childbearing years, we classify childless women as those who have never had children. Table 1 displays the summary statistics of the sample for mothers and childless women. Notably, the income at $T = -1$ (one year prior to the first childbirth) is approximately \$500 higher for childless women compared to mothers. Following the first childbirth, the average income gap increases to \$5,737. Mothers exhibit a higher likelihood of

⁹ From 1979 through 1988, the NLSY did not distinguish between biological children and step or adopted children. Using Danish administrative data, Kleven et al. (2021) found the same child penalties from biological and adoptive children in the long run.

¹⁰ We also perform a parallel analysis with fathers as the control group, although this aids more in exploring the gender gap than the family gap.

being married and they are less inclined to have an education beyond high school. Additionally, both groups exhibit similar characteristics concerning racial background and age.

5 Results

All our findings are derived from a balanced sample including childless women and mothers whose first childbirth occurred after the age of 20.¹¹ This sample is observed for a duration of two years: one year prior ($T = -1$) and one year subsequent ($T = 1$) to the birth (or placebo birth) of their first child.¹² First, we illustrate the observed income distributions of mothers and childless women in Panel (a) of Fig. 1. This figure shows the income distribution at $T = -1$, representing the year prior, and at $T = 1$, representing the year following the first childbirth.¹³ Panel (a) in Fig. 1 shows that mothers and childless women have similar income distributions at time $T = -1$. However, the distribution of annual income for mothers at time $T = 1$ is positioned on the left side of the income distribution for childless women, indicating relatively lower income levels for mothers at $T = 1$. Additionally, childless women display a slight rightward shift in their income distribution between $T = -1$ and $T = 1$, signifying a slight increase in their income over time. The disparity between these distributions is most pronounced at lower income levels and diminishes at higher income levels. Notably, approximately 27% of mothers exhibit zero income at $T = 1$.

Panel (b) of Fig. 1 displays the income change distribution over time for both mothers and childless women. These graphs solely illustrate the income difference between $T = -1$ (one year before the first childbirth) and $T = 1$ (one year after the first childbirth) without relying on any identifying assumptions. Notably, approximately 26% of mothers exhibit a higher income after their first childbirth, while childless women are less inclined to encounter a significant income decline and are more likely to increase their income over this period.

Subsequently, we estimate the counterfactual income distribution for mothers if they had never had any children, using the CIC methodology outlined in Eq. (4). Two plots the observed and estimated counterfactual income distribution for mothers. The counterfactual income distribution for mothers is positioned to the right of their observed income distribution at $T = 1$. The gap between these two distributions is larger in the lower part of the income distribution. However, it's important not to conclude that the motherhood penalty is the largest for mothers at the lower part of the income distribution, as mothers can change their ranks in the income distributions after having their first child.

After obtaining the counterfactual income distribution for mothers, we need to create a pair of $(Y(1), Y(0))$ for each mother and compute the gap, representing our

¹¹ Less than 4% of the entire sample of mothers experienced their first childbirth before age 20.

¹² We conducted our analysis across various event times, consistently obtaining similar results. Consequently, we opted for the period with the highest volume of observations, namely $T = -1$ and $T = 1$.

¹³ As discussed in Section 3 and Appendix 8.1, to define these time periods for childless women, we follow the Kleven et al. (2018) framework and assign placebo births to childless women within cells of their birth cohort and education, based on the observed distribution of age at first birth among mothers.

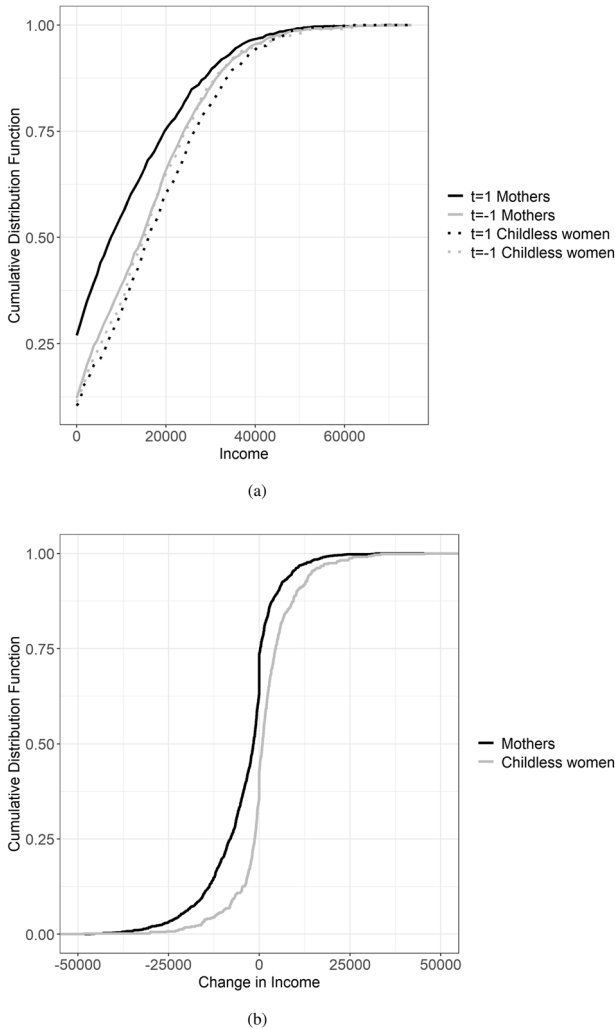


Fig. 1 Marginal distribution of total income from wages and salary. Panel (a) represents plots of the income distributions at $T = -1$ (a year before the birth of the first child) and $T = 1$ (a year after the birth of the first child) for mothers and childless women. Panel (b) plots the distribution of the change in income over time for mothers and childless women

parameter of interest. Figure 3 plots the heterogeneous distributional effects of the first childbirth on mothers' income, measured here by the quantile of the difference between $Y(1)$ and $\hat{Y}(0)$. This is the distribution of the difference between the income of mothers relative to their income had they not had any children while maintaining the same rank in the income distribution as before childbirth. On average, women lose around \$4120 of their annual income after the birth of their first child, but these effects are quite heterogeneous across women. Figure 3 shows that around 27% of women experience an increase in their income following their first childbirth, relative to what they would have earned if they had never become a mother and maintained the same rank in the income distribution as they had before giving birth. Conversely,

