Parenthood Timing and Gender Inequality

Julius Ilciukas*

November 20, 2024

Link to Latest Version

Abstract

I study how parenthood affects women's labor market outcomes and gender inequality. I introduce a method to quantify the impacts that simultaneously addresses selective parenthood timing and parenthood timing-dependent effects. The method leverages quasi-experimental variation in the success of assisted conception procedures throughout women's entire treatment histories. Using administrative Dutch data, I find that parenthood persistently reduces women's work hours and income by 9 to 24 percent, accounting for up to half of post-child gender inequality in these outcomes. I also disentangle and quantify how selective timing and timing-dependent effects bias conventional estimators, providing insight into the conflicting findings in the literature. My approach is applicable to other settings where individuals are quasi-experimentally assigned to one state but may enter others either through direct selection or by opting into quasi-experimental reassignment.

*University of Amsterdam. I thank Francesco Agostinelli, Monique de Haan, Phillip Heiler, Artūras Juodis, Hessel Oosterbeek, Erik Plug, Benjamin Scuderi, Arthur Seibold, Giuseppe Sorrenti, Mel Stephens, Bas van der Klaauw, Yun Xiao, Basit Zafar, Alminas Zaldokas, Lina Zhang, Yang Zhong, conference participants at AFEPOP, Berlin School of Economics Gender Workshop, COMPIE, EALE, EEA-ESEM, ESPE, Luxembourg Gender and Economics Workshop, Warwick Economics PhD Conference, and seminar participants at the University of Michigan and the University of Pennsylvania. The non-public micro data used in this paper is available via remote access to the Microdata services of Statistics Netherlands (CBS). All URLs accessed November 20, 2024.

1 Introduction

The differential impact of parenthood on the careers of women and men is widely considered a major contributor to gender disparities in the labor market (Goldin, 2014; Blau & Kahn, 2017; Bertrand, 2020; Cortés & Pan, 2023; Kleven et al., 2024). Quantifying this impact is crucial for understanding the sources of gender inequality and designing effective policies to address it. However, this has proven difficult due to two factors commonly encountered in applied economic research: selection and dynamic effects. Selection arises when the decision to have children or the timing of parenthood is related to labor market outcomes independent of fertility. Dynamic effects occur when the impacts vary depending on the timing of parenthood.

I propose a method to quantify the effects of parenthood that is simultaneously robust to selective parenthood timing and dynamic effects. The method leverages quasi-experimental variation in the success of assisted conception procedures (ACP). It compares labor market outcomes between women who become mothers through their first ACP and those who remain childless after its failure. This comparison ensures robustness to dynamic effects by eliminating differences in parenthood timing. The key challenge the method addresses is that the group of women who remain childless after their first ACP fails may be selective. To overcome this, it first leverages women's complete ACP histories, accounting for births resulting from subsequent ACPs. I show that a weighting scheme, which assigns greater weight to childless women with more failed ACPs, can be used to identify the average childless labor market outcomes for women reliant on ACPs to conceive. After leveraging all quasi-experimental variation in fertility due to ACPs, the method applies a bounding procedure to account for non-ACP births. I show that the share of women reliant on ACPs to become mothers can be identified. Then, it can be used to construct worst-case bounds for average motherhood labor market outcomes of this group by trimming the tails of the outcome distribution among women whose first ACP succeeded. The only crucial assumption for my method is that the success of each ACP a woman chooses to undergo is as good as random, conditional on observables.

The identified bounds are sharp, meaning no effect within them can be ruled out without additional assumptions or data. To tighten the baseline bounds, I assume that women who have non-ACP children after a successful ACP would have had at least one child if ACPs had failed. This assumption is consistent with the idea that families are more determined to have at least one child than to have additional children. To estimate the bounds and justify asymptotic inference, I build on the double/debiased machine learning approach developed by Semenova (2023). This allows me to account for uncertainty in the second step of my approach, which trims the tails of the outcome distribution.

I apply my estimator to administrative Dutch data, linking detailed labor market information from tax records with comprehensive hospital medical records. My analysis focuses on couples attempting to conceive their first child through intrauterine insemination, also

known as artificial insemination. I find that, in the first three years of parenthood, women experience reductions in both work hours and income between 6% and 33%. I then impose the auxiliary assumption regarding additional non-ACP births. The bounds leveraging this assumption indicate yearly reductions in women's work hours between 10% and 25% and decreases in income between 9% and 29%. These bounds remain stable for at least seven years into parenthood. The bounds for men are similar in length but centered around zero. During this period, parenthood causes between 36% and 54% of gender inequality in work hours and up to 46% in income.

Currently, two methods dominate the debate on the career impacts of parenthood; each addresses one of the two identification challenges, yet they provide conflicting evidence. The event study approach, popularized by Kleven et al. (2019, 2024), compares labor market outcomes between women with and without children. While this method accommodates dynamic effects, it assumes that fertility timing is not selective. The instrumental variable approach, introduced by Lundborg et al. (2017, 2024), addresses selective fertility by focusing on women undergoing in-vitro fertilization—one type of assisted conception procedure. This method assumes that the effects do not depend on parenthood timing. Event study estimates based on Danish data attribute most of the gender inequality in earnings to parenthood, while instrumental variable estimates suggest that parenthood has little impact on gender inequality (Lundborg et al., 2024). Moreover, event study estimates from the in-vitro fertilization sample are nearly identical to those from the general population. This complicates attributing differences in results between the two methods to sample differences.

I introduce procedures to assess the potential bias in the leading methods. I begin by implementing the instrumental variable and event study approaches on Dutch data. I find that their estimates differ substantially, mirroring findings from Denmark (Lundborg et al., 2024). Moreover, neither the event study nor instrumental variable estimates fall within my bounds. However, these differences do not necessarily imply bias because the three methods target different sub-populations, even when applied to the same dataset. To quantify the potential bias in the instrumental variable approach, I modify my baseline method to bound the effects of delaying parenthood. My results suggest that dynamic effects could either have a limited impact on the instrumental variable estimates or lead to a substantial underestimation of the career cost of motherhood; neither possibility can be ruled out without additional assumptions. For the event study approach, I use the timing of failed ACPs as a proxy for selective fertility timing. I then conduct a placebo event study to quantify its relationship with labor market outcomes in the absence of children. I find that selective fertility may lead the event study estimates to overstate both the career costs of motherhood and the benefits of fatherhood, resulting in an overestimation of parenthood's impact on gender inequality.

My work is most closely related to two recent working papers that exploit women's first in-vitro fertilization procedure and carefully address dynamic effects: Bensnes et al.

(2023) and Gallen et al. (2023). These studies rely on parametric assumptions about effect heterogeneity across women and the structure of dynamic effects. Such assumptions allow them to use short-run instrumental variable estimates to correct the bias that emerges as more women become mothers. Compared to this approach, the primary advantage of my method is that it does not require parametric assumptions. Additionally, my method does not rely on longitudinal data structure, allowing to estimate impacts on outcomes observed irregularly or even only once. Leveraging my identification results, I also develop a joint test for the two parametric assumptions. I find that these assumptions are rejected in my setting. This procedure can be further applied to test alternative assumptions that enable identification.

My work is also linked to the broader literature on the effects of children on gender inequality in the labor market (see Bertrand (2011); Blau & Kahn (2017) for a detailed overview). This includes studies that focus on the extensive fertility margin rather than parenthood itself (Rosenzweig & Wolpin, 1980; Bronars & Grogger, 1994; Angrist & Evans, 1996; Jacobsen et al., 1999; Iacovou, 2001; Cruces & Galiani, 2007; Maurin & Moschion, 2009; Hirvonen, 2009; Vere, 2011). It also includes studies on the effects of parenthood that rely on assumptions about dynamic effects but address selection by exploiting miscarriages (Hotz et al., 2005), infertility shocks (Agüero & Marks, 2008; Cristia, 2008), variation in access to abortion (Miller, 2011; Brooks & Zohar, 2021), and contraceptive failures (Gallen et al., 2023). Finally, it relates to the growing literature on the effects of parenthood that addresses dynamic effects but relies on arguably stronger assumptions about selective fertility, including studies leveraging differences in fertility timing (Fitzenberger et al., 2013; Angelov et al., 2016; Chung et al., 2017; Bütikofer et al., 2018; Kleven et al., 2019; Eichmeyer & Kent, 2022; Melentyeva & Riedel, 2023) and structural approaches (Adda et al., 2017).

My primary contribution to the literature on the career impacts of parenthood is providing estimates that are simultaneously robust to selective fertility and dynamic effects. I demonstrate that these factors may substantially bias leading estimators, and that accounting for them can reconcile conflicting findings in the literature. My secondary contribution concerns the external validity of my estimates. This study is the first to use intrauterine insemination and the first to use any ACP data outside of Scandinavia. In doing so, it mitigates several concerns raised about prior studies that focus solely on in-vitro fertilization and use data from Denmark or Sweden. Intrauterine insemination is the primary ACP in most countries and is less invasive and more accessible than in-vitro fertilization. This addresses concerns about sample selectivity, (mental) health side effects, and relevance for countries where in-vitro fertilization access is limited. Additionally, Dutch family policies align with the OECD average, making my results more relevant for common policy settings. To address mental health side effects further, I adapt my method to bound career impacts specifically for women who do not uptake antidepressants after failing to conceive. I further address sample selectivity by introducing a procedure to estimate effects for non-ACP par-

ents by imputing their childless career trajectories from families who attempted but failed to conceive via ACPs at a similar moment. Estimates from both methods remain consistent with my baseline results.

Methodologically, my approach builds on ideas from two distinct branches of literature. The first step of my approach, which accounts for selection into parenthood via subsequent ACPs, leverages insights from the extensive biostatistics literature on dynamically assigned treatments (see Hernán & Robins (2020) for an overview). In economics, it is most closely related to a procedure developed by Van den Berg & Vikström (2022), which explicitly incorporates treatment assignment eligibility. These methods are inapplicable in my setting because individuals may take up treatment without assignment, conceiving via non-ACP means. The second step of my approach, which addresses selection into parenthood through non-ACP means, relates to the extensive literature on bounds for treatment effects, beginning with Manski (1989, 1990). It is most closely related to a procedure typically used to account for sample selection, introduced by Zhang & Rubin (2003) and further developed by Lee (2009) (henceforth ZRL). While sample selection is conceptually distinct from dynamic effects, the ZRL method can be adapted to bound the effects of parenthood in my setting. My approach, which leverages women's entire ACP histories before resorting to bounding, offers two advantages. First, it allows to bound the effects for a more general group—women reliant on ACPs to conceive, rather than only those reliant on their first ACP. Second, it results in mechanically narrower bounds. I demonstrate that, in my application, ZRL bounds are at least several times wider, making them uninformative. I present a detailed discussion of how my approach relates to and differs from these methods in Section 3.4.

My primary contribution to the methodological literature is an approach to bound treatment effects in settings with quasi-experimental treatment assignment but imperfect compliance. Particularly, in cases where individuals transition from the originally assigned state either by undergoing multiple quasi-experimental assignments or through entirely selective pathways. While in my application these states represent entering parenthood at different points in time, they could also correspond to entirely different treatments. Examples include educational programs with multiple admission cycles, job training programs where unassigned individuals can reapply for alternative programs, legal settings where sanctions are applied non-deterministically and initially unsanctioned individuals may later be sanctioned, and clinical trials in the extension phases where participants can enroll in other trials or pursue alternative therapies. My secondary contribution is addressing another limitation of the instrumental variable approach when estimating impacts on one outcome over time. Specifically, the group for which effects are identified may change, complicating interpretation even when impacts do not depend on the timing of treatment. After introducing my method, I demonstrate how it can be used to quantify effects over time for a stable group.

The rest of the paper is structured as follows. Section 2 introduces the potential outcomes framework. Section 3 demonstrates the identification challenge, presents intuition

for the bounding approach, states the formal results, and discusses relations to existing methodological literature. Section 4 introduces the estimator. Section 5 describes the institutions, ACPs, and the data, and presents support for the identification assumptions. Section 6 presents the main estimates of the effects of parenthood on women's labor market outcomes and gender inequality. Section 7 covers extensions, including assessing the bias in existing approaches, addressing external validity concerns related to mental health, quantifying the effects over time for a stable group of individuals, and assessing the effects of parenthood in the general population. Section 8 concludes.

2 Model

In this section, I introduce a static model sufficient to demonstrate the identification challenge and the bounding approach without loss of generality. I generalize it to a dynamic version to present extensions in Section 7. The model is a modified version of the local average treatment effect (LATE) framework (Angrist & Imbens, 1995). D is a treatment indicator, which may be selective, representing whether a woman has any children. Z_1 is the treatment assignment indicator, representing whether a woman's first ACP succeeded. The population consists of women who underwent ACP for their first child.

When Z_1 is as good as randomly assigned and affects the outcome Y only through D (and affects D for at least some women), it is a valid instrument for D. In the context of parenthood, a concern is that Z_1 may affect Y through parenthood timing independent of parenthood status because women who become mothers through their first ACP have children earlier than those who become mothers after their first ACP fails. To formalize this concern, I distinguish three potential outcomes. $Y(1)$ is the potential outcome in the case that a woman becomes a mother at the first ACP; I refer to it as the *treated outcome*. $Y(0)$ is the potential outcome in the case that a woman remains childless; I refer to it as the *control outcome*. $Y(2)$ is the potential outcome in the case that a woman becomes a mother after the first ACP fails; I refer to it as the *later-treated outcome*. The key feature complicating identification is that $Y(1)$ may differ from $Y(2)$; the exact meaning or uniqueness of $Y(2)$ is irrelevant for the bounding approach I propose.[1] Formally, the relationship between realized and potential outcomes is:

$$Y = Y(0)(1 - D) + Y(1)DZ_1 + Y(2)D(1 - Z_1).$$

Most women whose first ACP fails undergo ACPs again, and my method leverages variation in parenthood resulting from the outcomes of these subsequent ACPs. To formalize this, I extend the LATE framework to describe how parenthood status depends not only

[1] The approach does not use realized later-treated outcomes; I require SUTVA for $Y(1)$ and $Y(0)$.

on the outcome of a woman's first ACP but also on the outcomes of subsequent ACPs.[2] I characterize each woman by two unobserved variables. First, $W \in \{1, \ldots, \overline{w}\}$, which is the total number of ACPs a woman would undergo for her first child if all previous ACPs failed. I refer to W as the willingness to undergo ACPs, although it only describes women's behavior in the scenario that all ACPs fail and does not require any broader interpretation. W may be related to potential outcomes. Second, R, which indicates if a woman would remain childless if all W ACPs failed. I refer to R as reliance on ACPs. It may be related to both the potential outcomes and the willingness to undergo ACPs. I refer to women with $R = 1$ as *reliers*, meaning that they are reliant on ACPs to have children, and women with $R = 0$ as *non-reliers*, meaning that they would have children even if all ACPs failed.

Reliers are the focus of this paper. They are the most general group of women whose parenthood status depends on ACP success, and such dependency is essential for obtaining informative results.[3] Reliers are closely related to compliers in the LATE framework, which are the women who would remain childless if their first ACP failed. However, reliers are a more general group, meaning that compliers are a subset of reliers. This is because, in addition to compliers, reliers also include women who would become mothers through a subsequent ACP if the first ACP failed but would remain childless if all ACPs failed; such women are part of the always-takers in the LATE framework. There are no never-takers or defiers because few women who conceive via ACPs end up childless, but all results can be extended to a setting with never-takers.

The observed indicator for the success of ACP j is Z_j. It takes the value 1 if the ACP succeeded, and 0 either if the ACP failed or if a woman did not undergo ACP j. To simplify notation, this only includes ACPs that occur before the first child. All women who experience ACP success have at least one child, which implies that indicators for the success of ACPs following another successful ACP take the value 0. In the case that all previous ACPs fail, women undergo at most as many ACPs as they are willing to. Formally:

$$Z_j = 0 \text{ for all } j \text{ such that } (Z_l = 1 \text{ for any } l < j) \text{ or } (W < j).$$

The total number of ACPs a woman undergoes is A, and it is also observed. A woman undergoes ACPs either until one succeeds or until she reaches the maximum number of ACPs she would undergo if all previous ACPs failed. Formally:

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

By definition, among women who never experienced ACP success, the realized number of

[2] More generally, control group individuals can selectively enter a sequence of quasi-experiments, gaining additional chances for treatment assignment, and may also obtain treatment without assignment.

[3] When outcomes have bounded support, effects for all women, including non-reliers, can be bounded using Horowitz & Manski (2000), but these bounds may be uninformative. In my application, seven years post-conception, such bounds on hours range from an 80% reduction to a 70% increase.

ACPs equals the number of ACPs a woman would undergo if all ACPs failed. A woman is a mother either if her last ACP succeeded or if she is a non-relier and had a child after all ACPs failed. Formally, the outcome of the last ACP a woman underwent is Z_A, and the relationship between the parenthood indicator and ACP success is:

$$D = Z_A + (1 - Z_A)(1 - R).$$

The main effect I focus on is becoming a parent at the first ACP relative to remaining childless. I discuss other effects in Section 7. The individual-level treatment effect is defined as the difference between the treated and control outcomes:

$$\tau = Y(1) - Y(0).$$

The average treatment effect (ATE) is:

$$\tau_{ATE} = \mathbb{E}[\tau].$$

The central parameter I focus on is the average treatment effect for reliers (ATR):

$$\tau_{ATR} = \mathbb{E}[\tau | R = 1].$$

For comparing my approach to the instrumental variable method, the local average treatment effect (LATE) is:

$$\tau_{LATE} = \mathbb{E}[\tau | C = 1],$$

where C is the complier indicator, which takes the value 1 if a woman would remain childless if her first ACP failed, and 0 otherwise.

3 Identification

In this section, I first describe the limitations of the intstrumental variable (IV) approach. Then, I present the intuition behind my approach, followed by formal results. Afterward, I discuss how my approach relates to the existing methodological literature.

To demonstrate the intuition, I leverage the (unconditional) sequential unconfoundedness assumption:

Assumption 1 (Sequential Unconfoundedness)**.**
$(Y(k), R, W) \perp\!\!\!\perp Z_j \mid A \geq j$, *for all* j, k.

It states that, among women who enter ACP j, the outcome of ACP j is as good as random; specifically, it is independent of potential outcomes and type. The intuitive idea behind this assumption is that among women whose previous ACPs have failed and who undergo an

additional insertion of embryos or sperm into the uterus, whether or not they get pregnant from this insertion is essentially random. This assumption concerns only the last stage of each ACP and not the earlier stages that may lead up to it, such as hormonal stimulation or, in the case of IVF, embryo fertilization in the lab. It does not restrict the relationship between potential outcomes and selection into subsequent ACPs, nor does it limit the relationship between potential outcomes and selection into parenthood via non-ACP means. To simplify exposition, I do not distinguish between IVF and intrauterine insemination in this section, but the main identification method accounts for procedure-dependent success rates and selection into different procedures. I present empirical evidence to support the sequential unconfoundedness assumption in Section 5.3.

3.1 Bias in the Instrumental Variable Approach

The IV approach uses the success of women's first ACP as an instrument for parenthood. It starts with the reduced form, which is the difference in average outcomes between those whose first ACP succeeded and those whose first ACP failed. This means a group of women who conceived on their first ACP is compared to a mixed group consisting of childless women (the compliers) and women who had children later (the always-takers). Under (sequential) unconfoundedness, the reduced form identifies a linear combination of two effects. First, the average treatment effect for compliers, and second, the effect of conceiving earlier versus later for always-takers:

$$\begin{aligned}\mathbb{E}[Y|Z_1=1]-\mathbb{E}[Y|Z_1=0] =& \,\mathbb{E}[Y(1)-Y(0)|D=0,Z_1=0]\Pr(D=0|Z_1=0)\\ &+\mathbb{E}[Y(1)-Y(2)|D=1,Z_1=0]\Pr(D=1|Z_1=0).\end{aligned}$$

Scaling the reduced form by the difference in the share of mothers between the two groups—the first stage—yields:

$$\frac{\mathbb{E}[Y|Z_1=1]-\mathbb{E}[Y|Z_1=0]}{\mathbb{E}[D|Z_1=1]-\mathbb{E}[D|Z_1=0]} = \tau_{LATE}+\mathbb{E}[Y(1)-Y(2)|C=0]\frac{\Pr(C=0)}{\Pr(C=1)}.$$

The second term on the right-hand side is the average effect of becoming a mother earlier relative to later for always-takers, scaled by the always-taker-to-complier ratio. When the outcomes do not depend on the moment of becoming a mother, meaning $Y(1)=Y(2)$, the second term drops out, and τ_{LATE} is identified. In the standard Rubin (1974) model with only one motherhood outcome, this assumption is covered by the no-multiple-versions-of-treatment (SUTVA). Otherwise, the second term biases the IV estimator of τ_{LATE}.[4]

[4]Another way to describe this bias is by using the "negative weights" terminology popularized by the recent literature on difference-in-differences (see Roth et al. (2023) for on overview). For example, when the always-taker-to-complier ratio is 3, the IV estimator can be thought of as assigning a weight of 4 to τ_{ATE} and a weight of -3 to the average effect of delayed parenthood for always-takers, $\mathbb{E}[Y(2)-Y(0)|C=0]$. Difference-in-differences methods are inapplicable in this setting because parenthood timing may be selective.

In the context of parenthood, the theoretical direction of the bias is ambiguous. On the one hand, it may understate the career costs of motherhood if women who have children later face greater child care demands at the peak of their careers, when their work hours and earning potential are highest. On the other hand, it could overstate the costs if women who become mothers earlier miss crucial career-building years, have more children, or lack the resources to access formal child care, all of which may negatively affect their long-term career trajectory. Even when the effect of becoming a mother earlier is small relative to the treatment effect, the bias may be large due to the relative size of the always-taker group. In practice, four years after the first procedure, the always-taker-to-complier ratio is 3.

3.2 Intuition for Bounding Approach

In this section, I present the intuition behind my bounding approach. I separately explain how I identify the relier average control outcome and bound their average treated outcome, how I leverage pre-ACP covariates to make the bounds sharp, and how I tighten them using additional assumptions.

3.2.1 Control Outcome

To demonstrate how the relier average control outcome can be identified, I first express it as a weighted average of childless outcomes among reliers with different willingness to undergo ACPs:

$$\mathbb{E}[Y(0)|R=1] = \sum_{w=1}^{\overline{w}} \mathbb{E}[Y(0)|R=1, W=w] \Pr(W=w|R=1).$$

I next describe how each element in the sum can be identified. First, the average outcome among women who underwent exactly w ACPs and remained childless identifies the average control outcome for reliers willing to undergo exactly w ACPs:

$$\mathbb{E}[Y|A=w, D=0] = \mathbb{E}[Y(0)|W=w, R=1].$$

This result holds because, among childless women, the control outcome is observed, and women who underwent exactly w ACPs and remained childless are a random subsample of reliers willing to undergo exactly w ACPs. This follows from two key observations. First, all women who undergo exactly w ACPs and remain childless must be reliers willing to undergo exactly w ACPs, as non-reliers would have had children, and reliers willing to undergo additional ACPs would have done so. Second, conditional on being willing to undergo w ACPs and being a relier, whether a woman undergoes w ACPs and remains childless is effectively random: it depends solely on whether any ACP up to w succeeds, and the outcome of each ACP is as good as random.

The shares of different types can be identified following similar arguments. First, women

who experience at least w failed ACPs are a random subsample of women willing to undergo at least w ACPs, hence, the share of such women initiating a subsequent ACP identifies the share of women willing to undergo at least $w+1$ ACPs in this group:

$$\Pr(A \geq w+1 \mid A \geq w, Z_w = 0) = \Pr(W \geq w+1 \mid W \geq w). \tag{1}$$

Second, women who do not undergo an additional ACP after their previous w ACPs fail are a random subsample of women willing to undergo exactly w ACPs, hence, the share of such women who remain childless identifies the share of women reliant on ACPs in this group:

$$\Pr(D = 0 \mid A = w, Z_w = 0) = \Pr(R = 1 \mid W = w). \tag{2}$$

Combining these probabilities allows to construct $\Pr(W = w, R = 1)$ for all w, meaning that the shares of all types are identified.

An important special case arises when women conceive solely through ACPs, making all women reliers. In this scenario, the above approach identifies the average control outcome. Combining it with the average treated outcome identified from women whose first ACP succeeded allows to point-identify τ_{ATE}. However, if some women become mothers independent of ACP outcomes, point-identifying τ_{ATE} without additional assumptions becomes impossible, as control outcomes for such women can never be observed. Moreover, since treated outcomes are only observed among women who conceived through their first ACP and it cannot be determined which of these women are reliers, the relier average treated outcome cannot be identified either. This prevents the point identification of τ_{ATR}. Next, I describe how the relier average treated outcome can be bounded to obtain bounds on τ_{ATR}.

3.2.2 Treated Outcome

I bound the relier average treated outcome using the distribution of outcomes among women whose first ACP succeeded. Since the success of the first procedure is as good as random, this distribution reflects the treated outcomes of all women entering ACPs. Combined with the relier share identified in the previous step, this allows to construct worst-case bounds for the relier average treated outcome by assuming they either have the lowest or highest treated outcomes of all women entering ACPs.

To illustrate the intuition, suppose there are 100 women whose first ACP succeeded, and the first step identifies that 80% of women are reliers. Then, by unconfoundedness, there are approximately 80 reliers among the 100 women, and their expected outcome is the same as the relier average treated outcome. While it is not known which 80 out of the 100 women are the reliers, the upper bound on their average treated outcome can be constructed by selecting the 80 women with the highest outcomes. Panel A in Figure 1 graphically demonstrates this intuition using 100 women whose treated outcomes are

uniformly distributed between 1 and 100. The left graph illustrates excluding the 20 women with the highest outcomes to identify the lower bound by averaging outcomes among the remaining 80 women. The right graph repeats this for the upper bound. The true location of the 80 reliers must be between these two extremes, meaning their average treated outcome must be between the averages of the two trimmed distributions.

3.2.3 Narrowing Bounds with Pre-ACP Covariates

When ACP success is as good as random conditional on pre-ACP covariates, these covariates can be used to narrow the bounds without additional assumptions. To illustrate the intuition, suppose there are 100 women whose first ACP succeeded, and that before ACP, 80 of them had low levels of education, while 20 were highly educated. Since ACP outcomes are as good as random conditional on observables, the baseline approach allows to identify the relier share within each pre-ACP education group. Suppose the first step identifies that in each groups, 80% of the women are reliers. This means that out of the 80 reliers among the 100 women whose ACP succeeded, 16 should be highly educated, and 64 should have low levels of education. Constructing the bounds without accounting for this information may lead to selecting women in a way that is inconsistent with these shares, making the baseline bounds overly conservative. The new lower bound is constructed by selecting the corresponding number of women with the lowest outcomes from each education group. This can only result in a weakly higher lower bound compared to selecting the 80 women with the lowest outcomes, without taking education into account. Panel B in Figure 1 illustrates this intuition graphically, where, for simplicity, I assume all highly educated women have higher outcomes than all women with low levels of education. The left panel trims 4 women with the highest outcomes from the highly educated group and 16 with low levels of education, corresponding to 20% non-reliers in each group. This results in trimming fewer women with high outcomes and, in turn, a higher average outcome among the remaining women than in the scenario where education is ignored. The right panel applies the same reasoning to the upper bound, resulting in a lower upper bound.

