PARENTHOOD TIMING AND GENDER INEQUALITY

JULIUS ILCIUKAS*

Amsterdam School of Economics, University of Amsterdam

July 9, 2025

Abstract

This paper develops a new methodology to estimate the labor market effects of parenthood, using quasi-experimental variation from intrauterine insemination procedures. Conventional instrumental variable estimates may be biased because childbirth often follows failed procedures, meaning that success may affect outcomes through both the incidence and timing of parenthood. I leverage women's complete procedure histories to address this bias and estimate effects for a broader, policy-relevant population. Using Dutch administrative data, I find that parenthood leads to persistent reductions in women's earnings (10–28%) and work hours (10–22%), causing up to half of the within-couple gender inequality observed after childbirth.

Keywords: parenthood, gender inequality, treatment effects, dynamic non-compliance

*I thank Jérôme Adda, Francesco Agostinelli, Douglas Almond, Monique de Haan, Christian Dustmann, Phillip Heiler, Christine Ho, Artūras Juodis, Jura Liaukonyte, 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, the Berlin School of Economics Gender Workshop, COMPIE, EALE, EEA-ESEM, ESPE, the Luxembourg Gender and Economics Workshop, SEHO, the Warwick Economics PhD Conference, and seminar participants at the Chinese University of Hong Kong, Monash University, Peking University HSBC Business School, Singapore Management University, Tilburg University, the University of Amsterdam, the University of Lausanne, the University of Michigan, and the University of Pennsylvania. The data used in this paper is available through the Microdata services of Statistics Netherlands. First version: January 27, 2024. All URLs accessed July 9, 2025.

1 Introduction

The differential impact of parenthood on the careers of women and men is widely regarded as a key driver of gender disparities in the labor market (Goldin, 2014; Bertrand, 2020; Cortés & Pan, 2023). Quantifying this impact is essential for understanding gender inequality and informing policy. Yet doing so is challenging due to selection: individuals who become parents may differ systematically from those who remain childless in their labor market outcomes, independent of parenthood itself.

To address selection, researchers have used quasi-experimental variation in fertility from the success of assisted conception procedures, such as in vitro fertilization (IVF) (Lundborg et al., 2017; Bensnes et al., 2023; Gallen et al., 2023; Lundborg et al., 2024) or from the failure of long-acting contraceptives, such as intrauterine devices (IUDs) (Gallen et al., 2023). In the IVF setting, plausibly exogenous variation arises from whether a woman conceives during her first procedure; in the IUD setting, from unintended pregnancies among women who had recently received the device.

Yet leveraging such variation is challenging. Researchers aim to compare women who conceive for quasi-random reasons with otherwise similar women who do not. However, many women who do not conceive during their first IVF procedure or shortly after receiving an IUD become mothers later, meaning that the initial quasi-experimental variation affects both the likelihood and timing of parenthood. Existing studies address this complication by assuming that labor market effects do not vary with timing—such as whether motherhood begins earlier or later in a woman's career, or how long she has been a mother. These assumptions are difficult to justify and may result in biased estimates if violated.

This paper contributes methodologically and empirically. I develop a general approach to estimate treatment effects in quasi-experimental settings with dynamic non-compliance, where individuals may receive treatment (e.g., become parents) through repeated quasi-experimental assignments (e.g., multiple IVF cycles or intercourse with an IUD in place) or through non-random pathways (e.g., adoption or natural conception following failed IVF or IUD removal). I apply this method to estimate the career effects of parenthood using variation from intrauterine insemination (IUI)—a common, minimally-invasive, and low-cost assisted conception procedure not previously used for this purpose.

The method compares outcomes between individuals who receive treatment during their first assignment and those who remain untreated thereafter, thereby eliminating differences in treatment timing. A key concern is that the untreated group may be selective, as some individuals pursue additional assignments or receive treatment through non-random pathways. I first show that a weighting scheme—assigning greater weight to individuals with more unsuccessful assignments—can address selection into treatment through further assignments. I then apply a bounding procedure to account for treatment obtained without assignment, assuming the most extreme selection consistent with the data.

The approach requires no assumptions about treatment effect heterogeneity over time or across individuals. Its core assumption is that each assignment outcome is as good as random, conditional

on observables—an assumption imposed in methods using the first assignment but extended here to repeated assignments. The identified bounds are sharp, meaning no effect within them can be ruled out without further assumptions or data. Crucially, the bounds need not contain point estimates obtained by restricting timing-dependent effects. Instead, they quantify not only potential bias but actual bias when such estimates fall outside the bounds.

I apply the method to a novel Dutch administrative dataset linking labor market information from tax records with hospital records. I focus on couples trying to conceive their first child through IUI, also known as artificial insemination. I find that parenthood persistently reduces women's annual work hours by 10%–22% and earnings by 10%–28%, with effects lasting at least six years after first birth. For men, the bounds are similar in length but centered near zero. Over this period, parenthood causes 32%–53% of the within-couple gender gap in work hours and 10%–45% in earnings after childbirth. I demonstrate that existing methods either understate or overstate the effects, depending on which assumptions about timing-dependent effects they impose.

In addition to potential bias, existing evidence based on IVF and IUD failures may suffer from limited generalizability. First, IVF and IUD samples are drawn from selective populations: IVF is pursued by women willing to undergo invasive treatment to conceive, while IUD failures occur among those who chose an invasive method to prevent pregnancy. Second, instrumental variable methods used in earlier studies identify effects for narrow subpopulations, such as women whose fertility depends entirely on the outcome of the first IVF procedure. Third, prior studies rely exclusively on data from Scandinavian countries with exceptionally generous family policies, limiting the applicability of findings to broader institutional contexts.

This study complements existing evidence and mitigates key concerns about generalizability. First, by focusing on IUI, a minimally-invasive procedure typically used as the first-line treatment for infertility, I provide evidence from a population that is arguably less selective than those examined in prior studies. A modest 4.6% of first-time Dutch mothers in 2017 had undergone IUI. However, because eligibility typically requires a year of unsuccessful attempts to conceive naturally, and infertility is usually unpredictable, IUI users are likely representative of a much larger group of mothers who would have pursued IUI had natural conception not occurred first.[1] I demonstrate that, prior to childbirth, IUI couples have similar labor market outcomes as those in the general population. Second, by leveraging full procedure histories, my method quantifies effects for a broader group than just those whose fertility depends on the first procedure. This includes women whose fertility depends on the entire sequence of procedures, raising the relevant sample share from 20% to 43%. Third, by using data from the Netherlands, where parental leave duration and childcare use are close to the OECD averages, I provide evidence that is arguably more relevant for typical institutional settings.

[1]Most women begin treatment around age 32, implying they have been trying to conceive since at least age 31. Given that the probability of conceiving naturally within a year at age 31 is nearly 80%, for every woman who undergoes IUI at 32, at least four others likely conceived naturally before becoming eligible. This suggests that the 4.6% who use IUI represent at least 23% of first-time mothers. Results are similar when using complete age profiles of women who undergo IUI.

Another factor that may limit generalizability and complicate interpretation is the physical and psychological side effects associated with IVF and IUDs (see Bögl et al. (2024) for a discussion of IVF side effects). The limited invasiveness of IUI helps mitigate concerns about adverse effects directly caused by the procedure. However, certain mental health consequences may be inseparable from the effects of parenthood, as the relevant comparisons involve either failing to achieve a wanted pregnancy (as with failed IVF or IUI) or experiencing an unwanted one (as with IUD failure). To assess their relevance, I examine antidepressant uptake and check for sharp declines in labor market outcomes following failed conception attempts. I also apply my bounding method to assess effects specifically for women who do not initiate antidepressant use after failing to conceive. Taken together, these analyses suggest that, in the context of IUI, mental health effects are unlikely to drive the observed labor market impacts. Nonetheless, my estimates may constitute a lower bound on the effects of parenthood relative to voluntary childlessness, which avoids the potential negative consequences of failed conception.

In contrast to IVF-based studies in Scandinavian contexts—which report that motherhood reduces earnings by around 10%, with effects declining over time (Lundborg et al., 2017; Bensnes et al., 2023; Lundborg et al., 2024; Gallen et al., 2023)—I estimate larger and more persistent losses. My findings are closer to the second set of results in Gallen et al. (2023), which show stable 20% declines five years after unintended births due to IUD failure. These estimates fall near the midpoint of my bounds, which range from 10 to 28% under the most conservative assumptions. Given the arguably more representative sample, more typical institutional context, and more robust methodology, my results suggest that motherhood may have a larger impact on most women's labor market outcomes than indicated by prior studies.

My work also contributes to the broader literature on children and gender inequality in the labor market (see Bertrand (2011), Blau & Kahn (2017), and Olivetti et al. (2024) for an overview).[2] A key insight in this literature, demonstrated by Kleven et al. (2024), is that much of the gender inequality in Western labor markets would disappear if early parents had the same labor market outcomes as those who have children later. Building on this insight, I propose a new method to decompose how much of the gap between early and later parents is caused by parenthood itself. I use the timing of the first IUI as a proxy for fertility timing. Leveraging variation in procedure success, I compare career trajectories between early and later parents in the absence of children. I find that early mothers progress more slowly, while early fathers fare better, suggesting that a modest share of the gender inequality associated with early parenthood reflects selection rather than the effects of parenthood itself.

Methodologically, my approach builds on and contributes to three strands of literature. The

[2]This includes studies on the extensive fertility margin (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); studies exploiting various quasi-experiments (Hotz et al., 2005; Agüero & Marks, 2008; Cristia, 2008; Miller, 2011; Brooks & Zohar, 2021); timing differences (Fitzenberger et al., 2013; Angelov et al., 2016; Chung et al., 2017; Bütikofer et al., 2018; Eichmeyer & Kent, 2022; Melentyeva & Riedel, 2023); and structural methods (Adda et al., 2017).

first step, leveraging women's procedure histories, draws on the biostatistics literature on dynamically assigned treatments (see Hernán & Robins (2020)). In economics, it is closest to the approach developed in Van den Berg & Vikström (2022), which incorporates assignment eligibility duration. These methods are unsuitable for settings with selective treatment uptake (such as natural conception in my setting) and are not directly applicable when individuals selectively pursue treatment assignment (such as deciding to initiate additional procedures). As a secondary methodological result, I show how sequential quasi-experimental assignment can be used in such settings to point-identify effects for a larger share of the sample, under assumptions equivalent to those used in conventional instrumental variable methods.

The second step of my approach, addressing selectively obtained treatment, relates to the literature on bounds for treatment effects, starting with Manski (1989, 1990), and is most similar to methods developed by Zhang & Rubin (2003) and Lee (2009) to address sample selection. While these methods can directly address timing-dependent effects, leveraging sequential assignment before resorting to bounding yields substantially tighter bounds: existing methods produce bounds that are 4.5 to 8 times wider in each year after childbirth and fail to rule out large positive and negative effects.

Finally, my approach complements recent work on dynamic non-compliance by Ferman & Tecchio (2023) and Angrist et al. (2024), who point-identify duration-dependent effects under the assumption that these effects are independent of the moment of treatment initiation. In contrast, my method yields sharp bounds when both the timing and duration of treatment matter. My empirical results indicate that the effects of motherhood depend on both the timing of childbirth and the duration since it occurred. Restricting either dimension leads to biased estimates.

The remainder of the paper is structured as follows. Section 2 introduces the model. Section 3 demonstrates the limitations of existing methods, presents intuition for my approach, states the formal results, and outlines estimation. Section 4 describes the institutions, assisted conception procedures, and the data, and presents support for the assumptions. Section 5 presents the main estimates of the effects of parenthood on women's labor market outcomes and gender inequality. Section 6 covers comparisons with existing methods, generalizability, and concerns about mental health and relationship stability. Section 7 concludes.

2 Model

Section 2.1 presents the model. Section 2.2 discusses interpretation details and limitations relevant for the application.

2.1 Setup

The model adapts the local average treatment effect (LATE) framework (Angrist & Imbens, 1995), formalizing how the treatment—parenthood—depends not only on initial quasi-experimental assignment but also on decisions to pursue subsequent assignments and their outcomes. These assignments may represent a natural conception following IUD failure or a successful IVF conception. While the method is applicable to both settings, I use assisted conception procedures (ACPs), such

as IVF and IUI, as an illustrative example.

Potential outcomes represent labor market outcomes, such as earnings or work hours, at a given moment (e.g., year) since a woman's first ACP. I define $Y_1(1)$ as the outcome if a woman's first ACP succeeds and she has a child (the *treated* outcome); $Y_0(1)$ as the outcome if her first ACP fails but she has a child later (the *later-treated* outcome); and $Y_0(0)$ as the outcome if she remains childless after the first ACP fails (the *control* outcome).

Since, by definition, all women whose first ACP succeeds have a child, I do not define an outcome for remaining childless in such cases. However, the method can be extended to accommodate them. While parenthood may involve multiple children, this study focuses on the extensive margin: treated outcomes reflect the consequences of having whatever number of children follow a successful first ACP.

The primary parameter of interest is the *effect of parenthood*, defined as the difference in potential outcomes between conceiving during the first ACP and remaining childless after its failure: $\tau = Y_1(1) - Y_0(0)$. A key challenge arises from the *effect of parenthood timing*, defined as the difference in outcomes between conceiving during a successful first ACP and conceiving later: $\delta = Y_1(1) - Y_0(1)$. Timing-dependent effects encompass all reasons why $\delta \neq 0$. These include, but are not limited to, effects that vary with the career stage at the onset of motherhood, the duration of motherhood at a given moment, the number of children (which may depend on when motherhood begins), or any interaction among these factors.

To formally leverage variation in ACP success across women's procedure histories, I characterize each woman by two latent variables. First, $W \in \{1, \ldots, \overline{w}\}$ is the total number of ACPs she would undergo for her first child if all prior attempts failed, up to an upper bound $\overline{w}$; I refer to W as *willingness* to undergo ACPs. Second, $R \in \{0, 1\}$ indicates whether she would remain childless if all W ACPs failed; I refer to R as reliance on ACPs. Women with $R = 1$ (*reliers*) depend on ACPs to have children, whereas those with $R = 0$ (*non-reliers*) would conceive even if all ACPs failed.

The observed indicator for the success of the jth ACP is Z_j, which equals 1 if the procedure succeeded and 0 if it failed or was not undertaken. To simplify notation, I count only ACPs occurring before the first child, meaning that $Z_j = 0$ for all subsequent procedures. The realized number of ACPs is $A = \min(\{j : Z_j = 1\} \cup \{W\})$, with last ACP outcome Z_A.