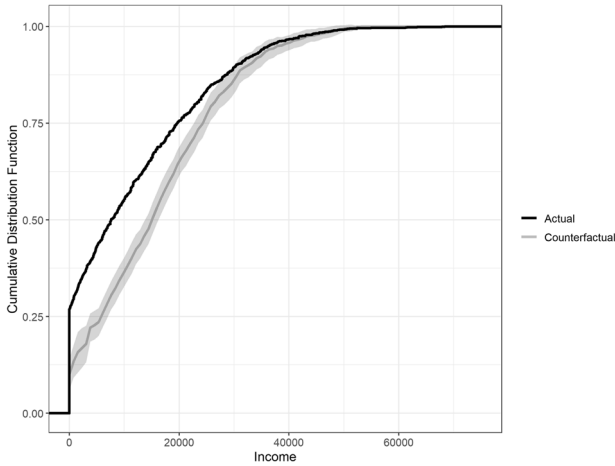


Fig. 2 Counterfactual distribution of income from wages and salary. Plots of the actual and counterfactual distribution of income for mothers using childless mothers as the control group. The counterfactual distribution also includes a 95% confidence interval computed using the empirical bootstrap with 1000 iterations

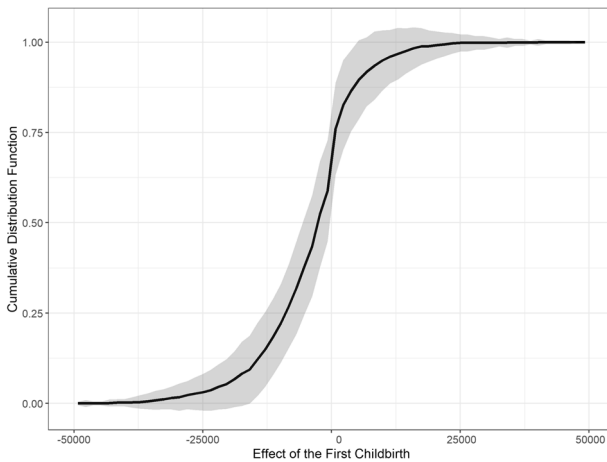


Fig. 3 The quantile of the treatment effect on the treated group. This distribution plots the quantile of the difference between $Y(1)$ and $Y(0)$ for mothers. It also includes a 95% confidence interval computed using the empirical bootstrap with 1000 iterations

3% of mothers experience an annual income reduction of at least \$25,000 relative to their hypothetical income if they hadn't had children and had retained their previous income rank. Similarly, 22% of mothers encounter an income decline of at least \$10,000 compared to what they would have earned if they hadn't become mothers and had preserved their prior income rank.

The primary finding from our analysis indicates a substantial heterogeneity in the impact of the first childbirth among mothers. While the negative effects of motherhood are largely concentrated among 10–20% of mothers, a fraction of mothers still

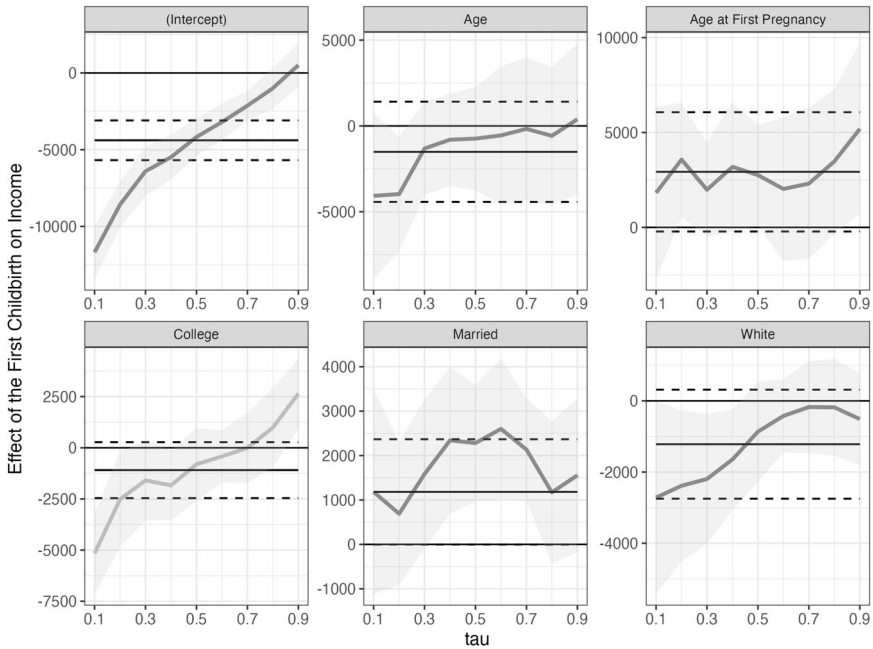


Fig. 4 Quantile regression estimates of the first childbirth effects on covariates. Quantile regression estimates of the effect of covariates on the quantiles of the effect of the first childbirth. The covariates include Age (26 years old and above), Col (education beyond high school), Married, Age at 1st birth (26 years and older), and White. The solid horizontal line provides OLS estimates of the effect of each covariate, and the horizontal dashed line contains a 90% confidence interval. The other lines provide quantile regression estimates of the effect of each covariate at particular quantiles from 0.1, 0.2, ..., 0.9. The shaded area contains pointwise 90% confidence intervals from the quantile regressions

experience an increase in their income compared to what they would have earned if they had never given birth and maintained the same rank in the income distribution as they did before they gave birth.

Lastly, we utilize quantile regression to explore how the motherhood penalty varies across mothers with different observable characteristics at different quantiles of the treatment effect distribution. We present the results in Fig. 4. When we study the heterogeneity effects with respect to observed characteristics, the outcome of interest is the *motherhood penalty*, which is measured by the gap between $Y(1)$ and $\hat{Y}(0)$. $Y(1)$ is the observed income after the birth of the first child and $\hat{Y}(0)$ is the estimated income that an individual would have had if they never gave birth to any child and if they had maintained their rank in the income distribution over time. Therefore, the outcome of interest in Fig. 4 is $Y(1) - \hat{Y}(0)$, which represents the income loss/gain for mothers after giving birth. We use quantile regression to see how covariates affect this conditional distribution at different quantiles. Since the outcome shows the gap between $Y(1)$ and $\hat{Y}(0)$, the largest negative effects happen at lower quantiles (i.e., the 10th quantile) and small negative effects or even positive effects of the first childbirth happen at higher quantiles (i.e., the 90th). Among those who appear to be most negatively affected (at lower quantiles) by the birth of their first child and tend to experience the most substantial impacts.

We plot nine distinct quantile regression estimates for τ , ranging from 0.1 to 0.9, as solid curves for each of the six coefficients. These point estimates represent the impact of a one-unit change in the covariate on the *distribution of* $Y(1) - \hat{Y}(0)$, while holding other covariates constant. Therefore, the horizontal axis depicts the quantiles (taus), while the vertical axis, measured in dollars, illustrates the covariate's effect on $Y(1) - \hat{Y}(0)$. In Fig. 4, the intercept can be interpreted as the estimated conditional quantile regression of the effect of the first childbirth on income for a non-white, unmarried mother younger than 26 years old, with education up to high school, and who experienced her first birth before the age of 25.