3.2.4 Narrowing Bounds with Post-ACP Outcomes

The bounds can be narrowed further with additional information on which women whose first ACP succeeded are (or are not) reliers. One such piece of information might be women's fertility outcomes beyond the first birth. For instance, it could be reasonable to assume that women who had non-ACP children after their first ACP succeeded would have also had at least one non-ACP child if all ACPs had failed. To illustrate the intuition, suppose there are 100 women whose first ACP succeeded and, for simplicity, abstract from pre-ACP covariates discussed in the previous section. Further suppose that in addition to identifying that 80% of the 100 women are reliers, as before, 10 of the 100 are observed to have a second non-ACP child. It is guaranteed that these 10 women are not reliers and they can

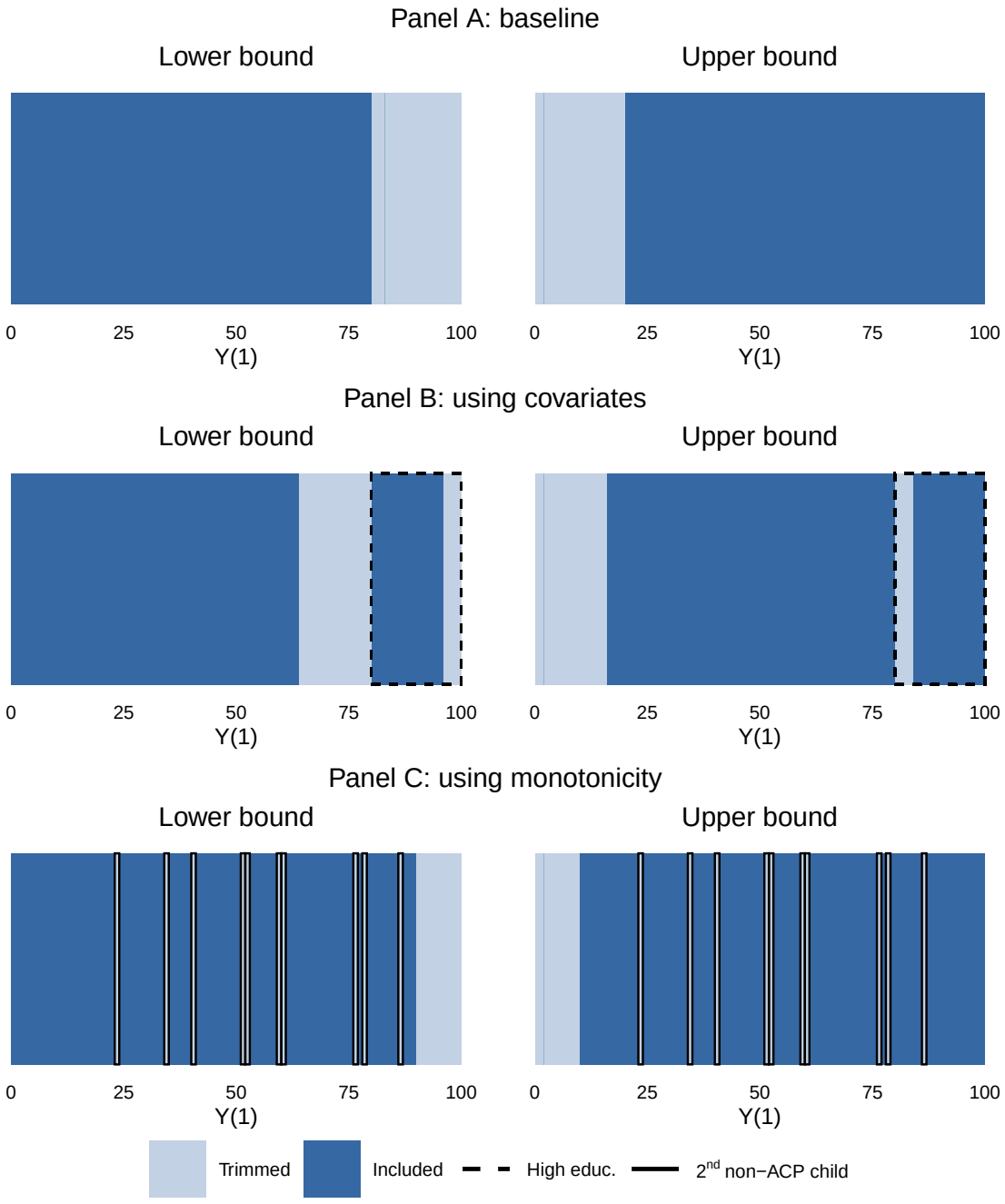

Figure 1: Intuition for Bounds

be excluded before selecting the 80 potential reliers to construct the bounds. Panel C in Figure 1 demonstrates this intuition graphically. In both the left and right figures, the 10 women who had a second child without ACPs are excluded first. In the left panel, an additional 10 women with the highest outcomes are excluded to construct the lower bound by averaging the outcomes among the remaining 80 women. The right panel repeats this process for the upper bound. The new bounds are strictly narrower than the baseline because the exclusion of the 20 women is less extreme: 10 of the same women are excluded for both the lower and upper bounds.

Formally, R^+ is an indicator for a woman's reliance on ACPs for additional children after becoming a mother through her first ACP. Specifically, R^+ takes the value 1 if, in the case that her first ACP succeeds, a woman would have only ACP children, and 0 otherwise. I refer to women who rely on ACPs for all subsequent children as *subsequent reliers*. R^+ is assumed to be independent of the success of a woman's first ACP in the same way as R. D^+ is an indicator for having at least one non-ACP child, defined as:[5]

$$D^+ = Z_A(1 - R^+) + (1 - Z_A)(1 - R).$$

In words, a woman has at least one non-ACP child either if an ACP succeeded and she is not a subsequent relier, or if all ACPs failed and she is not a relier.[6]

Assumption 2 (Monotonicity)**.**

$\Pr(R^+ \geq R) = 1.$

The monotonicity assumption states that women who would have additional non-ACP children in the scenario that their first procedure succeeds would also have at least one child in the scenario that all ACPs fail. This is consistent with families being more determined to have at least one child than to have additional children after the first one. This assumption might be violated when the success of the first ACP causes some couples to stay together instead of separating or prevents some women from becoming depressed, which leads to more effort to conceive and results in non-ACP births that would not have happened otherwise. I relax the assumption to address such violations in Section 7.3 and provide empirical support for both versions of the assumptions afterward.[7]

3.3 Sharp Bounds on Relier Average Treatment Effect

In this section, I formalized and combine ideas introduced in Section 3.2 to bound τ_{ATR}. Before stating the formal results, I relax the sequential unconfoundedness assumption to its

[5]To minimize notation, I do not distinguish between reliance on ACPs for subsequent children after becoming a mother through the first or subsequent ACPs; only the former is relevant for my analysis.

[6]It implies that women with at least one non-ACP child also have at least one child ($D \geq D^+$), and if ACPs fail, having a child is equivalent to having a non-ACP child ($D = D^+ \mid Z_A = 0$).

[7]The remaining theoretical results are presented assuming monotonicity; re-defining R^+ to always take the value 1 and D^+ to take the value 0 when $Z_1 = 1$ makes it equivalent to the case without monotonicity.

conditional counterpart:

Assumption 3 (Conditional Sequential Unconfoundedness)**.**
$(Y(k), R^+, R, W) \perp\!\!\!\perp Z_j \mid X_j$ *for all* j, k*, and* $X_j \in \mathcal{X}_j^1 = \{x \in \mathcal{X}_j : 1_{\{A \geq j\}} = 1\}$.

Where X_j are covariates at the time of ACP j, with support $\mathcal{X}_j$. They include an indicator for whether the woman has undergone at least j ACPs, $1_{\{A \geq j\}}$. Covariates specific to ACP j are set to 0 if the woman does not undergo ACP j.[8] In words, the success of ACP j is independent of potential outcomes and type, conditional on undergoing at least j ACPs and covariates at the time of ACP j. The next assumption provides regularity conditions. Let $e_j(x) = \Pr(Z_j = 1 \mid X_j = x)$.

Assumption 4 (Regularity)**.**

1. $0 < \underline{e} < e_j(x) < \overline{e} < 1$ *for all* j *and* $x \in \mathcal{X}_j^1$*, for some fixed* $\underline{e}$ *and* $\overline{e}$.

2. Y *has a probability density function for* $Z_1 = 1$*,* $D^+ = 0$*, and all* $x \in \mathcal{X}_1$.

The regularity assumption contains two parts. First, the probability of ACP success conditional on undergoing the procedure and covariates at the time differs from 0 and 1. Second, Y is a continuous random variable conditional on the first ACP succeeding, having only ACP children, and every value of X_1. Note that since an ACP cannot succeed unless a woman initiates the procedure, $e_j(x) = 0$ for all j and $x \in \mathcal{X}_j \setminus \mathcal{X}_j^1$.

The bounding procedure begins with identifying several nuisance functions involved in the trimming step. First, the covariate-conditional relier share is identified using the weighted share of women without children among those whose ACPs failed:

$$r(x) = \mathbb{E}\left[\frac{(1 - D^+)\prod_{j=1}^{\overline{w}}(1 - Z_j)}{\prod_{j=1}^{\overline{w}}(1 - e_j(X_j))}\middle| X_1 = x\right].$$

Since $e_j(x_j)$ takes values above zero only for women who undergo ACP j, larger weights are given to women who underwent more ACPs. This accounts for the fact that women willing to undergo more ACPs are less likely to not experience ACP success, making them underrepresented in this group. Next, the covariate-conditional share of subsequent reliers is identified from the share of women having only ACP children among those whose first ACP succeeded:

$$r^+(x) = \mathbb{E}\left[1 - D^+ \mid Z_1 = 1, X_1 = x\right].$$

The covariate-specific share of reliers among subsequent reliers is then given by:

$$p(x) = \frac{r(x)}{r^+(x)}.$$

[8] For $j > 1$, X_j also includes covariates from previous ACPs.

The covariate-conditional quantile function of the treated outcome distribution among subsequent reliers is identified from the outcome distribution among women whose first ACP succeeded and who have only ACP children:

$$q(u,x) = \inf\left\{q : u \leq \Pr(Y \leq q \mid X_1 = x, Z_1 = 1, D^+ = 0)\right\}.$$

Finally, $q(p(x),x)$ and $q(1-p(x),x)$ identify the covariate-conditional $p(x)$-th and $1-p(x)$-th quantiles of the treated outcome distribution among subsequent reliers. These quantiles will be used to trim the tails of the outcome distribution and select reliers in the scenarios where they have either the lowest or the highest treated outcomes.

The nuisance functions are combined with the data to construct the following moments:

$$m^L(G,\eta^0) = Y(1-D^+)1_{\{Y<q(p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D^+)\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}$$

$$m^U(G,\eta^0) = Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D^+)\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))},$$

where G is a vector containing all observed variables and η^0 contains the nuisance functions:

$$\eta^0(x_1,\ldots,x_{\overline{w}}) = \{r^+(x_1), r(x_1), q(p(x_1),x_1), q(1-p(x_1),x_1), e_1(x_1),\ldots,e_{\overline{w}}(x_{\overline{w}})\}.$$

The first term in $m^L(G,\eta^0)$ assigns positive weights to outcomes of women whose first ACP succeeded, who have only ACP children, and whose outcomes fall below the covariate-conditional trimming threshold $q(p(x),x)$. Higher weights are given to women whose first ACP was less likely to succeed, accounting for their under-representation in this group. This term will be used to identify the average relier treated outcome in the scenario that they have the lowest treated outcomes. The second term assigns positive weights to outcomes of childless women whose ACPs failed. Larger weights are given to women who underwent more ACPs to account for the fact that reliers willing to undergo more ACPs are less likely to not experience ACP success, making them underrepresented in this group. This term will be used to identify the average relier control outcome. $m^U(G,\eta^0)$ mirrors this for the scenario that reliers have the highest treated outcomes. Finally, the moments are scaled by the relier share.

Theorem. *Under assumptions 2, 3, and 4, sharp lower and upper bounds on τ_{ATR} are θ_L and θ_U, where:*

$$\theta_L = \frac{\mathbb{E}[m^L(G,\eta^0)]}{\mathbb{E}[r(X_1)]}$$

$$\theta_U = \frac{\mathbb{E}[m^U(G,\eta^0)]}{\mathbb{E}[r(X_1)]}.$$

3.4 Relation to Methodological Literature

My method extends and integrates ideas from two distinct branches of methodological literature. The first step of my approach, which addresses selection via subsequent ACPs, builds on the literature on evaluating time-varying treatments (see Hernán & Robins (2020) for an overview). These methods are typically designed for settings where individuals are exposed to sequences of treatment regimes, with assignment to each subsequent treatment being quasi-random, conditional on the outcome and treatment history at the assignment moment. They are most applicable in experimental settings where the researcher controls the treatment assignment mechanism. In my setting, treatment assignment is equivalent to conceiving via ACPs; however, women may also become mothers through potentially selective non-ACP means, which renders these methods unsuitable.

Even in the absence of such selective treatment, another subtle but important difference lies in the model of the treatment assignment mechanism. Specifically, a central feature of my model is the potentially endogenous decision to enter each subsequent ACP, which, upon entry, induces quasi-random treatment assignment. This implies that women who do not initiate an additional ACP cannot be assigned treatment. It distinguishes my setting from one in which all individuals have a non-zero probability of being assigned to different regimes. In this regard, my model is most related to the one considered by Van den Berg & Vikström (2022), where individuals start in an eligibility state and have a chance to be assigned treatment each period until they either receive it or permanently exit the eligibility state (and remain untreated). My model differs in that treatment is not assigned at specific moments among those still eligible; instead, the likelihood of assignment depends on when and how many times individuals choose to pursue it (e.g., undergo ACPs). As a result, there is no clear duration variable to account for selection; instead, the number of applications, their timing, and the covariates at the moment of application are the essential factors.

The second component of my approach, which addresses selection via non-ACP means, is methodologically closely related to the Zhang & Rubin (2003) and Lee (2009) procedure to handle unobserved outcomes in quasi-experimental settings. While dynamic effects present a conceptually different challenge from unobserved outcomes, the ZRL approach can be used to bound τ_{LATE} under dynamic effects by treating outcomes among women who become mothers after their first ACP fails as unobserved. This amounts to identifying the complier control outcome and share using women whose first ACP failed, and trimming this share from the tails of the outcome distribution among women whose first ACP succeeded to bound complier average treated outcome.

My approach differs from ZRL in two significant ways. First, I bound effects for reliers rather than compliers, which is not possible using the ZRL approach, as it only exploits the first treatment assignment moment. This is advantageous not only because reliers represent a more general group—since all compliers are reliers, but not vice versa—but also because it results in tighter bounds. This occurs because the width of the bounds decreases with

the size of the group for which the control outcome is identified, and my method identifies it for reliers rather than compliers.[9]

The second distinction of my approach is the use of a unique monotonicity assumption to tighten the bounds. The ZRL approach employs a different monotonicity assumption, specifically that, in a setting where outcomes are unobserved in both the treated and control groups, the treatment has a monotonic effect on having an observed outcome. In my setting, however, the treated outcomes for all women whose first ACP succeeds are observed. Instead, I use an auxiliary variable—non-ACP conceptions of additional children—to infer which women are non-reliers. While the economic idea behind my assumption differs, it can be connected to ZRL model, which enables the use of estimation methods developed for the ZRL procedure by Semenova (2023) and Heiler (2024), after modifying them to incorporate the first step of my identification procedure.

4 Estimation

The bounds on τ_{ATR} can be estimated using the sample averages of m^L and m^U after plugging in the estimated nuisance parameter. In the case where only a few discrete covariates are used, the nuisance parameter and the bounds can be estimated jointly using GMM, and asymptotic normality can be demonstrated by building on results from Lee (2009). However, incorporating continuous covariates may be crucial for obtaining narrow bounds, which requires estimating the nuisance parameter non-parametrically. In this case, plugging in an estimate of η^0 into m^L and m^U will complicate the asymptotic distribution of their sample averages. To justify asymptotic inference, I build on the estimation approach for the ZRL procedure introduced by Semenova (2023).

The method by Semenova (2023) involves two key components. The first is orthogonalization, where the baseline moments are modified by including additional terms that ensure their expectations evaluated at the true nuisance parameter remain unchanged but become insensitive to small changes in the nuisance parameter. The second is sample splitting, where the nuisance parameter used for each observation is estimated without that observation. Together, these components ensure that the asymptotic distribution of the averaged moments is not impacted by the estimation of the nuisance parameter for a wide class of non-parametric estimators. This enables inference using standard methods as if the true nuisance parameter was known.

I modify the Semenova (2023) moments to include the first step of my identification approach. First, I replace the terms for the complier control outcome and share with corresponding terms for the relier control outcome and share. These match how the two parameters are identified in the Theorem, specifically, using outcomes and parenthood in-

[9]To see this, assume that among 100 women whose first ACP succeeded, 80 are compliers and 90 are reliers. The lower bound obtained by averaging outcomes among 90 women with the lowest outcomes is mechanically higher than when averaging among 80 women with the lowest outcomes.

Table 1: Orthogonal Moment Functions

	Moment functions
$\psi^{L+}(G,\xi^0)$	$Y(1-D^+)1_{\{Y<q(p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D^+)\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}$
	$+q(p(X_1),X_1)\Big[\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}(1-D^+-r_1(X_1))$
	$-\frac{Z_1}{e_1(X_1)}p(X_1)(1-D^+-r^+(X_1)) - \frac{Z_1}{e_1(X_1)}(1-D^+)(1_{\{Y<q(p(X_1),X_1)\}}-p(X_1))\Big]$
	$-\frac{Z_1-e_1(X_1)}{e_1(X_1)}z^{L+}(1,X_1)r_1(X_1) + \sum_{k=1}^{\overline{w}}1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}[r_k(X_k)\beta_k(X_k)$
	$+q(p(X_1),X_1)(r_1(X_1)-r_k(X_k))]$
$\psi^{U+}(G,\xi^0)$	$Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D^+)\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}$
	$+q(1-p(X_1),X_1)\Big[\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}(1-D^+-r_1(X_1))$
	$-\frac{Z_1}{e_1(X_1)}p(X_1)(1-D^+-r^+(X_1)) - \frac{Z_1}{e_1(X_1)}(1-D^+)(1_{\{Y>q(1-p(X_1),X_1)\}}-p(X_1))\Big]$
	$-\frac{Z_1-e_1(X_1)}{e_1(X_1)}z^{U+}(1,X_1)r_1(X_1) + \sum_{k=1}^{\overline{w}}1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}[r_k(X_k)\beta_k(X_k)$
	$+q(1-p(X_1),X_1)(r_1(X_1)-r_k(X_k))]$
$\psi^{-}(G,\xi^0)$	$Y(1-D^+)\frac{Z_1}{e_1(X_1)}p(X_1) - Y(1-D^+)\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}$
	$-\beta^+(X_1)\Big[\frac{Z_1}{e_1(X_1)}\frac{(1-D^+-r^+(X_1))}{r^+(X_1)}r_1(X_1) - \Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}(1-D^+-r_1(X_1))\Big]$
	$-\frac{Z_1-e_1(X_1)}{e_1(X_1)}\beta^+(X_1)r_1(X_1) + \sum_{k=1}^{\overline{w}}1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}\big[r_k(X_k)\beta_k(X_k)$
	$+\beta^+(X_1)(r_1(X_1)-r_k(X_k))\big]$
$\psi^{R}(G,\xi^0)$	$r_1(X_1) + (1-D^+-r_1(X_1))\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}$
	$+\Sigma_{k=1}^{\overline{w}}1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{(1-Z_j)}{1-e_j(X_j)}\frac{(e_k(X_k)-Z_k)}{1-e_k(X_k)}[r_1(X_1)-r_k(X_k)]$
	Nuisance functions
$\xi^0(x_1,\ldots,x_{\overline{w}})$	$\{e_1(x_1),\ldots,e_{\overline{w}}(x_{\overline{w}}), r_1(x_1),\ldots,r_{\overline{w}}(x_{\overline{w}}), r^+(x_1), q(p(x_1),x_1), q(1-p(x_1),x_1),$ $\beta_1(x_1),\ldots,\beta_{\overline{w}}(x_{\overline{w}}), \beta^+(x_1), z^{U+}(x_1), z^{L+}(x_1)\}$
$r_k(x)$	$\mathbb{E}[(1-D^+)/(\Pi_{j=k+1}^{A}(1-e_j(X_j))) \mid X_k=x, Z_A=0]$
	$\mathbb{E}[\Pi_{j=k+1}^{A}(1-e_j(X_j)) \mid X_k=x, Z_A=0]$
$\beta_k(x)$	$\mathbb{E}[Y/(\Pi_{j=k+1}^{A}(1-e_j(X_j))) \mid X_k=x, D=0]$
	$\mathbb{E}[\Pi_{j=k+1}^{A}(1-e_j(X_j)) \mid X_k=x, D=0]$
$\beta^+(x)$	$\mathbb{E}[Y \mid X_1=x, Z_1=1, D^+=0]$
$z^{U+}(x)$	$\mathbb{E}[Y \mid X_1=x, Z_1=1, D^+=0, Y\geq q(1-p(x),x)]$
$z^{L+}(x)$	$\mathbb{E}[Y \mid X_1=x, Z_1=1, D^+=0, Y\leq q(p(x),x)]$

dicators among women who never experience ACP success, with weights that depend on the number of ACPs and the propensity score at each ACP, $e_j(X_j)$. Since this makes the moments sensitive to the propensity scores, I add additional terms to correct for it. These terms account for the use of the scores in estimating the relier average control outcome and the relier share. The new moments for the lower and upper bounds $\psi^{L+}(G,\xi^0)$ and $\psi^{U+}(G,\xi^0)$, and the nuisance parameter ξ^0 are given in Table 1.

The new moments identify the same parameters as the baseline moments:

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

However, the old moments are sensitive to small errors in the nuisance parameter, whereas the new moments are not. For example, for some j, let $\widehat{e}_j(x_j)$ be an estimate of the propensity score $e_j(x_j)$ such that $\widehat{e}_j(x_j) \neq e_j(x_j)$ for $x_j \in \mathcal{X}_j^1$. Define $r \in [0,1) \to \psi^{U+}(G,r) \equiv$

$\psi^{U+}(G, \xi_r)$, where:

$$\xi_r = \{e_1(x_1), \ldots, e_l(x_l, r), \ldots, e_{\overline{w}}(x_{\overline{w}}), r_1(x_1), \ldots, r_{\overline{w}}(x_{\overline{w}}), r^+(x_1), q(p(x_1), x_1),$$
$$q(1 - p(x_1), x_1), \beta_1(x_1), \ldots, \beta_{\overline{w}}(x_{\overline{w}}), \beta^+(x_1), z^{U+}(x_1), z^{L+}(x_1)\},$$

and where $e_l(x_l, r) = e_l(x_l) + r(\widehat{e_l}(x_l) - e_l(x_l))$, meaning that $e_l(x_l, 0) = e_l(x_l)$. Then for the new moment:

$$\partial_r \, \mathbb{E}[\psi^{U+}(G, \xi_r)|X_l]|_{r=0} = 0 \; a.s.,$$

while for the original moment:

$$\partial_r \, \mathbb{E}[m^U(G, \eta_r)|X_l]|_{r=0} \neq 0 \; a.s.,$$

meaning that the old moment is sensitive to the estimation error in $\widehat{e}_j(X_j)$, whereas the new one is not. I present the derivation of these results in Appendix A2.

A challenge that may arise when implementing the bounds under a non-trivial monotonicity assumption is that, when the relier and subsequent relier shares for some values of X_1 are very close, the estimated shares reverse order. In my main specification, I treat such cases as if the two shares were equal, with the corresponding moment $\psi^-(G, \xi^0)$ given in Table 1. I discuss this in detail, introduce an alternative approach that allows the direction of monotonicity to vary with covariates following Semenova (2023), and present results based this approach in Appendix A3. I also replace the denominator moment with an orthogonal counterpart $\psi^S(G, \xi^0)$, given in Table 1.

The estimators for the lower and upper bounds are:

$$\widehat{\theta_L} = \frac{\sum_i \left(\psi^{L+}(G_i, \widehat{\xi_i}) 1_{\{p(X_1) \leq 1\}} + \psi^-(G_i, \widehat{\xi_i}) 1_{\{p(X_1) > 1\}}\right)}{\sum_i \psi^R(G_i, \widehat{\xi_i})}$$
$$\widehat{\theta_U} = \frac{\sum_i \left(\psi^{U+}(G_i, \widehat{\xi_i}) 1_{\{p(X_1) \leq 1\}} + \psi^-(G_i, \widehat{\xi_i}) 1_{\{p(X_1) > 1\}}\right)}{\sum_i \psi^R(G_i, \widehat{\xi_i})},$$

where G_i is the data for observation i and $\widehat{\xi_i}$ is the nuisance parameter for observation i, estimated on a subsample that excludes observation i. I discuss implementation in detail in Appendix A4.[10] In Appendix A5, I also introduce a new method to estimate non-sharp

[10] I estimate the propensity scores using logistic regressions that include quadratic functions of each partner's age at the time of the procedure, interacted with procedure type and higher education dummies, based on women who initiate the respective ACP. I consider the first ten ACPs that women undergo, treating conceptions through other ACPs as non-ACP conceptions. I estimate remaining nuisance functions using Generalized Random Forests for conditional expectations and quantiles (Athey et al., 2019); covariates include all those in the propensity scores up to the current ACP, as well as women's and their partner's pre-ACP income and work hours. Following Heiler (2024), I base confidence intervals for the bounds on Stoye (2020).

bounds that leverages continuous covariates without requiring non-parametric estimation. The estimates from this method are very similar to my main results. This new procedure can also be applied to leverage continuous covariates in standard ZRL setting.

5 Institutions, Assisted Conception Procedures, and Data

In this section, I first describe Dutch family policies and the labor market context. Then, I discuss IVF and intrauterine insemination, and the differences between them. Afterward, I overview the data, provide empirical support for the sequential unconfoundedness assumption, and compare the ACP sample to the general population.

5.1 Family Policies in the Netherlands

Dutch women are entitled to 4 to 6 weeks of pregnancy leave before the estimated due date and at least 10 weeks of maternity leave after giving birth. The total leave must sum up to at least 16 weeks.[11] During this leave, women receive 100% of their wage from the unemployment insurance agency (up to a maximum daily limit). Fathers are entitled to one week of leave within the first four weeks after birth at a 100% replacement rate, which is paid by the employer.[12] Children may be enrolled in private daycare centers as young as three months. In 2022, 72% of children under two were enrolled in formal child care. Among enrolled children, the average time spent in child care was 20 hours per week (OECD, 2023a). After turning four and starting elementary education, children become eligible for out-of-school care. In 2023, the average child care cost per family was 8,950 euros, and parents were reimbursed for 64% of that amount. This translates to an average net cost per family that is equivalent to 10% of the median disposable household income.[13]

Compared to other OECD countries, the Netherlands has average family policies. Paternity and maternity leave durations are slightly below the OECD averages of 2.5 and 21 weeks, respectively (OECD, 2023c). While the Netherlands has the highest formal child care enrollment rate for children under two among OECD countries, the average time spent in child care is the lowest (OECD, 2023a). After age four, enrollment rates and average hours for out-of-school care are similar to those in other OECD countries (OECD, 2022).