The parenthood (*treatment*) indicator is $D = Z_A + (1 - Z_A)(1 - R)$, where a woman is a mother if an ACP succeeds or if she is a non-relier who would conceive even if all ACPs failed. A woman's realized labor market outcome is $Y = Y_1(1)Z_1 + (1 - Z_1)DY_0(1) + (1 - Z_1)(1 - D)Y_0(0)$.

I also leverage information on non-ACP births among women whose first ACP succeeds. $R^+ \in \{0, 1\}$ indicates whether a woman is reliant on ACPs for all additional children after conceiving her first child via ACP. Women with $R^+ = 1$ (*subsequent reliers*) would have only ACP-conceived children after conceiving their first child via ACP, whereas those with $R^+ = 0$ would also have one or more non-ACP children. The indicator for having at least one non-ACP child is $D^+ = Z_A(1 - R^+) + (1 - Z_A)D$, where a woman has at least one non-ACP child if an ACP succeeded

and she is not a subsequent relier, or if all ACPs failed and she had a child regardless.[3]

The main effect I focus on is the average treatment effect on reliers, $\tau_{ATR} = \mathbb{E}[\tau \mid R = 1]$. To assess timing effects and their contribution to bias in conventional estimates, I will also consider the average timing effect among non-reliers, $\mathbb{E}[\delta \mid R = 0]$.

2.2 Discussion

In this section, I relate my framework to the standard LATE framework and clarify the interpretation of willingness and reliance.

Reliers are the focus of my framework and are closely related to compliers in the LATE framework. Compliers ($C = 1$) are women whose parenthood status depends entirely on the outcome of their first ACP—they have a child if it succeeds and remain childless if it fails. In contrast, always-takers ($C = 0$) have a child regardless of the first ACP outcome. Reliers include all compliers and some always-takers—specifically, those always-takers who would conceive through later ACPs if the first failed but would remain childless if all ACPs failed.[4] Focusing on reliers not only allows estimation for a more general population but, crucially, permits substantially narrower bounds than those attainable for compliers using existing methods (Zhang & Rubin, 2003; Lee, 2009).

As with compliance, willingness and reliance are formal constructs used to define the parameters under study and need not have an economic interpretation. Willingness refers to the number of ACPs a researcher would observe a woman undergo following repeated failures, while reliance describes her fertility behavior if all observed ACPs fail. Both may be random from the woman's perspective or under her direct control and correlated with potential outcomes. Their sole purpose is to formalize what can be learned by leveraging randomness in ACP success, without imposing assumptions about how women determine the number of ACPs or how non-ACP fertility is realized.

In practice, not all ACPs may be observed by the researcher. Classifying births resulting from unobserved ACPs as non-ACP births does not introduce bias, as these cases are addressed in the bounding step by assuming worst-case selection. However, the more births that are handled through bounding, the wider the resulting bounds. Conversely, the more births that can be attributed to ACPs—either due to more complete data or because few births occur without ACPs—the fewer cases require bounding, resulting in tighter bounds.

Like compliance, reliance and willingness are defined cross-sectionally: at any given point after the first ACP, reliers are those who would remain childless up to that point if all ACPs failed. Appendix SA1 describes how the method can be used to quantify effects for a consistent population across periods—namely, women who would remain childless until the end of the sample period. This also addresses another key concern with conventional methods, related to anticipation of parenthood, where estimates may be biased because control outcomes are identified using women who eventually have children. The results remain similar to my baseline estimates.

[3] For berevity, I do not distinguish between subsequent reliance after the first child is conceived via the first or a later ACP; this is without loss of generality, as only the former scenario is relevant.

[4] Formally defining this requires additional notation for the potential outcomes of subsequent ACPs. When the first ACP fails, the relationship is given by $C(1 - Z_1) = R\Pi_j^A(1 - Z_j)$

3 Method

Section 3.1 describes the bias in conventional methods. Section 3.2 presents the intuition behind my approach. Section 3.3 formalizes the bounding procedure. Section 3.4 outlines the estimation strategy.

To demonstrate the intuition, I leverage the sequential unconfoundedness assumption:

Assumption 1 (Sequential unconfoundedness). $(Y_z(d), R, W) \perp\!\!\!\perp Z_j \mid A \geq j$, *for all* z, d, j.

It states that, among women who undergo ACP j, the outcome of ACP j is effectively random—independent of potential outcomes and type. This aligns with the standard unconfoundedness assumption in previous studies using IVF: among women undergoing embryo insertion into the uterus, pregnancy resulting from the procedure is essentially random. Unlike prior studies, this assumption covers not only the first, but also subsequent procedures women undergo.

It is worth highlighting that sequential unconfoundedness assumption concerns only the success of each individual procedure (Z_j) once it occurs ($A \geq j$), not the decision to undergo the procedure (A). This assumption differs from the standard sequential unconfoundedness used in the literature on dynamically assigned treatments, which in my setting would imply that parenthood is as good as randomly assigned in each period among all women. Instead, it applies only to conception via ACPs among women who actually undergo the procedure.

The assumption does not restrict how success rates may vary across procedures. To simplify exposition, I do not distinguish between IVF and IUI when discussing the intuition. The main method in Section 3.3 accounts for selection into procedure type, procedure-specific success rates, and other observed factors that influence success, such as age at the time of the procedure. Empirical support for sequential unconfoundedness is discussed in Section 4.4.

3.1 Bias in Existing Methods

This section describes limitations of existing methods. Section 3.1.1 discusses the conventional IV approach used by Lundborg et al. (2017) and Lundborg et al. (2024); Section 3.1.2 discusses the recursive IV approach used by Bensnes et al. (2023) and Gallen et al. (2023).

The key issue can be summarized as follows. Both methods may yield biased estimates when the effects of parenthood depend on timing, though they impose different restrictions. The IV approach allows effects to vary with life-cycle moment but requires that they do not depend on motherhood duration: for example, effects are assumed to be the same at age 35 whether a woman became a mother at 25 or 30. In contrast, the recursive IV approach allows effects to depend on motherhood duration but not on life-cycle moment: for example, effects are assumed to be the same after five years of motherhood whether a woman is now age 30 or 35. It also assumes that effects are similar for compliers and always-takers. Both sets of assumptions are difficult to justify, as effects may vary with both the life-cycle moment and the duration of motherhood. For instance, becoming a mother at age 25 may have fundamentally different career consequences than becoming a mother at age 30.

3.1.1 Bias in the Instrumental Variable Approach

The IV approach uses the success of a woman's first ACP as an instrument for parenthood. It starts with the reduced form: the difference in average outcomes between those whose first ACP succeeded and those whose first ACP failed: $\tau_{RF} = \mathbb{E}[Y \mid Z_1 = 1] - \mathbb{E}[Y \mid Z_1 = 0]$. The reduces form compares women who conceived at their first ACP to a mixed group of childless women (compliers) and women who had children later (always-takers). Following Angrist & Imbens (1995), under (sequential) unconfoundedness, the reduced form identifies a combination of two effects: the average treatment effect for compliers and a timing effect for always-takers:

$$\begin{aligned}\tau_{RF} &= \mathbb{E}[Y_1(1) - Y_0(0) \mid C = 1] \Pr(C = 1) + \mathbb{E}[Y_1(1) - Y_0(1) \mid C = 0] \Pr(C = 0) && (1)\\ &= \mathbb{E}[\tau \mid C = 1] \Pr(C = 1) + \mathbb{E}[\delta \mid C = 0] \Pr(C = 0). && (2)\end{aligned}$$

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

$$\frac{\mathbb{E}[Y \mid Z_1 = 1] - \mathbb{E}[Y \mid Z_1 = 0]}{\mathbb{E}[D \mid Z_1 = 1] - \mathbb{E}[D \mid Z_1 = 0]} = \mathbb{E}[\tau \mid C = 1] + \mathbb{E}[\delta \mid C = 0] \frac{\Pr(C = 0)}{\Pr(C = 1)}. \tag{3}$$

The standard IV exclusion restriction implies that effects do not depend on timing, meaning $\delta = 0$. In this case, the second term on the right hand side of (3) drops out and the average treatment effect for compliers is identified. Otherwise, the second term biases the IV estimator.

In the context of parenthood, the direction of bias is ambiguous. It may lead to an underestimation of career costs if women who have children later experience the most intensive caregiving demands from young children at the peak of their careers ($Y_0(1) < Y_1(1)$), or to overestimation if early motherhood permanently hinders career progression ($Y_0(1) > Y_1(1)$). Because 75% of women whose first IVF fails eventually have children (Lundborg et al., 2017), $\Pr(C = 0) = 0.75$ and $\Pr(C = 0)/\Pr(C = 1) = 3$. The large scaling factor means that even small timing effects in equation (3) are amplified and can introduce sizable bias.[5]

3.1.2 Bias in the Recursive Instrumental Variable Approach

The IV approach may yield biased results if the effects of parenthood at a given life-cycle moment depend on when a woman become a mother. One concern is that effects may depend on the age of the children, and women who have children later mechanically have a younger first child at any given point in time. Bensnes et al. (2023) and Gallen et al. (2023) address this concern using a recursive IV approach, also discussed in methodological work on dynamic non-compliance by Ferman & Tecchio (2023) and Angrist et al. (2024).

As this method relies on panel data, formal exposition requires substantial additional notation. For brevity, I highlight only the key assumptions and refer to Bensnes et al. (2023) and Gallen et al. (2023) for detailed discussion and Appendix SA2 for technical details. The core idea is to

[5]This bias can be described using the negative weights terminology from the difference-in-differences literature. With an always-taker-to-complier ratio of 3, the IV estimator assigns a weight of 4 to $\mathbb{E}[\tau]$ and -3 to $\mathbb{E}[Y_0(1) - Y_0(0) \mid C = 0]$—the effect of delayed parenthood for always-takers. Difference-in-differences methods are not applicable here due to selective parenthood timing.

estimate the effects of having a young child by applying the IV approach to data from just after the first procedure, when most women in the control group do not yet have a child. These estimates are then used to correct long-run IV estimates, when the treatment group has an older first child and some in the control group have a younger one. This is justified under the assumption that the effects of having a young child earlier in life for compliers are similar to those of having a young child later in life for always-takers.

If, conditional on the age of the first child, effects vary across the life cycle or across women, estimates may be biased. In the ACP context, pre-treatment outcomes indicate that always-takers are positively selected on earnings and work hours, challenging the assumption of homogeneous effects. Moreover, even when child age is held constant, earlier motherhood may lead to different outcomes due to career-stage-specific effects (e.g. having a five-year-old at age 30 versus 35), the cumulative impact on career progression (e.g. having been a parent since age 25 versus 30), or differences in the number of children. To address these concerns, I develop a method that allows effects to vary both across the life-cycle and with parenthood duration.

3.2 Intuition for the Method

In this section, I present the intuition behind my approach. I separately explain how I identify the relier average control outcome and bound their average treated outcome, and how I use additional information to tighten the bounds.

3.2.1 Relier Average 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, and then explain how each term in this expression is identified:

$$\mathbb{E}[Y_0(0) \mid R = 1] = \sum_{w=1}^{\overline{w}} \mathbb{E}[Y_0(0) \mid R = 1, W = w] \Pr(W = w \mid R = 1). \tag{4}$$

I will argue that women who underwent exactly w ACPs and remained childless form an as good as random sample of reliers willing to undergo w ACPs, allowing identification of the average control outcome for such reliers using the average observed outcome in this group:

$$\mathbb{E}[Y \mid A = w, D = 0] = \mathbb{E}[Y_0(0) \mid W = w, R = 1]. \tag{5}$$

This is because, first, the observed women must be reliers willing to undergo exactly w ACPs, since non-reliers would have children, and women willing to pursue more than w ACPs would have done so. Second, for such reliers, experiencing w failed ACPs is effectively random: since they all have the same willingness, this is determined solely by the success or failure of the first w procedures, each of which is as good as random.

The shares of different types can be identified following a similar argument. Women who experience at least w failed ACPs form an as good as random subsample of those willing to undergo at least w ACPs. Thus, the share of these women initiating an additional ACP identifies the share

willing to undergo at least $w+1$ ACPs among those willing to undergo at least w ACPs:

$$\Pr(A \geq w+1 \mid A \geq w, Z_w = 0) = \Pr(W \geq w+1 \mid W \geq w). \tag{6}$$

Similarly, women who do not undergo an additional ACP after w failed ACPs form an as good as random subsample of those willing to undergo exactly w ACPs. Thus, the share of these women who remain childless identifies the share of reliers willing to undergo w ACPs:

$$\Pr(D = 0 \mid A = w, Z_w = 0) = \Pr(R = 1 \mid W = w). \tag{7}$$

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

3.2.2 Relier Average Treated Outcome

I bound the relier average treated outcome using the outcome distribution among women whose first ACP succeeded. Since ACP outcomes are as good as random, this distribution represents the full distribution of treated outcomes in the ACP sample. Using the relier share identified in the previous step, worst-case bounds can be constructed by assuming that reliers are those with the highest or lowest treated outcomes. For instance, suppose 100 women had a successful first ACP, and the relier share is 80%. It is not know which 80 are reliers, but the upper (lower) bound on their average outcome can be constructed by taking the average of the top (bottom) 80 outcomes.

These bounds can be refined by incorporating pre-ACP covariates. Suppose that after splitting the sample by education group, each group is estimated to have an 80% relier share. Selecting the bottom 80% of outcomes among women whose first IUI succeeded without considering education may yield a selection of potential reliers whose education levels are inconsistent with the education-specific relier shares. Since this yields the most conservative lower bound, any alternative selection can only increase it. The refined bounds are constructed by selecting the lowest and highest treated outcomes within each education group.

Further narrowing is possible after imposing assumptions on which women are (not) reliers. For example, it may be reasonable to assume that women who have a second or third child without ACPs after their first ACP succeeded would have conceived even if all ACPs had failed–or, equivalently, that women who are reliant on ACPs for their first child are also reliant for subsequent children. Formally:

Assumption 2 (Monotonicity). $R^+ \geq R$.

Monotonicity reflects the idea that families are more determined to have a first child than additional ones. From a fertility choice perspective, it rules out couples who prefer multiple children but would rather have none than only one. This restriction may be reasonable, as couples who initiate ACPs choose to have at least one child and are likely aware that they may not be able to have more than one.

To see how monotonicity helps narrow the bounds, consider the previous example. If 10 of the 100 women whose first ACP succeeded subsequently have a non-ACP child, they can be excluded

from the pool of potential reliers, as they are certainly not reliant on ACPs. Selecting the 80 lowest (highest) outcomes from the remaining 90 can only yield a higher (lower) average than selecting from the full set of 100.