On average, the impact of childbirth appears to be relatively more adverse for older mothers (26 years old and above); however, the effect exhibits considerable heterogeneity. Among mothers experiencing substantial negative effects following their first childbirth (i.e., those positioned at the lower end of the treatment effect distribution), older mothers tend to encounter significantly more pronounced negative effects compared to younger mothers. The age of mothers at their first birth seems to exert a uniform effect across the entire spectrum of the distribution, without displaying substantial heterogeneity. The heterogeneity of the motherhood penalty among highly educated mothers (beyond high school) and less-educated mothers is substantial, particularly at the left tail of the distribution. At the 10th percentile of the conditional distribution, the motherhood penalty for highly educated mothers is roughly \$5000 more negative compared to less-educated mothers. On the other hand, for individuals that are toward the top of the distribution of the treatment effect (i.e., experienced a smaller drop or even positive changes in their income), highly educated mothers tend to have larger benefits than less-educated women. These findings align with the research of Anderson et al. (2003) and Pal and Waldfogel (2016), who also observed a larger wage penalty for college graduates.

On average, married mothers tend to encounter positive changes in their income post-birth compared to single/divorced mothers, which is inconsistent with the findings of Glauber (2007) and Budig and England (2001). From a distributional standpoint, married mothers display a relatively consistent effect across the entire spectrum of the distribution. Among mothers experiencing median effects (positioned at the middle of the effect distribution), the impact is more positive for married mothers than for their single/divorced counterparts. The favorable income trend for married mothers might stem from potential access to additional family resources, such as their spouse's income, allowing for shared childcare responsibilities and reduced career interruptions.

Moreover, there is evidence suggesting that white mothers, particularly those most affected by the first childbirth, encounter more negative effects. This aligns somewhat with previous studies (e.g., Waldfogel, 1997; Anderson et al., 2003), which reported a comparatively smaller wage penalty for black mothers. White mothers may trade off higher-paid jobs to care for their children and experience more considerable penalties in their income. Conversely, among those less affected by childbirth, the disparity between the effects for white mothers and other racial groups is marginal. In summary, among mothers most adversely affected by the birth of their first child, older, single/divorced, white, and highly educated (beyond high school) individuals tend to experience the most substantial impacts.

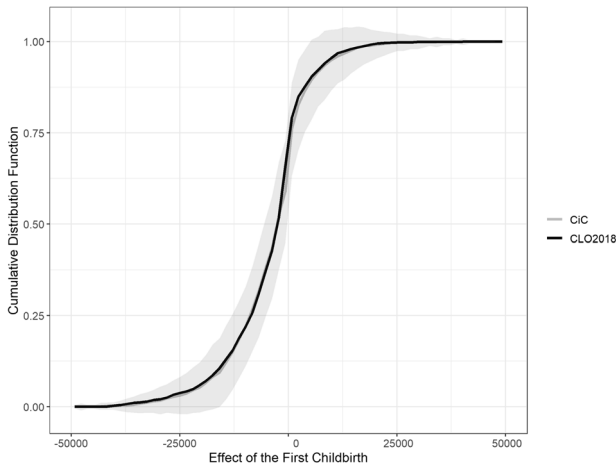


Fig. 5 CLO (2018) method as an alternative estimator. This figure contains our original estimates of the distribution of the effect of the first childbirth (with a 95% uniform confidence band) along with the analogous estimate using CLO (2018) as an alternative estimator

Finally, we examine the variation in the impact of the first childbirth across mothers concerning what their income would have been if they had not become parents (i.e., $\hat{Y}(0)$). The quantile regression estimates resulting from regressing $Y(1) - \hat{Y}(0)$ on $\hat{Y}(0)$ are depicted in Fig. 10 in Appendix 8.2. The visualization indicates that, on average, the effect of the initial childbirth tends to be more negative for mothers who would have attained higher income levels had they not had any children and had maintained the same income rank as before childbirth.

6 Robustness checks

6.1 Callaway, Li, and Oka approach

As an alternative approach to estimating the counterfactual income distribution for mothers and assessing the distributional treatment effect of the first childbirth on mothers' income, we use the framework proposed by Callaway et al. (2018) (CLO hereafter). The CLO methodology relies on two key identifying assumptions. First, the distribution of change in the untreated potential outcome over time conditional on a set of covariates is independent of treatment status (similar to the distributional parallel trend assumption). Second, it assumes that the conditional dependence (copula) of changes in the untreated potential outcome and the initial level of the untreated outcome remains consistent for both the treated and untreated groups. By utilizing quantile regression to obtain conditional distributions, the CLO approach enables the identification of the income distribution for mothers if they had not had any childbirth.

Figure 5 plots the distributions of the effect of the first childbirth on women's income, comparing both the CIC and CLO estimates. The treatment effect distributions derived from both methodologies demonstrate remarkable similarity,

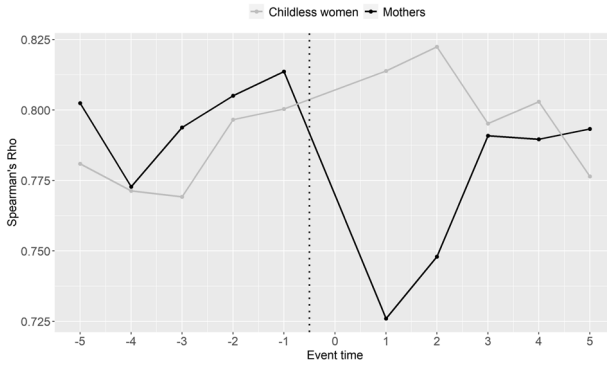


Fig. 6 The Spearman's Rho for mothers and childless women. The Figure plots the computed Spearman's Rho between Y_t and $LagY_t$ across event time dummies (T) for mothers and childless women

making direct comparison challenging. Initially, our estimation indicates that approximately 73% of mothers encounter a decline in their income following childbirth. Similarly, under the CLO assumptions, roughly 73% of mothers would experience lower income after their first childbirth. Specifically, for mothers positioned at the 5th percentile of the childbirth effect distribution (those most adversely affected), our initial estimate suggests a loss of \$21,004 in their annual income. Meanwhile, employing the CLO approach yields an estimated loss of \$21,734 for this percentile. In contrast, for mothers experiencing modest or positive effects post-birth (at the 90th quantile), our estimation illustrates an income increase of \$5548. Under CLO assumptions, this group is estimated to earn approximately \$5079 more per year.