While the employment rate for mothers, fathers, and non-parents in the Netherlands is above the OECD average, part-time work is much more common, making the Netherlands average in terms of hours worked (OECD, 2023b). In 2021, the maternal employment rate was around 80%, compared to the OECD average of 71%.[14] However, in 2023, 52% of women and 18% of men worked part-time, defined as less than 30 hours per week, which

[11] In the case of multiple births, women are eligible for 20 weeks of total leave.

[12] A reform in 2020 allowed fathers to request up to five additional weeks of leave within the first six months after birth at a 70% replacement rate; most of the births in the data occurred before this.

[13] www.cbs.nl/nl-nl/nieuws/2024/30/ouders-betaalden-gemiddeld-3-210-euro-aan-kinderopvang-in-2023, longreads.cbs.nl/materiele-welvaart-in-nederland-2024/inkomen-van-huishoudens/.

[14] The EU average was 75% (OECD, 2024).

is more than double the respective OECD averages (OECD, 2023d). Among two-parent families, only 14% had both partners working full-time, 52% had a full-time working man and a part-time working woman, and another 12% had both partners working part-time.[15]

5.2 Assisted Conception Procedures

I use two types of ACPs: first, IVF, which has previously been used to study the career impacts of parenthood in Denmark and Sweden (Lundborg et al., 2017; Bensnes et al., 2023; Gallen et al., 2023; Lundborg et al., 2024), and second, intrauterine insemination (IUI), which has not been used to study the career impacts of parenthood before. In both procedures, the first stage may involve hormonal stimulation to improve egg production. IVF is a surgical procedure where eggs are retrieved through the vaginal wall using a specialized needle and fertilized in the lab. In the last stage, the developed embryos are transferred into the uterus. IVF is relatively invasive, performed under sedation or anesthesia, and has a success rate (after the embryo transfer) of approximately 25%. In IUI, sperm is injected directly into the uterus using a catheter. IUI has a lower success rate of approximately 10%; however, it is significantly less costly and invasive than IVF—a procedure may take as little as 5 minutes and is generally not painful. IUI is the first-line infertility treatment in most countries. Dutch couples without a specific infertility diagnosis are typically required to undergo IUI six times before attempting IVF. The compulsory health insurance in the Netherlands covers unlimited IUI and up to three IVF procedures. In 2022, the price of each additional IVF cycle was 4,000 euros; however, since multiple embryos can be fertilized and frozen in a single cycle, additional attempts may only involve unfreezing and inserting the embryos, costing around 1,000 euros.

5.3 Data

I use administrative data from Statistics Netherlands, which cover all individuals residing in the country. The data on ACPs cover the period from 2012 to 2017 and are derived from the Diagnosis-Treatment Combination information system, which Dutch hospitals are mandated to report to. The main variables for my analysis are the type of procedure—IVF or IUI—and the date of sperm or embryo insertion into the uterus. I define ACP success as having a child born within 10 months after the insertion without any subsequent insertions in between. This definition has been validated against medical records by Lundborg et al. (2017).

The data on labor market outcomes span the period from 2011 to 2023. The main outcomes I use are annual work hours and gross labor earnings, both derived from tax records. Work hours include paid maternity leave. Similarly, gross labor earnings include maternity pay. While including leave pay accurately reflects women's financial situation and income share, incorporating leave duration complicates the interpretation of work hours.

[15] www.cbs.nl/en-gb/news/2024/10/fewer-and-fewer-families-in-which-only-the-father-works

To account for this additional uncertainty surrounding actual work hours, I first introduce maximum-leave-adjusted hours, scaling women's reported hours in each year they give birth (including subsequent births) by 36/52. This corresponds to the 16 weeks of unreported pregnancy and maternity leave that most women could take. In my main analysis, I estimate the upper bound of the effect using reported work hours and the lower bound using adjusted hours, which helps ensure that the effect on actual work hours lies within these bounds. Since existing methods that point-identify the effects do not naturally accommodate such adjustments, I use the leave-adjusted work hours in supplementary analyses. Using either measure only meaningfully affects the results in the year following women's first ACP and does not impact the comparison of different methods or estimates of the bias.

I also use several demographic variables, including an indicator for completing higher education, number of children, birth dates, and cohabitation status. My main sample consists of childless couples who were cohabiting before the woman's first IUI procedure. To ensure the first observed ACP is their actual first, I follow Lundborg et al. (2017) and exclude those whose first observed procedure took place in the first year of the data, as they likely had prior ACPs. I also exclude those whose first ACP occurred in the last year to avoid misattributing births from unobserved ACPs in the following year to failed ACPs in the previous year. These restrictions have little impact on my results. My main sample consists of 15,523 couples. To compare the ACP sample with the general population, I use women who were cohabiting with a male partner when they got pregnant with their first child between 2013 and 2017, without prior ACPs. This group consists of 376,157 couples.

Table 2 compares average characteristics between couples whose first ACP succeeded (column 1) and couples whose first ACP failed (column 2). Labor market and education outcomes are measured in the year preceding women's first ACP. The two groups had similar average annual earnings. However, women whose first ACP succeeded were working 30 more hours per year, were almost 2 percentage points more likely to be employed, and were slightly more educated, on average. A similar education gradient in IVF success has been documented in Denmark by Groes et al. (2024). The results for partners are similar. Most notably, women and partners whose first ACP succeeded were almost 9 months younger on average, which is not surprising because age is potentially the most important factor in ACP success. Following Lundborg et al. (2024), the last column in the table presents the differences between the two groups after accounting for education and age, which makes the remaining differences in pre-ACP labor market outcomes negligible. Excluding education does not change these results. This supports the assumption that the success of women's first ACP is conditionally independent of their potential labor market outcomes.

Table 3 presents balance results for subsequent ACPs up to the tenth. Since these ACPs also include IVF, I additionally control for each partner's age interacted with treatment type. This ensures that ACP success only needs to be as good as random among women who undergo the same procedure (and are of similar age), allowing for selection into IUI or

Table 2: First ACP Outcomes and Descriptives

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

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

Table 3: Balance in Later ACPs

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

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

IVF based on the woman's type and potential outcomes. Overall, the results indicate no systematic differences in pre-ACP outcomes between those with successful and unsuccessful subsequent ACPs, supporting the conditional sequential unconfoundedness assumption.

Unlike in many other quasi-experimental settings, there is little opportunity for women

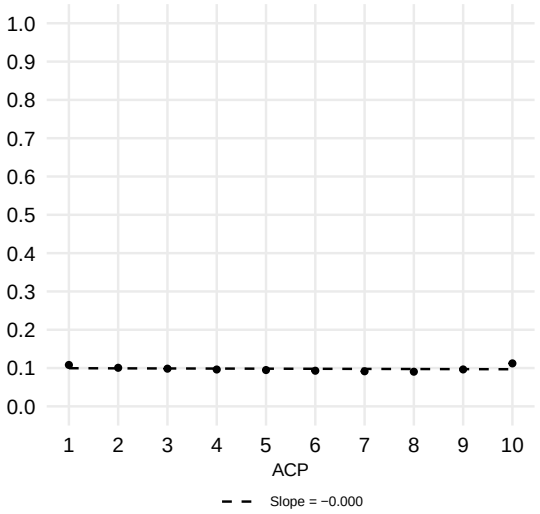

Figure 2: Estimated Success Probabilities

to directly manipulate the outcomes of their ACPs. This means that the primary threat to the sequential unconfoundedness assumption is that ACP success depends on underlying health factors, which also affect women's labor market outcomes. Since health-related differences can be expected to also be reflected in pre-ACP labor market outcomes, balance on these outcomes provides relatively strong support for the assumption.

Another threat to the sequential unconfoundedness assumption is that some women choose to undergo additional ACPs because they have information suggesting the procedures are more likely to work for them. This may result in the willingness to undergo additional ACPs, W, being correlated with the ex-ante likelihood of success in previous ACPs. To test this, I examine how the likelihood of ACP success varies across procedures. To account for the fact that this likelihood decreases with age, which could obscure any patterns, I hold covariates fixed at their average levels from the first procedure. The results, presented in Figure 2, suggest that the conditional likelihood of success in subsequent ACPs among women who choose to undergo them remains similar to that of all women at their first ACP. This supports the assumption that the likelihood of ACP success is independent of the willingness to undergo ACPs.

Before comparing the ACP sample to the non-ACP parents, I present some descriptive statistics on women's ACP histories and non-ACP fertility. On average, a woman whose first ACP fails undergoes an additional 4.1 procedures. The estimated average willingness to undergo ACP in case of failure, W, is 7.3. Three years after the first ACP, the estimated relier share is 0.8, decreasing to 0.45 after seven years. For compliers, these shares are 0.4 and 0.25, respectively. The estimated correlation between eventual relier status and the willingness to undergo ACP is close to zero (although it is not required for the bounding

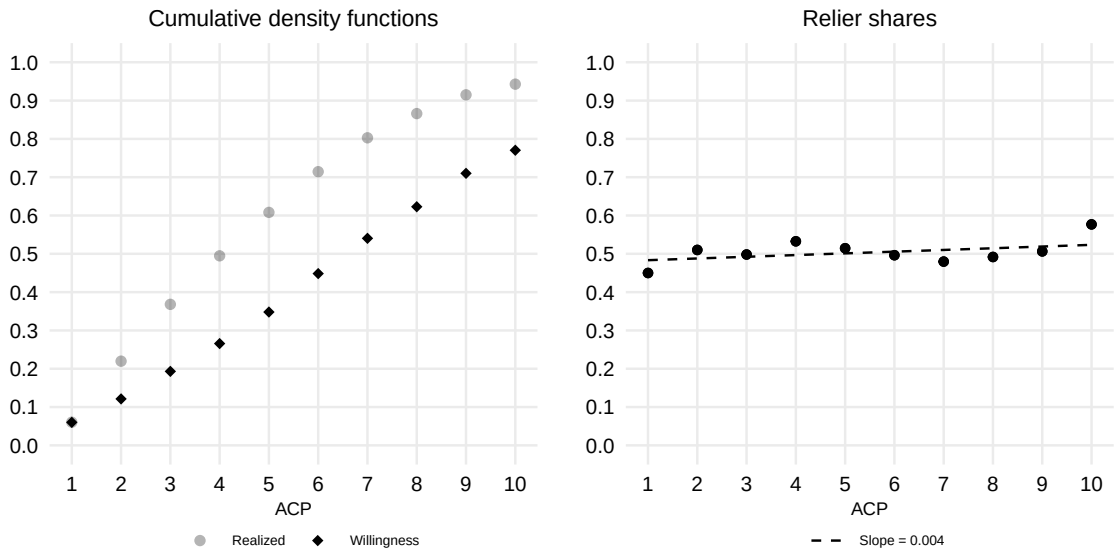

Figure 3: ACP Histories and Reliance

approach). Figure 3 presents the realized distribution of the number of ACPs women undergo, along with the estimated distributions of willingness to undergo ACP and relier shares, all measured seven years after the first ACP.

Table 4 compares the main sample to a representative sample of mothers, weighted to match the distribution of birth years for the first child among women whose first ACP succeeded. Prior to motherhood, women in the representative sample were less likely to work, worked fewer hours on average, had lower income, and were less educated. The differences between fathers are similar. After parenthood, both groups have similar average completed fertility, at 1.8 children. While women whose first ACP succeeded were substantially more likely to give birth to twins (7% compared to 1.5% in the representative sample), multiple births were uncommon in absolute terms in both groups.

Although women entering ACPs differ from the representative sample, my analysis focuses specifically on reliers. Their average pre-ACP characteristics can be identified similar to their average control outcomes. Column 3 in Table 4 presents estimated average characteristics for women who remained reliers seven years after their first ACP. The last column compares this group to the representative sample. While reliers were still more likely to work and had higher income than the representative sample, the differences were substantially smaller than when compared to the whole ACP sample. One notable exception is age, with relier women being, on average, five years older than those in the representative sample. This is expected, as women who try to become mothers at an older age are more likely to remain childless.

Table 4: Full Sample, Reliers, and Representative Sample

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

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

6 Results

Figure 4 presents the estimated effects on women's annual work hours and income for each year, starting from conception.[16] In the first year, the bounds indicate a reduction in women's work time between 10 and 130 hours, or 1% to 11%, relative to the point-identified relier average control outcome. The impact on income in the first year is negligible. In years two and three, the bounds suggest a reduction in work hours between 80 and 400 hours, or 7% to 34%, and a reduction in income between 1,800 and 9,200 euros, or 6% to 32%. The bounds widen over time, making it impossible to rule out a zero effect on both work hours and income in the fourth year.

Figure 5 presents the estimated effects on women's outcomes leveraging the assumption that those who conceived additional non-ACP children after their first successful ACP would have conceived at least one child even if all ACPs had failed. The bounds for the effect on work hours are stable from year 3 to year 7, indicating reductions between 90 and 290 hours, or 8% to 26%. The bounds for income widen slightly over time; by year 7, they suggest a reduction between 1,500 and 10,800 euros, or 5% to 34%. Since the bounds without

[16]The group of reliers changes over time, similar to how the group of compliers changes in the LATE framework. I extend my method to bound the effects for a stable relier group and present the corresponding estimates in Section 7.4.

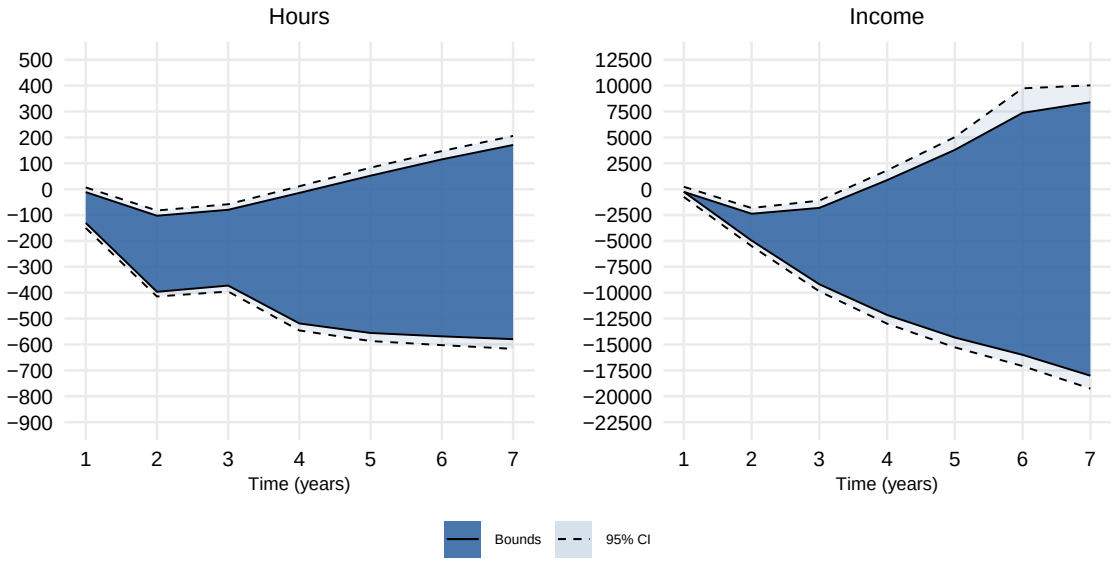

Figure 4: Effects on Women (without Monotonicity)

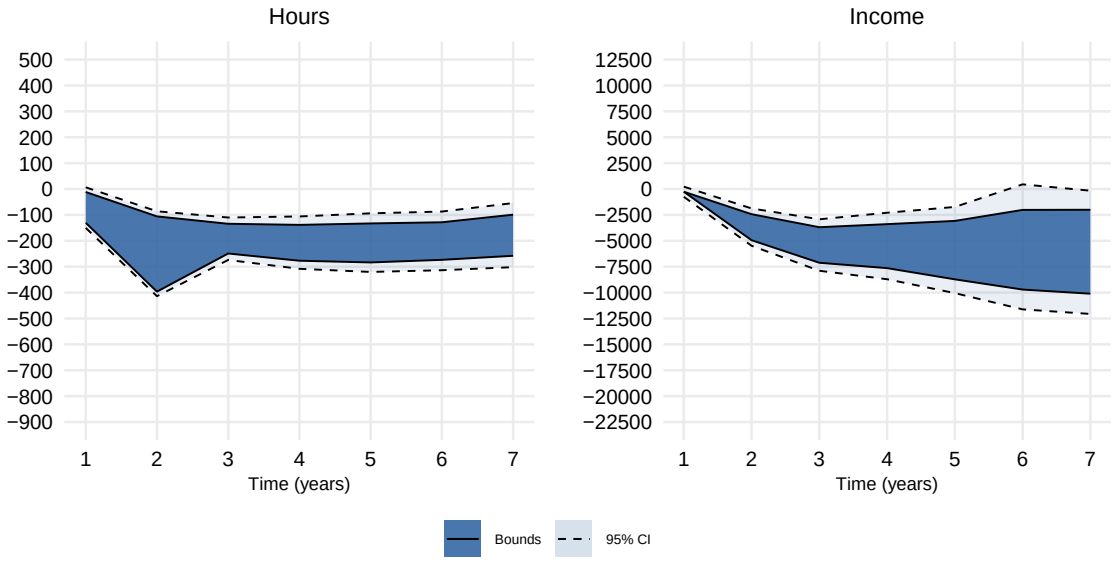

Figure 5: Effects on Women

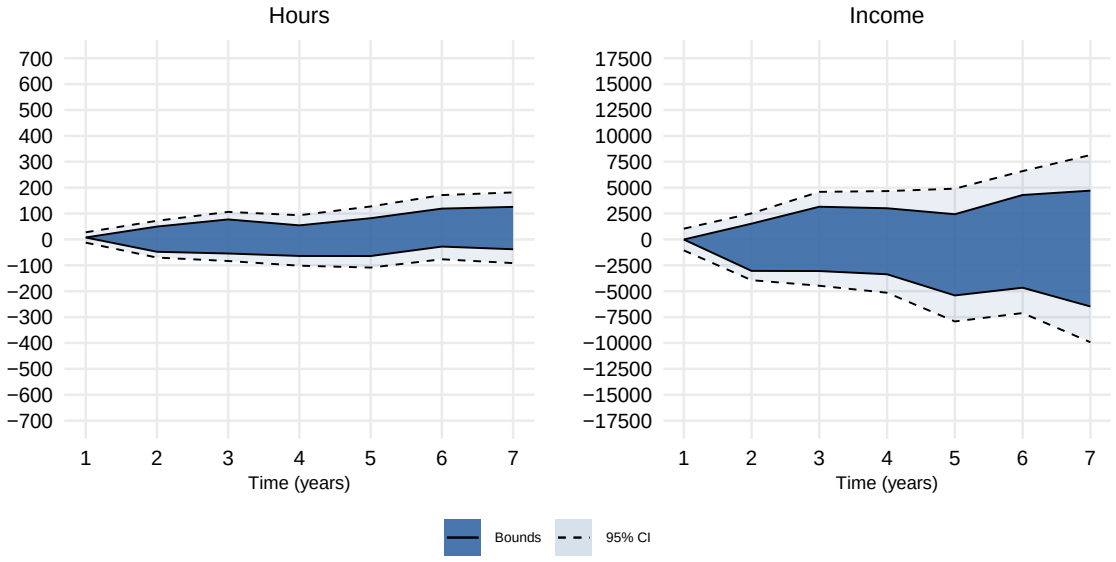

Figure 6: Effects on Men

the monotonicity assumption are only slightly narrower in the early years but remain more informative later on, I present the remaining results leveraging the monotonicity assumption.

Figure 6 presents the estimated effects on men's outcomes. The bounds are similar in width to those for women but are centered around zero. Seven years into parenthood, the estimates rule out reductions in men's work hours exceeding 4% and reductions in income exceeding 16%.

Figure 7 plots the estimated effects on the gaps in outcomes between men and women, relative to the gaps in their average treated outcomes—in other words, the share of gender inequality caused by parenthood.[17] The results indicate that between year three and year seven, parenthood caused between 26% and 60% of the gender inequality in annual work hours and up to 50% of the gender inequality in annual income. The upper bounds for both outcomes remain stable from year three onward.

Aggregating the upper and lower bounds across periods results in non-sharp bounds on the cumulative effects. This is because per-period bounds do not account for within-woman and within-couple outcome relationships over time. To obtain sharp bounds on the share of gender inequality caused by parenthood during the first seven years, I use cumulative hours and income over this period as the outcomes in year seven. The results suggest that parenthood caused between 36% and 54% of gender inequality in work hours, and between 5% and 46% of the inequality in income during this period.

[17] Using the differences between male and female outcomes ensures the bounds are sharp (see Semenova (2023)). Notably, they are narrower than those obtained by combining separate lower and upper bounds for each group, which overlook the within-couple relationship between outcomes. The bounds on the ratio are calculated using the formula $1 - a/b$, where a is the point-identified control outcome, and b is either the lower or upper bound on the treated outcome, all estimated using orthogonal moments.

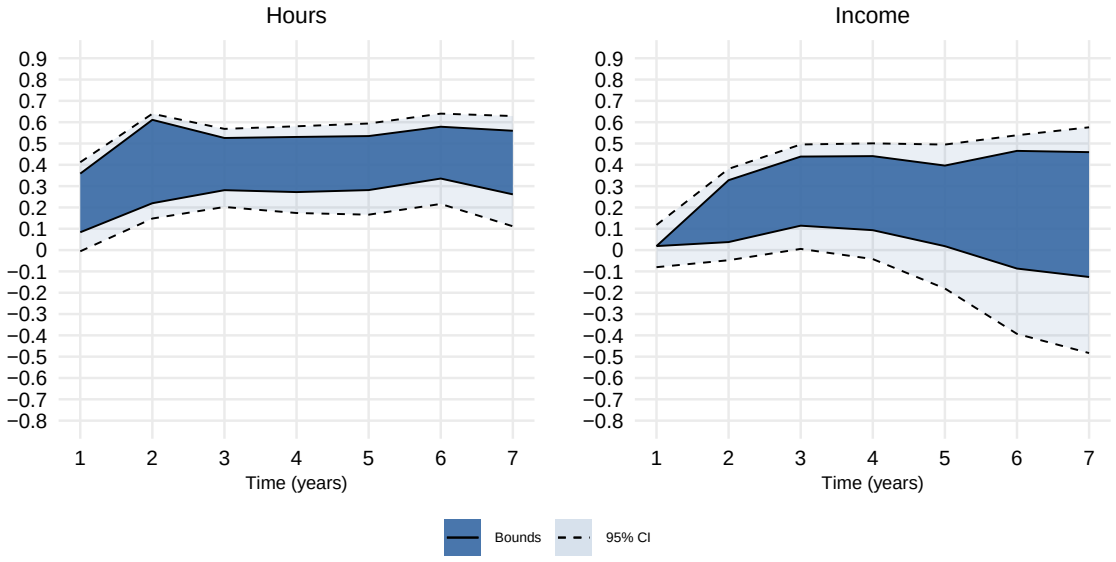

Figure 7: Share of Gender Inequality Caused by Parenthood

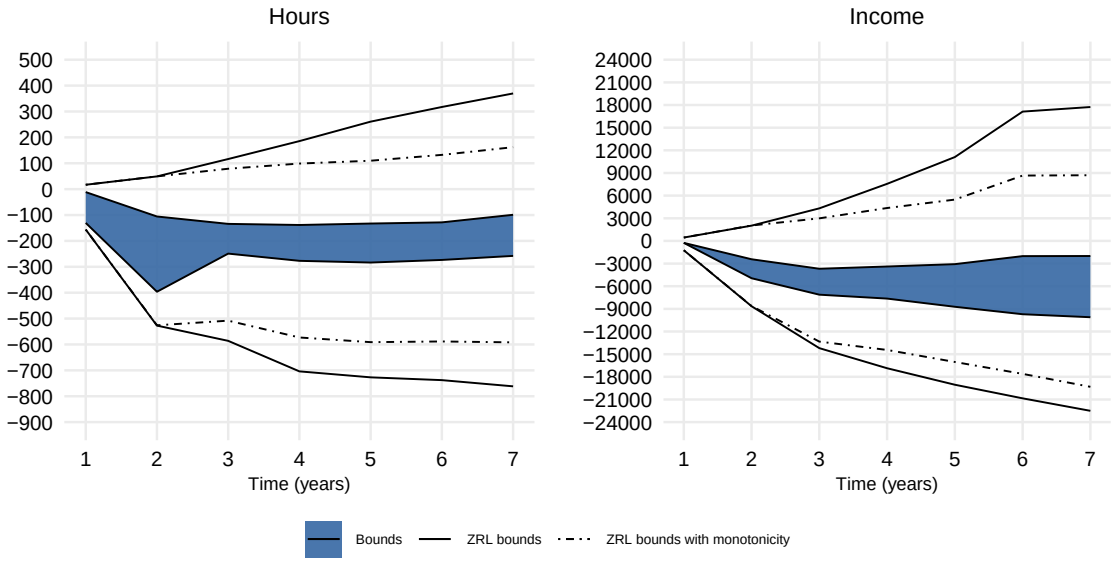

Figure 8: Comparison with ZRL Bounds for Effects on Women

I present technical sensitivity analyses for the main results in the appendix. These analyses include applying an alternative monotonicity assumption following Semenova (2023) (Appendix A3), using a GMM estimator that does not rely on double/debiased machine learning (Appendix A5), and adjusting for the age difference between partners when estimating the share of gender inequality caused by parenthood (Appendix A6). The estimates remain largely unchanged.

Before turning to extensions, I compare my bounds with those that rely solely on

women's first ACP, which are equivalent to the Zhang & Rubin (2003) and Lee (2009) bounds. Figure 8 presents the estimated effects on women's labor market outcomes.[18] The ZRL bounds are considerably wider, failing to rule out both large positive and negative effects on women's income and work hours, both in the short and medium run. Seven years into parenthood, they are 7.1 and 4.7 times wider than my bounds for hours and income, respectively. The figure also includes bounds that leverage women's first ACP together with the monotonicity assumption. These bounds also fail to rule out substantial positive and negative effects. Seven years into parenthood, they remain 5 and 3.5 times wider than my bounds for hours and income, respectively.

7 Extensions

In this section, I present extensions to address internal and external validity concerns regarding my estimates and to assess the bias in the leading estimators. Section 7.1 generalizes the model to a dynamic setting. Section 7.2 introduces procedures for assessing the bias in the IV and event study (ES) estimators. Section 7.3 discusses concerns related to mental health and relationship breakdowns. Section 7.4 presents extensions to ensure that the estimates from different periods cover the same sub-population. Section 7.5 discusses heterogeneity by effort to conceive. Finally, Section 7.6 introduces an extrapolation procedure for quantifying the effects in the non-ACP population.

7.1 Dynamic Model

The model introduced in Section 2 is sufficient to demonstrate the identification challenge and the bounding approach without loss of generality. The objective of generalizing it to a dynamic version is threefold: first, to highlight the nuances in comparing estimates across different periods; second, to precisely formalize how outcomes depend on the timing of parenthood, rather than merely distinguishing between conception at first ACP and later conceptions; and third, to connect my method to the ES approach, which relies on the panel data structure. These three features are required to formalize extensions that ensure estimates from different periods cover the same sub-population, quantify the effects of delaying parenthood, and assess the magnitude of selective fertility, respectively.