While monotonicity may be plausible for many couples, it may not hold universally, as fertility may not be entirely determined by choice. A particular concern is that conceiving a first child may improve relationship stability or mental health relative to failing to conceive, leading to more natural conception attempts. This may lead to non-ACP births after a successful first ACP that would not have occurred had the first ACP failed, thereby violating monotonicity. I relax the assumption and address such concerns in Section 6.4. Appendix SA1 formally tests both versions of the monotonicity assumption and presents results under a relaxed specification that allows its direction to vary with covariates. Monotonicity is not rejected, and the estimates under the alternative assumption remain similar. The remaining results assume monotonicity. In settings without auxiliary variables that can credibly help identify reliers, sharp bounds without monotonicity (i.e., under trivial monotonicity) can be obtained by redefining $R^+ = 1$ and $D^+ = (1 - Z_A)D$, thus treating all women as subsequent reliers.

3.2.3 Point Identification for Reliers Assuming Exclusion

Before presenting the formal result for the bounds, I show how τ_{ATR} can be point-identified under assumptions equivalent to the IV approach. Let $Y_0^* = Y_0(0)R + Y_0(1)(1 - R)$ denote the outcome when all ACPs fail. For reliers, this is the control outcome; for non-reliers, it is the later-treated outcome.

Similar to Section 3.2.1, the willingness-conditional average of Y_0^* can be identified using women who underwent w unsuccessful ACPs: $\mathbb{E}[Y \mid A = w, Z_w = 0] = \mathbb{E}[Y_0^* \mid W = w]$. Averaging over W using identified type shares gives $\mathbb{E}[Y_0^*]$. Subtracting this from the average treated outcome (identified using women whose first ACP succeeded), and scaling by the relier share yields:

$$\frac{\mathbb{E}[Y_1(1) - Y_0^*]}{\Pr(R = 1)} = \tau_{ATR} + \mathbb{E}[\delta \mid R = 0]\frac{\Pr(R = 0)}{\Pr(R = 1)}. \tag{8}$$

I refer to this as the *sequential IV* approach, since it uses a sequence of ACP attempts to identify a parameter analogous to IV in equation (3) of Section 3.1.1.

Under the IV exclusion restriction ($\delta = 0$), equation (8) identifies τ_{ATR}. When $\delta \neq 0$, the expression no longer identifies τ_{ATR}, but the bias is attenuated relative to standard IV. This is because non-reliers are a subset of always-takers, implying that $\Pr(R = 0)/\Pr(R = 1) \leq \Pr(C = 0)/\Pr(C = 1)$, meaning that the timing effect receives a smaller weight. Comparing the point estimate to the bounds offers a test for exclusion violations: under the restriction, sequential IV estimates should fall within the bounds; otherwise, $\mathbb{E}[\delta \mid R = 0] \neq 0$. Since non-reliers are a subset of always-takers, this comparison also sheds light on the extent of bias in conventional IV estimates arising from exclusion violations in this group.

3.3 Bounds on the Relier Average Treatment Effect

In this section, I formalize and combine ideas introduced in Section 3.2 to bound the relier average treatment effect τ_{ATR}. Before stating the formal results, I introduce the conditional sequential unconfoundedness assumption:

Assumption 3 (Conditional sequential unconfoundedness).
$(Y_z(d), R^+, R, W) \perp\!\!\!\perp Z_j \mid X_j, A \geq j$ *for all* z, d, j, *and* $X_j \in \mathcal{X}_j$.

Where $X_j \in \mathcal{X}_j$ are covariates measured at ACP j, which may include age at the time of the procedure and procedure type. In words, the success of ACP j is independent of potential outcomes and type, conditional on undergoing at least j ACPs and on these covariates. The next assumption provides regularity conditions. Let $e_j(x) = \Pr(Z_j = 1 \mid X_j = x, A \geq j)$.

Assumption 4 (Regularity).

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

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

It 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 any value of X_1. In practice, adding a negligible amount of continuously distributed noise to Y is sufficient to avoid ties in trimming without meaningful bias.

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

$$r(x) = \mathbb{E}\left[\frac{(1 - D^+)\Pi_{j=1}^{A}(1 - Z_j)}{\Pi_{j=1}^{A}(1 - e_j(X_j))}\middle| X_1 = x\right]. \tag{9}$$

Larger weights are given to women who underwent more ACPs to account 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 $r^+(x) = \mathbb{E}\left[1 - D^+ \mid Z_1 = 1, X_1 = x\right]$. Under monotonicity, the covariate-conditional share of reliers among subsequent reliers is then $p(x) = r(x)/r^+(x)$.

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

$$q(u, x) = \inf\left\{q : u \leq \Pr(Y \leq q \mid X_1 = x, Z_1 = 1, D^+ = 0)\right\}. \tag{10}$$

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 are used

to trim the outcome distribution and select reliers in scenarios where they have either the lowest or highest treated outcomes.

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

$$m^L(G,\eta^0) = Y(1-D^+)1_{\{Y<q(p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D^+)\Pi_{j=1}^{A}\frac{(1-Z_j)}{(1-e_j(X_j))} \tag{11}$$

$$m^U(G,\eta^0) = Y(1-D^+)1_{\{Y>q(1-p(X_1),X_1)\}}\frac{Z_1}{e_1(X_1)} - Y(1-D^+)\Pi_{j=1}^{A}\frac{(1-Z_j)}{(1-e_j(X_j))}, \tag{12}$$

where vector G contains observed variables and η^0 contains the nuisance functions:

$$\eta^0(x_1,\ldots,x_A) = \{r^+(x_1), r(x_1), q(p(x_1),x_1), q(1-p(x_1),x_1), e_1(x_1),\ldots,e_{\overline{w}}(x_{\overline{w}})\}. \tag{13}$$

The first term in $m^L(G,\eta^0)$ is used to bound the relier average treated outcome. It assigns positive weights to women whose first ACP succeeded, who have only ACP children, and whose outcomes fall below the covariate-conditional trimming threshold $q(p(x),x)$. The second term, used to identify the relier average control outcome, assigns positive weights to outcomes of childless women. 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. $m^U(G,\eta^0)$ mirrors this for the scenario where reliers have the highest treated outcomes. The moments are then scaled by the relier share.

Theorem. *Under Assumptions 2, 3, and 4, sharp lower and upper bounds on τ_{ATR} are given by $\theta_L = \mathbb{E}[m^L(G,\eta^0)]/\mathbb{E}[r(X_1)]$ and $\theta_U = \mathbb{E}[m^U(G,\eta^0)]/\mathbb{E}[r(X_1)]$.*

Proof. See Appendix A1.

3.4 Estimation

In principle, the bounds can be estimated by replacing the expectations in the theorem with sample averages and plugging in an estimate of the nuisance parameter η^0. With few discrete covariates, asymptotic normality can be established following Lee (2009). However, tighter bounds may require incorporating continuous covariates, which complicates inference due to the need for nonparametric estimation of the nuisance function. To address this, I build on the procedure of Semenova (2023) for estimating the Zhang & Rubin (2003) and Lee (2009) bounds with high-dimensional covariates, using orthogonalization and sample splitting. I adapt it to incorporate the first step of my identification strategy, which leverages sequential treatment assignment, and construct orthogonal moments. The moments presented in Appendix A2 meet the key conditions under which Semenova (2023) establishes the validity of standard GMM inference.

I highlight only the most relevant implementation choices and refer to Appendix A2 for further details. I use 3-fold cross-fitting, estimating the nuisance function for each third of the sample using the remaining two-thirds. Propensity scores are estimated using logistic regressions that include quadratic terms for each partner's age at the procedure, interacted with procedure-type and education dummies, based on women who initiated the respective ACP. Following Heiler (2024),

I estimate the remaining nuisance functions using Generalized Random Forests (Athey et al., 2019), incorporating all propensity score covariates up to the current ACP, along with pre-ACP earnings and work hours for both partners. Confidence intervals are constructed following Stoye (2020). I use data on the first ten ACPs, which makes the bounds only negligibly wider than if additional ACPs were included, while avoiding the need to estimate nuisance functions on small subsamples of women who undergo more than ten (fewer than 7% of the sample). This means that reliers are women who would not have a child if their first ten observed ACPs failed. Varying the number of ACPs used between 8 and 12 has little impact on the estimates.

4 Institutions, Procedures, and Data

Section 4.1 describes Dutch family policies and the labor market context. Section 4.2 discusses IVF and IUI and the differences between them. Section 4.3 overviews the data and compares the IUI sample to a representative sample. Section 4.4 provides empirical support for the sequential unconfoundedness assumption.

4.1 Family Policies in the Netherlands

Dutch women are entitled to 4 to 6 weeks of pregnancy leave before the due date and at least 10 weeks of maternity leave following childbirth, totaling a minimum of 16 weeks. In the case of multiple births, the total entitlement increases to 20 weeks. During this period, mothers receive full wage replacement from the unemployment insurance agency, subject to a daily maximum. In the sample period, fathers are entitled to one week of fully paid leave within the first four weeks after childbirth, financed by the employer.

Children can enroll in private daycare from three months old. In 2022, 72% of children under two attended formal child care, averaging 20 hours per week (OECD, 2023a). After turning four and starting elementary school, they become eligible for out-of-school care. In 2023, families using child care paid an average of 8,950 euros, of which 64% was reimbursed by the government, resulting in a net cost equivalent to 10% of median disposable household income.[6]

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

While employment rates for mothers, fathers, and non-parents in the Netherlands exceed their respective OECD averages, part-time work is far more common, making average hours worked comparable to the OECD average (OECD, 2023b). In 2021, the maternal employment rate was 80%, compared to the OECD average of 71%. However, in 2023, 52% of women and 18% of men worked part-time (less than 30 hours per week), more than twice the respective OECD averages

[6] 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/.

(OECD, 2023d). Among two-parent families, only 14% had both parents working full-time, 52% had a full-time working father and a part-time working mother, and 12% had both parents working part-time.[7]

4.2 Assisted Conception Procedures

I use two types of ACPs: IVF, previously used to study the career impact of parenthood in Denmark, Norway, and Sweden (Lundborg et al., 2017; Bensnes et al., 2023; Gallen et al., 2023; Lundborg et al., 2024), and IUI, which has not been used for this purpose. Both procedures may begin with cycle tracking and hormonal stimulation to enhance egg production. IVF is a surgical procedure where eggs are retrieved through the vaginal wall, fertilized in the lab, and transferred as embryos into the uterus. It is relatively invasive, performed under sedation or anesthesia, and has a success rate of about 25% per embryo transfer. IUI involves injecting sperm directly into the uterus via a catheter, mimicking natural conception by facilitating fertilization within the body. With a lower success rate of about 10%, IUI is significantly less invasive—lasting about five minutes and generally painless—and is the first-line infertility treatment in most countries.

In the Netherlands, couples without a specific infertility diagnosis are typically required to undergo six IUI cycles before becoming eligible for IVF. Compulsory health insurance covers an unlimited number of IUI cycles and up to three IVF procedures. In 2022, each additional IVF cycle costs approximately 4,000 euros, but since multiple embryos can be frozen per cycle, subsequent transfers may cost 1,000 euros or less.

4.3 Data and Sample Characteristics

I use administrative data from Statistics Netherlands, covering all residents. ACP records from 2012–2017 are drawn from the Diagnosis-Treatment Combination system, which Dutch hospitals are required to maintain. The main variables are the procedure type—IUI or IVF—and the date of sperm or embryo insertion, which marks the key moment for my analysis. From this point onward, outcomes can no longer be directly influenced by doctors or patients, introducing potentially quasi-random variation in fertility outcomes. ACP success is defined as having a child born within ten months of insertion with no subsequent procedures, a definition validated against medical records by Lundborg et al. (2017).

Labor market data span 2011–2023 and include annual work hours and gross earnings derived from tax records. Reported work hours include maternity leave, and earnings include maternity pay. In Appendix SA1, I replicate the main analyses using adjusted hours that account for the maximum potential duration of unobserved leave. This adjustment affects estimates only in the first year after childbirth. I also use several demographic variables, including an indicator for higher education attainment, number of children, year and month of birth, and cohabitation status.

My main sample includes 12,734 Dutch-born women who underwent IUI to conceive their first child, had no prior ACPs, and were cohabiting with a male Dutch-born partner at the time of

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

Table 1: First ACP Outcomes and Descriptives

	Success (1)	Fail (2)	Dif. (1)-(2)	IPW dif. (1)-(2) cond.	Rep. (5)	Suc. vs rep. (1)-(5)
Work (W)	0.912	0.916	-0.004	-0.009	0.936	-0.024
	[0.283]	[0.277]	(0.008)	(0.008)	[0.244]	(0.007)
Work (P)	0.894	0.885	0.009	0.002	0.897	-0.002
	[0.307]	[0.319]	(0.009)	(0.009)	[0.304]	(0.008)
Hours (W)	1300.012	1298.876	1.136	-1.951	1310.923	-10.911
	[547.832]	[558.316]	(15.730)	(16.119)	[544.468]	(14.554)
Hours (P)	1513.337	1494.541	18.796	3.345	1497.603	15.734
	[635.121]	[656.050]	(18.457)	(19.041)	[651.043]	(17.403)
Earn. 1000s EUR (W)	29.358	29.648	-0.290	0.203	26.555	2.803
	[18.000]	[18.911]	(0.531)	(0.561)	[15.989]	(0.427)
Earn. 1000s EUR (P)	38.082	38.060	0.022	0.322	33.862	4.220
	[25.425]	[26.525]	(0.745)	(0.774)	[24.148]	(0.646)
Bachelor deg. (W)	0.512	0.494	0.018		0.518	-0.007
	[0.500]	[0.500]	(0.014)		[0.500]	(0.013)
Bachelor deg. (P)	0.425	0.410	0.014		0.430	-0.005
	[0.494]	[0.492]	(0.014)		[0.495]	(0.013)
Age (W)	31.373	32.060	-0.687		28.840	2.533
	[3.889]	[4.265]	(0.119)		[3.896]	(0.104)
Age (P)	34.088	34.856	-0.768		31.415	2.673
	[4.968]	[5.500]	(0.154)		[4.803]	(0.128)
Observations	1,411	11,323			171,180	
Joint p-val.			0.001	0.955		0.000

Note: *Success* – average among women whose first ACP succeeded; *Fail* – average among women whose first ACP failed; *Dif.* – difference between *Success* and *Fail*; *IPW dif.* – difference adjusted for age and education using inverse probability weights from the baseline specification; *Rep.* – average in representative sample of women who conceived their first child without ACPs; *Suc. vs rep* – difference between *Success* and *Rep.*. Reference year: year of first ACP (ACP sample); 9 months before first birth (representative sample). ACP sample: women who underwent intrauterine insemination for their first child between 2013 and 2016, with no prior assisted conception procedures, cohabiting with a male partner in the year prior to the reference year. Representative sample: women with no ACPs before first birth, cohabiting with a male partner in the year prior to the reference year, with reference year between 2013 and 2016. Labor market outcomes measured in the year before the reference year; age measured in the reference year. *Bachelor deg.* – indicator for completing a bachelor's degree. *Earn.* – earnings, (W) – woman, (P) – partner. Standard deviations in brackets. Standard errors in parentheses.

the procedure. Throughout the analysis, I refer to these men as the partners. I address potential complications related to separation in Section 6.4. Following Lundborg et al. (2017), I exclude women whose first observed procedure occurred in the first data year to avoid including women with unobserved prior IUIs, and those whose first observed IUI occurred in the last data year to avoid misclassifying failed procedures as successful due to unobserved subsequent procedures. For comparison with the general population, I use 171,180 women who conceived their first child without prior ACPs between 2013 and 2016 while cohabiting with a male partner. Conception dates are approximated as nine months before birth. All analyses use the full samples, with hours and earnings set to zero for individuals not in paid employment.