6.2 Placebo tests for treatment effect heterogeneity

Equation (4) is one of the main equations that help us to estimate the whole distribution of individual treatment effects. Even if mothers do not maintain their positions within the income distribution as they did prior to giving birth, our findings would still reflect a certain degree of heterogeneity in the treatment effect. In this section, we propose several placebo tests to analyze if there is any significant heterogeneity in the treatment effect besides the violation of mothers keeping their rank in the income distribution. Following the methodology proposed in Azadikhah Jahromi and Callaway (2022) (with extensions due to access to multiple periods), we conduct Spearman's Rho computations between Y_t and $LagY_t$ across various event time dummies (T)¹⁴ for both mothers and childless women. This analysis aims to evaluate rank dependency over time. If Spearman's Rho equals 1, it suggests that rank invariance over time holds, signifying that any heterogeneity in our results stems solely from the varied impact of the first childbirth. In Fig. 6, we present the Spearman's Rhos computed for both groups. Notably, at $T \leq 0$, the computed Spearman's Rho values for both mothers and childless women are relatively high,

¹⁴ This spans from $T = -5$ (5 years prior to the birth of the first child) to $T = 5$ (5 years after).

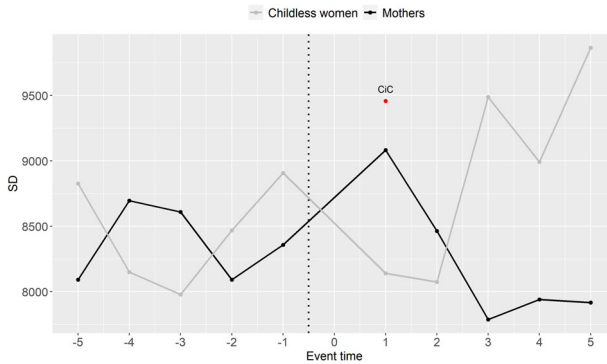


Fig. 7 The standard deviation of $Y(0) - \tilde{Y}(0)$ for mothers and childless women. The Figure plots the computed value of $Y(0) - \tilde{Y}(0)$ for mothers and childless women across event time dummies (T)

surpassing 0.76. This implies a substantial dependency between Y_t and $LagY_t$ for both cohorts during these time periods.

It is intriguing to note that the Spearman's Rho values for childless women at $T = 1$ (one year after the first childbirth) remained consistently high, while there was a notable sudden decline in dependency for mothers immediately after childbirth. This observed discrepancy between the rho values peaks at time $T = 1$. These findings suggest compelling evidence for the existence of meaningful heterogeneity beyond the violation indicated by Eq. (4). However, it's important to exercise caution in interpreting the rho values at $T > 2$, as long-term studies might encompass additional factors (such as the birth of subsequent children) that could influence these observations.

Second, we test the validity of Eq. (4) for childless women by comparing the observed income with the imputed income at time T that corresponds to the same rank as in period $T - 1$; $\tilde{Y}(0) = Q_{Y(0)|gtx}(F_{Y(0)|gt-1x})$. If Childless women maintain their rank in the income distribution as they did before placebo childbirth, then the discrepancy between the observed income, $Y(0)$, and the imputed income values, $\tilde{Y}(0)$, should ideally be zero. In Fig. 7, the gray line illustrates the standard deviation of $Y(0) - \tilde{Y}(0)$ for childless women at various event time dummies. As the standard deviation values are non-zero, we compute the $\tilde{Y}(0)$ as well as standard deviations for mothers to gain a better understanding of the rank invariance over time.

The black line in Fig. 7 represents the standard deviation of observed and imputed income values for mothers at different event time dummies. The "CiC dot" in Fig. 7 refers to the standard deviation of the $Y(0) - \tilde{Y}(0)$ for mothers at $T = 1$ (one year after the first childbirth) where $\tilde{Y}(0)$ is the income distribution for mothers if they had not had any children and maintained their rank in the income distribution as they had before giving birth (estimated as in Eq. (4)). For mothers, the standard deviation of the gap at $T = 1$ is \$9083 and for childless women, this value is \$8141. Once more, this suggests that there exists additional heterogeneity attributed to childbirth, which is not solely a result of the assumption of rank invariance over time.

To see if the heterogeneity of the treatment effects after the first childbirth is different from the heterogeneity of the placebo effects, we display the distributions of $|Y(1) - \hat{Y}(0) - \hat{\mu}1|$ (where $\hat{\mu}1$ denotes the average difference between $Y(0)$ and

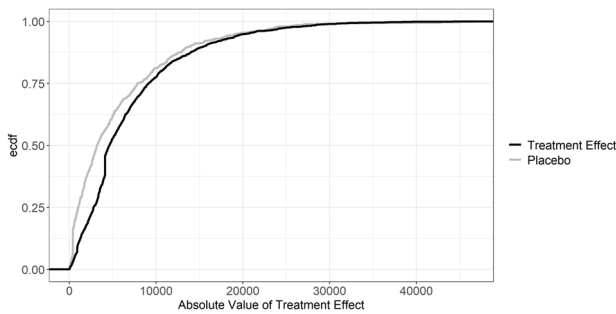


Fig. 8 Placebo treatment effect heterogeneity. The Figure plots the treatment effect heterogeneity under rank invariance over time for mothers (black line) and placebo treatment effect heterogeneity under rank invariance over time for childless women (gray line)

$\hat{Y}(0)$ for mothers), alongside the distribution of placebo effects for childless women in Fig. 8. Our findings indicate a greater level of treatment effect heterogeneity for mothers compared to childless women. Considering the robustness checks conducted, our outcomes suggest the existence of substantial diversity beyond the breach of Eq. (4).

6.3 Placebo test for regression to the mean

Based on the data depicted in Fig. 10 in Appendix 8.2, we infer that among mothers most adversely affected by childbirth, those with higher incomes tend to experience more negative effects compared to those with lower incomes. Nevertheless, these findings might arise due to a regression to the mean phenomenon rather than the inherent heterogeneity associated with childbirth. This indicates that high-income mothers might encounter an income decline even in the absence of childbirth. In essence, when regression to the mean exists, our results may exaggerate the effects for high-income mothers and downplay the effects for low-income ones (in the absence of childbirth). Hence, the seemingly larger effects for high-income mothers in Fig. 10 in Appendix 8.2 could stem from regression to the mean rather than diverse treatment effects across income levels.

To address this concern, we adopt the method outlined in Azadikhah Jahromi and Callaway (2022), utilizing childless women in our sample by randomly assigning half of them as “treated.” We then replicate our estimations to explore the impact of simulated “first childbirths” across varying income levels for childless women. This process is repeated 1000 times, and the outcomes, in conjunction with the original estimate, are presented in Fig. 9. Our findings indicate that the simulated “placebo births” among childless women induce more substantial declines in income at higher imputed income levels. This confirms the occurrence of regression to the mean even without actual childbirth. However, the effects observed in our original estimates are significantly greater than those in the simulated scenarios, indicating the presence of treatment effect heterogeneity distinct from the regression to the mean phenomenon.