All women start ACP for their first child at time $t = 1$, ticking up to T. W_t is the total number of ACPs a woman would undergo by period t to conceive her first child, assuming all previous procedures fail. R_t takes the value 1 if a woman would be childless in period t if all ACPs up to that period failed, and 0 otherwise. R_t^+ takes the value 1 if a woman would have only ACP children in period t if her first ACP succeeded, and 0 otherwise. C_t takes the value 1 if a woman would be childless in period t if her first ACP failed, and 0 otherwise. The realized number of ACPs a woman undergoes for her first child by period t

[18] The alternative bounds are estimated using the baseline approach, ignoring all ACPs after the first one.

is A_t:

$$A_t = \min\left(\{j : Z_j = 1, j \leq W_t\} \cup \{W_t\}\right).$$

I do not index ACP outcomes by time, implicitly assuming that the outcomes of past ACPs cannot change. D_t takes the value 1 if a woman has any children in period t, and 0 otherwise, defined as:

$$D_t = Z_{A_t} + (1 - Z_{A_t})(1 - R_t).$$

D_t^+ takes the value 1 if a woman has any non-ACP children in period t, and 0 otherwise, defined as:

$$D_t^+ = Z_{A_t}(1 - R_t^+) + (1 - Z_{A_t})(1 - R_t).$$

$Y_t(0)$ is the potential outcome in period t if a woman remains childless. $Y_t(k)$ is the potential outcome if a woman becomes a mother in period k. A woman's realized labor market outcome in period t is Y_t, and the relationship between potential and realized outcomes is given by:

$$Y_t = Y_t(0)1_{\{D_T=0\}} + Y_t(1)1_{\{D_1=1\}} + \sum_{k=2}^{T} Y_t(k)1_{\{D_k=1, D_{k-1}=0\}}.$$

I also introduce an auxiliary variable $K_t \in \{0, 1\}$ that describes some behavior in the scenario that all ACPs up to period t fail. For example, K_t might take the value 1 if a woman would remain with her partner from her first ACP until period t if all ACPs up to period t failed, and 0 if she would separate from her partner in this scenario.

I impose three assumptions.

Assumption 5 (Irreversibility)**.**

$R_t \geq R_{t+1}, R_t^+ \geq R_{t+1}^+, W_{t+1} \geq W_t$ for all t.

The irreversibility assumption states that, first, women can only transition from being reliers to non-reliers and from being subsequent reliers to non-subsequent-reliers, and second, the willingness to undergo ACP can only increase over time. This assumption implies that parenthood is irreversible.

Assumption 6 (No Anticipation)**.**

$Y_t(k) = Y_t(0)$ for all $k > t$.

The no anticipation assumption states that outcomes before becoming a mother do not depend on having children in the future. This assumption is plausible regarding conception because the success of future procedures is unknown. It may be less plausible regarding

adoption, which could be anticipated; however, adoptions are extremely rare in my application.[19]

Assumption 7 (Dynamic Conditional Sequential Unconfoundedness)**.**

$(Y_t(k), R_l^+, R_l, K_l, W_l) \perp\!\!\!\perp Z_j \mid X_{tj}$ *for all* j, k, l, t, *and* $X_{tj} \in \mathcal{X}_{tj}^1 = \{x \in \mathcal{X}_{tj} : 1_{\{A_t \geq j\}} = 1\}$.

Where X_{tj} are covariates the the time of ACP j in period t with support $\mathcal{X}_{tj}$. The dynamic conditional sequential unconfoundedness assumption follows from a stronger assumption that, among women who enter a specific ACP, the outcome of that ACP, conditional on covariates at the time, is independent of type, potential outcomes, and behavior in different scenarios across all periods. I also assume that SUTVA holds for all potential outcomes and types.

The individual-level treatment effect in period t is:

$$\tau(t) = Y_t(1) - Y_t(0).$$

The average treatment effect in period t is:

$$\tau_{ATE}(t) = \mathbb{E}[\tau(t)].$$

The average treatment effect in period t for reliers in period t is:

$$\tau_{ATR}(t) = \mathbb{E}[\tau(t) \mid R_t = 1].$$

The average treatment effect in period t for compliers in period t is:

$$\tau_{LATE}(t) = \mathbb{E}[\tau(t) \mid C_t = 1].$$

Comparing τ_{ATR} and τ_{LATE} between periods is complicated because the groups of reliers and compliers can change, meaning that changes in τ_{ATR} or τ_{LATE} may reflect compositional changes rather than changes in the effects over time for a fixed group. To address this, I define the average treatment effect in period t for women who remain reliers until the last period T:

$$\tau_{ATRL}(t) = \mathbb{E}[\tau(t) \mid R_T = 1].$$

To simplify exposition, I introduce additional notation for the effects of delaying parenthood. Let R^* be the period in which a woman would become a mother if all ACPs failed; R^* takes the value 0 if she would remain childless. Similarly, let C^* be the period in which a woman

[19] Each year, there are around 40 domestic adoptions in the Netherlands, and the share of foreign-born children in my sample is less than 1% (including those who were not adopted but whose mothers were abroad at the time of childbirth).

would become a mother if her first ACP failed; C^* also takes the value 0 if she would remain childless.[20] The average effect of becoming a mother after all ACPs fail, relative to becoming a mother at the first ACP in period t for non-reliers in period t, is:

$$\delta_R(t) = \mathbb{E}[Y_t(R^*) - Y_t(1) \mid R_t = 0],$$

I refer to this as the average effect of delaying parenthood for non-reliers. Similarly, the average effect of becoming a mother after the first procedure fails, relative to becoming a mother at the first procedure in period t for always-takers in period t, is:

$$\delta_C(t) = \mathbb{E}[Y_t(C^*) - Y_t(1) \mid C_t = 0],$$

and I refer to it as the average effect of delaying parenthood for always-takers.

7.2 Bias in Existing Methods

In this section, I introduce methods to assess the bias in the ES and IV estimators due to selective fertility and dynamic effects, respectively. This bias cannot be quantified by comparing corresponding estimates to the bounds on τ_{ATR} because different methods identify effects for different sub-populations. Specifically, the IV targets τ_{LATE}, while the ES, when implemented using women whose first ACP succeeded, targets a covariate-weighted τ_{ATE}.[21] Both of these parameters may differ from τ_{ATR}. After comparing estimates from different methods, I introduce procedures to explicitly isolate the contributions of selective fertility and dynamic effects.

Figure 9 presents estimates from different methods.[22] The ES estimates suggest substantial negative impacts on women's labor market outcomes, while the IV estimates indicate much smaller effects. The ES estimates based on a representative sample align closely with those based on the ACP sample. These results mirror the finding from Denmark (Lundborg et al., 2024). Relative to the bounds, the ES estimates indicate a larger cost of motherhood, while the IV estimates are generally within the bounds but suggest a larger reduction in hours and income in the fourth year. In the medium run, the bonds do not rule out substantially larger negative effects than those suggested by the IV estimates.

While my method addresses the bias in the IV and ES estimators, it does so at the expense of point identification—a disadvantage to researchers willing to impose stronger assumptions regarding selection or dynamic effects. However, my method compensates with enhanced precision. I discuss this in Appendix A7, where I demonstrate that the 95% confidence intervals for my bounds closely match those for IV and ES point estimates.

[20] Formally: $R^* = \max(\min\{j : R_j = 0\} \cup \{0\})$, $C^* = \max(\min\{j : C_j = 0\} \cup \{0\})$.

[21] The two methods identify weighted versions of the two parameters, with observations with $e_1(X_1)$ closer to 0.5 receiving higher weights. However, this has little impact because variation in $e_1(X_1)$ is minimal.

[22] Hours refer to leave-adjusted work hours; implementation details are presented in Appendix A4.

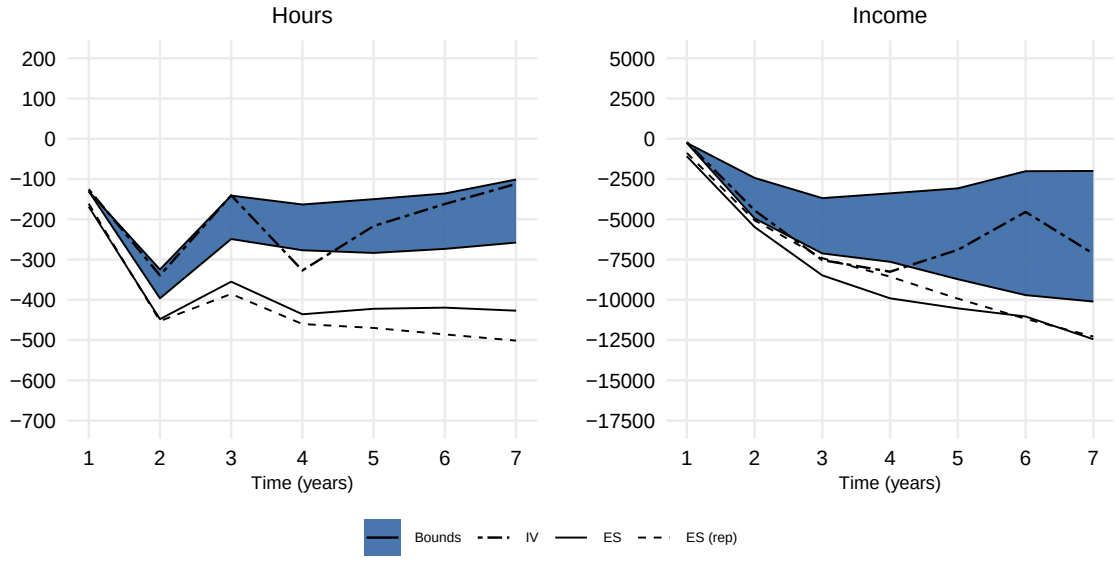

Figure 9: Naive Comparison of Different Methods

7.2.1 Instrumental Variable, Dynamic Effects, and Delayed Parenthood

In this section, I quantify the effects of delaying parenthood, which may bias the IV estimator of τ_{LATE}. I also discuss how τ_{ATR} can be point-identified under the assumption of static effects and how it can be used to test the assumptions employed in the methods proposed by Bensnes et al. (2023) and Gallen et al. (2023).

Under assumptions 6 and 7 the IV identifies:

$$\tau_{LATE}(t) - \delta_C(t)\frac{\Pr(C_t = 0)}{\Pr(C_t = 1)},$$

where the second term is the average effect of delaying parenthood for always-takers, scaled by their relative share.

To assess the importance of dynamic effects, I bound the average effect of delaying parenthood for non-reliers, δ_R. Focusing on δ_R is appealing because it allows for comparison with τ_{ATR}, similar to the comparison between τ_{LATE} and δ_C. Additionally, leveraging the monotonicity assumption provides more informative bounds for δ_R than for δ_C. The bounding procedure mirrors that for τ_{ATR}, but focuses on women who conceive after all ACPs fail, rather than those who remain childless after all ACPs fail.[23]

[23] In the first step, I identify the average later-treated outcome for non-reliers with a specific willingness to undergo ACP using the average outcome among women who get their first child without ACP after undergoing a specific number of ACPs: $\mathbb{E}[Y_t|A_t = w, Z_A = 0, D_t = 1] = \mathbb{E}[Y_t(R^*)|W_t = w, R_t = 0]$. The argument follows similar steps as for the relier average control outcome. Then, I bound the average treated non-relier outcome by trimming the tails of the outcome distribution among women whose first ACP succeeded using the identified non-relier share. Opposite to the baseline method, monotonicity is leveraged by always including women who get non-ACP children after the first ACP succeeds.

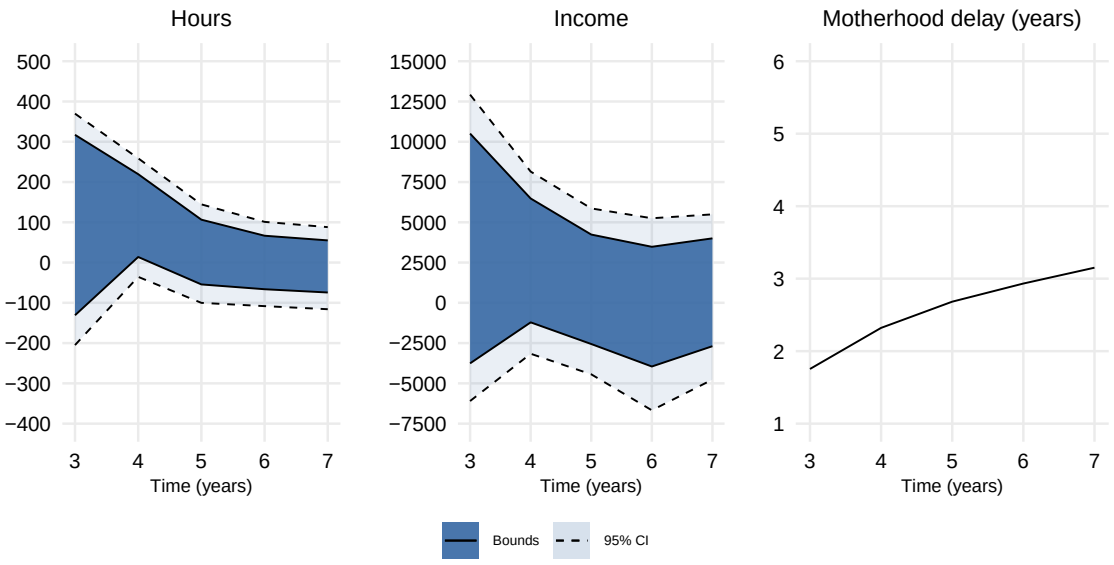

Figure 10: Effects of Delaying Motherhood

Figure 10 presents the results. The last graph shows the effects of ACP failure on parenthood timing. Among women who would conceive independently of ACP failure within four years, failure delays fertility by an average of 2.3 years. For those who would conceive independently within seven years, the average delay is 3.1 years. The estimated effects of delaying parenthood on women's work hours allow to rule out small negative contemporaneous impacts and even suggest at least a small positive effect in the fourth year. This impact is even larger when work hours are not adjusted for possible parental leave. Seven years after the first ACP, the bounds for both income and hours are relatively narrow and centered around zero.

The positive contemporaneous effect of delaying parenthood on work hours is surprising. Delaying fertility implies having a younger first child, which is generally expected to have a negative contemporaneous effect on women's labor market outcomes due to higher care requirements. Because of this, IV estimates are typically assumed to understate the career cost of parenthood (Lundborg et al., 2024), whereas my results suggest the opposite. One possible explanation for this concerns differences in total fertility. Women who enter motherhood earlier tend to have more children, which increases the need for care. This can explain why those who delay motherhood work more even after becoming mothers. The data support this explanation: women whose first ACP succeeds have 0.2 children more than women who become mothers after the first ACP fails, on average. This difference is stable from 2 to 7 years after the first ACP.

While the bounds suggest, at most, modest effects of delaying motherhood in the medium run, this may translate to substantial bias of the IV estimator because of the large always-taker share, which converges to 75%. Assuming that the effects of delaying

motherhood for always-takers are similar to the effects for non-reliers, IV estimates may understate the career impacts of motherhood on work hours and income in year seven by as much as 70%.

When the effects of parenthood are static, τ_{ATR} can be point-identified using a method proposed by Leuven et al. (2024). Alternatively, since under static effects $\mathbb{E}[Y_t(R^*)|R_t = 0] = \mathbb{E}[Y_t(1)|R_t = 0]$ and since $\mathbb{E}[Y_t(R^*)|R_t = 0]$ can be identified similar to $\mathbb{E}[Y_t(0)|R_t = 1]$, the relier average treated outcome can be backed out from the overall average treated outcome as:

$$\mathbb{E}[Y_t(1)|R_t = 1] = \frac{\mathbb{E}[Y_t(1)] - \Pr(R_t = 0)\,\mathbb{E}[Y_t(R^*)|R_t = 0]}{\Pr(R_t = 1)}.$$

Then, it can be compared to the relier average control outcome $\mathbb{E}[Y_t(0)|R_t = 0]$ identified using the baseline procedure. Identifying τ_{ATR} this way serves as a starting point for testing the parametric assumptions used in the approaches of Bensnes et al. (2023) and Gallen et al. (2023). Specifically, if heterogeneity among women is limited and if the effect depends solely on parenthood duration, then using either τ_{ATR} or τ_{LATE} identified in the short run to correct the bias in the longer run should yield the same results. I present the formal argument and the empirical results in Appendix A8. They indicate substantial violations of the parametric assumptions. Gallen et al. (2023) discuss how the two assumptions could be relaxed to allow for heterogeneity with respect to pre-ACP covariates; the relaxed assumptions can be tested using a similar approach.

7.2.2 Event Study and Selective Fertility

Next, I quantify the extent of selective fertility, which may bias the ES estimator. I begin by mapping the ES approach into my model. Then, using women's first ACP as a proxy for their fertility decision and leveraging ACP failures, I quantify the differences in average childless career trajectories among women who choose to become mothers at different times. Since the ES estimates based on the ACP and representative samples are indistinguishable, as shown in Figure 9, quantifying the extent of selective fertility in the ACP sample may also shed light on its role in the ES estimates for the non-ACP sample.

The most popular ES variation uses women who are one year away from becoming mothers as a control group for similarly aged women who already have children. To formally illustrate the ES approach in the context of my model, I first focus on women who conceive through their first ACP. The parallel trends assumption states that conditional on age and calendar time, control outcomes t periods after becoming a mother are the same as control outcomes in the period before becoming a mother, on average:

Assumption 8 (Parallel Trends)**.**

$\mathbb{E}[Y_t(0)|age_t = a, year_t = y, Z_1 = 1] = \mathbb{E}[Y_0(0)|age_0 = a, year_0 = y, Z_1 = 1]$, *for all* t, a, y.

Where age_t and $year_t$ are the woman's age and calendar year in period t, respectively. This assumption allows for unbiased predictions of childless outcomes in period t based on women's age and calendar year in period t, using outcomes of women who were of the same age and in the same calendar year in period 0—just before becoming mothers. Comparing the realized treated outcomes in period t with these predictions gives the average treatment effect:[24]

$$\tau_{ATE}(t) = \mathbb{E}[Y_t|Z_1 = 1] - \mathbb{E}[\mathbb{E}[Y_0|age_0 = age_t, year_0 = year_t, Z_1 = 1]|Z_1 = 1].$$

When the parallel trends assumption does not hold, the bias term for $\tau_{ATE}(t)$ is:

$$\mathbb{E}[Y_t(0)|Z_1 = 1] - \mathbb{E}[\mathbb{E}[Y_0(0)|age_0 = age_t, year_0 = year_t, Z_1 = 1]|Z_1 = 1].$$

It measures the difference in age- and year-specific average childless outcomes between women who have their first child earlier versus later, reflecting selective fertility timing. The theoretical direction of the bias is ambiguous. On one hand, it might lead to an overestimation of the career costs of parenthood if women with the highest career potential delay fertility. On the other hand, it could result in an underestimation if women who would succeed in their careers regardless of fertility choose to have children earlier. Additionally, women may adjust their fertility plans in anticipation of career shocks, such as job loss or promotion, which could introduce substantial bias even if the long-term trends between the two groups are similar.

I quantify the extent of selective timing by comparing relier average childless career trajectories identified using my baseline approach with those predicted using pre-ACP outcomes of similarly aged reliers, as in the ES approach. These predictions are obtained by modifying my baseline approach to identify age- and calendar-year-specific average pre-ACP relier outcomes.[25] Intuitively, I perform a placebo event study using reliers who remain childless after ACP failure, treating their first ACP as the event, and identify how much the ES approach would overstate the effects of parenthood for this group:

$$\mathbb{E}[Y_t(0)|R_t = 1] - \mathbb{E}[\mathbb{E}[Y_0(0)|age_0 = age_t, year_0 = year_t, R_t = 1]|R_t = 1].$$

Quantifying the extent of selective fertility specifically among reliers allows for a comparison

[24] The inner expectation is over Y_0, the outer expectation is over age_t and $year_t$. The inner expectation may not be well defined because the support of age_0 and $year_0$ differs from that of age_t and $year_t$; I abstract from this in demonstrating the intuition and address it under implementation in Appendix A4.

[25] Formally, this procedure involves two steps. First, I identify the relier average control outcomes in the period before their first ACP, conditional on age and calendar year: $\mathbb{E}[Y_0(0) \mid age_0 = a, year_0 = y, R_t = 1]$. The identification procedure follows the same steps as for the relier average control outcome $\mathbb{E}[Y_t(0) \mid R_t = 1]$, except that all expectations are conditioned on pre-ACP age and calendar year, and the realized labor market outcome in period t is replaced by that in period 0. Then, as in the ES approach, I use these conditional average control outcomes in period 0 to construct age- and calendar-year-specific predictions for control outcomes after t periods.

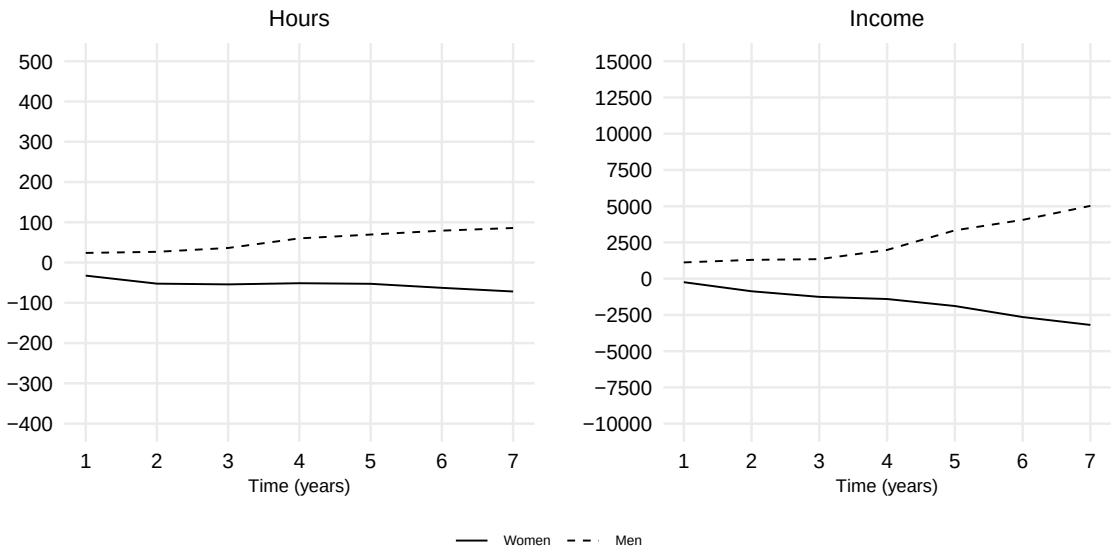

Figure 11: Placebo Event Study

with τ_{ATR}, making it possible to distinguish how much of the gender inequality associated with parenthood is driven by the effect of parenthood itself versus selective timing. My empirical specification follows Kleven et al. (2019); I discuss implementation details in Appendix A4.

Figure 11 presents separate estimates for men and women. They suggest that women who choose to have children earlier would experience worse career outcomes than those who choose to have children later, even in the absence of children, while the opposite is true for men. The positive relationship between fatherhood and labor market outcomes aligns with the descriptive literature documenting a "fatherhood premium" (Lundberg & Rose, 2000). ES estimates also suggest the presence of this premium around the world, including the Netherlands, (Kleven et al., 2024).

The main results, presented in Figure 12, demonstrate how much of the gender inequality among reliers is jointly explained by the causal effect of parenthood and selective fertility timing. In year 7, the two factors account for between 70% and 85% of inequality in work hours and 50% to 80% of inequality in income.[26] These results are consistent with the baseline ES estimates, which capture both factors jointly. They suggest that at least 34% of the gender inequality in work hours and 42% of the inequality in incomes associated with parenthood is due to selective fertility timing, rather than the effects of parenthood itself.

In the context of my model, these results imply that selective fertility causes the ES estimates to substantially overstate the impact of parenthood on women's labor market outcomes and gender equality. However, my model considers a particular counterfactual scenario for not having children, specifically, one in which women try and fail to conceive.

[26]The share is even larger if differences in age between partners are accounted for.

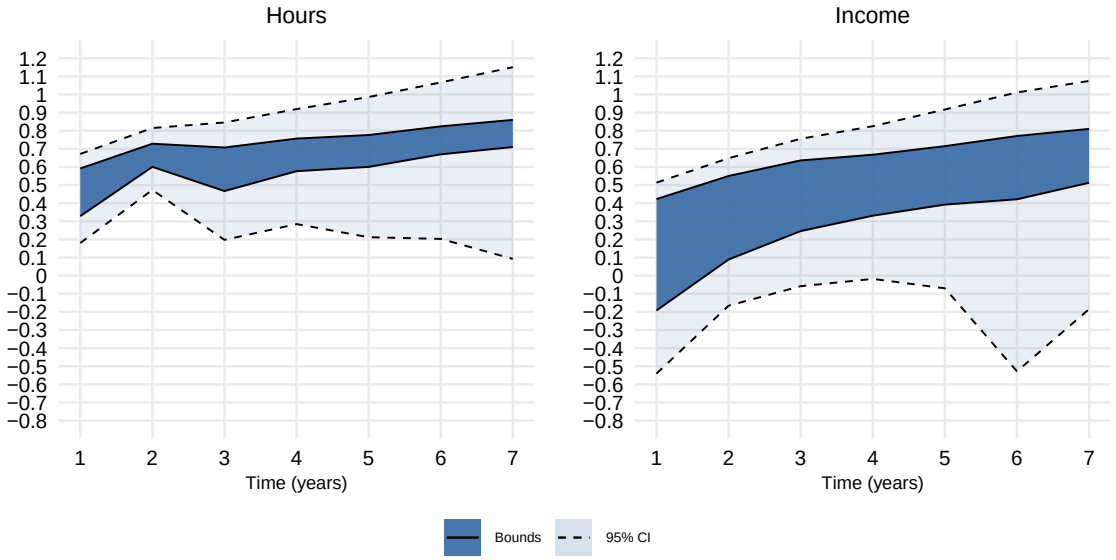

Figure 12: Share of Gender Inequality Explained by Effects of Parenthood and Selective Fertility

Women's outcomes in this scenario may differ from a scenario in which they choose not to have children, and the latter scenario may be more relevant from both a policy perspective and the perspective of families making career and fertility decisions. One reason these scenarios may differ is that failure to conceive (either in general or specifically via ACPs), relative to not trying to conceive, may impact women's (mental) health and/or relationship stability, and in turn, labor market outcomes. If such effects are substantial, then the placebo event study estimates may not reflect selective fertility but rather an important component of parenthood that is overlooked by my baseline approach. I address these effects in the next section.

7.3 Mental Health and Relationship Stability

Women who remain childless after ACP failure may experience negative (mental) health outcomes and/or relationship breakdowns. This is important for my analysis for two main reasons. First, it may limit the external validity of my findings, particularly when extrapolating to a setting where women choose to remain childless rather than fail to conceive. If these effects are especially strong among ACP families, it may also impact extrapolation to non-ACP families. Second, it threatens the internal validity of my estimates because relationship breakdowns and mental health issues may lead to fewer attempts to conceive without ACPs, which may violate the monotonicity assumption used to tighten the bounds. In this section, I first discuss the potential relevance of these effects, then introduce a procedure to address them, and present empirical evidence.