Table 1 compares average characteristics of couples whose first IUI succeeded (column 1) and failed (column 2), measured in the year before the procedure. Because age strongly predicts IUI success, these comparisons are descriptive and not intended to test unconfoundedness. Despite this,

the groups are similar in earnings, employment rates, and education. The only notable difference is age: both women and their partners whose first IUI succeeded were nearly nine months younger, consistent with age being the key determinant of procedure success.

Table 1 also compares women whose first IUI succeeded to a representative sample of mothers (column 5). The two groups had similar education and work hours, and women in the IUI group were only slightly less likely to be working. However, two differences stand out: women in the IUI group were on average 2.5 years older and earned about 2,800 euros more annually. Patterns for their partners are similar. The age difference is not surprising. As in most countries, Dutch couples must usually attempt natural conception for at least a year before becoming eligible for medical assistance, and IUI is often not initiated immediately thereafter. Moreover, the representative sample likely includes women with unplanned births, which tend to occur at younger ages. While the earnings difference is arguably small relative to the standard deviation of 18,000, the next section further shows that it is largely explained by the age difference. On average, women whose first IUI succeeded have 1.84 children, compared to 1.92 among mothers in the representative sample (not shown). However, they are more likely to have multiple births (7% vs. 1.3%).

4.4 Unconfoundedness and Procedure Profiles

Since women have limited direct control over ACP outcomes, the main threat to conditional sequential unconfoundedness is that success may depend on health factors that also influence labor market outcomes. As such factors could be expected to also affect pre-ACP outcomes, unconfoundedness can be assessed by comparing pre-ACP characteristics between women whose procedures succeed and those whose procedures fail.

To this end, column 3 in Table 1 reports average covariate differences between women whose first IUI succeeded and those whose first IUI failed, after adjusting for age and education (following Lundborg et al. (2017)) using inverse probability weights from the main specification. This adjustment renders the remaining gaps negligible. Excluding education has no effect. Appendix SA3 presents equivalent results for subsequent ACPs. All remaining analyses account for differences in success rates by age at the time of each procedure, education, and procedure type.

Another reason sequential unconfoundedness may fail is if women pursue additional ACPs based on information suggesting a higher success likelihood, inducing correlation between willingness and procedure success. A similar concern arises in the IV approach, which assumes complier status is independent of first ACP success, though it may depend on later procedure success. To assess this, I examine success rates across procedures. If women with a higher willingness are more likely to succeed, one might expect higher success rates at later procedures. Since success likelihood declines with age, which might obscure any pattern, I first estimate age-conditional IUI success rates using the baseline specification. I then compare these rates holding age fixed at the first procedure average. Panel A in Figure 1 shows that success rates are similar across procedures, suggesting limited correlation between willingness and success.

Panels B–D in Figure 1 further document the distribution ACP attempts and non-ACP births. Panel B plots complier and relier shares over time, estimated using the baseline specification. Two

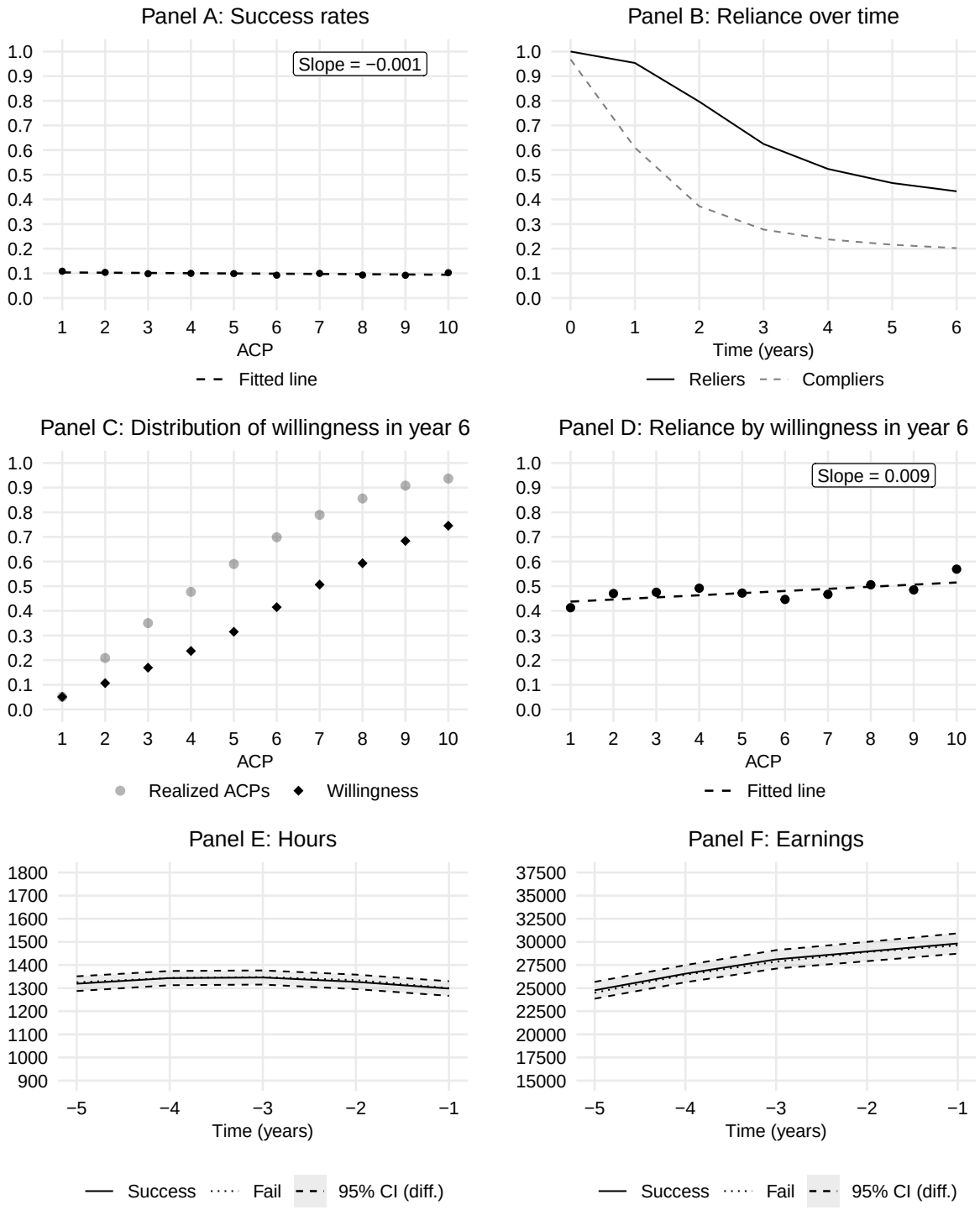


Figure 1: Procedure Descriptives and Pre-trends

Note: Panel A: age-conditional intrauterine insemination (IUI) success rates, holding covariates at their average values at the first procedure. Panel B: estimated complier and relier shares over time. Panel C: cumulative distribution of the number of assisted conception procedures pursued after the first failure and the number women would pursue if the first ten failed, up to six years after the first IUI. Panel D: relier shares by willingness, six years after the first IUI. Panels E and F: pre-IUI work hour and earnings trajectories for women whose first IUI succeeded versus those whose first procedure failed. Sample includes all women who underwent IUI for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure. All panels use inverse probability weights from the baseline specification. Panels C and D use the first 11 ACPs to illustrate the share of women exceeding 10 ACPs; all other panels use only the first 10 ACPs. Time is measured relative to the first procedure.

years after the first IUI, relier and complier shares are 0.8 and 0.38, falling to 0.43 and 0.2 by year six. Thus, my estimates cover a group roughly twice as large as in conventional IV estimates.

Panel C shows the cumulative distribution of the number of procedures women undergo after first failure and the number they would undergo if the first ten failed, estimated using the baseline specification. Women undergo on average 4.1 additional procedures, and their average willingness is 6.7. 43% of women whose first IUI fails eventually undergo IVF (not shown). Panel D plots the relier share six years after the first IUI by willingness, indicating no correlation.

Panels E and F compare work hour and earnings trajectories before the first IUI between women whose first IUI succeeds and those whose first IUI fails. The comparison uses weights from the baseline specification. Work hour profiles are relatively flat, while earnings profiles are increasing. Differences between the groups are small and statistically insignificant at the 5% level. From three to one year before the first IUI, women's earnings increase by 1,700 euros on average, suggesting that the earnings gap relative to the representative sample is largely explained by age differences.

5 Results

Panels A and B in Figure 2 present effects on women's annual work hours and earnings, estimated using the baseline specification. In the conception year, the bounds indicate negligible impacts on either outcome. Between the second and sixth years of motherhood, the average upper and lower bounds imply reductions in annual work hours of 120 and 260, respectively. These correspond to declines of 10% to 22% relative to the point-identified relier average control outcome. For earnings, the average upper and lower bounds indicate reductions of 3,400 and 9,300 euros, or 10% to 28%.

Panels C and D in Figure 2 present effects on men's outcomes. The bounds are similar in width to those for women but are centered near zero. Six years into parenthood, they rule out reductions in work hours greater than 2% and in earnings greater than 14%.

Panels E and F in Figure 2 plots the share of within-couple gender inequality among parents caused by parenthood in each year since childbirth—that is, the average effect on the gender gap relative to the average difference between outcomes for treated men and women. Between the second and sixth years of parenthood, parenthood causes 21–58% of the gender gap in annual work hours and up to 48% of the gap in annual earnings. Aggregating these effects across years yields non-sharp bounds for cumulative impacts, as per-period estimates do not account for within-woman or within-couple correlations over time. Using cumulative hours and earnings over the seven years instead, I find that parenthood causes 32–53% of inequality in work hours and 10–45% in earnings.

Because outcomes are measured by years since childbirth, men's and women's outcomes may be measured at systematically different ages, limiting the relevance of the estimates for lifetime inequality. Appendix SA1 addresses this by measuring outcomes for both partners at the same age. The results remain similar, consistent with the small age differences between partners.

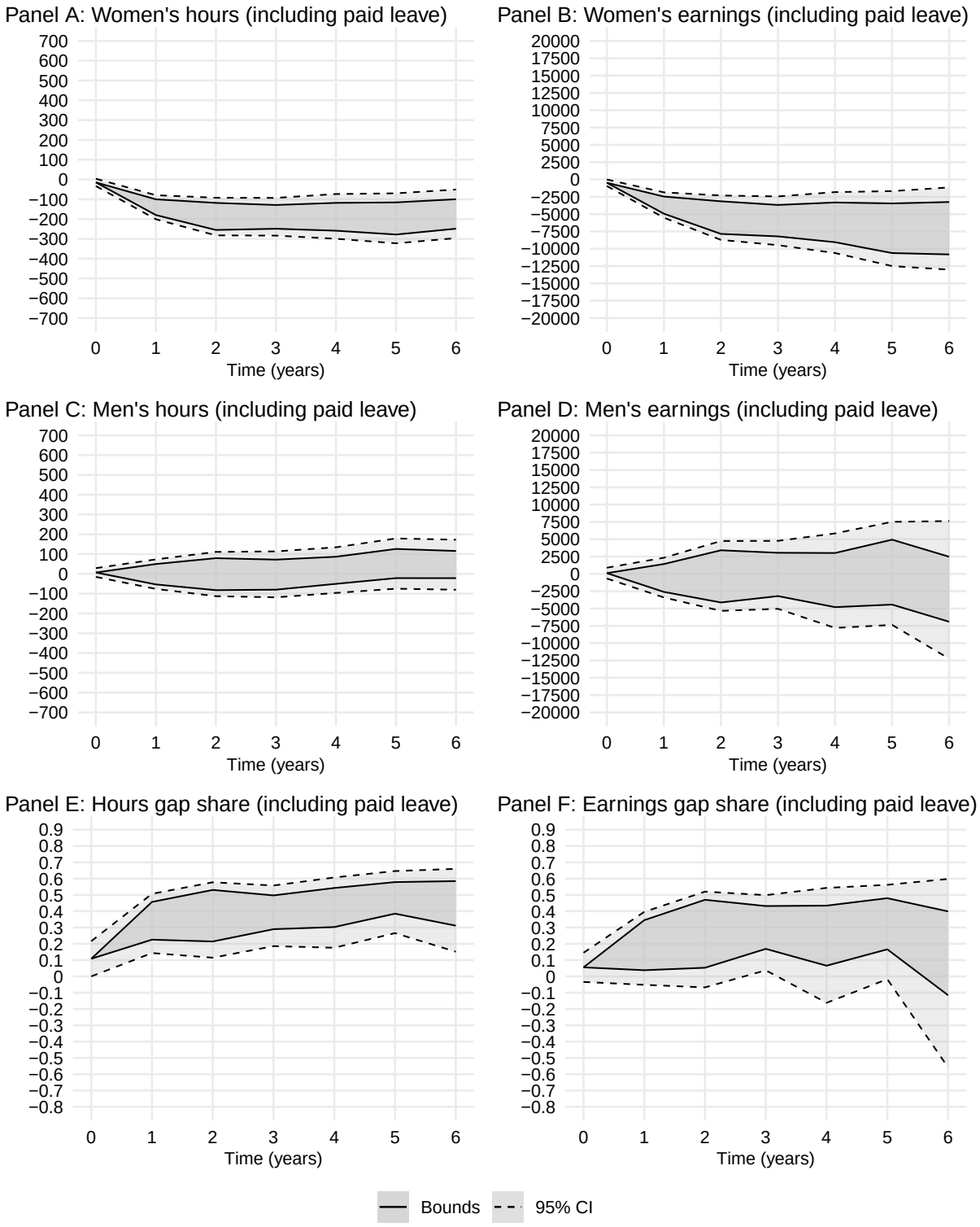


Figure 2: Effects of Parenthood

Note: Panels A though D: effects of parenthood on annual work hours and earnings (in EUR), estimated using the baseline specification. Panels E and F: Share of within-couple gender inequality among parents caused by parenthood, calculated as $1 - a/b$, where a is the average gap in the control outcome and b is the lower or upper bound for the average treated outcome, both estimated using orthogonal moments from the baseline specification. Confidence intervals based on the Delta method. Time relative to first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure, as well as these partners.