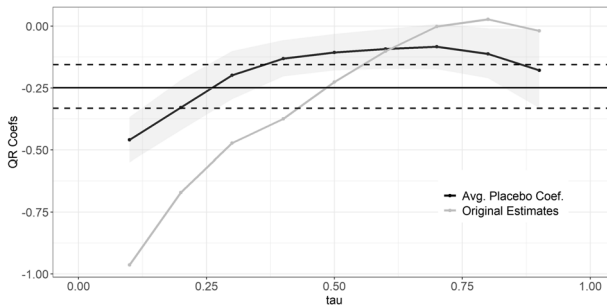


Fig. 9 Placebo test for regression to the mean. The black line shows quantile regression estimates of the effect of $\hat{Y}(0)$ on the effects of the first childbirth (placebo birth) on childless women's income over 1000 placebo estimates. The horizontal lines provide OLS estimates of the effect of $\hat{Y}(0)$ with 1000 placebo estimates and the horizontal dashed line contains a 90% confidence band. The gray line provides our original quantile regression estimates of the effect of $\hat{Y}(0)$ at particular quantiles

7 Conclusion

This study explores the heterogeneous effects of first childbirth on women's income, utilizing data from the National Longitudinal Survey of Youth from 1979 to 2018. We estimate the full distribution of these effects by comparing the observed income of mothers to what their income would have been if they had not had children and maintained their rank in the conditional income distribution. Our analysis accounts for changes in income distribution over time. Specifically, under the identification assumptions of the Changes-in-Changes model, we first construct a counterfactual income distribution for mothers and then compare their actual income to a transformed version of their income from the previous period. It is important to note that our results are specific to women in the United States who completed their reproductive age approximately between 1997 and 2004, and the motherhood penalty may differ in other countries or different times due to varying social, economic, and policy environments.

Our findings reveal that approximately 73% of mothers experienced an income penalty following the birth of their first child while maintaining their previous rank in the income distribution. On average, these women lost around \$4120 in annual income after the birth of their first child; however, these effects exhibit significant heterogeneity among mothers, with the adverse effects of motherhood on income being particularly pronounced among only 10–20% of mothers. For example, at the 5th quantile, the estimated annual income of mothers is \$21,005 lower than it would have been in the absence of children, while at the 95th quantile, their annual income is estimated to be \$9887 higher than their counterfactual income. Furthermore, we observe variations in the motherhood penalty among mothers with distinct observable characteristics, with the most negatively affected group comprising older, single/divorced, white, more educated mothers, who had higher counterfactual incomes compared to what they would have earned had they not given birth. The captured heterogeneity in the motherhood penalty highlights the need for policy-makers to consider these differences when designing and implementing policies such as parental leave. Focusing solely on the average effect of first childbirth on women's

income may not adequately address the diverse experiences and needs of mothers across different socioeconomic backgrounds and income levels. Future research and policy discussions should take these nuanced effects into account to develop more targeted and effective strategies for supporting mothers in the workforce.

Acknowledgements We thank the Washington Center for Equitable Growth for their support of our research. We also would like to thank Jane Waldfogel, Teresa Ghilarducci, Brantly Callaway, Cynthia Bansak, and Jonathan Fisher for their helpful feedback and suggestions. We likewise thank audience members of the Washington Center for Equitable Growth's Grantee Conference and seminar participants at Widener University.

Author contributions Corresponding Authors: A.A.J. and W.H. Both authors contributed to the manuscript equally. Both authors read and approved the final manuscript.

Conflict of interest The authors declare no competing interests.

8 Appendix

8.1 Assigning placebo birth to childless women

Since childless women have never given birth to any children, we do not have a specific time of treatment for them. We use the method described by Kleven et al. (2019) to assign placebo births to childless women (i.e., the control group) and follow the procedure outlined below:

- We approximate the age at first child birth within each birth cohort (c) and education level cell (e) using a log-normal distribution. We categorize birth cohorts into four distinct groups (1957–1958, 1959–1960, 1961–1962, 1963–1964) and define education based on the highest level attained within four cells (less than high school, high school, some college, college).
- This distribution, denoted as $\mathcal{A}_{c,e}$, assumes a log-normal form $LN(\hat{\mu}_{c,e}, \hat{\sigma}_{c,e}^2)$ where $\hat{\mu}_{c,e}$ and $\hat{\sigma}_{c,e}^2$ represent the estimated mean and variance. These parameters are derived from the actual distributions observed within each cohort-education cell.
- We employ a random assignment technique by allocating placebo age at the first birth to childless women sampled from the estimated conditional distribution of mothers' age.

8.2 The heterogeneity of the treatment effect with respect to $\hat{Y}(0)$

We examine the variation in the impact of the first childbirth across mothers concerning what their income would have been if they had not become parents (i.e., $\hat{Y}(0)$). The quantile regression estimates resulting from regressing $Y(1) - \hat{Y}(0)$ on $\hat{Y}(0)$ are depicted in Fig. 10 in Appendix 8.2. The visualization indicates that, on average, the effect of the initial childbirth tends to be more negative for mothers who would have attained higher income levels had they not had any children and had maintained the same income rank as before childbirth. Among mothers experiencing the most

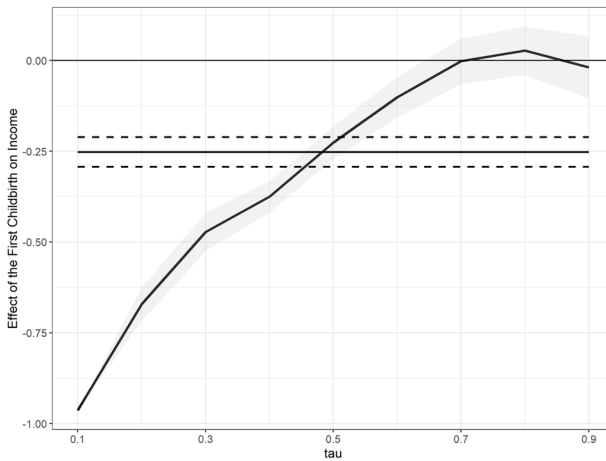


Fig. 10 Quantile regression estimates of the first childbirth effects on $Y(0)$. Quantile regression estimates of the effect of what mothers' income would have been if they had not had any children on the quantiles of the effect of the first childbirth (treatment effect). The solid horizontal line shows OLS estimates of the effect of $Y(0)$, and the horizontal dashed line contains a 90% confidence interval. The other line represents quantile regression estimates of the effect of $Y(0)$ at 0.1, 0.2, ..., 0.9 quantiles. The shaded area contains pointwise 90% confidence intervals from the quantile regressions

substantial negative effects (situated at the lower quantile), those who would have achieved higher income in the absence of childbirth and retained the same income ranking face notably more adverse effects. Conversely, for mothers experiencing slight negative impacts (or even positive effects) post-birth, there exists no distinction between the effects for those who would have attained high income versus those who would have had lower income in the absence of children.