Mental health issues and relationship breakdowns are arguably less concerning for the external validity of my results when they arise from unmet fertility goals. In this case,

these factors act as potential mechanisms through which parenthood influences labor market outcomes in both ACP and non-ACP populations, regardless of whether individuals choose to have children. In contrast, the ES approach, which uses families that may not have tried or may not yet want to have children as a control group, might overlook these impacts. Nonetheless, a remaining concern for extrapolating my results to non-ACP families, in this case, is that such impacts may be especially strong in ACP families due to their particularly high desire for children.

Mental health issues, in particular, raise concerns for extrapolating my estimates to a scenario where women choose not to have children, when these issues result from the failure to conceive or emerge as a side effect of ACPs, rather than from the absence of children. This concern has been raised about studies using IVF, especially because IVF is relatively invasive (Bögl et al., 2024). As a result, undergoing additional procedures after an initial failure may impact women's mental health, and consequently, their labor market outcomes, independent of fertility or the desire to have children.

Both mental health issues and relationship breakdowns also threaten the internal validity of my estimates that rely on the monotonicity assumption. Improved mental health and relationship stability after a successful conception may lead to more attempts to conceive via non-ACP means. Consequently, success in the first ACP could result in some women having non-ACP children they would not have had if ACPs had failed, thereby violating the monotonicity assumption.

Before addressing these concerns formally, I discuss their potential empirical relevance. One strength of my setting compared to studies focused on IVF is that IUI is the primary ACP that couples undergo. As discussed in Section 5.2, IUI is significantly less invasive than IVF, which helps mitigate concerns about the potential side effects these procedures may have on women's mental and physical health.[27] It is also worth noting that while some women may experience ACP side effects, their impact on labor market outcomes is likely negligible within the context of parenthood. Lundborg et al. (2024) present conservative back-of-the-envelope calculations, drawing on economic research on severe health shocks and medical literature documenting the prevalence of ACP side effects. They conclude that these effects are unlikely to meaningfully influence women's career trajectories. Finally, a naive approach to quantify the importance of these mental health effects is to estimate the impact of ACP failure on severe mental health outcomes, proxied by antidepressant uptake. The estimates, presented in Appendix A9, are precise and indistinguishable from zero. This provides suggestive evidence that ACPs do not have large effects on women's mental health outcomes. However, this does not fully address the external validity concerns. This is because these effects are identified relative to a scenario where women have children, which may also negatively impact their mental health compared to the scenario in which they choose not to have children.

[27] This does not completely resolve the concerns because around a third of women whose first IUI procedure fails eventually turn to IVF.

Next, I introduce a new method to address the external validity concerns related to the effects of ACPs on mental health and relationship stability. I adapt my baseline approach to bound the effects specifically for women who, in the event of ACP failure, would remain with their partners and not experience severe mental health issues, as proxied by the onset of antidepressant medication. This ensures that my estimates are not driven by women whose mental health or relationships may be negatively impacted specifically by ACPs, nor by those who face the most significant consequences of childlessness. Additionally, this addresses internal validity concerns related to the monotonicity assumption by allowing for violations among women who might separate from their partners or experience mental health issues after ACP failure.

The formal identification result amounts to treating only childless women who do not uptake antidepressants and remain with their original partner as reliers.[28] Let K_t take the value 1 if, in period t and in the scenario that all ACPs fail, a woman would not suffer severe mental health issues and would remain with her partner; otherwise, let it take the value 0. I refer to this group as *resilient* ($K_t = 1$). I bound the average treatment effect specifically for resilient reliers $\mathbb{E}[\tau(t) \mid R_t = 1, K_t = 1]$. I also relax the monotonicity assumption to:

Assumption 9 (Partial Monotonicity)**.**

$\Pr(R_t^+ \geq R_t \mid K_t = 1) = 1.$

It states that monotonicity holds for resilient women. This implies that women who conceive additional children without ACP after their first ACP succeeds would, in the scenario that all ACPs fail, either have at least one non-ACP child, experience mental health issues, or separate from their partners. This allows for the unconditional monotonicity violations considered in this section. I discuss empirical support for the original and the partial monotonicity assumptions in Appendix A9.

Figure 13 presents the results. In the first few years of parenthood, the bounds are similar to those using the baseline approach. Up to the fourth year, the estimated share of reliers that are resilient is above 90%. The bounds widen in the later years; by year 7, the estimated share of reliers that are resilient is 85%. Nonetheless, this has only a modest effect on the main conclusions. In the most extreme scenario, the share of gender inequality in work hours and income caused by parenthood increases by less than 10 percentage points.

It is important to note that the procedure I use to address mental health and relationship stability concerns is conservative for two reasons. First, it excludes from the comparison women who would experience poor mental health or relationship breakdowns regardless of fertility or attempts to conceive. This makes the bounds wider than if such women were included. Second, it tackles both mental health and relationship breakdowns simultaneously. The bounds presented in Appendix A9, which address these concerns separately, are even

[28] Redefining D^+ to take the value 1 in cases where a woman has no children but either uptakes antidepressants or separates from her partner, and following a similar argument as for the baseline bounds, gives the result.

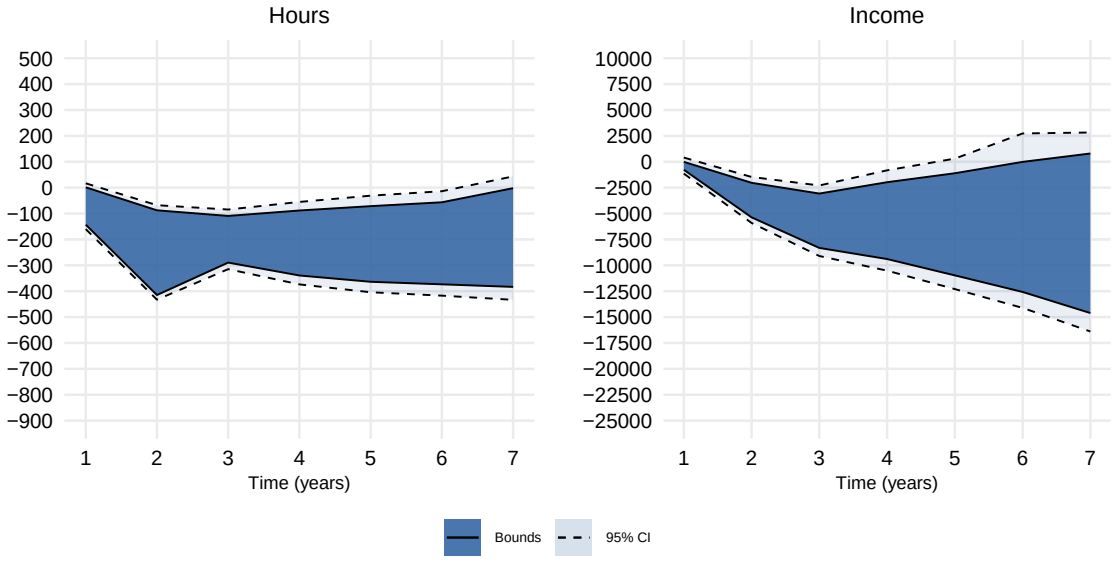

Figure 13: Effects on Resilient Women

closer to the baseline estimates.

These results are important for several reasons. Primarily, they support the extrapolation of my findings to a counterfactual scenario where families choose not to have children, thus avoiding potential mental health issues or relationship breakdowns associated with unsuccessful conception attempts and ACPs. Additionally, they support extrapolation to non-ACP families, where the impacts of failing to conceive may be less pronounced. Finally, they suggest that my findings are not driven by relationship breakdowns or negative mental health consequences following a failure to conceive, including the side effects of ACPs.

7.4 Effects Over Time and Stable Relier Group

The changes in my main estimates over time reflect a combination of two factors. First, how the effect of parenthood evolves with time spent in parenthood. Second, since the group of reliers shrinks over time, how effects differ between women who remain reliers for a different duration. An equivalent concern regarding changing compliers applies to the IV estimates. This means that my main results provide limited insight into how the effects evolve with time spent in parenthood. To address it, I adapt my approach to bound the effects for a stable group of reliers.

I bound τ_{ATRL}, which is the effect for women who remain reliant until the last period. The key assumption enabling this is irreversible fertility. It ensures that reliers who are childless in the current period were also childless in previous periods. This allows to identify their average control outcomes in earlier periods similar to their average control outcome in the current period. It also allows to bound their average treated outcome in previous periods by trimming the current relier share from the tails of the outcome distribution in

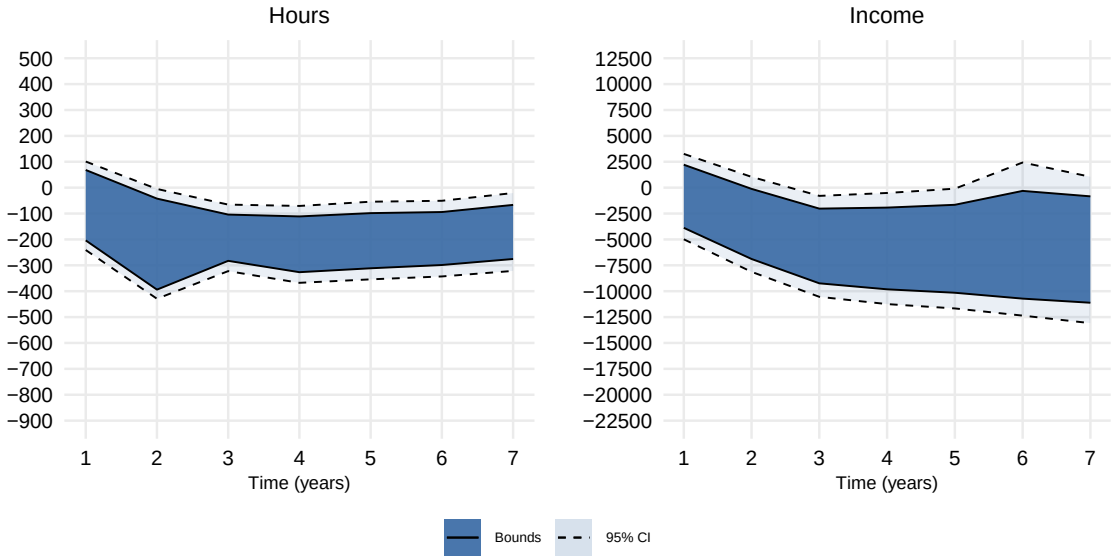

Figure 14: Effects on a Stable Group of Women

earlier periods among current subsequent reliers.[29] Figure 14 presents the results, which remain similar to the baseline estimates and allow to rule out large changes in effects over time.

7.5 Heterogeneity by Effort to Conceive

An interesting question in its own right, and a concern when extrapolating my results to the non-ACP population, is whether families that exert more effort to have children face a lower cost of parenthood. To explore this relationship, I proxy potential effort by the willingness to undergo subsequent ACPs and examine heterogeneity in the effects along this dimension. I build on a method introduced by Leuven et al. (2024) to assess such heterogeneity in a setting without dynamic effects.

Intuitively, I leverage the fact that some women successfully conceive through subsequent ACPs immediately after the first one fails, meaning their fertility timing closely resembles what it would have been if their first ACP had succeeded. Since each ACP outcome is as good as random, these women are a representative sample of those willing to undergo multiple ACPs in quick succession, allowing to bound the treatment effect for this group.

[29] Without monotonicity, these bounds are mechanically at least as wide as the bounds on τ_{ATR} because the trimmed non-relier share can only increase over time. Under monotonicity is that bounds on τ_{ATRL} may be narrower than bounds on τ_{ATR}. Intuitively, this can occur because the width of the bounds depends on the difference between the subsequent relier share and the relier share, and the decrease in the subsequent relier share over time may outweigh the increase in the relier share. For instance, if all women who would eventually have children without ACP do so shortly after their first ACP fails, the relier share will remain stable over time. In contrast, women whose first ACP succeeds can only have a non-ACP child after their first ACP birth, meaning the subsequent relier share may decrease gradually. In this sense, information on future fertility helps obtain narrower bounds on past effects.

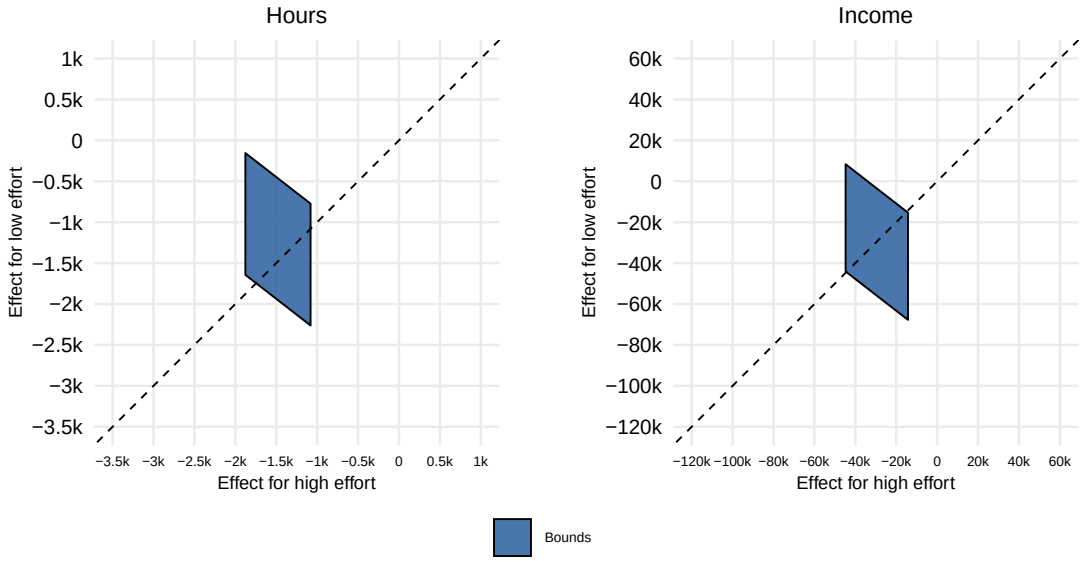

Figure 15: Effects on Women's Cumulative Outcomes 7 Years into Parenthood by Willingness to Undergo ACPs in Year 1

The average control outcoems for different types are identified using the baseline method. I present a formal argument in Appendix A10.

The bounds for a subgroup of reliers with a high willingness to undergo ACPs can be compared to the aggregate bounds obtained using the baseline method. However, this serves as a conservative test for effect heterogeneity because it ignores the fact that some points in the aggregate bounds may be inconsistent with those in the subgroup bounds.[30] Instead, for each possible average effect in the high willingness subgroup, bounds for the low willingness subgroup can be backed out from the aggregate bounds.

To present the relationship between bounds for two groups in a single figure, I focus on cumulative hours and income over the first 7 years of parenthood. The horizontal axis in Figure 15 presents the effect for women with a median-or-above willingness to undergo ACPs in the first period. The vertical axis presents the effect for those whose willingness to undergo ACPs in the first period is below the median. The results do not allow to rule out homogeneous effects between the two groups, as indicated by the dashed line. In the most extreme scenario, women with below-median willingness to undergo ACP experience reductions in hours that are no more than twice those experienced by women with above-median willingness, and reductions in income that are no more than 3.5 times as large. I address the effect heterogeneity between ACP and non-ACP families in the next section.

[30] For example, consider a scenario where the lower bound for the subgroup is below the upper bound for the aggregate group. A naive comparison would fail to rule out homogeneous effects, but it overlooks the fact that the aggregate upper bound may not be attainable when the effect for the subgroup matches its lower bound.

7.6 Effects for Non-ACP Families

In this section, I provide suggestive evidence to address the concern that my results have limited relevance for non-ACP families because ACP families face a lower career cost of parenthood. It is worth noting that this concern may be less pronounced in my setting compared to studies using IVF. As discussed in Section 5.2, IUI is less invasive and more accessible than IVF, making couples who undergo the procedure more comparable to the general population. To address this concern formally, I propose a procedure to identify the effects for non-ACP mothers using their realized treated outcomes and control outcomes imputed using childless women who tried to have children at a similar moment using ACPs. My approach involves two steps. First, I estimate the relier average childless outcomes conditional on pre-ACP education and age using my baseline method. Then, I use these estimates to predict outcomes for non-ACP women based on their age and education prior to conception. The full argument is presented in Appendix A11. The primary concern to address is that the childless career trajectories imputed using ACP women may systematically differ from those among non-ACP women.

I benchmark my method against the ES approach. The main piece of evidence used to justify the ES approach is that women who become mothers earlier have similar career trajectories up to the moment of parenthood as women who become mothers later. However, a concern is that differences in childless career trajectories might emerge after some women choose to become mothers while others decide to remain childless. One reason this may occur is that the decision to have children is influenced by anticipated worse career trajectories. In contrast, the main appeal of my method is that it uses women who choose to have children at a similar moment, albeit through different means. Assuming my approach also performs well in replicating women's childless career trajectories up to motherhood, there is arguably less reason to worry that differences would emerge after motherhood.

Figure 16 plots the average childless career trajectories for women in the non-ACP sample, imputed using the two methods. Year 0 serves as the baseline for the ES approach, meaning that the ES estimates mechanically match the average control outcomes in this period. The average outcomes imputed using the new approach align almost perfectly. This is the strongest support for the validity of my method. Its ability to replicate pre-parenthood career trajectories among non-ACP mothers using only age and education is not mechanical, as the average outcomes for the ACP and non-ACP samples can differ substantially. While the new approach suggests a relatively flat control hours profile after parenthood, the ES approach suggests a rising one. Since almost all of these women have completed their education by $t = 0$, the steady increase in average work hours implied by the ES approach is harder to explain. Similarly, while both methods suggest rising income over time, the trajectory based on the new approach is flatter.

As an additional validation exercise for imputation from the ACP sample to the non-ACP sample, I impute treated career trajectories for non-ACP women using women whose

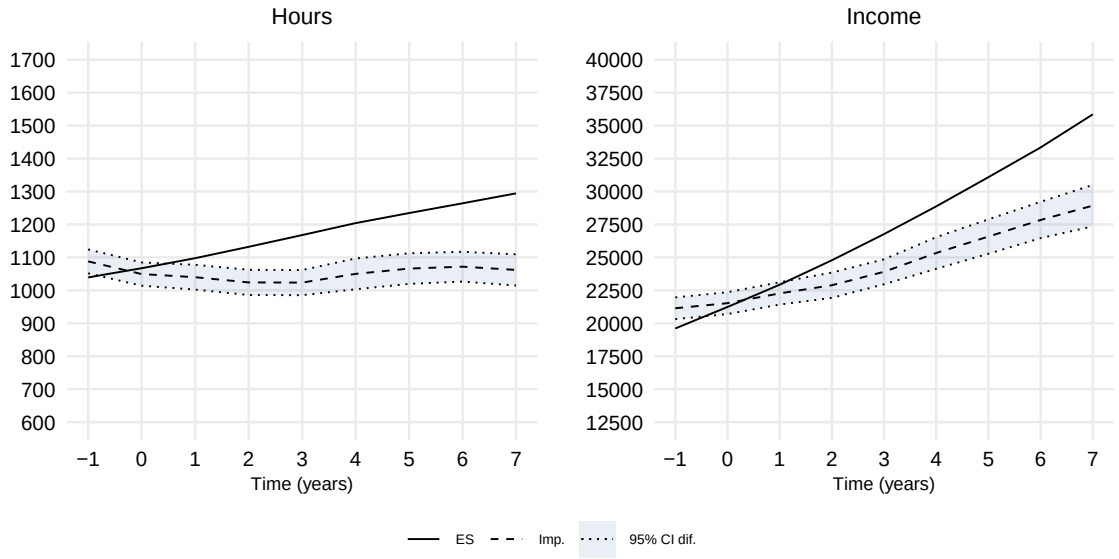

Figure 16: Imputed Childless Career Trajectories for Non-ACP Mothers

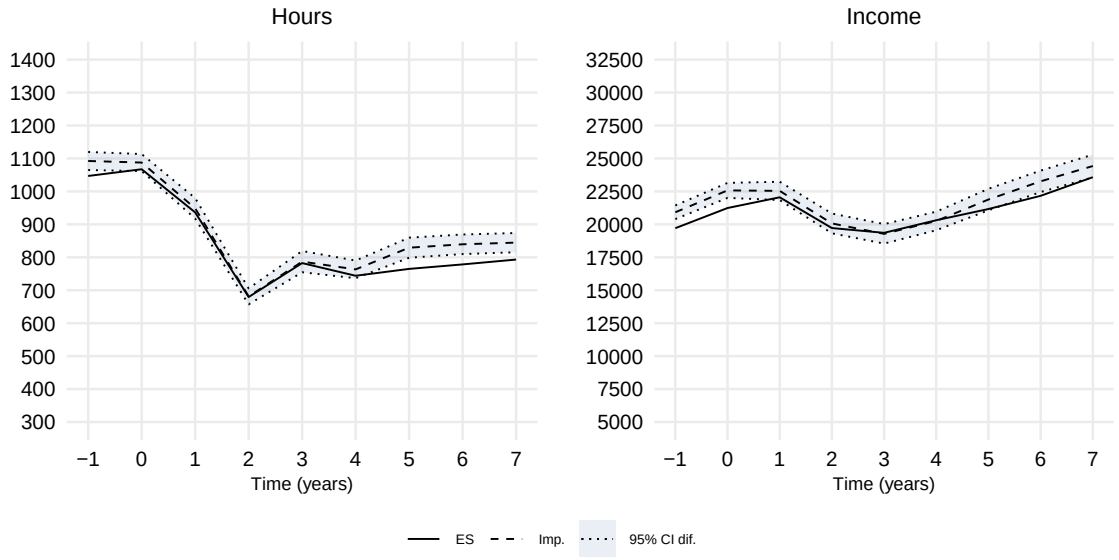

Figure 17: Imputed Motherhood Career Trajectories for Non-ACP Mothers

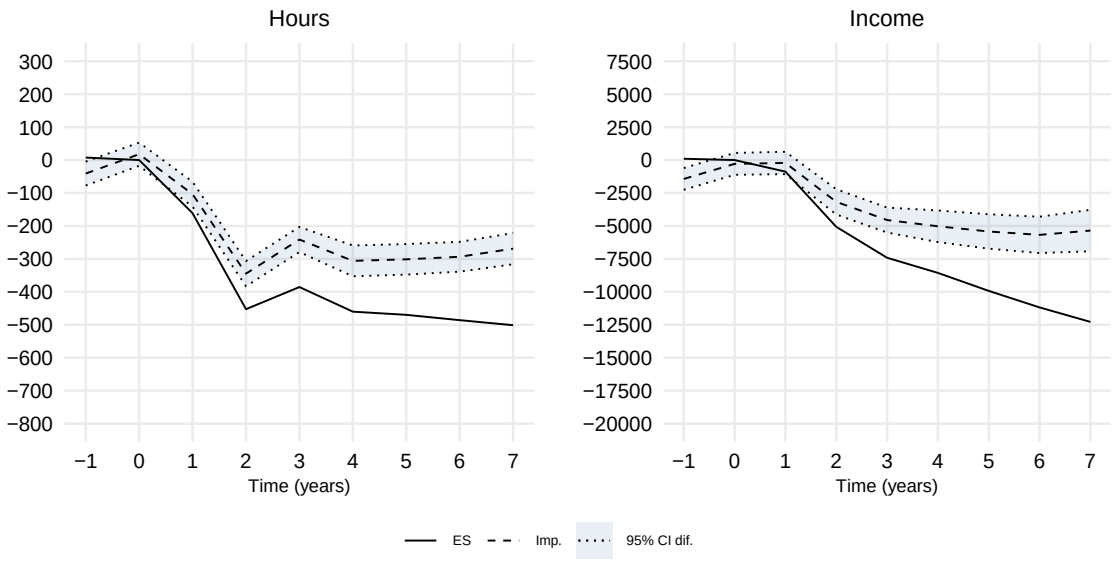

Figure 18: Effects of Parenthood Based on Imputed Childless Career Trajectories for Non-ACP Mothers

first ACP succeeded. The results are presented in Figure 17. The ES estimates match the treated career trajectories exactly because they correspond to OLS fitted values. The trajectories imputed using the new approach follow them closely.[31]

Finally, Figure 18 presents estimates of the average career impacts of motherhood in the non-ACP sample using the new approach, alongside the ES estimates. The ES estimates suggest that the career cost of motherhood increases with time, while estimates using the new approach suggest a more stable impact. Seven years into motherhood, the ES estimates indicate a 38% reduction in hours and a 33% reduction in income. Estimates using the new approach suggests a substantially smaller reduction of 27% in hours and 17% in income. These results for the non-ACP population are consistent with my bias decomposition estimates for the ACP sample, suggesting that ES estimates may overstate the career cost of motherhood.

8 Conclusion

Parenthood can explain most of the observed gender inequality in labor force participation in the Western world (Kleven et al., 2024). However, providing causal evidence of its effects has proven challenging for two main reasons. First, the decision to have children may be related to labor market outcomes independent of fertility. Second, the effects of parenthood may depend on its timing. In this paper, I propose a method that leverages women's ACP

[31] Average completed fertility in both groups is 1.8, suggesting that fertility differences do not confound the comparison.

histories to bound the effects of parenthood, ensuring robustness to both selective fertility and dynamic effects.

It is worth emphasizing that the procedure I introduce can be used to bound treatment effects in various settings where individuals are quasi-experimentally assigned to one state but may enter others, either by undergoing quasi-experimental assignment multiple times or through entirely selective means. Such settings include educational programs with multiple admission cycles, job training programs with multiple entry pathways, legal settings with quasi-experimental assignment to judges, and clinical trials in the extension phase, allowing the control group to access alternative therapies. The procedure can also be applied to sequential experiments where outcome data is missing for some individuals.

Applying my method to estimate the career impacts of parenthood in the Netherlands, I find persistent reductions in women's yearly work hours, between 10% and 25%, and in their income, between 9% and 29%. Despite that, I find that at least half of the observed post-child gender inequality in these outcomes is not caused by parenthood. I also provide evidence suggesting that estimates based on the ES approach may substantially overstate the career impacts of parenthood due to selective fertility. In contrast, I find that IV estimates may understate these impacts due to dynamic effects. Moreover, I demonstrate that accounting these biases can reconcile the conflicting results in the literature.

My analysis also addresses several external validity concerns raised about existing studies on the effects of parenthood, particularly those relying on IVF and Scandinavian data. By focusing on IUI, which is significantly less invasive and more accessible than IVF, I mitigate concerns related to the mental health impacts of IVF and the sample selectivity of families undergoing this procedure. To further address concerns about the mental health effects associated with failing to conceive, I adapt my method to ensure my estimates are based on women who do not take antidepressants in this scenario. Finally, using data from the Netherlands, where family-friendly policies are relatively average compared to other OECD countries, makes my results more applicable to countries with less generous policies than those in the Scandinavian context.

My findings have important implications for understanding the causes of gender inequality in the labor market and for identifying remedies to alleviate it. Existing research leveraging short-run variation in family-friendly policies suggests that they often have, at best, modest impacts on gender inequality (Cortés & Pan, 2023). My results indicate that this may be partly because a large share of gender inequality is not caused by parenthood per se. Moreover, studies focusing on short-run variations may overlook how family-friendly policies influence women's behavior by lowering the anticipated cost of parenthood, potentially leading to greater educational and career investment. Consequently, such policies may still have substantial impacts on gender inequality in the long run, independent of realized fertility and its effects.