6 Existing Methods and Generalizability

Section 6.1 compares my estimates to those based on existing partial identification approaches. Section 6.2 compares them to estimates from instrumental variable and recursive instrumental variable methods that restrict timing-dependent effects. Section 6.3 quantifies the extent to which gender inequality is explained by the effects of parenthood and by differences in career progression between early and late parents in the absence of children. Section 6.4 addresses concerns related to mental health and relationship breakdowns. For brevity, I focus on point estimates.[8]

6.1 Comparison with Conventional Bounding Methods

I compare my bounds to those based only on the first IUI without leveraging monotonicity, equivalent to the approaches in Zhang & Rubin (2003) and Lee (2009) (ZRL). Figure 3 shows that, in every year following childbirth, ZRL bounds are 4.5 to 8 times wider and do not rule out large positive or negative effects on women's earnings and work hours. Even when monotonicity is leveraged, bounds based solely on the first IUI remain 3.8 to 5.7 times wider than mine. These results demonstrate that exploiting sequential quasi-experimental assignment before resorting to bounding yields substantially more informative estimates.[9]

6.2 Comparison with Instrumental Variable Methods

I compare my estimates to those based on the IV approach used by Lundborg et al. (2017, 2024) and the recursive IV approach used by Bensnes et al. (2023) and Gallen et al. (2023). Implementation details are provided in Appendix SA2.

Figure 4 presents the estimates for women's outcomes. In most years, conventional IV estimates are either consistent with the largest negative effects implied by the bounds or even more negative. These bounds are attainable only under extreme negative selection of ACP-reliant women with respect to treated labor market outcomes. While this does not imply bias—since the IV identifies effects for compliers, whereas the bounds target reliers—it suggests that average effects may be smaller when considering a broader population.

To assess the potential bias in conventional IV estimates, I estimate effects for reliers using the sequential IV approach described in Section 3.2.3. The results are similar to the conventional IV estimates. Since both the bounds and the sequential IV approach estimate effects for reliers, the difference between these estimates suggests that timing-dependent effects among non-reliers bias the sequential IV estimates. Moreover, since non-reliers are a subset of always-takers, as discussed

[8] While efficiency is not the primary focus of this study, leveraging subsequent ACPs substantially improves precision. I show in Appendix SA4 that 95% confidence intervals for conventional IV estimates are very similar to those for my bounds, suggesting that the precision gains offset the lack of point identification.

[9] ZRL bounds are wider because leveraging only the first ACP allows for the identification of the average control outcome only for compliers rather than the broader group of reliers. Bounding treated outcomes for smaller groups requires more conservative assumptions, allowing for the possibility that outcomes lie entirely in the tails, which is not possible when the group is larger.

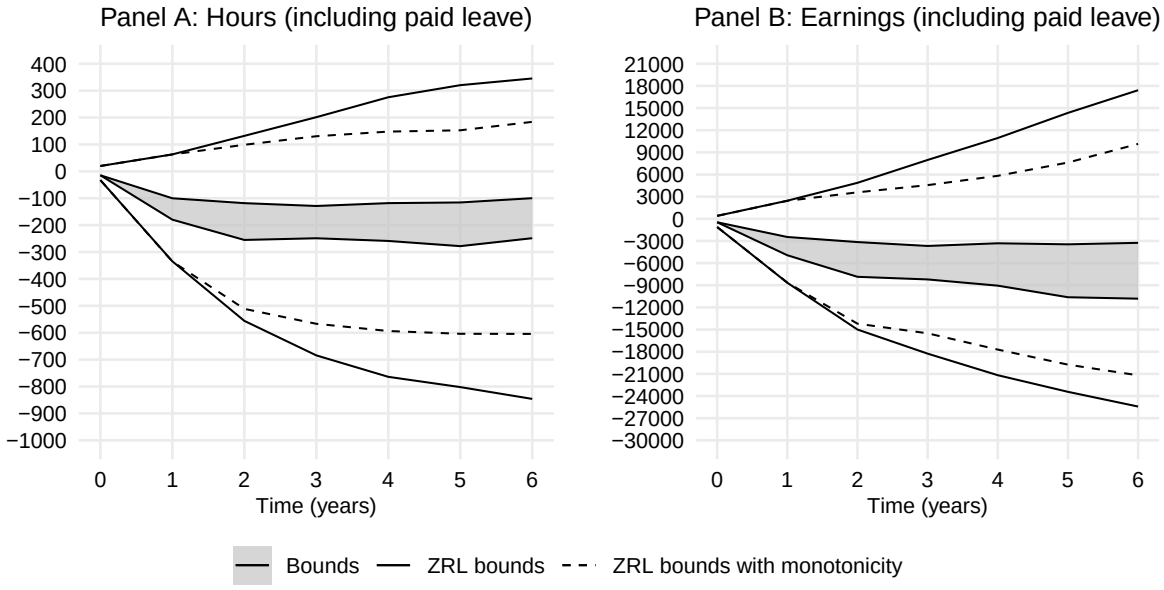


Figure 3: Comparison with ZRL Bounds for Effects on Women

Note: Effects of motherhood on women's annual work hours and earnings (in EUR). *Bounds* – baseline specification; *ZRL bounds* – baseline specification using only the first IUI and no information on non-ACP births; *ZRL bounds with monotonicity* – baseline specification using only the first IUI. Time relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

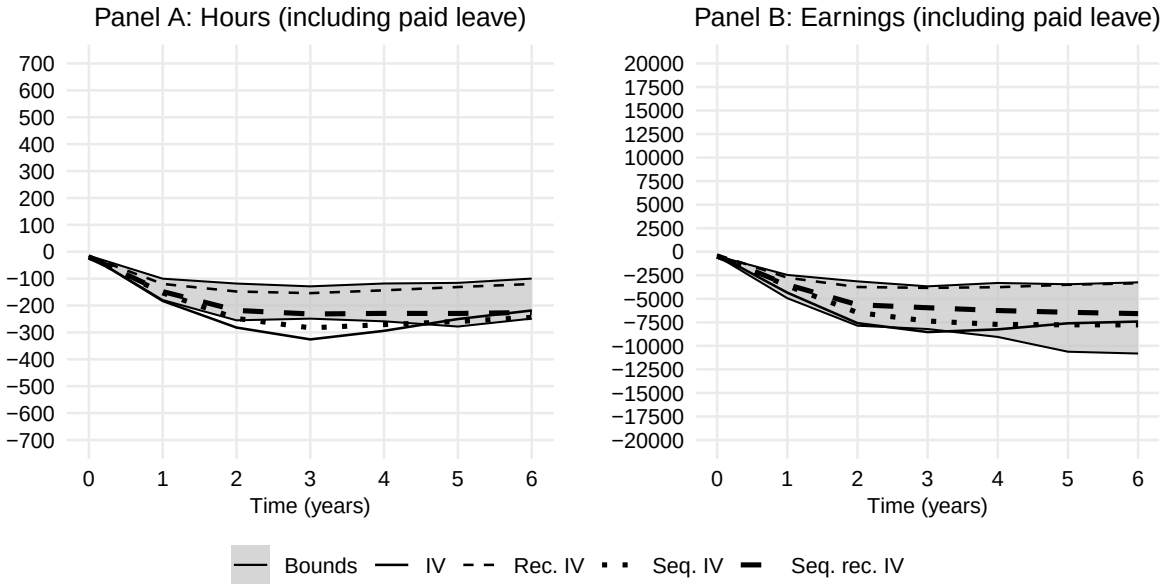


Figure 4: Comparison of Different Methods for Effects on Women

Note: Effects of motherhood on women's annual work hours and earnings (in EUR), based on different estimation methods. Bounds estimated using the baseline specification. *IV* – instrumental variable; *Rec. IV* – recursive instrumental variable; *Seq. IV* – sequential instrumental variable; *Seq. Rec. IV* – sequential recursive instrumental variable. See Appendix SA2 for implementation details. Time relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

in Section 3.2.3, conventional IV estimates likely suffer from the same bias.

As discussed in Section 2, the direction of bias in the IV estimates is informative about the effect of delaying parenthood. The point estimate falling below the bounds indicates that women who became mothers earlier experience worse outcomes than those who did so more recently. This is surprising, as one might expect the opposite: women who conceive after IUI failure do so, on average, 3.1 years later, meaning their first child is younger and likely requires more care. One possible explanation is differences in total fertility: women whose first IUI succeeds have, on average, 0.2 more children, potentially increasing caregiving demands. Another possibility is that entering motherhood at an earlier career stage has persistent adverse effects on labor market outcomes, independent of child age.

Figure 4 also presents recursive IV estimates for women's outcomes. In contrast to the conventional IV estimates, these estimates are only consistent with the smallest negative effects implied by the bounds. Such bounds are attainable only under extreme positive selection of ACP-reliant women with respect to treated labor market outcomes. Given that the recursive IV approach assumes homogeneous treatment effects, these results suggest bias unless selection is as strong as theoretically possible.

Figure 4 also presents recursive estimates constructed using sequential IV estimates (sequential recursive IV; see Appendix SA2 for details). The two sets of recursive estimates differ markedly: those based on sequential IV lie near the lower bound for hours worked and closer to the center of the bounds for earnings, whereas those based on conventional IV are near the upper bound for both outcomes. Under the maintained assumptions, both methods should yield similar estimates. The difference is statistically significant at the 1% level from year 3 onward, providing further evidence that recursive estimates may be biased due to effect heterogeneity across women or timing-dependent effects.

This comparison highlights the advantages of the proposed method. The results not only quantify the potential for bias in existing approaches, but also allow the corresponding estimates to be ruled out. Since this applies to both methods that restrict how effects vary with parenthood duration and those that restrict how they vary with the moment of becoming a parent, the results suggest that the effects of parenthood depend on timing through both channels. Moreover, the bounds establish a relatively narrow range for the effects. In several periods, they are tighter than the gap between estimates produced by existing methods, demonstrating that informative conclusions can be reached without imposing restrictions on timing-dependent effects.

6.3 Gender Inequality Without Children

Kleven et al. (2024) show that in most Western countries, gender inequality in labor force participation would be largely eliminated if mothers worked as much as women who delay childbearing. Rabaté & Rellstab (2022) show that in the Netherlands, this pattern holds for both work hours and earnings. This raises a central question: to what extent is the gap between early and late mothers driven by the causal effect of parenthood, as opposed to unobserved differences between them. In principle, one could assess this by comparing my causal estimates to the observed gap

in a representative sample (or the sample of IUI mothers). Yet this approach is problematic, as women reliant on ACPs may differ systematically from the broader population, affecting both the magnitude of parenthood effects and how much they explain the gap.

To address this, I instead assess selection and causal effects within a consistent sample of women who rely on ACPs. Using the timing of a woman's first IUI as a proxy for fertility decisions, I estimate how career trajectories would differ between early and late IUI mothers if none had children. I then combine these estimates with bounds on the causal effect of parenthood for this sample, quantifying the share of the gender gap explained jointly by the causal effect of parenthood and other differences related to fertility timing.

I adapt the event-study specification from Kleven et al. (2024):

$$Y_{it}^g = \sum_{k \neq 0} \alpha_k^g 1_{\{T_{it}=k\}} + \sum_a \beta_a^g 1_{\{\text{Age}_{it}=a\}} + \sum_s \gamma_s^g 1_{\{t=s\}} + \nu_{it}^g, \tag{14}$$

where Y_{it}^g is the labor market outcome of individual i of gender g in period t. The first term reflect time relative to the event: the year before pregnancy in Kleven et al. (2024), or the year before first IUI in my analysis. The second and third terms control for age and calendar year. The parameter of interest, α_k^g, measures average outcome differences k years after the event, relative to similarly aged individuals a year before the event.

I estimate (14) using women who remain childless through the end of the sample period and their partners.[10] Since none of these women have children, α_k^g captures differences in outcomes between reliers who chose to have children k years ago and those who will choose to do so in a year, in the absence of children.

Figure 5 presents average control outcomes since the first IUI for women who remain reliers until the end of the sample period, and their partners. For both hours and earnings, career trajectories evolve smoothly, and gender gaps remain stable. The figure also shows outcomes extrapolated from similarly aged reliers who delay childbearing, obtained by subtracting α_k^g. Extrapolated and realized profiles align closely in the early years, when based on individuals with comparable fertility timing, but diverge over time as timing differences grow. Men's realized earnings exceed those extrapolated from later fathers, while work hours remain similar. For women, both earnings and hours fall short of extrapolated profiles. This pattern suggests positive selection among early fathers and negative selection among early mothers. Extrapolating from those who delay childbearing may thus understate gender gaps in the absence of children.

I next ask how much of the gender inequality among parent's reliant on ACPs can be jointly explained by the causal effect of parenthood and by differences in outcomes between late and early

[10] I weight observations by $1/(\Pi_{j=1}^{A}(1-e_j(X_j)))$ to ensure that reliers with higher willingness are not underrepresented, thereby maintaining comparability with my main estimates. I use a balanced panel covering one year before the first IUI up to six years after.