8.3 Using alternative control group

Following existing literature on the motherhood penalty (like Blau & Kahn, 2017; Correll et al., 2007; Glauber, 2007 among others), we consider fathers as an alternative control group and repeat our analysis.

The summary statistics for fathers and mothers are presented in Table 2. Similarly, we estimate the counterfactual income distribution for mothers with three observed income distributions as in Eq. (2); income distribution of fathers one year after the birth of their first child, income distribution of fathers one year prior to the birth of their first child, and income distribution of mothers one year prior to their first childbirth.

Likewise, Fig. 11 shows the counterfactual distribution of income for mothers using fathers as the control group. Notably, when we use fathers as the control group, the estimated counterfactual income distribution for mothers represents higher values than our original estimate (i.e., childless women as the control group). In other words, we estimate that mothers would have had higher levels of income in the absence of childbirth compared to the case in which we consider childless women as the control group. This is not surprising as in Eq. (2), we use the observed income distributions of fathers at $T = 1$ and $T = -1$ to estimate the counterfactual income distribution for mothers. Since fathers,

Table 2 Summary statistics—mothers and fathers

	Mothers	Fathers	Difference	P val on Difference
Income at T=-1	15.74	19.5	-3.767	0
Income at T=1	11.78	21.32	-9.546	0
Change in Income	-3.96	1.82	-5.778	0
White	0.65	0.57	0.083	0
Married	0.74	0.67	0.073	0
Education beyond High School	0.49	0.34	0.15	0
Age	25.41	25.38	0.03	0.8
Age at 1st Birth	25.61	26.05	-0.446	0
N	1729	1919		

Source: National Longitudinal Survey of Youth 1979

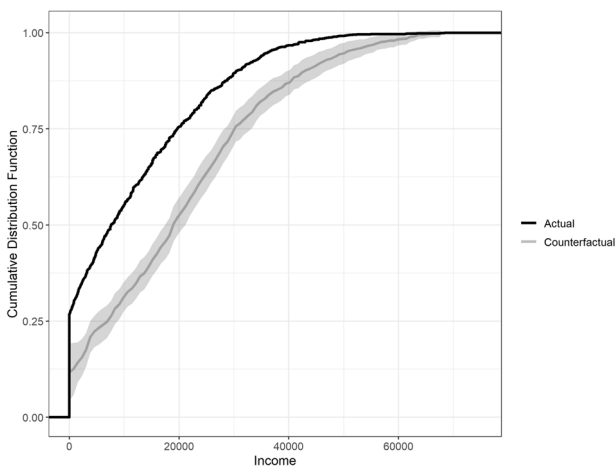


Fig. 11 The counterfactual distribution of income—fathers as the control group. Plots of the actual and counterfactual distribution of income for mothers using fathers as the control group. The counterfactual distribution also includes a 95% confidence interval computed using the empirical bootstrap with 1000 iterations

on average, have higher income values; therefore, the estimated counterfactual income distribution for mothers would sit on the right side of the actual distribution.

In Fig. 12 we also plot the distributional effects of the first childbirth on mothers' income as in Eq. (3) using fathers as the control group. Around 76% of mothers experience a drop in their income after the birth of their first child, and around 9% of mothers lose at least \$25,000 of their income.

Figure 13 contains the quantile regression estimates of the first childbirth effects on a set of covariates, using fathers as the control group. The results are consistent with our original estimates, except we capture higher heterogeneity for married mothers. Considering fathers as the control group, we can conclude that among mothers who appear to be most negatively affected by the first childbirth, older, married, white, more educated, and those who have postponed pregnancies tend to be the most affected.

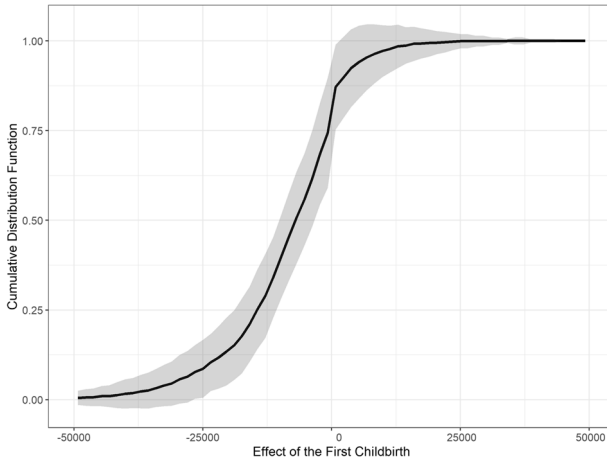


Fig. 12 The quantile of the treatment effect on the treated group- fathers as the control group. The black line plots the actual income distribution of mothers and the gray line plots the counterfactual distribution of income for mothers using fathers as the control group

8.4 Heterogeneity following Kleven et al. (2019)

To investigate the heterogeneity of the income gap between mothers and childless women, we start our analysis by following the event study approach proposed by Kleven et al. (2019). In their model, they looked into the sharp changes in the labor market outcomes around the birth of their first child to exploit within-person variation in the timing of first births. We run a quantile regression of the following equation at specific quantiles:

$$Y_{ist}^g = \sum_{j \neq -1} \alpha_j^g \cdot \mathbf{I}[j = t] + \sum_k \beta_k^g \cdot \mathbf{I}[k = age_{is}^g] + \sum_y \gamma_y^g \cdot \mathbf{I}[y = s] + \nu_{ist}^g \quad (8)$$

Where Y_{ist}^g is the income for individual i of group g in year s . The event time t is equal to zero at the time of the first childbirth and all other times were indexed relative to that year. We omit the event time dummy at $t = -1$, so estimated coefficients summarize the income responses to the childbirth relative to the base period of 1 year prior to the first birth. We include year-fixed effects and age-fixed effects. We convert the estimated level effect to the percentage by calculating $P_t^g \equiv \frac{\hat{\alpha}_t^g}{E[\tilde{Y}_{ist}^g]}$ Where \tilde{Y}_{ist}^g is the predicted income when we omit the contribution of event dummies. This means P_t^g captures the impacts of the first birth as a percentage of the no-child counterfactual at event time t .

Panel (a) in Fig. 14 plots the median impact of the first child at different event times. At $T=1$, mothers at the $\tau = 50th$ quantile of the income distribution experience about a 50% drop in their income compared to a year before the first child’s arrival. Figure 15 The gap also persists even ten years after birth.¹⁵ Panel (b) contains the impacts of the first birth on income for mothers (black line) and childless women (gray line) at $T=1$ relative to the year just before the first childbirth ($T=-1$) at

¹⁵ Figure 15 shows the impacts of the first childbirth on income for mothers and childless women at some other specific quantiles ($\tau = 0.3, 0.5, 0.7, 0.9$). The impacts are different at various quantiles of income, representing the heterogeneity of the impacts of the first childbirth on mothers’ income.