References

Adda, J., Dustmann, C., & Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, *125*(2), 293–337.

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

Angelov, N., Johansson, P., & Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of labor economics*, *34*(3), 545–579.

Angrist, J., & Evans, W. N. (1996). Children and their parents' labor supply: Evidence from exogenous variation in family size. *National bureau of economic research*.

Angrist, J., & Imbens, G. (1995). Identification and estimation of local average treatment effects. *National Bureau of Economic Research*.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *arXiv preprint arXiv:1610.01271*.

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

Bertrand, M. (2011). New perspectives on gender. In *Handbook of labor economics* (Vol. 4, pp. 1543–1590). Elsevier.

Bertrand, M. (2020). Gender in the twenty-first century. *AEA Papers and proceedings*, *110*, 1–24.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, *55*(3), 789–865.

Bögl, S., Moshfegh, J., Persson, P., & Polyakova, M. (2024). The economics of infertility: Evidence from reproductive medicine. *National Bureau of Economic Research*.

Bronars, S. G., & Grogger, J. (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, 1141–1156.

Brooks, N., & Zohar, T. (2021). Out of labor and into the labor force? the role of abortion access, social stigma, and financial constraints. *CEMFI Working Paper No. 2111*.

Bütikofer, A., Jensen, S., & Salvanes, K. G. (2018). The role of parenthood on the gender gap among top earners. *European Economic Review*, *109*, 103–123.

Chung, Y., Downs, B., Sandler, D. H., Sienkiewicz, R., et al. (2017). The parental gender earnings gap in the united states. *Unpublished manuscript*.

Cortés, P., & Pan, J. (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature*, *61*(4), 1359–1409.

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

Cruces, G., & Galiani, S. (2007). Fertility and female labor supply in latin america: New causal evidence. *Labour Economics*, *14*(3), 565–573.

Eichmeyer, S., & Kent, C. (2022). Parenthood in poverty. *Centre for Economic Policy Research*.

Fitzenberger, B., Sommerfeld, K., & Steffes, S. (2013). Causal effects on employment after first birth—a dynamic treatment approach. *Labour Economics*, *25*, 49–62.

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

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American economic review*, *104*(4), 1091–1119.

Groes, F., Houštecká, A., Iorio, D., & Santaeulàlia-Llopis, R. (2024). The unequal battle against infertility: Theory and evidence from ivf success. *Centre for Economic Policy Research*.

Heiler, P. (2024). Heterogeneous treatment effect bounds under sample selection with an application to the effects of social media on political polarization. *Journal of Econometrics*, *244*(1), 105856.

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

Hirvonen, L. (2009). The effect of children on earnings using exogenous variation in family size: Swedish evidence. *Swedish Institute for Social Research Working Paper*.

Horowitz, J. L., & Manski, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society*, 281–302.

Horowitz, J. L., & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American statistical Association*, *95*(449), 77–84.

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

Iacovou, M. (2001). Fertility and female labour supply. *ISER Working Paper Series*.

Jacobsen, J. P., Pearce III, J. W., & Rosenbloom, J. L. (1999). The effects of childbearing on married women's labor supply and earnings: using twin births as a natural experiment. *Journal of Human Resources*, 449–474.

Kleven, H., Landais, C., & Leite-Mariante, G. (2024). The child penalty atlas. *Review of Economic Studies*, rdae104.

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

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

Leuven, E., Oosterbeek, H., van der Klaauw, B., & Ketel, N. (2024). Using principal stratification to analyze repeated treatment assignment. *Unpublished working paper*.

Lundberg, S., & Rose, E. (2000). Parenthood and the earnings of married men and women. *Labour Economics*, *7*(6), 689–710.

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

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

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, 343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, *80*(2), 319–323.

Maurin, E., & Moschion, J. (2009). The social multiplier and labor market participation of mothers. *American Economic Journal: Applied Economics*, *1*(1), 251–272.

Melentyeva, V., & Riedel, L. (2023). Child penalty estimation and mothers' age at first birth. *ECONtribute Discussion Paper*.

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

OECD. (2022). *Out-of-school-hours services.*

OECD. (2023a). *Enrolment in childcare and pre-school.*

OECD. (2023b). *Oecd employment database.* Retrieved from https://stats.oecd.org/Index.aspx?DatasetCode=AVE_HRS

OECD. (2023c). *Parental leave system*.

OECD. (2023d). *Part-time employment rate (indicator)*. Retrieved from https://www.oecd.org/en/data/indicators/part-time-employment-rate.html

OECD. (2024). *Society at a glance 2024*. Retrieved from https://www.oecd-ilibrary.org/content/publication/918d8db3-en

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.

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

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

Stoye, J. (2020). A simple, short, but never-empty confidence interval for partially identified parameters. *arXiv preprint arXiv:2010.10484*.

Van den Berg, G. J., & Vikström, J. (2022). Long-run effects of dynamically assigned treatments: A new methodology and an evaluation of training effects on earnings. *Econometrica*, *90*(3), 1337–1354.

Vere, J. P. (2011). Fertility and parents' labour supply: new evidence from us census data: Winner of the oep prize for best paper on women and work. *Oxford Economic Papers*, *63*(2), 211–231.

Zhang, J. L., & Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, *28*(4), 353–368.

A1 Proof of Theorem

Lemma. *For any for l s.t. $1 \leq l \leq \overline{w}$ and any measurable function $g(M_l)$, where:*

$$M_l = (Y(1), Y(0), R^+, R, W, Z_1, \ldots, Z_l, X_1, \ldots, X_l)$$

under assumptions 3 and 4:

$$\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))}\middle| X_l\right] = \mathbb{E}\left[g(M_l)|X_l\right].$$

Proof of lemma. Since $Z_j = 0|A < j$, and since X_j includes $1_{\{A \geq j\}}$, $Z_1, \ldots, Z_{j-1}$, and $X_1, \ldots, X_{j-1}$, assumption 3 implies:

$$(Y(1), Y(0), R^+, R, W, Z_1, \ldots, Z_{j-1}, X_1, \ldots, X_{j-1}) \perp\!\!\!\perp Z_j | X_j \text{ for all } j > 1. \tag{3}$$

Assumption 4 ensures that $1 - e_j(x_j) > 0$ for all j and $x_j \in \mathcal{X}_j$. Then, w.l.o.g. for some l s.t. $l < \overline{w}$:

$$\begin{aligned}
&\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))}\middle| X_l\right] = \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))}\middle| X_l\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))}\middle| X_{\overline{w}}, X_l\right]\middle| X_l\right] &(4)\\
&= \mathbb{E}\left[\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))}\middle| X_{\overline{w}}\right]\middle| X_l\right] &(5)\\
&= \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}-1} \frac{(1-Z_j)}{(1-e_j(X_j))}\,\mathbb{E}\left[\frac{1-Z_{\overline{w}}}{1-e_{\overline{w}}(X_{\overline{w}})}\middle| X_{\overline{w}}\right]\middle| X_l\right] &(6)\\
&= \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}-1} \frac{(1-Z_j)}{(1-e_j(X_j))}\middle| X_l\right] &(7)\\
&= \mathbb{E}\left[g(M_l)|X_l\right], &(8)
\end{aligned}$$

where (4) holds by law of iterated expectations, (5) holds because X_j includes X_l for $j \geq l$, (6) holds by 3, (8) holds by assumption 3 because:

$$\begin{aligned}
\mathbb{E}\left[\frac{1-Z_{\overline{w}}}{1-e_{\overline{w}}(X_{\overline{w}})}\middle| X_{\overline{w}}\right] &= \frac{1-\mathbb{E}\left[Z_{\overline{w}}|X_{\overline{w}}\right]}{1-e_{\overline{w}}(X_{\overline{w}})} \\
&= \frac{1-e_{\overline{w}}(X_{\overline{w}})}{1-e_{\overline{w}}(X_{\overline{w}})} \\
&= 1,
\end{aligned}$$

and where (8) follows by steps similar to (4) through (6) for X_j for j s.t. $l < j < \overline{w}$. □

Proof of theorem. I demonstrate the result for the upper bound, the result for the lower bound is symmetric. First, I demonstrate that $\mathbb{E}\left[Y(1-D^+)\prod_{j=1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))}\right] / \mathbb{E}[r(X_1)] =$

$\mathbb{E}[Y(0)|R=1]$. Using that $D^+ = D|Z_A = 0$, $Y = Y(0)|D = 0$, and $D = R|Z_A = 0$:

$$\mathbb{E}\left[Y(1-D^+)\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}\right] = \mathbb{E}\left[\mathbb{E}\left[Y(0)R\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}\middle|X_1\right]\right] \tag{9}$$

$$= \mathbb{E}\left[\mathbb{E}\left[Y(0)R\frac{1-Z_1}{1-e_1(X_1)}\middle|X_1\right]\right] \tag{10}$$

$$= \mathbb{E}\left[\mathbb{E}\left[Y(0)R|X_1\right]\mathbb{E}\left[\frac{1-Z_1}{1-e_1(X_1)}\middle|X_1\right]\right] \tag{11}$$

$$= \mathbb{E}\left[Y(0)R|X_1\right] \tag{12}$$

$$= \mathbb{E}\left[Y(0)|R=1\right]\Pr(R=1), \tag{13}$$

where (9) holds by law of iterated expectations, (10) holds by Lemma, and (11) and (12) hold by assumption 3. Moreover since $D+ = R|Z_A = 0$ and $1 - Z_A = \prod_{j=1}^{\overline{w}}(1-Z_j)$:

$$\mathbb{E}\left[(1-D^+)\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}\middle|X_1\right] = \mathbb{E}\left[R\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}\middle|X_1\right] \tag{14}$$

$$= \mathbb{E}\left[R\frac{1-Z_1}{1-e_1(X_1)}\middle|X_1\right] \tag{15}$$

$$= \Pr(R=1|X_1), \tag{16}$$

where (15) holds by Lemma and (16) holds by assumption 3. Since $\mathbb{E}[\Pr(R=1|X_1=x)] = \Pr(R=1)$, $\mathbb{E}\left[Y(1-D^+)\prod_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j(X_j))}\right] / \mathbb{E}[r(X_1)] = \mathbb{E}[Y(0)|R=1]$ holds.

Remains to show that $\mathbb{E}\left[Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)}\right] / \mathbb{E}[r(X_1)]$ is a sharp upper bound for $\mathbb{E}[Y(1)|R=1]$. I first demonstrate that $p(x) = \Pr(R=1|D^+=0, Z_1=1, X_1=x)$. Assumption 3 together with $D^+ = 1-R^+|Z_1=1$ implies that $r^+(x) = \Pr(R^+=1|X_1=x)$. Under assumption 2, $\Pr(R=1|X_1=x) = \Pr(R=1, R^+=1|X_1=x)$. Applying the definition of conditional probability gives $p(x) = \Pr(R=1|R^+=1, X_1=x)$. Assumption 3 together with $D^+ = 1-R^+|Z_1=1$ gives $\Pr(R=1|D^+=0, Z_1=1, X_1=x) = \Pr(R=1|R^+=1, X_1=x)$, which implies the result.

The remainder of the proof is similar to Lee (2009). Let $\gamma_x = \mathbb{E}[Y|Z_1=1, D^+=0, Y \geq q(1-p(X_1), X_1), X_1=x]$. I next demonstrate that γ_x is a sharp upper bound for $\mathbb{E}[Y(1)|X_1=x, R=1]$. Using that $p(x) = \Pr(R=1|D^+=0, Z_1=1, X_1=x)$, Corollary 4.1 Horowitz & Manski (1995) gives, $\gamma_x \geq \mathbb{E}[Y|Z_1=1, D^+=0, R=1, X_1=x]$. Using that $D^+ = 0|R=1$ and $Y = Y(1)|Z_1=1$ and by assumption 3, $\mathbb{E}[Y|Z_1=1, D^+=0, R=1, X_1=x] = \mathbb{E}[Y(1)|X_1=x, R=1]$, meaning that γ_x is an upper bound bound for $\mathbb{E}[Y(1)|X_1=x, R=1]$. Since $p(x)$ is identified, Corollary 4.1 Horowitz & Manski (1995) also implies the bound is sharp.

Let $f_{x|R=1}(x)$ be the p.d.f. of X_1 conditional on $R=1$. Applying Bayes rule to $\Pr(R=1|X_1=x)$ identified by $r(x)$ and p.d.f. of X_1 is identified directly gives $f_{x|R=1}(x)$,

making $\int_{\mathcal{X}_1} \gamma_x f_{x|R=1}(x)dx$ the sharp upper bound for $\mathbb{E}[Y(1)|R=1]$.

The last step is to show that

$$\int_{\mathcal{X}_1} \gamma_x f_{x|R=1}(x)dx = \mathbb{E}\left[Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)}\right] / \mathbb{E}[r(X_1)].$$

By the law of iterated expectations:

$$\mathbb{E}\left[Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)}\right] = \mathbb{E}\left[\frac{1}{e_1(X_1)}\mathbb{E}[Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}Z_1|X_1]\right].$$

Applying the definition of conditional probability:

$$\mathbb{E}[Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}Z_1|X_1] = \\ \mathbb{E}[\gamma_{X_1}|X_1]\Pr(D^+=0,Z_1=1,Y>q(1-p(X_1),X_1)|X_1).$$

Applying the definition of conditional probability twice:

$$\Pr(D^+=0,Z_1=1,Y>q(1-p(X_1),X_1)) = \\ \Pr(Y>q(1-p(X_1),X_1)|D^+=0,Z_1=1,X_1)\Pr(D^+=0|Z_1=1,X_1)\Pr(Z_1=1|X_1).$$

Using the definitions of $p(X_1)$, $r^+(X_1)$, and $e_1(X_1)$, the term on the right-hand side is $p(X_1)r^+(X_1)e_1(X_1)$, and from definition of $p(X_1)$ it simplifies to $r(X_1)e_1(X_1)$, which gives:

$$\mathbb{E}\left[Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)}\right] = \mathbb{E}\left[\frac{1}{e_1(X_1)}\mathbb{E}[\gamma_{X_1}|X_1]r(X_1)e_1(X_1)\right] \\ = \mathbb{E}[\gamma_{X_1}r(X_1)].$$

Applying Bayes rule for densities:

$$\mathbb{E}[\gamma_{X_1}r(X_1)] = \int_{\mathcal{X}_1}\gamma_x\Pr(R=1|X_1=x)f_x(x)dx \\ = \int_{\mathcal{X}_1}\gamma_x f_{x|R=1}(x)dx\Pr(R=1).$$

Since $\mathbb{E}[r(X_1)] = \Pr(R=1)$, the statement holds.

□

A2 Orthogonality

I demonstrate orthogonality of $\psi^{U+}(G,\xi^0)$ with respect to one of the propensity scores $e_l(x_l)$ for l s.t. $1<l<\overline{w}$. The arguments for other parameters involve first applying the Lemma to eliminate the dependence of the conditional expectation of the moment function on propensity scores $e_j(x_j)$ for $j>1$. Afterward, the steps are similar to those in Semenova

(2023). The approach for other moments follows a similar process.

Define $r \in [0,1) \to \psi^{U+}(G,r) \equiv \psi^{U+}(G,\xi_r)$, where:

$$\xi_r = \{e_1(x_1), \ldots, e_l(x_l, r), \ldots, e_{\overline{w}}(x_{\overline{w}}), r_1(x_1), \ldots, r_{\overline{w}}(x_{\overline{w}}), r^+(x_1), \\ q(p(x_1), x_1), q(1 - p(x_1), x_1)), \beta_1(x_1), \ldots, \beta_{\overline{w}}(x_{\overline{w}}), \beta^+(x_1), z^{U+}(x_1), z^{L+}(x_1)\}$$

and where $e_l(x_l, r) = e_l(x_l) + r(\widehat{e}_l(x_l) - e_l(x_l))$, with $\widehat{e}_j(x_j)$ such that $\widehat{e}_j(x_j) \neq e_j(x_j)$ for $x_j \in \mathcal{X}_j^1$. Note that since $e_j(x_j) = 0$ is known for $x_j \in \mathcal{X}_j \backslash \mathcal{X}_j^1$, it follows that $\widehat{e}_j(x_j) = e_j(x_j)$, implying that $e_l(x_l, r) = e_l(x_l)$ for such x_l.

I demonstrate that $\partial_r \mathbb{E}[\psi^{U+}(G,\xi_r)|X_l]|_{r=0} = 0$, a.s.. Since the moment does not depend r when $A < l$ (because $1_{\{A \geq l\}} = 0$ and because $e_l(X_l, r) = e_l(X_l)$ in such cases) it is sufficient to show that $\partial_r \mathbb{E}[\psi^{U+}(G,\xi_r)|X_l]|_{r=0} = 0$ for values of X_l s.t. $A \geq l$; the rest of the argument assumes X_l satisfies this condition.

For $k \geq l$ define $S_k \equiv \{1, \ldots, k\} \setminus \{l\}$. Using that $Z_j = 0, e_j(X_j) = 0 | A < l$ $\mathbb{E}[\psi^{U+}(G,\xi_r)|X_l]$ simplifies to:

$$\begin{aligned}
&\mathbb{E}[\psi^{U+}(G,\xi_r)|X_l] = \mathbb{E}\Bigg[- Y(0) R \Pi_{j \in S_{\overline{w}}} \frac{(1 - Z_j)}{(1 - e_j(X_j))} \frac{(1 - Z_l)}{(1 - e_l(X_l, r))} \\
&+ q(p(X_1), X_1)[\Pi_{j \in S_{\overline{w}}} \frac{(1 - Z_j)}{(1 - e_j(X_j))} \frac{(1 - Z_l)}{(1 - e_l(X_l, r))} (R - r_1(X_1))] \\
&+ \sum_{k=l+1}^{\overline{w}} 1_{\{A \geq k\}} \Pi_{j \in S_{k-1}} \frac{(1 - Z_j)}{(1 - e_j(X_j))} \frac{(1 - Z_l)}{(1 - e_l(X_l, r))} \frac{e_k(X_k) - Z_k}{1 - e_k(X_k)} [r_k(X_k) \beta_k(X_k) \\
&+ q(p(X_1), X_1)(r_1(X_1) - r_k(X_k))] \\
&+ 1_{\{A \geq k\}} \Pi_{j \in S_l} \frac{(1 - Z_j)}{(1 - e_j(X_j))} \frac{e_l(X_l, r) - Z_l}{1 - e_l(X_l, r)} [r_k(X_k) \beta_k(X_k) \\
&+ q(p(X_1), X_1)(r_1(X_1) - r_k(X_k))] \Bigg| X_l \Bigg].
\end{aligned}$$

Define:

$$\begin{aligned}
f_k^l(X_k) \equiv 1_{\{A \geq k\}} \Pi_{j \in S_{k-1}} &\frac{(1 - Z_j)}{(1 - e_j(X_j))} \frac{1 - Z_l}{(1 - e_l(X_l, r))} [r_k(X_k) \beta_k(X_k) \\
&+ q(p(X_1), X_1)(r_1(X_1) - r_k(X_k))].
\end{aligned}$$

For $k > l$:

$$\mathbb{E}\left[f_k^l(X_k)\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}\middle|X_l\right]=\mathbb{E}\left[\mathbb{E}\left[f_k^l(X_k)\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}\middle|X_k,X_l\right]\middle|X_l\right] \tag{17}$$

$$=\mathbb{E}\left[f_k^l(X_k)\,\mathbb{E}\left[\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}\middle|X_k\right]\middle|X_l\right] \tag{18}$$

$$=0, \tag{19}$$

where (17) holds by law of iterated expectations, (18) holds because X_k contains X_l, and (19) holds because by (3):

$$\mathbb{E}\left[\frac{e_k(X_k)-Z_k}{1-e_k(X_k)}\middle|X_k\right]=0.$$

Moreover, by (3):

$$\mathbb{E}\Bigg[-Y(0)R\Pi_{j\in S_{\overline{w}}}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{(1-Z_l)}{(1-e_l(X_l,r))}$$
$$+q(p(X_1),X_1)[\Pi_{j\in S_{\overline{w}}}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{(1-Z_l)}{(1-e_l(X_l,r))}(R-r_1(X_1))]\Bigg|X_l\Bigg]$$
$$=\mathbb{E}\Bigg[-Y(0)R\Pi_{j\in S_l}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{(1-Z_l)}{(1-e_l(X_l,r))}$$
$$+q(p(X_1),X_1)[\Pi_{j\in S_l}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{(1-Z_l)}{(1-e_l(X_l,r))}(R-r_1(X_1))]\Bigg|X_l\Bigg]$$
$$=\Pi_{j\in S_l}\frac{(1-Z_j)}{(1-e_j(X_j))}\,\mathbb{E}[-Y(0)R$$
$$+q(p(X_1),X_1)[(R-r_1(X_1))]|X_l]\,\mathbb{E}\left[\frac{(1-Z_l)}{(1-e_l(X_l,r))}\middle|X_l\right] \tag{20}$$
$$=\Pi_{j\in S_l}\frac{(1-Z_j)}{(1-e_j(X_j))}[-\beta_l(X_l)r_l(X_l)+q(p(X_1),X_1)(r_l(X_l)-r_1(X_1))]\,\mathbb{E}\left[\frac{(1-Z_l)}{(1-e_l(X_l,r))}\middle|X_l\right] \tag{21}$$
$$\equiv\mu_l(X_l)\,\mathbb{E}\left[\frac{(1-Z_l)}{(1-e_l(X_l,r))}\middle|X_l\right], \tag{22}$$

where (20) also holds by (3), and where (21) holds by applying (3) to the definitions of $\beta_l(.)$ and $r_l(.)$, after noting that we are consider values of X_l s.t. $1_{\{A>l\}}=1$.

Similarly, applying (3) gives:

$$\mathbb{E}\left[\Pi_{j\in S_l}\frac{(1-Z_j)}{(1-e_j(X_j))}\frac{e_l(X_l,r)-Z_l}{(1-e_l(X_l,r))}[r_l(X_l)\beta_l(X_l)+q(p(X_1),X_1)(r_1(X_1)-r_l(X_l))]\middle|X_l\right]$$
$$=\Pi_{j\in S_l}\frac{(1-Z_j)}{(1-e_j(X_j))}[r_l(X_l)\beta_l(X_l)+q(p(X_1),X_1)(r_1(X_1)-r_l(X_l))]\,\mathbb{E}\left[\frac{e_l(X_l,r)-Z_l}{(1-e_l(X_l,r))}\middle|X_l\right].$$

Combining the above, $\mathbb{E}[\psi^{U+}(G,\xi_r)|X_l]$ simplifies to:

$$\begin{aligned}\mathbb{E}[\psi^{U+}(G,\xi_r)|X_l] &= \mu_l(X_l)\left[\mathbb{E}\left[\frac{(1-Z_l)}{(1-e_l(X_l,r))}\middle|X_l\right]-\mathbb{E}\left[\frac{e_l(X_l,r)-Z_l}{(1-e_l(X_l,r))}\middle|X_l\right]\right]\\ &= \mu_l(X_l)\left[\frac{1-e_l(X_l)}{1-e_l(X_l,r)}-\frac{e_l(X_l,l)-e_l(X_l)}{1-e_l(X_l,r)}\right]\\ &= \mu_l(X_l),\end{aligned}$$

meaning that $\partial_r\,\mathbb{E}[\psi^{U+}(G,\xi_r)|X_l]|_{r=0}=0$ a.s.. Meanwhile, for the baseline moment:

$$\begin{aligned}\partial_r\,\mathbb{E}[m^U(G,\eta_r)|X_l]|_{r=0} &= \partial_r\mu_l(X_l)\frac{1-e_l(X_l)}{1-e_l(X_l,r)}\bigg|_{r=0}\\ &= \mu_l(X_l)\frac{1-e_l(X_l)}{(1-e_l(X_l,r))^2}(\widehat{e_l}(X_l)-e_l(X_l)).\end{aligned}$$

meaning that $\partial_r\,\mathbb{E}[m^U(G,\eta_r)|X_l]|_{r=0}\neq 0$ a.s..

A3 Relaxing Monotonicity Following Semenova (2023)

A challenge that may arise when implementing the bounds under a non-trivial monotonicity assumption is that, for some values of X_1, such as x_1^*, the estimated relier share is greater than the estimated subsequent relier share. This may occur because the monotonicity assumption is violated; however, since the two shares are estimated on different groups, this may also occur by chance. In either case, it results in the estimated trimming share $p(x_1^*)$ taking values above one, and since such values are not valid inputs for the quantile function $q(p(x_1^*),x_1^*)$, the bounds become ill-defined. To address the equivalent problem in the Lee (2009) setting, Semenova (2023) relaxes the monotonicity assumption, allowing the direction of monotonicity to vary with X_1. In my setting, this implies that all women with certain pre-ACP covariates who had a non-ACP child after ACP failure would have also had a non-ACP child if their first ACP had succeeded. This assumption is harder to justify from an economic perspective; one plausible interpretation could be that some families would like to have at least two children but would rather remain childless than have only one. Even if the assumption can be justified, implementing this approach using adapted

moments is complicated because they involve weighted quantile functions estimated on small groups of women who enter subsequent ACPs. Because of this, in my main specification, I maintain the original monotonicity assumption and treat cases where the estimated relier share exceeds the subsequent relier share as if the two were equal. If the reversal of the estimated shares occurs because the true shares are very close, treating them as equal or following Semenova (2023) should yield practically identical results. Under sequential unconfoundedness, the expectation of the moment for treating them as equal, $\psi^{-}(G, \xi^0)$, given in Table 1, identifies the difference between the conditional subsequent relier average treated outcome and the conditional relier average control outcome:

$$\frac{\mathbb{E}[\psi^{-}(G, \xi^0) \mid X_1 = x_1^*]}{\mathbb{E}[r(X_1) \mid X_1 = x_1^*]} = \mathbb{E}[Y(1) \mid R^+ = 1, X_1 = x_1^*] - \mathbb{E}[Y(0) \mid R = 1, X_1 = x_1^*].$$

When the shares of the two types are equal, monotonicity implies that the two groups are the same, and the difference between the two terms is $\mathbb{E}[\tau \mid R = 1, X_1 = x_1^*]$.

To test the sensitivity of my result, I allow for the direction of monotonicity to vary with covariates following Semenova (2023). Define $\mathcal{X}_{help} \equiv \{x : r^+(x) \geq r(x)\}$ and $\mathcal{X}_{hurt} \equiv \mathcal{X}_1 \setminus \mathcal{X}_{help}$. The relaxed monotonicity assumption is that $\forall x \in \mathcal{X}_{help}$ $R^+ \geq R$ *a.s.*, and $\forall x \in \mathcal{X}_{hurt}$ $R^+ < R$ *a.s.*. Table A5 describes the moments for the case when $X_1 \in \mathcal{X}_{hurt}$. The new estimator of the lower bound is:

$$\frac{\sum_i \left(\psi^{L+}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) \leq 1\}} + \psi^{L-}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) > 1\}}\right)}{\sum_i \left(\psi^{R}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) \leq 1\}} + \psi^{R+}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) > 1\}}\right)}.$$

The new estimator of the upper bound is:

$$\frac{\sum_i \left(\psi^{U+}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) \leq 1\}} + \psi^{U-}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) > 1\}}\right)}{\sum_i \left(\psi^{R}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) \leq 1\}} + \psi^{R+}(G_i, \widehat{\zeta}_i) 1_{\{p(X_1) > 1\}}\right)}.$$

I implement it following the baseline approach, but since a weighted generalized quantile forests estimator is not available, I estimate all nuisance functions involving expectations and quantiles using OLS and quantile regressions, respectively. Using regression only for the quantile function has little impact on the estimates. Figure A19 presents the results for women's outcomes. Overall, they are similar to the baseline estimates.