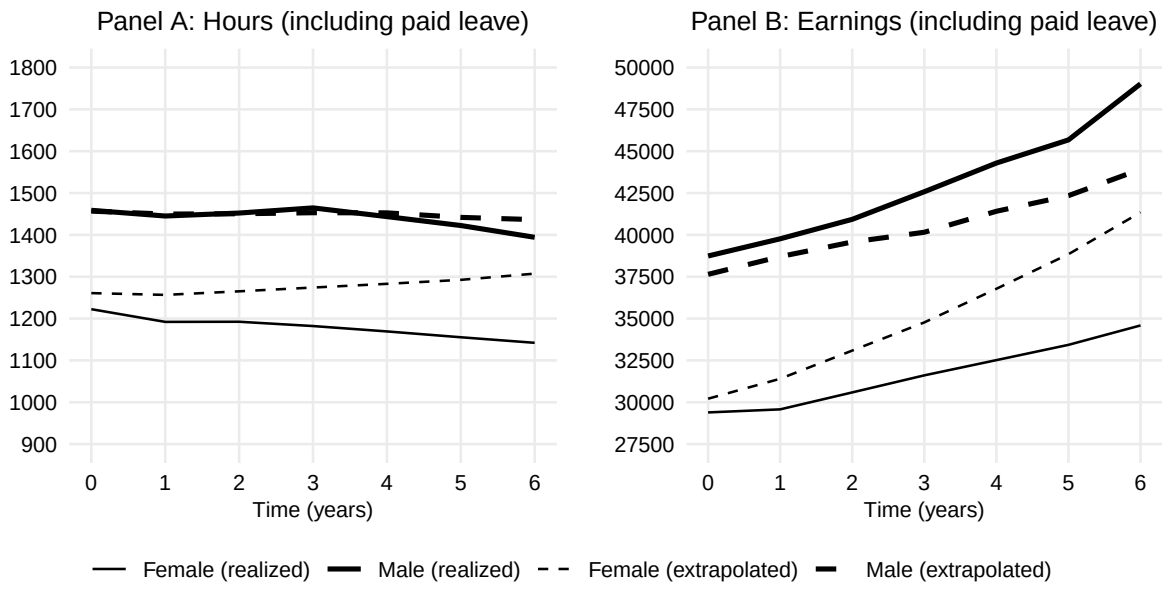


Figure 5: Career Progression in the Absence of Children

Note: Estimated annual work hours and earnings (in EUR) for ACP-reliant couples in the absence of children. *Realized* – estimated using couples who remain childless the end of the sample period; *Extrapolated* – constructed using similarly aged individuals in couples one year away from their first IUI; see Section 6.3 for procedure details. Time relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016, were cohabiting with a male partner in the year before the first procedure, and their partners.

parents in the absence of children.[11] Figure 6 presents the results, showing that toward the end of the sample period, the two factors together explain 62 to 77% of the gap in work hours and 70 to 84% of the gap in earnings. In contrast, baseline estimates indicate that the causal effect of parenthood alone accounts for at most 58% and 40% of the respective gaps. These findings suggest that a sizable share of the gender inequality associated with early parenthood reflects selection.

6.4 Mental Health and Relationship Stability

Women who remain childless after ACP failure may experience mental health deterioration or relationship breakdowns, potentially affecting labor market outcomes. Bögl et al. (2024) document such effects in Sweden in the context of IVF. Focusing on IUI—a less invasive procedure than IVF—may mitigate concerns about mental health deterioration due to physical side effects. However, 43% of women whose first IUI fails eventually undergo IVF. Moreover, the experience of attempting but failing to conceive may itself affect mental health or relationship stability, limiting the generalizability of the estimates to a setting in which women choose to remain childless. Likewise, successful conception via the first IUI may improve mental health and relationship stability,

[11] I start with a point in the bounds on the effect of parenthood on the gender gap for reliers: $\mathbb{E}[Y_{1k}^{gap}(1) - Y_{0k}^{gap}(0) \mid R_k = 1]$, where $Y_{zk}^{gap}(d)$ represents the within-couple gap between male and female potential outcomes k years after the first IUI when the outcome of the first IUI is z and the parenthood status is d. I then add the difference in childless outcomes between early and late relier fathers, α_k^{male}, and subtract the corresponding difference for relier mothers, α_k^{female}. I divide the result by the gender gap in the case of parenthood, $\mathbb{E}[Y_{1k}^{gap}(1) \mid R_k = 1]$.

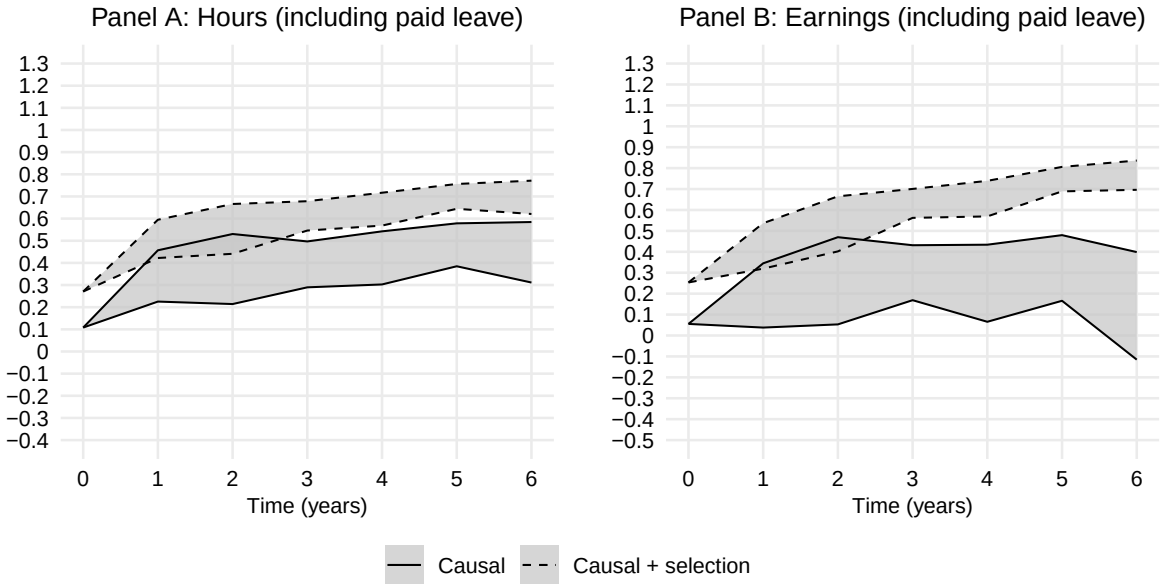


Figure 6: Share of Within-couple Gender Inequality Explained by Effects of Parenthood and Selection

Note: Share of within-couple gender inequality in annual work hours and earnings (in EUR) explained by parenthood. *Causal* refers to the causal effect of parenthood alone (baseline estimates); *Causal + selection* includes both the causal effect and differences between early and late parents in the absence of children. See Section 6.3 for procedural details. Time is measured relative to the first intrauterine insemination. The sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016, were cohabiting with the male partner in the year before the first procedure, and their partners.

increasing the likelihood of subsequent natural conception attempts. As a result, these women may go on to have more children naturally, whereas they might have remained childless had IUI failed—violating monotonicity.

To provide suggestive evidence on potential mental health effects, I estimate the impact of conceiving during the first IUI versus failing to conceive on antidepressant uptake. These estimates, presented in Appendix SA1, are precise and indistinguishable from zero. Yet, this does not directly address the generalizability concern, as both conceiving and failing to conceive may worsen mental health relative to not attempting conception.

To directly address potential mental health and relationship impacts, I modify the baseline approach to bound effects for women who, in the event of ACP failure, would (i) remain childless, (ii) continue cohabiting with their partners, and (iii) not initiate antidepressant use. In practice, this involves treating women who separate or begin antidepressant use after ACP failure as if they had non-ACP children. This adjustment widens the bounds but isolates the effect among women not facing severe consequences of failed conception. By restricting attention to this group, the approach also relaxes the monotonicity assumption, allowing violations among those experiencing poor mental health or separation. If the adjusted bounds remain close to baseline, it suggests that these severe consequences have limited empirical relevance for the labor market impacts.

Figure 7 presents the results for women's outcomes. Over the sample period, the adjustment reduces the relier group by up to 20%, primarily due to couples no longer cohabiting. In the early years, the bounds remain close to the baseline but widen over time. The share of within-

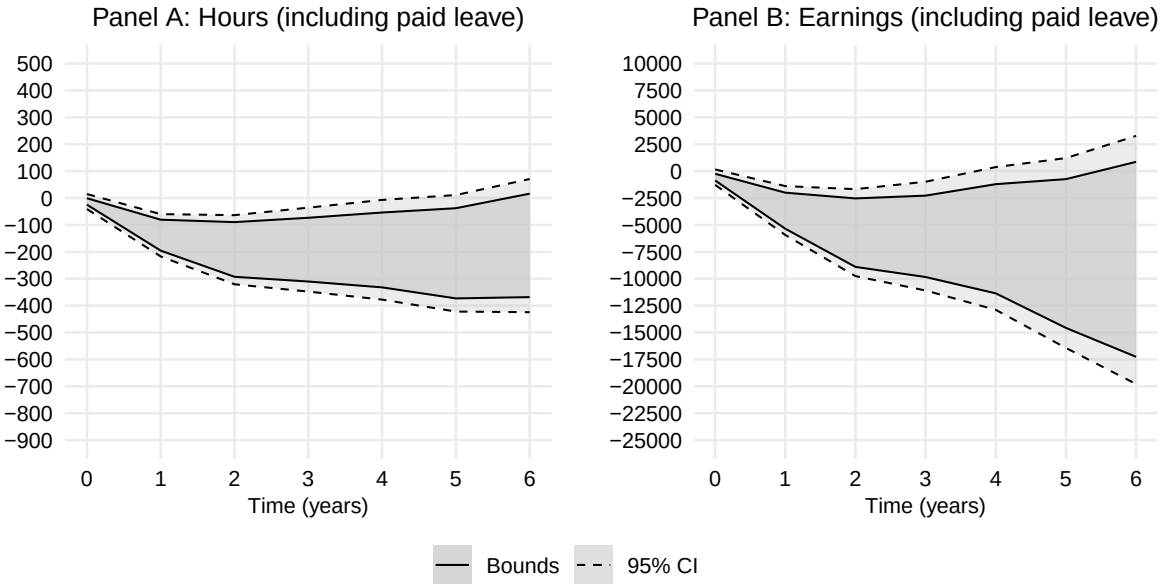


Figure 7: Effects on Women Less Affected by Failed Conception

Note: Effects of motherhood on women's annual work hours and earnings (in EUR), estimated using the baseline specification but excluding women who would initiate antidepressant use or separate from their partner after failing to conceive; see Section 6.4 for procedure details. Time relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

couple gender inequality possibly caused by parenthood increases to a maximum of 63% for work hours and 58% for earnings (not reported). Thus, even under a conservative adjustment, the results change only modestly, suggesting that severe mental health issues and separation may play a limited role in driving the estimates. Nonetheless, my results may serve as a lower bound on the effects of parenthood relative to voluntary childlessness, as mental health and relationship difficulties following failed conception—even if not severe enough to lead to antidepressant use or separation—may still influence labor market outcomes.

7 Conclusion

This paper develops a method for estimating treatment effects in quasi-experimental settings with dynamic non-compliance, where individuals may repeatedly opt into assignment or receive treatment through selective pathways. The approach requires no assumptions about how effects vary across individuals or with treatment timing. It addresses the bias in conventional IV methods that restrict timing-dependent effects and identifies impacts for a broader, policy-relevant population—those whose eventual treatment status depends on final rather than initial assignment.

The method can be applied in various settings settings, including educational programs with multiple admission cycles, job training initiatives where unassigned individuals can reapply, legal settings where individuals are assigned to authorities with varying propensities to sanction and where unsanctioned individuals may reoffend and face future sanctions, and clinical trials in extension phases where participants may enroll in other trials or pursue alternative therapies.

I apply the method to study the career impacts of parenthood, leveraging quasi-experimental variation in the success of intrauterine insemination. Using Dutch administrative data, I find that parenthood persistently reduces women's annual work hours by 10–22% and earnings by 10–28%. These effects account for up to half of the within-couple gender inequality observed after childbirth. I further demonstrate that estimates based on existing methods may be substantially biased, as they do not allow effects to vary with both the timing of parenthood and the duration since birth.

My findings suggest that there is considerable scope for policies to reduce gender inequality by mitigating the direct effects of parenthood. However, such interventions may be insufficient, as a substantial share of the inequality may persist even in the absence of children. Another parenthood-related factor that may contribute to gender inequality is the anticipation of having children, which may shape women's career trajectories independently of actual parenthood. Investigating the extent of these anticipatory responses and identifying effective policies to influence them remains an important direction for future research.

Appendix

A1 Proof of Theorem

To simply exposition, let $X_j^* \in \mathcal{X}_j^*$ contain X_j and $1_{A\geq j}$, and define $\mathcal{X}_j^{*1} = \{x \in \mathcal{X}_j^* : 1_{A\geq j} = 1\}$. Further, define $e_j^*(x) = \Pr(Z_j = 1|X_j^*)$ and $Z_l^* = (1 - Z_l)/(1 - e^*(X_l))$. Note that $e_j^*(x) = 0$ and $Z_j = 0$ if $1_{A\geq j} = 0$. Hence, Assumption 4 implies $\mathbb{E}[Z_l^*|X_l^*] = 1$. Moreover, by definition, $\Pi_j^A \frac{(1-Z_j)}{(1-e_j(X_j))} = \Pi_j^{\overline{w}} Z_j^*$.

Corollary. *Under Assumption 3:*

$$(Y_1(1), Y_0(0), R^+, R, W, Z_1, \ldots, Z_{j-1}, X_1^*, \ldots, X_{j-1}^*) \perp\!\!\!\perp Z_j^*|X_j^* \text{ for all } j > 1. \tag{15}$$

Proof of Corollary. X_j^* includes $1_{\{A\geq j\}}$, and when $A < j$, we have $Z_j = 0$, which covers the cases when $X_j^* \in \mathcal{X}_j^* \setminus \mathcal{X}_j^{*1}$. The remainder follows from Assumption 3, since $1_{\{A\geq j\}}$, $Z_1, \ldots, Z_{j-1}$, $X_1^*, \ldots, X_{j-1}^*$, and $e_j^*(X_j^*)$ are fixed given X_j^*. □

Lemma. *For any for l s.t. $1 \leq l \leq \overline{w}$ and any measurable function $g(M_l)$, where $M_l = (Y_1(1), Y_0(0), R^+, R, W, Z_1, \ldots, Z_l, X_1^*, \ldots, X_l^*)$, under Assumptions 3 and 4:* $\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} Z_j^* \middle| X_l^*\right] = \mathbb{E}\left[g(M_l)|X_l^*\right]$.