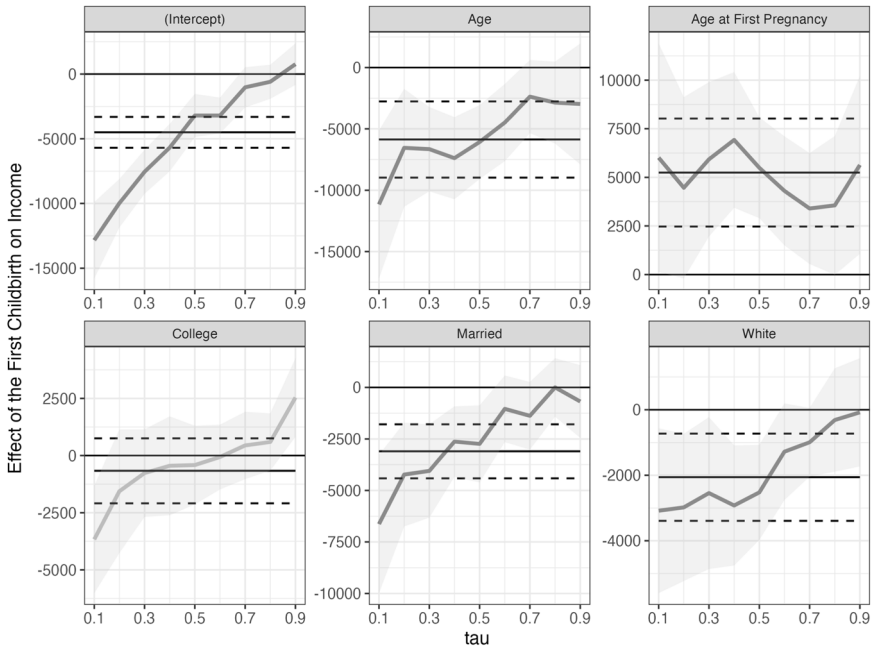


Fig. 13 Quantile regression estimates of the first childbirth effects on covariates—fathers as the control group. Quantile regression estimates of the effect of covariates on the quantiles of the effect of the first childbirth. The covariates include Age (26 years old and above), Col (beyond high school), Married, Age at first birth, and White. The solid red horizontal line provides OLS estimates of the effect of each covariate, and the horizontal dashed line contains a 90% confidence interval. The other lines provide quantile regression estimates of the effect of each covariate at particular quantiles from 0.1, 0.2, ..., 0.9. The shaded area contains pointwise 90% confidence intervals from the quantile regressions

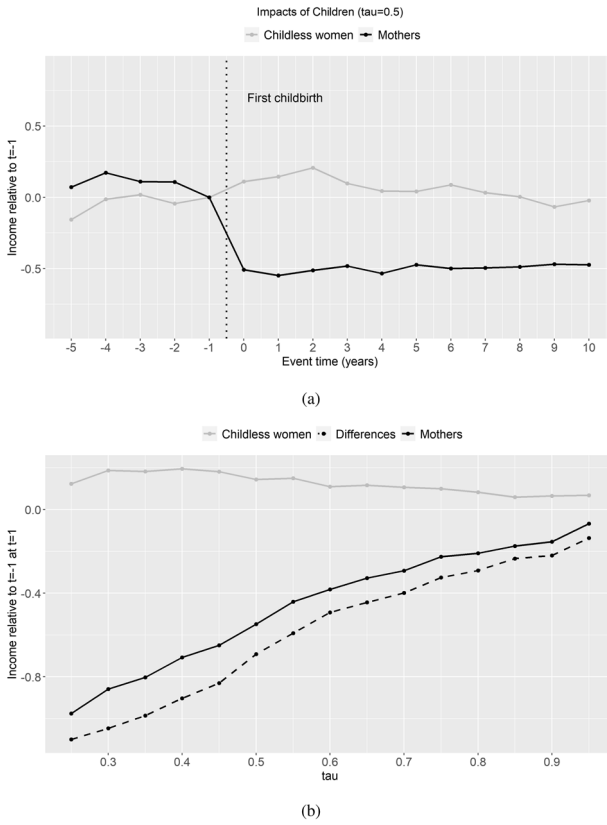


Fig. 14 Impacts of first childbirth among mothers and childless women. Panel (a) shows the median impacts of the first childbirth on income relative to $t=-1$ for mothers and childless women. Panel (b) plots the difference between event time coefficients at $T=1$ for mothers and childless women across different quantiles

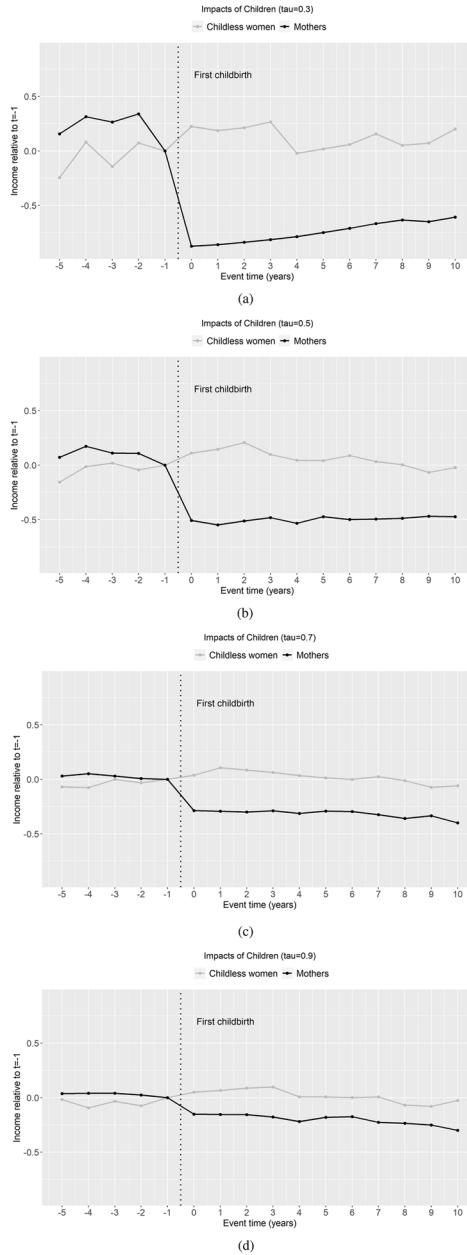


Fig. 15 Impacts of the first childbirth among mothers and childless women for different quantiles

several τ s. The dashed line provides the difference between these two lines. The plotted difference is much larger at lower quantiles as opposed to higher quantiles, suggesting the heterogeneity of the effect. In the next part, we use the distributional treatment effect analysis to better investigate the heterogeneous effects of birth on women's income.