Table A5: Moment Functions for Covariate-Conditional Monotonicity

	Moment functions
$\psi_L^-(W,\zeta_0)$	$\frac{Z_1}{e_1(X_1)}(1-D^+)Y - \Pi_{j=1}^{\overline{w}}\frac{1-Z_j}{1-e_j(X_j)}(1-D^+)Y1_{\{Y>q^0(1-1/p(X_1),X_1)\}}$ $-q^0(1-1/p(X_1),X_1)\Big[\frac{Z_1}{e_1(X_1)}(1-D^+-r^+(X_1))$ $-\Pi_{j=1}^{\overline{w}}\frac{1-Z_j}{1-e_j(X_j)}\frac{1}{p(X_1)}(1-D^+-r_1(X_1))$ $-\Pi_{j=1}^{\overline{w}}\frac{1-Z_j}{1-e_j(X_j)}(1-D^+)(1_{\{Y>q^0(1-1/p(X_1),X_1)\}}-1/p(X_1))\Big]$ $-\frac{Z_1-e_1(X_1)}{e_1(X_1)}\beta^+(1,X_1)r^+(X_1)$ $+\sum_{k=1}^{\overline{w}}1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{1-D_j}{1-e_j(X_j)}\frac{e_k(X_k)-D_k}{1-e_k(X_k)}$ $\times\Big[\Big(r_k(X_1)r_k^L(X_k)z_k^{L-}(X_k)+\frac{q^0(1-1/p(X_1),X_1)}{p(X_1)}(r_1(X_1)-r_k(X_1))\Big)$ $+q^0(1-1/p(X_1),X_1)r_k(X_1)(1/p(X_1)-r_k^L(X_k))\Big]$
$\psi_U^-(W,\zeta_0)$	$\frac{Z_1}{e_1(X_1)}(1-D^+)Y - \Pi_{j=1}^{\overline{w}}\frac{1-Z_j}{1-e_j(X_j)}(1-D^+)Y1_{\{Y<q^0(1/p(X_1),X_1)\}}$ $-q^0(1/p(X_1),X_1)\Big[\frac{Z_1}{e_1(X_1)}(1-D^+-r^+(X_1))$ $-\Pi_{j=1}^{\overline{w}}\frac{1-Z_j}{1-e_j(X_j)}\frac{1}{p(X_1)}(1-D^+-r_1(X_1))$ $-\Pi_{j=1}^{\overline{w}}\frac{1-Z_j}{1-e_j(X_j)}(1-D^+)(1_{\{Y<q^0(1/p(X_1),X_1)\}}-1/p(X_1))\Big]$ $-\frac{Z_1-e_1(X_1)}{e_1(X_1)}\beta^+(1,X_1)r^+(X_1)$ $+\sum_{k=1}^{\overline{w}}1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{1-D_j}{1-e_j(X_j)}\frac{e_k(X_k)-D_k}{1-e_k(X_k)}$ $\times\Big[\Big(r_k(X_1)r_k^U(X_k)z_k^{U-}(X_k)+\frac{q^0(1/p(X_1),X_1)}{p(X_1)}(r_1(X_1)-r_k(X_1))\Big)$ $+q^0(1/p(X_1),X_1)r_k(X_1)(1/p(X_1)-r_k^U(X_k))\Big]$
$\psi^{R+}(G,\zeta^0)$	$r^+(X_1)+(1-D^+-r^+(X_1))\frac{Z_1}{e_1(X_1)}$
	Nuisance functions
$\zeta^0(x_1,\ldots,x_{\overline{w}})$	$\{e_1(x_1),\ldots,e_{\overline{w}}(x_{\overline{w}}),r_1(x_1),\ldots,r^{\overline{w}}(x_{\overline{w}}),r^+(x_1),q(p(x_1),x_1),q(1-p(x_1),x_1),$ $\beta^1(x_1),\ldots,\beta^{\overline{w}}(x_{\overline{w}}),\beta^+(x_1),z^{U+}(x_1),z^{L+}(x_1),z_1^{U-}(x_1),\ldots,z_{\overline{w}}^{U-}(x_{\overline{w}}),q^0(1/p(x_1),x_1),$ $q^0(1-1/p(x_1),x_1),z_1^{L-}(x_1),\ldots,z_{\overline{w}}^{L-}(x_{\overline{w}}),r_1^L(x_1),\ldots,r_{\overline{w}}^L(x_{\overline{w}}),r_1^U(x_1),\ldots,r_{\overline{w}}^U(x_{\overline{w}})\}$
$q^0(u,x)$	$inf\{q:u\leq\mathbb{E}[1_{\{Y\leq q\}}/\Pi_{j=2}^{\overline{w}}(1-e_j(X_j))\mid X_1=x,D=0]/$ $\mathbb{E}[\Pi_{j=2}^{\overline{w}}(1-e_j(X_j))\mid X_1=x,D=0]\}$
$z_k^{L-}(x)$	$\mathbb{E}[Y/\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert Y\geq q^0\left(1-1/p(X_1),X_1\right),D=0,X_k=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert Y\geq q^0\left(1-1/p(X_1),X_1\right),D=0,X_k=x]$
$z_k^{U-}(x)$	$\mathbb{E}[Y/\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert Y\leq q^0\left(1/p(X_1),X_1\right),D=0,X_k=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert Y\leq q^0\left(1/p(X_1),X_1\right),D=0,X_k=x]$
$r_k^L(x)$	$\mathbb{E}[1_{Y>q^0(1-1/p(X_1),X_1)}/\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert D=0,X_k=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert Y\leq q^0\left(1/p(X_1),X_1\right),D=0,X_k=x]$
$r_k^U(x)$	$\mathbb{E}[1_{Y<q^0(1/p(X_1),X_1)}/\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert D=0,X_k=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j(X_j)\vert Y\leq q^0\left(1/p(X_1),X_1\right),D=0,X_k=x]$

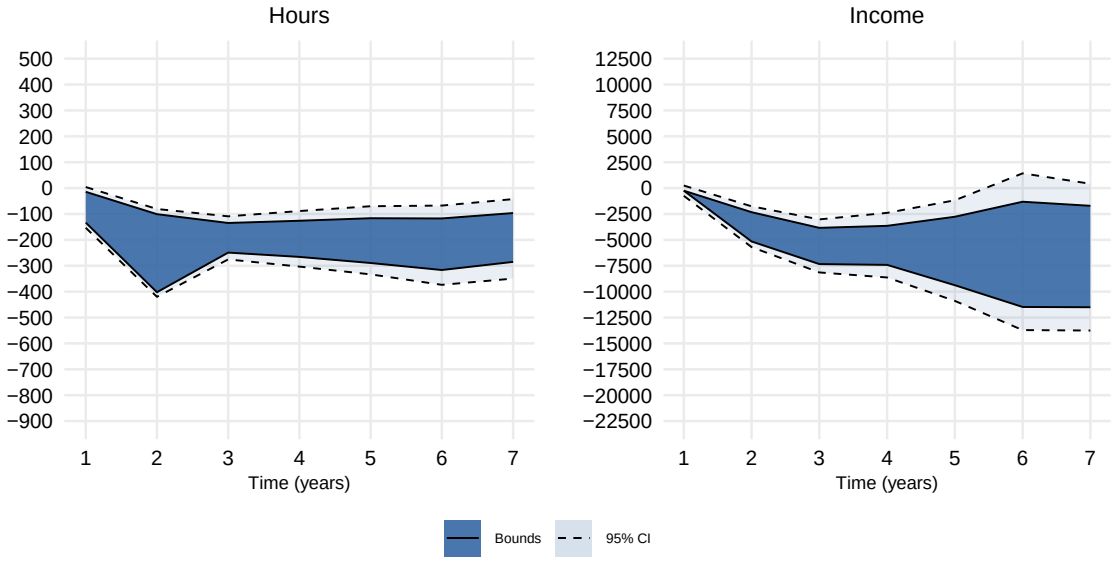

Figure A19: Effect on Women Under Relaxed Monotonicity Following Semenova (2023)

A4 Implementation Details

A4.1 Main Specification

I use 3-fold cross-fitting, meaning that in each sample split 2/3 of the observations are used to estimate the nuisance functions for the remaining observations. Because I assume that the propensity scores only include a few discrete covariates—age in years, dummy for higher education, and procedure type—they could be estimated non-parametrically using saturated fixed effects regressions. However, later propensity scores need to be estimated on small samples, and including many fixed effects makes them susceptible to outliers, this is especially undesirable because these scores are also used as weights to estimate other nuisance functions. Instead, in my main specification, I estimate them using logistic regressions. Specifically, for each ACP, I regress the outcome among women who entered that ACP on second-order polynomials of women's and partners' ages at the time of the procedure, interacted with treatment-type dummies (IUI or ACP), and separate dummies for each partner having at least a bachelor's degree.[32] To further avoid outlier weights, I only use the first 10 ACPs women undergo and treat conceptions through later ACPs as conceptions without ACPs; only 7% of women reach the tenth ACP. This means that, in my application, reliers are women who would remain childless in the scenario that their first 10 ACPs fail. Including up to 15 ACPs has little impact on my estimates. The remaining nuisance functions are estimated using Generalized Random Forests for conditional expectations and quantiles (Athey et al., 2019).[33] The covariates in X_1 include the woman's and their partner's income and work hours measured in the year before the woman's first ACP, and other covariates included in the first propensity score. The covariates in X_k additionally include those from the propensity scores at all ACPs up to and including ACP k. I modify work hours and income outcomes by adding a small amount of continuously distributed noise to ensure the new outcomes are continuous $u \sim U(0, 0.001)$.[34] Following Heiler (2024), my confidence intervals for the bounds are based on Stoye (2020).

Confidence intervals for the bounds on the effect scaled by the treated mean are also based on Stoye (2020), with covariance matrices obtained using delta method in these steps: (1) estimate $\widehat{\xi}_i$ using cross-fitting, (2) construct separate sample moments for the control mean and the upper and lower bounds for the treated mean evaluated at $\widehat{\xi}_i$ (m_1, m_2, and m_3, respectively), (3) compute the joint covariance matrix for the three sample moments, (4) obtain the joint covariance matrix for $(m_2 - m_1)/m_2$ and $(m_3 - m_1)/m_3$ using delta method.

[32] Using age-fixed effects and/or excluding the education dummies does not meaningfully impact the results.

[33] I estimate the truncated conditional expectation functions z_t^{U+} and z_t^{L+} by trimming data above or below the estimated quantiles and estimating conditional expectations. While this method may affect the asymptotic distribution of the bounds, simulations using data generated to approximate the distributions of the real data suggest such impacts are small.

[34] Lee (2009) procedure requires continuous outcomes only to avoid ties in the trimming procedure; adding a small amount of continuously distributed noise resolves this issue, allowing the effects to be appropriately bounded even in the case of discrete outcomes.

Placebo event study confidence intervals are obtained using Bayes bootstrap standard errors with weights $w_i \sim exp(1)$ and 150 draws.

A4.2 Extensions

A4.2.1 Instrumental Variable

I implement the IV following Lundborg et al. (2017), where the first stage specification is:

$$D_{it} = Z_{i1}\beta_t^{FS} + X_{i1}\chi_t^{FS} + \varepsilon_{it}^{FS},$$

and the second stage specification is:

$$Y_{it} = \widehat{D}_{it}\beta_t^{IV} + X_{i1}\chi_t^{IV} + \varepsilon_{it}^{IV},$$

where the parameters for the effect of parenthood in period t is β_t^{IV}.

A4.2.2 Even Study

I implement the ES following the fixed effect specification of Kleven et al. (2019):

$$Y_{it} = \beta_0^{ES} + \sum_{j\neq 0} \beta_j^{ES} 1_{\{t=j\}} + \sum_a \alpha_a 1_{\{age_{it}=a\}} + \sum_y \gamma_y 1_{\{year_{it}=y\}} + \upsilon_{it}, \tag{23}$$

where the parameters for the effect of parenthood in period t is β_t^{ES}. To make the ACP sample ES estimates as comparable to the IV estimates as possible, I restrict the sample to women whose first ACP succeeded.

A4.2.3 Placebo Event Study

I implement the placebo ES using (23) on a sample of women who remain childless 7 years after the first ACP with weights $w_i^w = 1/\Pi_{j=1}^{A_{i7}}(1 - e_j(X_{ji}))$. This ensures that β_t^{ES} corresponds to the bias that would arise in the ES approach due to selective fertility among reliers, which allows for comparison with τ_{ATR}(t). Standard errors are obtained using Bayes bootstrap in these steps: (1) draw weights $w_i \sim exp(1)$, (2) estimate $e_j(x_j)$ for all j with weights w_i to obtain an estimate of w_i^w, (3) estimate the placebo ES with estimated weights $w_i w_i^w$, (4) repeat steps 1-3 150 times to obtain a collection of bootstrap estimates (5) estimate the variance of the bootstrap estimates.

Share of gender inequality due to parenthood and selective fertility in year t estimated in the following steps: (1) construct separate sample moments for the control mean and the upper and lower bounds for the treated mean (a_1, a_2, and a_3, respectively), where Y is the female labor market outcome subtracted from the male labor market outcome in period t, (2) implement the placebo ES using the female labor market outcome and age, repeat it for the male labor market outcome and age, obtain the estimate for period t, a_4, by

subtracting the female estimate for period t from the male estimates estimate for period t, (3) construct the bounds $(a_2 - (a_1 + a_4))/a_2$ and $(a_3 - (a_1 + a_4))/a_3$. Confidence intervals are based on Stoye (2020), with covariance matrices obtained using Bayes bootstrap and delta method in these steps: (1) estimate $\widehat{\xi}_i$ using cross-fitting where Y is the female labor market outcome subtracted from the male labor market outcome, (2) draw weights $w_i \sim exp(1)$, (3) implement the placebo ES with weights w_i using the female labor market outcome and age, repeat it for the male labor market outcome and age, obtain the estimate a_4 for the difference between the male and the female estimates, (4) construct separate sample moments for the control mean and the upper and lower bounds for the treated mean evaluated at $\widehat{\xi}_i$ with weights w_i (a_1, a_2, and a_3, respectively), (5) repeat steps 2-4 150 times to obtain a collection of bootstrap estimates, (6) estimate the joint covariance matrix of a_1, a_2, a_3 and a_4, (7) obtain the joint covariance matrix for $(a_2 - (a_1 + a_4))/a_2$ and $(a_3 - (a_1 + a_4))/a_3$ using delta method.

A5 Using Continuous Covariates without Debiased Machine Learning

Here I introduce a new method to narrow the bounds by leveraging continuous covariates. For a known measurable function $g : \mathcal{X}_1 \to \mathbb{R}$, define $\varepsilon \equiv Y(1) - g(X_1)$. Intuitively, $g(X_1)$ can be thought of as OLS fitted values, and ε can be thought of as OLS residuals after regressing Y on X_1 among women whose first ACP succeeded. The idea behind the new approach is that the component of $\mathbb{E}[Y(1)|R = 1]$ explained by $g(.)$ can be identified. As a result, only the residual component needs to be bounded, and the distribution of ε can be tighter than the distribution of $Y(1)$, which results in narrower bounds. Formally, first, by definition, $\mathbb{E}[Y(1)|R = 1] = \mathbb{E}[g(X_1) + \varepsilon|R = 1]$. Second, since X_1 is observed, $\mathbb{E}[g(X_1)|R = 1]$ can be identified using women who remain childless similar to $\mathbb{E}[Y(0)|R = 1]$, specifically:

$$\mathbb{E}\left[g(X_1)\frac{(1-D)}{\prod_{j=1}^{\overline{w}}(1-e_j(X_j))}\right] \Big/ \mathbb{E}\left[\frac{(1-D)}{\prod_{j=1}^{\overline{w}}(1-e_j(X_j))}\right] = \mathbb{E}\left[g(X_1) \mid R = 1\right].$$

Since among women whose first ACP succeeds, $Y(1)$ and $g(X_1)$ are observed, ε is observed, meaning that $\mathbb{E}[\varepsilon|R = 1]$ can be bounded similar to how $\mathbb{E}[Y(1)|R = 1]$ is bounded using the baseline method without covariates. Then, it can be combined with the point-identified $\mathbb{E}[g(X_1)|R = 1]$ to obtain bounds on $\mathbb{E}[Y(1)|R = 0]$. As in the baseline method, combining point-identified $\mathbb{E}[Y(0)|R = 0]$ with the bounds on $\mathbb{E}[Y(1)|R = 0]$ gives bounds on τ_{ATR}. Note that theoretically, the bounds obtained using this approach need not be narrower and could even be wider than the baseline bounds that ignore covariates. To see this, consider a case where Y is constant. In this case, the baseline bounds collapse to a point, whereas the new bounds may not, since $g(X_1)$ need not be constant, meaning that ε is not constant either. In practice, however, the bounds can be substantially narrower than those that

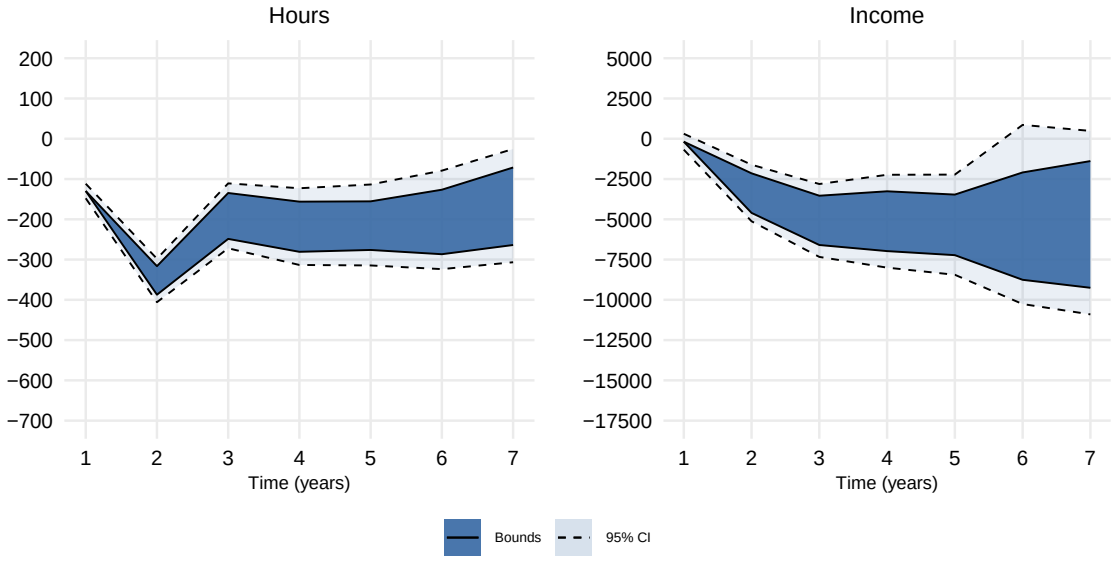

Figure A20: Effects on Women Using Reisualization Approach (Leave-Adjusted Hours)

do not leverage covariates, and in some cases, they can match the sharp bounds. Bounds based on different $g(.)$'s can also be compared empirically. This approach can also be used to leverage continuous covariates in the baseline Lee (2009) bounds and can also be combined with the method proposed by Lee (2009) to narrow the bounds using discrete covariates by estimating bounds for each discrete covariate cell before aggregating.

I implement the above approach in these steps: (1) estimate $e_j(x_j)$ for all j to obtain estimates of weights $w_i^w = Z_{1i}/e_1(X_{1i}) + (1-Z_{Ai})/\prod_{j=1}^{A_i}(1-e_j(X_{ji}))$, (2) estimate $g(x_1)$ by regressing Y on X_1 using women with $Z_1 = 1$ and estimates of weights w_i^w, (3) separately regress D on X_1 using women whose first ACP succeeded and women whose ACPs all failed, with estimates of weights w_i^w, (4) split the sample into quintiles based on differences in fitted values for the two regressions in step (3), (5) estimate bounds on the effect in each quintile using $Y - g(X_1)$ as the outcome with weights w_i^w, (6) aggregate across bins with weights proportional to the estimated relative share in each bin with estimates of weights w_i^w. Confidence intervals are based on Stoye (2020) with the covariance matrix obtained via Bayesian bootstrap with 150 draws and weights $w_i \sim \exp(1)$ used for step (1), and replacing w_i^w with $w_i w_i^w$ for other steps. Figure A20 presents the results for women's outcomes. Overall, are very close to the baseline estimates.

A6 Accounting for Age Difference Between Partners

My main estimates of the share of gender inequality caused by parenthood focus on the within-couple gender gap for each year after becoming parents. This gap also captures differences related to the within-couple age gap, which may distort the picture of aggregate

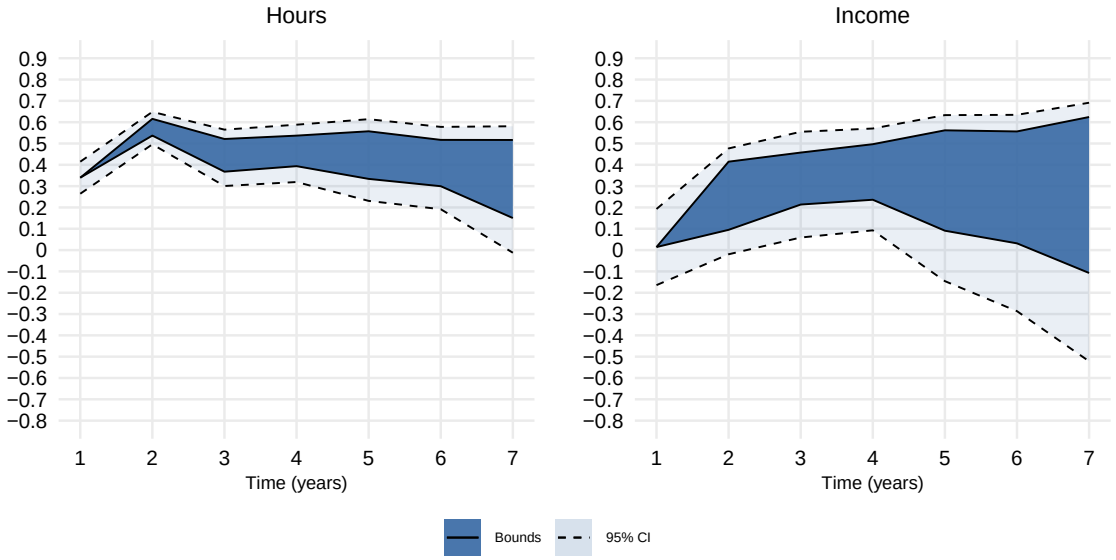

Figure A21: Share of Gender Inequality Caused by Parenthood Using Partner's Income Lagged to Match Woman's Age (Leave-Adjusted Hours)

gender inequality in the economy because men's outcomes are measured at systematically older ages. A particular concern is that if work hours and income increase with age, my estimates might understate the share of aggregate gender inequality caused by parenthood.

Ideally, using cumulative lifetime outcomes would directly address this issue, but since such data is unavailable, a different approach is required. One way to address this would be to correct for age differences parametrically, but this would require strong and potentially opaque assumptions. Instead, I opt for a simple approach that fits elegantly into my framework: I adjust the timing of when men's outcomes are measured based on the woman's age. For example, if a woman is two years younger than her male partner, I lag the male's outcome in each period by two years. This ensures that gender gaps in outcomes within couples are measured at the same point in their life cycle. The adjustment reduces my sample by 22%, as it excludes couples where the male partner is much older or younger, leaving me with 12,146 observations.

Figure A21 presents the results. The adjustment has little impact on the estimates; if anything, the upper bound for the share of gender inequality caused by parenthood in hours decreases, while that in income increases by no more than 10 percentage points.

A7 Confidence Intervals for Different Methods

While my method only partially identifies the effects, my estimates are substantially more precise. Figure A22 presents the width of 95% confidence intervals for my bounds, IV estimates, and ES estimates. The latter two methods are implemented as described in

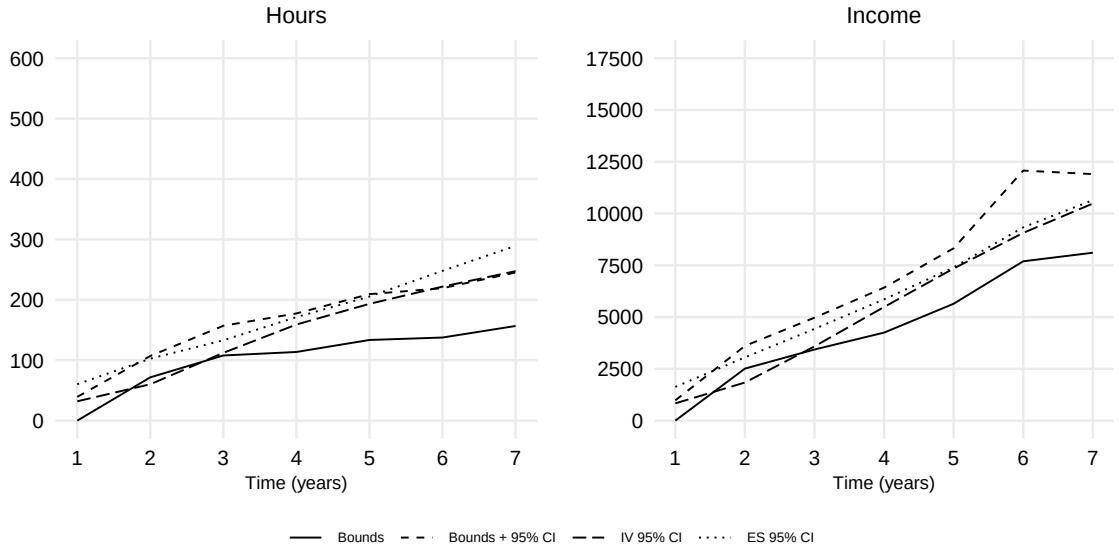

Figure A22: 95% Confidence Interval Width for Different Methods

Section A4.2. The confidence intervals for the three methods are almost identical. Most of the uncertainty in my estimates arises from identifying bounds rather than sampling variability in the estimation process. Similarly, the method introduced in Section A8, used to estimate τ_{ATR} under the assumption of static effects, provides a more precise alternative to the IV method. Intuitively, this improvement occurs because much of the uncertainty around IV estimates stems from scaling the reduced form by a low first stage. Leveraging women's complete ACP histories improves the first stage by expanding it from compliers to reliers, thereby reducing the amplification of noise.

A8 Testing Assumptions for Parametric Bias Correction

In this section, I introduce an estimator of $\tau_{ATR}(t)$ that parallels the Wald estimator of $\tau_{LATE}(t)$, as it identifies a linear combination of a group average treatment effect and a group average effect of delaying parenthood. I then demonstrate how each of the two estimators can be used to identify $\tau_{ATE}(t)$ under the parametric assumptions used by Bensnes et al. (2023) and Gallen et al. (2023). Since the assumptions imply that both methods should yield identical results, differing results allow for the rejection of these parametric assumptions.