Proof of Lemma. For any l s.t. $l < \overline{w}$:

$$\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} Z_j^* \middle| X_l^*\right] = \mathbb{E}\left[\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} Z_j^* \middle| X_{\overline{w}}^*\right] \middle| X_l^*\right] \tag{16}$$

$$= \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}-1} Z_j^* \, \mathbb{E}\left[Z_{\overline{w}}^* | X_{\overline{w}}^*\right] \middle| X_l^*\right] \tag{17}$$

$$= \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}-1} Z_j^* \middle| X_l^*\right] \tag{18}$$

$$= \mathbb{E}\left[g(M_l)|X_l^*\right], \tag{19}$$

where (16) holds by law of iterated expectations and because X_j^* includes X_l^* for $j \geq l$, (17) holds

by the Corollary, (18) holds because $\mathbb{E}[Z_l^*|X_l^*] = 1$ under Assumption 4, and where (19) follows from steps similar to (16) through (18) 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 $\mathbb{E}\left[Y(1-D^+)\Pi_{j=1}^{\overline{w}} Z_j^*\right] / \mathbb{E}[r(X_1^*)] = \mathbb{E}[Y_0(0)|R=1]$. Note that:

$$\begin{aligned}
\mathbb{E}\left[Y(1-D^+)\Pi_{j=1}^{\overline{w}} Z_j^*\right] &= \mathbb{E}[Y_0(0)R\Pi_{j=1}^{\overline{w}} Z_j^*] && (20)\\
&= \mathbb{E}\left[\mathbb{E}\left[Y_0(0)R\Pi_{j=1}^{\overline{w}} Z_j^* \middle| X_1^*\right]\right] && (21)\\
&= \mathbb{E}\left[\mathbb{E}\left[Y_0(0)RZ_1^* | X_1^*\right]\right] && (22)\\
&= \mathbb{E}\left[\mathbb{E}\left[Y_0(0)R | X_1^*\right]\mathbb{E}\left[Z_1^* | X_1^*\right]\right] && (23)\\
&= \mathbb{E}\left[Y_0(0)R\right] && (24)\\
&= \mathbb{E}\left[Y_0(0)|R=1\right]\Pr(R=1), && (25)
\end{aligned}$$

where (20) follows from the definitions, (21) holds by law of iterated expectations, (22) holds by Lemma, (23) holds by Assumption 3, and (24) holds because $\mathbb{E}[Z_l^*|X_l^*] = 1$ under Assumption 4. Next, note that:

$$\begin{aligned}
\mathbb{E}\left[r(X_1)|X_1\right] &= \mathbb{E}\left[R\Pi_{j=1}^{\overline{w}} Z_j^* \middle| X_1^*\right] && (26)\\
&= \mathbb{E}\left[RZ_1^* | X_1^*\right] && (27)\\
&= \Pr(R=1|X_1^*), && (28)
\end{aligned}$$

where (26) follows from definitions, (27) holds by Lemma, and where (28) is obtained using that $\mathbb{E}[Z_l^*|X_l^*] = 1$ under Assumption 4 and applying Assumption 3. Since $\mathbb{E}[\Pr(R=1|X_1^*=x)] = \Pr(R=1)$, the result holds.

It remains to show that $\mathbb{E}\left[Y(1-D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}} Z_1/e_1^*(X_1^*)\right] / \mathbb{E}[r(X_1^*)]$ is a sharp upper bound for $\mathbb{E}[Y_1(1)|R=1]$. 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 $\gamma_x = \mathbb{E}[Y|Z_1=1, D^+=0, Y \geq q(1-p(x),x), X_1^*=x]$. I next demonstrate that γ_x is a sharp upper bound for $\mathbb{E}[Y_1(1)|X_1^*=x, R=1]$. Using that $p(x) = \Pr(R=1|D^+=0, Z_1=1, X_1^*=x)$, Corollary 4.1 in 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 $Y = Y_1(1)|Z_1=1$ and by Assumption 3, $\mathbb{E}[Y|Z_1=1, D^+=0, R=1, X_1^*=x] = \mathbb{E}[Y_1(1)|X_1^*=x, R=1]$, meaning that γ_x is an upper bound for $\mathbb{E}[Y_1(1)|X_1^*=x, R=1]$. Since $p(x)$ is identified and $Y_1(1)$ is observed only among those whose first ACP succeeded Corollary 4.1 in Horowitz & Manski (1995) implies sharpness.

Let $f_{x|R=1}(x)$ be the p.d.f. of X_1^* conditional on $R=1$. Applying Bayes rule for densities

to $\Pr(R=1|X_1^*=x)$ identified by $r(x)$ and p.d.f. of X_1^* identified directly identifies $f_{x|R=1}(x)$, making $\int_{\mathcal{X}_1^*}\gamma_x f_{x|R=1}(x)dx$ the sharp upper bound for $\mathbb{E}[Y_1(1)|R=1]$.

The last step is to show that:

$$\int_{\mathcal{X}_1^*}\gamma_x f_{x|R=1}(x)dx = \mathbb{E}[Y(1-D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}}Z_1/e_1^*(X_1^*)]/\,\mathbb{E}[r(X_1^*)]. \tag{29}$$

Using the law of iterated expectations and the definition of conditional probability:

$$\mathbb{E}[Y(1-D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}}Z_1/e_1^*(X_1^*)] \tag{30}$$
$$= \mathbb{E}[\mathbb{E}[\gamma_{X_1^*}|X_1^*]\Pr(D^+=0,Z_1=1,Y>q(1-p(X_1^*),X_1^*)|X_1^*)/e_1^*(X_1^*)]. \tag{31}$$

Applying the definition of conditional probability twice and the definition of $p(X_1^*)$:

$$\Pr(D^+=0,Z_1=1,Y>q(1-p(X_1^*),X_1^*)) = r(X_1^*)e_1^*(X_1^*) \tag{32}$$

Thus:

$$\mathbb{E}[Y(1-D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}}Z_1/e_1^*(X_1^*)] = \mathbb{E}[\gamma_{X_1^*}r(X_1^*)]. \tag{33}$$

Applying Bayes rule for densities: $\mathbb{E}[\gamma_{X_1^*}r(X_1^*)] = \int_{\mathcal{X}_1^*}\gamma_x f_{x|R=1}(x)dx\Pr(R=1)$. Since $\mathbb{E}[r(X_1^*)] = \Pr(R=1)$, the statement holds.

□

A2 Estimating the Bounds

The moment functions given in Table A1 identify the same parameters as the baseline moments:

$$\mathbb{E}[\psi^{L+}(G,\xi^0)] = \mathbb{E}[m^L(G,\eta^0)],\ \mathbb{E}[\psi^{U+}(G,\xi^0)] = \mathbb{E}[m^U(G,\eta^0)]. \tag{34}$$

However, the original 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 $x_j^* \in \mathcal{X}_j^{*1}$ (see Appendix A1 for the definitions of $e_j^*(x_j^*)$ and X_j^*). Define $r \in [0,1) \to \psi^{U+}(G,r) \equiv \psi^{U+}(G,\xi_r)$, where:

$$\xi_r = \{e_1^*(x_1^*),\ldots,e_l^*(x_l^*,r),\ldots,e_{\overline{w}}^*(x_{\overline{w}}^*),r_1(x_1^*),\ldots,r_{\overline{w}}(x_{\overline{w}}^*),r^+(x_1^*),q(p(x_1^*),x_1^*), \tag{35}$$
$$q(1-p(x_1^*),x_1^*),\beta_1(x_1^*),\ldots,\beta_{\overline{w}}(x_{\overline{w}}^*),\beta^+(x_1^*),z^{U+}(x_1^*),z^{L+}(x_1^*)\}, \tag{36}$$

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, $\partial_r\,\mathbb{E}[\psi^{U+}(G,\xi_r)\mid X_l^*]\big|_{r=0} = 0$ almost surely, while for the original moment, $\partial_r\,\mathbb{E}[m^U(G,\eta_r)\mid X_l^*]\big|_{r=0} \neq 0$ almost surely. I use this property together with sample splitting to justify asymptotic inference as if the nuisance parameter were known, appealing to the argument in Semenova (2023), who formally establish the result for the Lee (2009) setting. A formal proof of orthogonality in my context is available upon request. It involves repeated application of the corollary and substitution of the nuisance function definitions to show that $\mathbb{E}[\psi^{U+}(G,\xi_r)\mid X_l^*]$ does not depend on $e_j^*(X_j^*)$.

Table A1: Orthogonal Moments

	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
$\xi^0(x_1^*,\ldots,x_{\overline{w}}^*)$	$\{e_1^*(x_1^*),\ldots,e_{\overline{w}}^*(x_{\overline{w}}^*), r_1(x_1^*),\ldots,r_{\overline{w}}(x_{\overline{w}}^*), r^+(x_1^*), q(p(x_1^*),x_1^*), q(1-p(x_1^*),x_1^*),$ $\beta_1(x_1^*),\ldots,\beta_{\overline{w}}(x_{\overline{w}}^*), \beta^+(x_1^*), z^{U+}(x_1^*), z^{L+}(x_1^*)\}$
$r_k(x)$	$\mathbb{E}[(1-D^+)/(\Pi_{j=k+1}^{A}(1-e_j^*(X_j^*))) \mid X_k^*=x, Z_A=0]$ $\mathbb{E}[\Pi_{j=k+1}^{A}(1-e_j^*(X_j^*)) \mid X_k^*=x, Z_A=0]$
$\beta_k(x)$	$\mathbb{E}[Y/(\Pi_{j=k+1}^{A}(1-e_j^*(X_j^*))) \mid X_k^*=x, D=0]$ $\mathbb{E}[\Pi_{j=k+1}^{A}(1-e_j^*(X_j^*)) \mid X_k^*=x, D=0]$
$\beta^+(x)$	$\mathbb{E}[Y \mid X_1^*=x, Z_1=1, D^+=0]$
$z^{U+}(x)$	$\mathbb{E}[Y \mid X_1^*=x, Z_1=1, D^+=0, Y\geq q(1-p(x),x)]$
$z^{L+}(x)$	$\mathbb{E}[Y \mid X_1^*=x, Z_1=1, D^+=0, Y\leq q(p(x),x)]$

The estimator for θ_b, for $b\in\{L,U\}$, is given by:

$$\widehat{\theta_b} = \left(\Sigma_i\left(\psi^{b+}(G_i,\widehat{\xi_i})1_{\{p(X_1^*)\leq 1\}} + \psi^{-}(G_i,\widehat{\xi_i})1_{\{p(X_1^*)>1\}}\right)\right) / \left(\Sigma_i\psi^{R}(G_i,\widehat{\xi_i})\right) \tag{37}$$

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. ψ^{b+} is the orthogonal counterpart of m^b, and ψ^R is the orthogonal counterpart of r. ψ^- covers the case where the estimated relier share exceeds the subsequent relier share. Under monotonicity, such reversals may occur in estimation when the true shares are close. ψ^- treats these instances as if the two shares were equal. Specifically, it is constructed such that under the assumptions in the theorem $\mathbb{E}[\psi^-(G,\xi)]/\mathbb{E}[r(X_1)] = \mathbb{E}[Y_1(1) \mid R^+=1] - \mathbb{E}[Y_0(0) \mid R=1]$. I address potential monotonicity violations further in Appendix SA1.

I use 3-fold cross-fitting, meaning that in each split, two-thirds of the sample are used to estimate the nuisance parameters. Propensity scores are estimated using logistic regressions, with second-order polynomials of women's and partners' ages at the time of the procedure, interacted

with treatment-type dummies (IUI or IVF), and separate dummies for each partner holding at least a bachelor's degree. The remaining nuisance functions are estimated using Generalized Random Forests (Athey et al., 2019). I estimate z_t^{U+} and z_t^{L+} by trimming the sample above or below estimated quantiles and estimating conditional expectations. While nonparametric results for truncated expectations exist (Olma, 2021), no public implementation is currently available. The covariates in X_1^* include the woman's and their partner's earnings 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. Outcomes are made continuous by adding small noise $u \sim U(0, 0.001)$ to avoid ties. Confidence intervals are constructed following Heiler (2024), based on Stoye (2020).

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., Ferman, B., Gao, C., Hull, P., Tecchio, O. L., & Yeh, R. W. (2024). Instrumental variables with time-varying exposure: New estimates of revascularization effects on quality of life. *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. *Handbook of Labor Economics*, *4*, 1543–1590.

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. *American Economic Review*, *84*(5), 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.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, *86*(5), 1867–1900.

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*.

Ferman, B., & Tecchio, O. (2023). Identifying dynamic lates with a static instrument. *arXiv preprint arXiv:2305.18114*.

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.

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*, 281–302.

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. *The Review of Economic Studies*, rdae104.

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

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

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

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

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *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

Olivetti, C., Pan, J., & Petrongolo, B. (2024). The evolution of gender in the labor market. *Handbook of Labor Economics*, *5*, 619–677.

Olma, T. (2021). Nonparametric estimation of truncated conditional expectation functions. *arXiv preprint arXiv:2109.06150*.

Rabaté, S., & Rellstab, S. (2022). What determines the child penalty in the netherlands? the role of policy and norms. *De Economist*, *170*(2), 195–229.

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.

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.

Supplementary Appendix for "Parenthood Timing and Gender Inequality"

SA1 Robustness and Extensions

Appendix SA1.1 presents an extension to bound the effects over time for a stable group. Appendix SA1.2 tests the monotonicity assumption, while Appendix SA1.3 reports the main estimates allowing the direction of monotonicity to vary with covariates. Appendix SA1.4 presents estimates of work hours adjusted for potential parental leave. Appendix SA1.5 addresses age differences between partners. Appendix SA1.6 presents estimates of the impact on antidepressant uptake.

SA1.1 Stable Relier Group and Anticipation

The evolution of my main estimates over time reflects two factors. First, how the effect of parenthood changes with time spent in parenthood. Second, how effects differ between women who remain reliers for different durations, as the relier group shrinks over time. To address this, I adapt my approach to bound effects for women who remain reliers through the final period. This is feasible because fertility is irreversible. Specifically, this implies two things. First, at each point in time, the average past control outcomes for reliers can be identified using the outcomes of women who are still childless at that time. Second, their average treated outcomes at each moment since the first ACP can be bounded by assuming that the remaining reliers were either at the top or bottom of the earlier treated outcome distribution. In practice, this amounts to estimating the baseline specification using fertility outcomes from the final period and labor market outcomes from previous periods. Because control outcomes are identified solely using women who remain childless, this also addresses concerns about baseline or conventional estimates being biased by the fact that women in the control group may anticipate future parenthood. Figure SA1 presents estimates for women who remain reliers six years after their first ACP, which are comparable to the baseline results.