References

- Abadie, A., Angrist, J., & Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, *70*(1), 91–117.
- Adda, J., Dustmann, C., & Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, *125*(2), 293–337.
- Ahn, N., & Mira, P. (2002). A note on the changing relationship between fertility and female employment rates in developed countries. *Journal of Population Economics*, *15*(4), 667–682.
- Albrecht, J., Van Vuuren, A., & Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the netherlands. *Labour Economics*, *16*(4), 383–396.
- Anderson, D. J., Binder, M., & Krause, K. (2003). The motherhood wage penalty revisited: Experience, heterogeneity, work effort, and work-schedule flexibility. *ILR Review*, *56*(2), 273–294.
- Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.
- Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, *74*(2), 431–497.
- Azadikhah Jahromi, A., & Callaway, B. (2022). The heterogeneous effects of job displacement on earnings. *Empirical Economics*, *62*, 213–245.
- Becker, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, *3*(1, Part 2), S33–S58.
- Bernhardt, E. M. (1993). Fertility and employment. *European Sociological Review*, *9*(1), 25–42.
- Bertrand, M., & Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, *94*(4), 991–1013.
- Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, *96*(4), 988–1012.
- Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865.
- Browning, M. (1992). Children and household economic behavior. *Journal of Economic Literature*, *30*(3), 1434–1475.
- Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American sociological review*, *66*, 204–225.
- Budig, M. J., & Hodges, M. J. (2010). Differences in disadvantage: Variation in the motherhood penalty across white women's earnings distribution. *American Sociological Review*, *75*(5), 705–728.
- Cain, G. G., & Dooley, M. D. (1976). Estimation of a model of labor supply, fertility, and wages of married women. *Journal of Political Economy*, *84*(4, Part 2), S179–S199.
- Callaway, B., & Li, T. (2019). Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics*, *10*(4), 1579–1618.
- Callaway, B., Li, T., & Oka, T. (2018). Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods. *Journal of Econometrics*, *206*(2), 395–413.
- Carrasco, R. (2001). Binary choice with binary endogenous regressors in panel data: Estimation the effect of fertility on female labor participation. *Journal of Business & Economic Statistics*, *19*(4), 385–394.
- Chernozhukov, V., & Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, *73*(1), 245–261.
- Chernozhukov, V., Fernandez-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268.
- Cooke, L. P. (2014). Gendered parenthood penalties and premiums across the earnings distribution in australia, the united kingdom, and the united states. *European Sociological Review*, *30*(3), 360–372.
- Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, *112*(5), 1297–1338.

- Cortes, P., & Pan, J. (2018) Occupation and gender. *The Oxford handbook of women and the economy*, 425–452.
- Cramer, J. C. (1980). Fertility and female employment: problems of causal direction. *American Sociological Review*, 45(2), 167–190.
- Davies, R., & Pierre, G. (2005). The family gap in pay in europe: A cross-country study. *Labour Economics*, 12(4), 469–486.
- England, P., Bearak, J., Budig, M. J., & Hodges, M. J. (2016). Do highly paid, highly skilled women experience the largest motherhood penalty? *American sociological review*, 81(6), 1161–1189.
- Gangl, M., & Ziefle, A. (2009). Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in britain, germany, and the united states. *Demography*, 46(2), 341–369.
- Glauber, R. (2007). Marriage and the motherhood wage penalty among african americans, hispanics, and whites. *Journal of Marriage and Family*, 69(4), 951–961.
- Glauber, R. (2018). Trends in the motherhood wage penalty and fatherhood wage premium for low, middle, and high earners. *Demography*, 55(5), 1663–1680.
- Goldin, C. (2014). A grand gender convergence: Its last chapter. *American economic review*, 104(4), 1091–1119.
- Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4), 487–535.
- Hill, M. S. (1979). The wage effects of marital status and children. *Journal of Human Resources*, 579–594.
- Iacovou, M. (2001). *Fertility and female labour supply* (No. 2001–19). ISER Working Paper Series.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1), 4–29.
- Jee, E., Misra, J., & Murray-Close, M. (2019). Motherhood penalties in the us, 1986–2014. *Journal of Marriage and Family*, 81(2), 434–449.
- Kalwaj, A. S. (2000). The effects of female employment status on the presence and number of children. *Journal of Population Economics*, 13(2), 221–239.
- Killewald, A., & Bearak, J. (2014). Is the motherhood penalty larger for low-wage women? a comment on quantile regression. *American sociological review*, 79(2), 350–357.
- Kleven, H., Landais, C., & Søgaaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181–209.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019, May). Child penalties across countries: Evidence and explanations. In *AEA Papers and Proceedings* (Vol. 109, pp. 122–126). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.
- Kleven, H., Landais, C., & Søgaaard, J. E. (2021). Does biology drive child penalties? evidence from biological and adoptive families. *American Economic Review: Insights*, 3(2), 183–98.
- Koenker, R. *Quantile regression*. volume 38. Cambridge University Press (2005).
- Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.
- Korenman, S., & Neumark, D. (1992). Marriage, Motherhood, and Wages. *The Journal of Human Resources*, 27(2), 233.
- Kuziemko, I., Pan, J., Shen, J., & Washington, E. (2018). *The mommy effect: Do women anticipate the employment effects of motherhood?* (No. w24740). National Bureau of Economic Research.
- Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2), 319–323.
- Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. *Review of Labor Economics*, 68(4), 543–572.
- Melly, B., & Santangelo, G. (2015). The changes-in-changes model with covariates. *Universität Bern, Bern*, 720, 721.
- Molina, J. A., & Montuenga, V. M. (2009). The motherhood wage penalty in spain. *Journal of Family and Economic Issues*, 30(3), 237–251.
- Moore, W. J., & Wilson, R. M. (1982). The influence of children on the wage rates of married women. *Eastern Economic Journal*, 8(3), 197–210.
- Pal, I., & Waldfogel, J. (2016). The family gap in pay: New evidence for 1967 to 2013. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(4), 104–127.
- Rosenzweig, M. R., & Wolpin, K. I. (1980). Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 88(2), 328–348.

- Shapiro, D., & Mott, F. L. (1979). Labor supply behavior of prospective and new mothers. *Demography*, 16(2), 199–208.
- Stanca, L. (2012). Suffer the little children: Measuring the effects of parenthood on well-being worldwide. *Journal of Economic Behavior & Organization*, 81(3), 742–750.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.
- Waldfogel, J. (1997). The effect of children on women's wages. *American Sociological Review*, 62(2), 209–217.
- Waldfogel, J. (1998). Understanding the “family gap” in pay for women with children. *Journal of Economic Perspectives*, 12(1), 137–156.
- Winder, K. L. (2008) fertility and the motherhood wage penalty. *The University of California, Merced*, <http://citeseerx.ist.psu.edu/viewdoc/download>.
- Zhang, X. (2009). *Earnings of women with and without children* (pp. 5–13). Ottawa: Statistics Canada.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.