I first introduce several functions that will be used to estimate τ_{LATE} and τ_{ATR} under

the assumption of of static effects:

$$
\begin{aligned}
g_a^{0+}(G) &= \gamma_a^{0,1+}(X_1) + (a - \gamma^{0,1+}(X_1))\Pi_{j=1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j(X_j))} \\
&\quad + \Sigma_{k=1}^{\overline{w}} \Big[1_{\{A \geq k\}} \Pi_{j=1}^{k-1} \frac{(1-Z_j)}{1-e_j(X_j)} \frac{(e_k(X_k) - Z_k)}{1-e_k(X_k)} [\gamma_a^{0,1+}(X_1) - \gamma_a^{0,k+}(X_k)] \Big] \\
g_a^0(G) &= \gamma_a^0(X_1) + (a - \gamma_a^0(X_1)) \frac{Z_1}{e_1(X_1)} \\
g_a^1(G) &= \gamma_a^1(X_1) + (a - \gamma_a^1(X_1)) \frac{1-Z_1}{1-e_1(X_1)},
\end{aligned}
$$

where $\gamma_a^1(X_1)$ is the OLS prediction of a given X_1 among observations with $Z_1 = 1$ with weights $1/e_1(X_1)$, $\gamma_a^0(X_1)$ is the OLS prediction of a given X_1 among observations with $Z_1 = 0$ with weights $1/(1-e_1(X_1))$, $\gamma_a^{0,k+}(X_1)$ is the OLS prediction of a at X_k given X_k among observations with $Z_1 = 0, A \geq k$ with weights $1/(\prod_{j=1}^{A}(1-e_j(X_j)))$.

$\mathbb{E}[g_{Y_t}^1(G) - g_{Y_t}^0(G)] / \mathbb{E}[g_{D_t}^1(G) - g_{D_t}^0(G)]$ corresponds to a Wald estimator of $\tau_{LATE}(t)$ where the reduced form and the first stage are both implemented in a doubly-robust manner to maximize precision. $\mathbb{E}[g_{Y_t}^1(G) - g_{Y_t}^{0+}(G)] / \mathbb{E}[g_{D_t}^1(G) - g_{D_t}^{0+}(G)]$ corresponds to the Wald-like estimator of $\tau_{ATR}(t)$, where the reduced form and the first stage are also implemented in a doubly-robust manner to maximize precision.

Following standard argument under assumptions 5,6,7, 10, and 4 for all t, gives:

$$
\frac{\mathbb{E}[g_{Y_1}^1(G) - g_{Y_1}^0(G)]}{\mathbb{E}[g_{D_1}^1(G) - g_{D_1}^0(G)]} = \tau_{LATE}(1),
$$

and similarly, using the Lemma and the standard argument gives:

$$
\frac{\mathbb{E}[g_{Y_1}^1(G) - g_{Y_1}^{0+}(G)]}{\mathbb{E}[g_{D_1}^1(G) - g_{D_1}^{0+}(G)]} = \tau_{ATR}(1).
$$

In periods after the first, both approaches may be biased due to dynamic effects, specifically in the second period:

$$
\begin{aligned}
\frac{\mathbb{E}[g_{Y_2}^1(G) - g_{Y_2}^0(G)]}{\mathbb{E}[g_{D_2}^1(G) - g_{D_2}^0(G)]} &= \tau_{LATE}(2) + \frac{\Pr(C_2 = 0, C_1 = 1)}{\Pr(C_2 = 1)} \mathbb{E}[Y_2(1) - Y_2(2) | C_2 = 0, C_1 = 1] \\
\frac{\mathbb{E}[g_{Y_2}^1(G) - g_{Y_2}^{0+}(G)]}{\mathbb{E}[g_{D_2}^1(G) - g_{D_2}^{0+}(G)]} &= \tau_{ATR}(2) + \frac{\Pr(R_2 = 0, R_1 = 1)}{\Pr(R_1 = 1)} \mathbb{E}[Y_2(1) - Y_2(2) | R_2 = 0, R_1 = 1].
\end{aligned}
$$

To correct the bias in the IV estimates Bensnes et al. (2023) and Gallen et al. (2023) assume:

Assumption 10 (Parametric Effects)**.**

1. $Y_t(1) - Y_t(0) = \tau_{ATE}(t)$,
2. $Y_t(k) - Y_t(0) = \tau_{ATE}(1 + t - k)$ *for all* $k \leq t$.

The first part restricts heterogeneity across individuals while the second part means that the effects only depend on time spent in parenthood but not the moment of becoming a parent. Under the two assumptions, the parameter identified by the Wald estimator in the second period simplifies to:

$$\frac{\mathbb{E}[g^1_{Y_2}(G) - g^0_{Y_2}(G)]}{\mathbb{E}[g^1_{D_2}(G) - g^0_{D_2}(G)]} = \tau_{ATE}(2) + \frac{\Pr(C_2 = 0, C_1 = 1)}{\Pr(C_1 = 1)}(\tau_{ATE}(2) - \tau_{ATE}(1))$$

Since under assumption 10 $\tau_{ATE}(1) = \tau_{LATE}(1)$, and since $\tau_{LATE}(1)$, $\Pr(C_2 = 0, C_1 = 1)$, and $\Pr(C_2 = 1)$ are identified, $\tau_{ATE}(2)$ can be backed out. Following similar reasoning for subsequent periods allows to back out $\tau_{ATE}(t)$ for all t.

My test for assumption 10 uses the fact that $\tau_{ATE}(t)$ can also be backed out using the Wald-like estimates of $\tau_{ATR}(t)$, and that when the assumptions 10 holds, the two approaches should give similar results. To ease exposition, define the pseudo-outcome:

$$\widehat{Y^l_t} = \begin{cases} Y_t, \text{ if } D_1 = 1 \text{ or } D_t = 0, \\ Y_t - \tau^l(k), \text{ otherwise, where } k = 1 + t - (min\{j : D_j = 1\}) \ , \end{cases}$$

for $l \in \{C, R\}$, where:

$$\tau^C(t) = \frac{\mathbb{E}[g^1_{\widehat{Y_t}}(G) - g^0_{\widehat{Y_t}}(G)]}{\mathbb{E}[g^1_{D_1}(G) - g^0_{D_1}(G)]},$$

and

$$\tau^R(t) = \frac{\mathbb{E}[g^1_{\widehat{Y_t}}(G) - g^{0+}_{\widehat{Y_t}}(G)]}{\mathbb{E}[g^1_{D_1}(G) - g^{0+}_{D_1}(G)]}.$$

For women who become mothers in later periods, the pseudo-outcome is the realized outcome adjusted by subtracting the effect of being a mother for their motherhood duration, which is identified in previous periods. Under assumptions 10, the pseudo-outcome equals their control outcome. $\tau^C(t)$ corresponds to how $\tau_{ATE}(t)$ is identified using the Gallen et al. (2023) method based on $\tau^{LATE}(t)$. $\tau^R(t)$ corresponds to how it can be identified using $\tau_{ATR}(t)$. Under assumptions 5,6,7, 10, and 4 for all t, $\tau^R(t) = \tau^C(t)$ for all t; if the two are not equal, at least one of the assumptions must be violated. Note that the only additional assumption that I require relative to Bensnes et al. (2023) and Gallen et al. (2023) is that the outcomes of subsequent ACPs are as good as random, conditional on observables.

Figure A23 presents the results for women's outcomes. Confidence intervals for the different between the estimates in each period are obtained using Bayes bootstrap with weights $w_i \sim exp(1)$ and 150 draws, where all parameters are estimated sequentially in each draw. Estimates of $\tau^C(t)$ suggest a substantially smaller career cost of motherhood than $\tau^R(t)$, which indicates that the parametric effects assumption is violated.

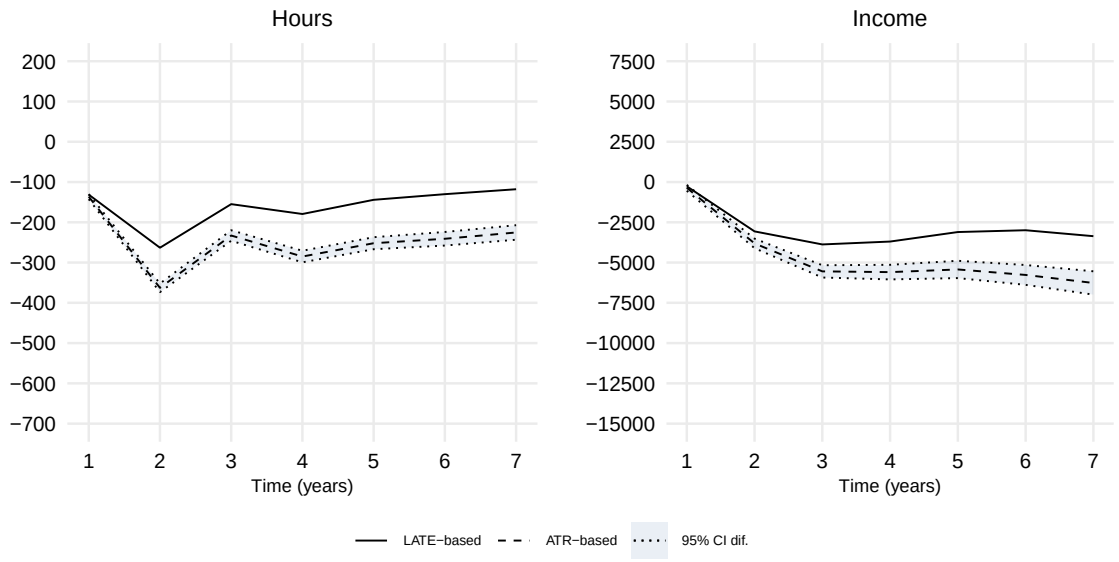

Figure A23: Estimates Using Parametric Bias Correction

A9 Mental Health, Relationship Stability, and Monotonicity

Here I estimate the effects on antidepressant uptake, discuss empirical support for the monotonicity assumption, and present bounds that address mental health and separation separately.

To maximize precision in estimating the impact on antidepressant uptake, I use the method described in section A8 with target parameter:

$$\frac{\mathbb{E}[g^1_{Y_1}(G) - g^{0+}_{Y_1}(G)]}{\mathbb{E}[g^1_{D_1}(G) - g^{0+}_{D_1}(G)]},$$

where the outcome is taking antidepressants in a given year. In the absence of dynamic effects, it identifies τ_{ATR}. Figure A24 presents the results, the effects are precisely estimated and indistinguishable from zero. Estimates based on an IV estimator are also not statically different from zero. Note that focusing on the reduced form effects would only make the estimates closer to zero.

Next, I discuss empirical support for the monotonicity assumption. Since it states that reliers are subsequent reliers with probability one ($\Pr(R^+ \geq R) = 1$), it implies that the relier share is at least as large as the subsequent relier share, $\mathbb{E}[r^+(X_1)] \geq \mathbb{E}[r(X_1)]$. Moreover, it implies that the subsequent relier share at each covariate value is at least as large as the relier share at that value, $r^+(X_1) \geq r(X_1)$. Since $r^+(X_1)$ and $r(X_1)$ can be estimated, comparing them allows to evaluate potential monotonicity violations.

The top left graph in Figure A25 plots the empirical distribution of the difference between the estimated conditional subsequent relier and the estimated relier shares in year 7.

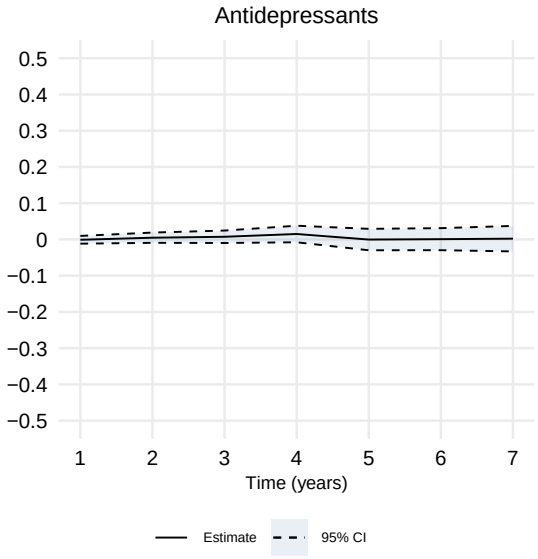

Figure A24: Effects on Antidepressant Uptake

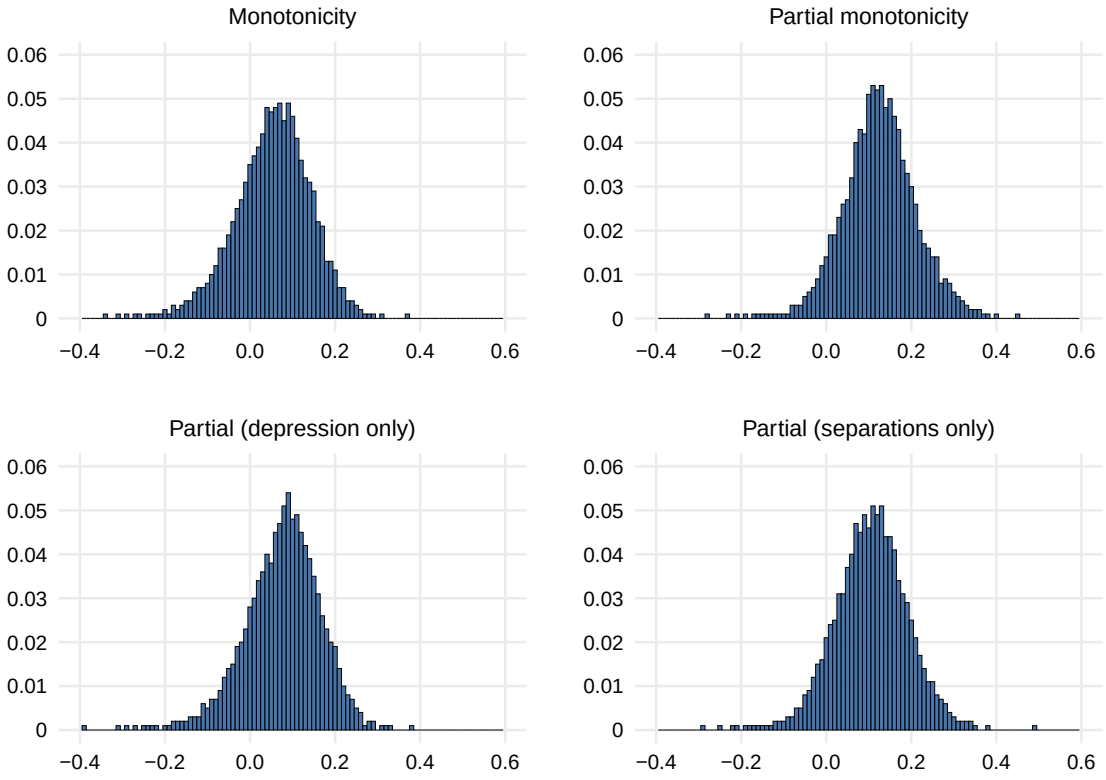

Figure A25: Histogram of estimated $r^+(X_1) - r(X_1)$

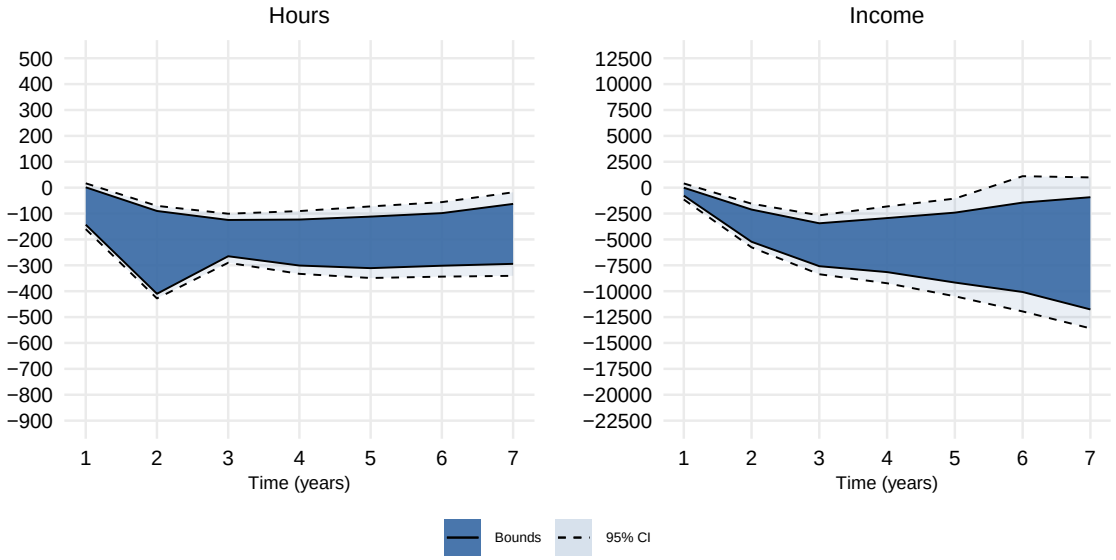

Figure A26: Effects on Resilient Women (Depression Only)

For 25% of observations, the difference is smaller than zero. While this does not support the monotonicity assumption, such inconsistencies may be due to estimation error in the two functions. Consistent with this explanation, in most of these cases, the difference between the two shares is very close to zero. For only 5% of the observations, the estimated difference is below -0.1, suggesting that violations of the monotonicity assumption are not readily apparent.

The right graph in Figure A25 repeats this for the relaxed partial monotonicity assumption introduced in section 7.3, which allows for monotonicity violations among women who would separate from their partners or uptake antidepressants after ACP failure. The estimated difference between the two shares is below zero for only 5% of observations, and it is below -0.1 for only 1% of observations, suggesting even stronger support for the partial monotonicity assumption. The two panels in the second row of Figure A25 repeat this when monotonicity is allowed to fail among women who would uptake antidepressants or separate from their partner separately, and the results remain similar. The equivalent results from earlier years are only more favorable for the monotonicity assumption.

Finally, Figures A26 and A27 present estimated labor market impacts of parenthood for reliers who would not uptake antidepressants after ACP failure and who would remain with their original ACP partner, respectively. Overall, the estimates are close to the baseline.

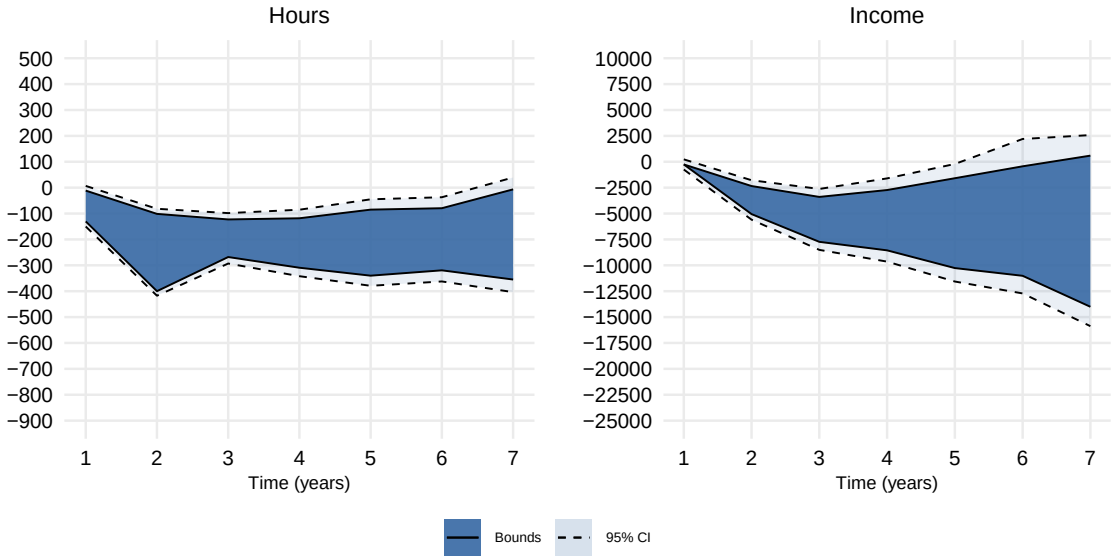

Figure A27: Effects on Resilient Women (Separation Only)

A10 Heterogeneity by Willingness to Undergo ACPs

$\mathbb{E}[Y(0)|R_t = 1, W_1 \geq k]$ can be identified following similar steps to the theorem. For brevity, I present how $\mathbb{E}[Y(1)|R_t = 1, W_1 \geq k]$ can be bounded in a setting with unconditional dynamic sequential unconfoundedness. Extending it to a setting with conditional dynamic sequential unconfoundedness amounts to weighting observations to account for differences in the likelihood of undergoing at least k failed ACPs.

The key feature enabling this analysis is that the dynamic model imposes that if women become mothers in the first period through subsequent ACPs, their treated outcomes are realized: $Y = Y(1) \mid A_1 = k, Z_k = 1$. This allows for a group of women whose treated outcomes are realized after undergoing multiple ACPs, simultaneously revealing information about their willingness to undergo ACPs and their treated outcomes.

An important nuance for narrowing the bounds by leveraging monotonicity in this analysis concerns the definition of non-ACP fertility, D^+. The original definition does not distinguish whether a woman would have a non-ACP child after her first ACP succeeded or if a subsequent ACP succeeded. This distinction does not matter for the baseline analysis, as only the former scenario is considered. Here, however, the scenarios in which women enter motherhood after subsequent ACPs are also considered. These different scenarios can be distinguished by defining *k-reliance*, which describes whether a woman would conceive any non-ACP children if she had her first child through her k'th ACP. Then, monotonicity should hold between reliance and k-reliance, similar to the relationship between reliance and subsequent reliance (or 1-reliance). For brevity, I preserve the original definition of D^+.

Formally:

$$\mathbb{E}[1_{\{Y_t \leq q\}} | Z_k = 1, A_1 = k, D^+ = 0] = \mathbb{E}[1_{\{Y_t(1) \leq q\}} | Z_k = 1, A_1 \geq k, R_t^+ = 1] \tag{24}$$

$$= \mathbb{E}[1_{\{Y_t(1) \leq q\}} | A_1 \geq k, W_1 \geq k, R_t^+ = 1] \tag{25}$$

$$= \mathbb{E}[1_{\{Y_t(1) \leq q\}} | Z_{k-1} = 0, A_1 \geq k-1, W_1 \geq k, R_t^+ = 1] \tag{26}$$

$$= \mathbb{E}[1_{\{Y_t(1) \leq q\}} | A_1 \geq k-1, W_1 \geq k, R_t^+ = 1] \tag{27}$$

$$= \mathbb{E}[1_{\{Y_t(1) \leq q\}} | W_1 \geq k, R_t^+ = 1] \tag{28}$$

$$= \Pr(Y_t(1) \leq q | W_1 \geq k, R_t^+ = 1), \tag{29}$$

where (24) holds because $D_1 = 1 | A_1 = k, Z_k = 1$, which implies $Y = Y(1) | A_1 = k, Z_k = 1$, and because $D_t^+ = 1 - R_t^+ | Z_k = 1, A_1 \geq k$. (25) holds by dynamic sequential unconfoundedness and because $W_1 \geq k | A_k \geq k$. (26) holds because $1_{\{A_1 \geq k\}} = 1_{\{A \geq k-1, Z_{k-1}=0\}} | W_1 \geq k$. (27) holds by dynamic sequential unconfoundedness, (28) is obtained by iteratively applying steps similar to (25) through (27), and (29) holds by definition. Following similar arguments, $\Pr(R_t = 1 | W_1 \geq k)$ and $\Pr(R_t^+ = 1 | W_1 \geq k)$ can be identified, which is sufficient to bound $\mathbb{E}[\tau | R_t = 1, W_1 \geq k]$.

Since under conditional sequential unconfoundedness, these bounds involve weighted quantile functions, I implement them in the following steps: (1) estimate $e_j(x_j)$ for $j \leq k$, (2) obtain bounds using the method described in Section A5 and the subsample of women who underwent $k+1$ ACPs in the first period, treating their $(k+1)$'th ACP as the first and implementing every step with estimates of weights $1/\Pi_{j=1}^{k}(1 - e_j(X_j))$.

A11 Imputation for Non-ACP Families

Consider a population consisting of women who conceive their first child without ACPs and women who undergo ACPs for their first child. Let 1^{REP} be an indicator that takes the value 1 if a woman belongs to the non-ACP group, and 0 otherwise. For non-ACP women, $t = 1$ represents the moment of conception for their first child, and their treated outcomes from this point onward are observed. The ES approach imputes childless outcomes for women in period t with covariates x_t^*, which include age and calendar year, using average control outcomes among women with the same covariates in period 0, $X_0^* = x_t^*$:

$$g_t^E(x_t) = \mathbb{E}\left[Y_0(0) \mid 1^{REP} = 1, X_0^* = x_t^*\right].$$

Instead, I propose to impute childless outcomes for women in period t with covariates x_t^* using average control outcomes in period t among women who remain reliers until the last

period and have with similar covariates in period t, $X_t^* = x_t^*$:[35]

$$g_t^I(x_t) = \mathbb{E}\left[Y_t(0) \mid R_T = 1, X_t^* = x_t^*\right].$$

The two covariates I leverage are age and pre-parenthood or pre-ACP education.

I estimate $g_t^I(x_t)$ by regressing Y_t on dummies for age in period t and pre-ACP higher education dummy using a sample of women who remain childless 7 years after the first ACP with weights $w_i^w = 1/\Pi_{j=1}^{A_{i7}}(1 - e_j(X_{ji}))$. To avoid empty bins, I group women aged 40 or older into a single category, as fewer than 5% of women are 40 or older in the year before their pre-ACP. To ensure the ES estimates are comparable, I apply the same approach; this has little impact, as few mothers in the non-ACP group reach this age within the sample period. Standard errors for the difference from the ES estimates obtained using Bayes bootstrap in the following steps: (1) draw weights $w_i \sim exp(1)$, (2) estimate $e_j(x_j)$ for all j with weights w_i to obtain an estimate of w_i^w, (3) estimate $g_t^I(x_t)$ with estimated weights $w_i w_i^w$, (4) estimate the ES with weights w_i, (5) compute sample average of ES predictions and $g_t^I(X_t)$ in the non-ACP population with weights w_i and take the difference between the two (6) repeat steps 1-5 150 times to obtain a collection of bootstrap estimates (7) estimate covariance matrix of the bootstrap estimates.

[35] Assuming covariates X_t^* are deterministic conditional on X_1^* (such as age and education in period 0), $g_t^I(X_t^*)$ can be identified using my baseline method. For both methods, the average counterfactual career trajectories for non-ACP families are obtained by evaluating the expectation of the imputation function in the non-ACP population: $\mathbb{E}[g_t^I(X_t^*)|1^{REP} = 1]$ and $\mathbb{E}[g_t^E(X_t^*)|1^{REP} = 1]$. The formal identification assumption for my approach is that reliers have similar control career trajectories to women in the representative sample, conditional on pre-parenthood or pre-ACP covariates $\mathbb{E}[Y_t(0)|R_T = 1, X_0^*] = \mathbb{E}[Y_t(0)|1^{REP} = 1, X_0^*]$ for all t.