SA1.2 Testing Monotonicity

Monotonicity states that all reliers are subsequent reliers, implying that the relier share is at least as large as the subsequent relier share: $\Pr(R^+ = 1) \geq \Pr(R = 1)$. Figure SA2 plots the estimated shares over time, showing that the subsequent relier share consistently exceeds the relier share, in line with monotonicity.

Monotonicity further implies that the relier share is at least as large as the subsequent relier share at each covariate value: $r^+(X_1^*) \geq r(X_1^*)$. Since the conditional shares are estimated non-parametrically, formally testing whether their differences allow rejecting monotonicity is not trivial, but comparing them offers some insight. The Panel A in Figure SA3 plots the empirical distribution of the difference between estimated conditional subsequent relier and relier shares in year 6. For 31% of observations, the difference is below zero. While this would contradict monotonicity if observed in the true shares, such deviations can result from estimation error when the shares are close. Namely, when all subsequent reliers are reliers $\Pr(R^+ = R) = 1$, the difference should be

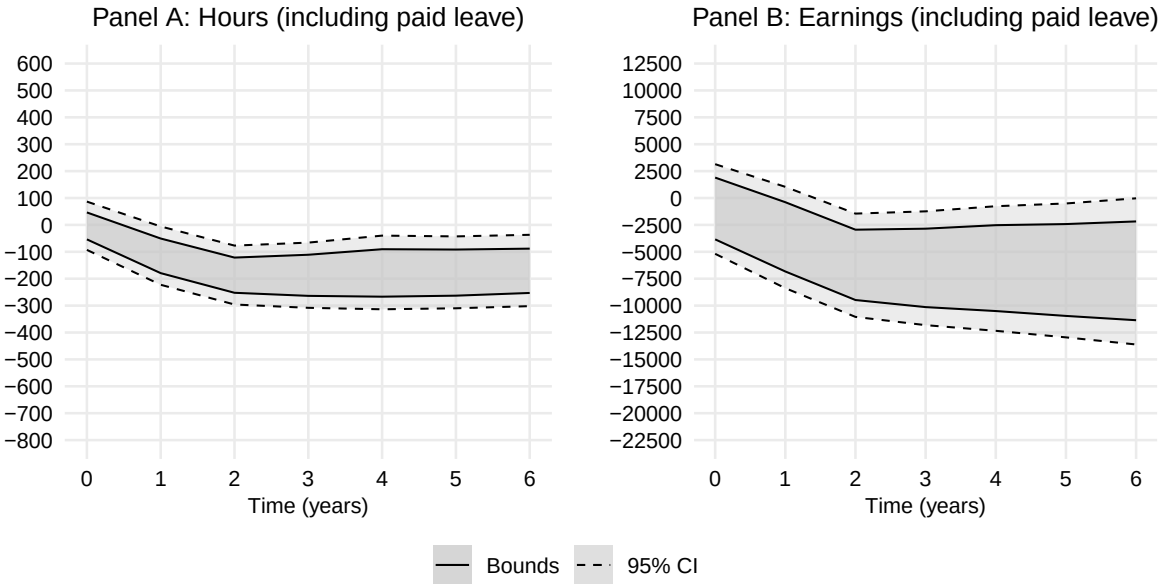


Figure SA1: Effects Women who Remain Childless

Note: Effects of motherhood on women's annual work hours and earnings (in EUR), estimated using the baseline specification. Time relative to the first intrauterine insemination; fertility outcomes measured at the end of the sample period. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

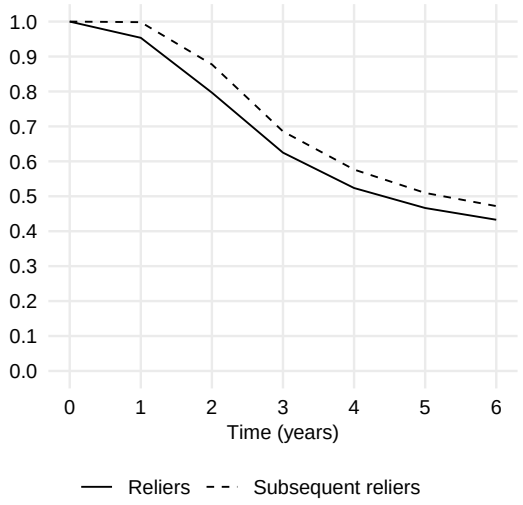


Figure SA2: Estimated Relier and Subsequent Relier Shares

Note: Relier and subsequent relier shares over time, estimated using the baseline specification. Time relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

below zero for 50% of observations. Consistent with this, the differences are generally small, with only 7% of observations below -0.1, suggesting no clear monotonicity violations.

To formally test monotonicity using covariates, I adapt the approach of Semenova (2023). I partition the sample into $J = 25$ discrete cells $\mathcal{X}_{1,j}^*$ based on quintiles of women's work hours and age

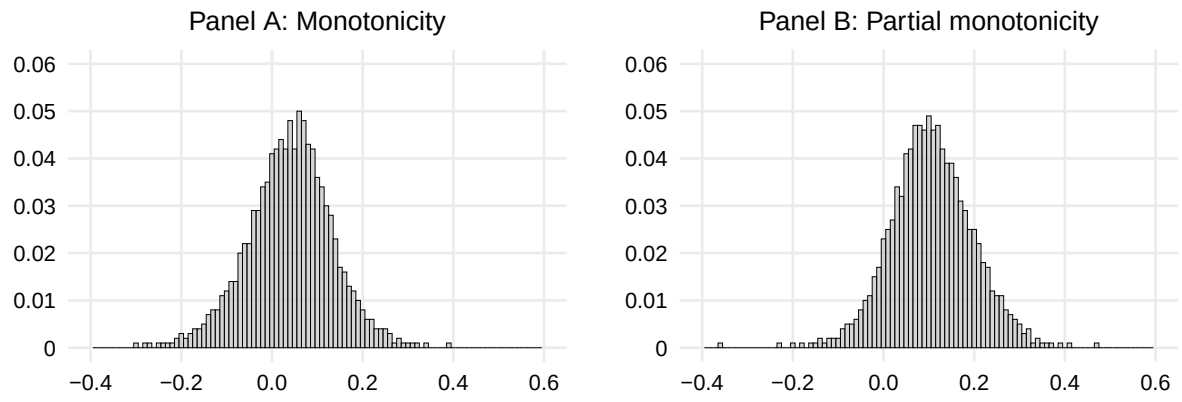


Figure SA3: Histogram of Difference in Subsequent Relier and Relier Shares

Note: Histogram of the difference in covariate-conditional subsequent relier and relier shares six years after the first intrauterine insemination. *Monotonicity* – estimated using the baseline specification. *Partial Monotonicity* – treating women who uptake antidepressants or separate from their partner after failing to conceive as if they had conceived naturally, and then applies the baseline specification. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

in the year prior to their first IUI. Since these two covariates are highly predictive of the remaining covariates used in the analysis, including additional ones results in small cells (e.g., almost no women have extremely high work hours while having extremely low earnings). Monotonicity implies that the subsequent relier share is at least as large as the relier share in each cell, meaning that each value in the vector $\mu = (\mathbb{E}[r^+(X_1^*) - r(X_1^*) \mid X_1^* \in \mathcal{X}_{1,j}^*])_{j=1}^J$ must be non-negative. The null hypothesis is $-\mu \leq 0$, and the test statistic is $\max_{j \in J}^* \frac{-\widehat{\mu}_j}{\widehat{\sigma}_j}$. The critical value is the self-normalized critical value of Chernozhukov et al. (2019). σ_j are estimated using multiplier bootstrap with 100 draws and weights $w_i \sim \exp(1)$ to account for the uncertainty in the estimation of propensity scores (results remain unchanged when treating scores as fixed). Consistent with the results in Figure SA3, in 28% of cells, $\widehat{\mu}_j$ in year 6 is negative. However, the p-value for the test statistic is 0.54, indicating that these differences are not statistically significant, failing to reject monotonicity. Using women's earnings instead of hours to partition the sample yields similar results.

The Panel B in Figure SA3 repeats the above when restricting focus to reliers who would remain cohabiting with their partners and not uptake antidepressants after failing to conceive (and treating those who separate from their partners or take up antidepressants after ACP failure as if they had children, excluding them from the population covered by the bounds). The estimated difference between the subsequent relier and reliers shares is below zero for only 11% of observations and below -0.1 for just 1%, providing stronger support for monotonicity in this group.

SA1.3 Relaxing Monotonicity

To address violations of monotonicity in the Lee (2009) setting, Semenova (2023) relaxes the assumption by allowing its direction to vary with X_1^*. To test the sensitivity of my results, I adopt an equivalent approach in my setting. If the reversal of the estimated relier and subsequent relier shares occurs because the true shares are very close, this method and the baseline approach should

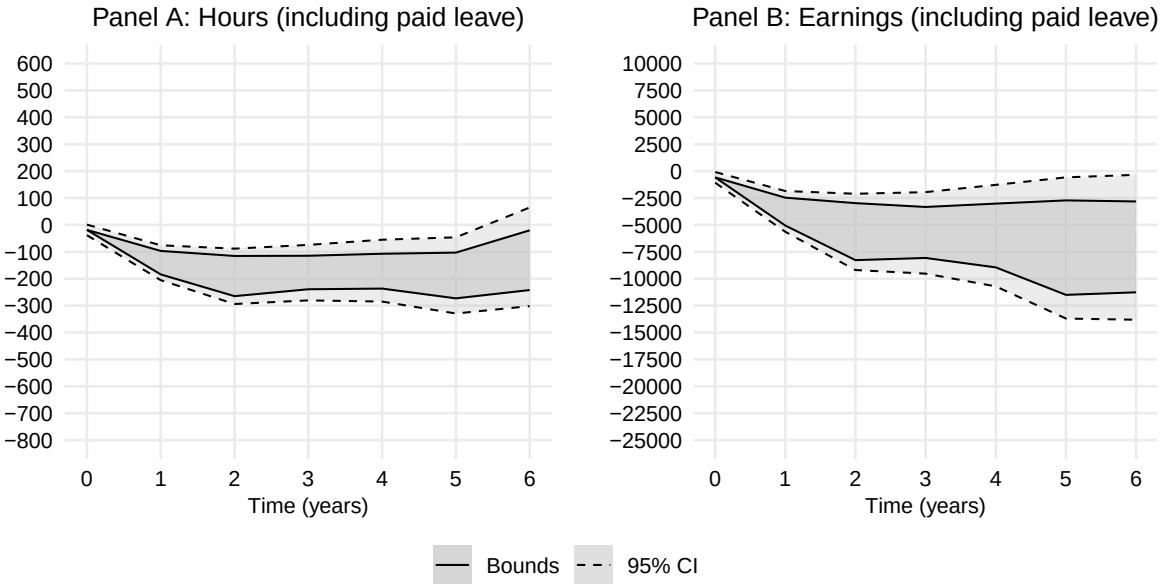


Figure SA4: Effect on Women Under Covariate-Conditional Monotonicity

Note: Effects of motherhood on women's annual work hours and earnings (in EUR), estimated allowing the direction of monotonicity to vary with covariates; see Appendix SA1.3. Time relative to first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

yield nearly identical results.

Define $\mathcal{X}^*_{help} \equiv \{x : r^+(x) \geq r(x)\}$ and $\mathcal{X}^*_{hurt} \equiv \mathcal{X}^*_1 \backslash \mathcal{X}^*_{help}$. 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 SA1 presents the moments for the case when $X^*_1 \in \mathcal{X}^*_{hurt}$. The estimator for θ_b, for $b \in \{L, U\}$, is then given by:

$$\widehat{\theta}_b = \frac{\sum_i \left(\psi^{b+}(G_i, \widehat{\zeta}_i) 1_{\{p(X^*_1)\leq 1\}} + \psi^{b-}(G_i, \widehat{\zeta}_i) 1_{\{p(X^*_1)>1\}}\right)}{\sum_i \left(\psi^{R}(G_i, \widehat{\zeta}_i) 1_{\{p(X^*_1)\leq 1\}} + \psi^{R+}(G_i, \widehat{\zeta}_i) 1_{\{p(X^*_1)>1\}}\right)}. \tag{38}$$

I implement it following the baseline approach. Since a weighted generalized quantile forests estimator is not available, I estimate q^0 using quantile regressions. Figure SA4 presents the estimates for women's outcomes, which closely resemble the baseline results.

SA1.4 Work Hours for Unobserved Leave

I define maximum-leave-adjusted hours by scaling women's reported work hours in each childbirth year by a factor of 36/52, accounting for up to 16 weeks of leave. Since the control group consists of women without children, this adjustment ensures that the estimated effect on actual hours worked falls within the intersection of the bounds obtained using reported and adjusted values. Figure SA5 shows the results for female work hours and the corresponding gender gap. The estimates change little beyond the first year of parenthood.

SA1.5 Age Difference Between Partners

My estimates of the share of within-couple gender inequality caused by parenthood focus on the within-couple gender gap in each year after becoming parents. This gap also reflects within-couple age differences, which may distort its relation to aggregate gender inequality, as men's outcomes

Table SA1: 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
$\zeta^0(x_1^*,\ldots,x_{\overline{w}}^*)$	$\{e_1^*(x_1^*),\ldots,e_{\overline{w}}^*(x_{\overline{w}}^*),r_1(x_1^*),\ldots,r^{\overline{w}}(x_{\overline{w}}^*),r^+(x_1^*),q(p(x_1^*),x_1^*),q(1-p(x_1^*),x_1^*),$ $\beta^1(x_1^*),\ldots,\beta^{\overline{w}}(x_{\overline{w}}^*),\beta^+(x_1^*),z^{U+}(x_1^*),z^{L+}(x_1^*),z_1^{U-}(x_1^*),\ldots,z_{\overline{w}}^{U-}(x_{\overline{w}}^*),q^0(1/p(x_1^*),x_1^*),$ $q^0(1-1/p(x_1^*),x_1^*),z_1^{L-}(x_1^*),\ldots,z_{\overline{w}}^{L-}(x_{\overline{w}}^*),r_1^L(x_1^*),\ldots,r_{\overline{w}}^L(x_{\overline{w}}^*),r_1^U(x_1^*),\ldots,r_{\overline{w}}^U(x_{\overline{w}}^*)\}$
$q^0(u,x)$	$inf\{q:u\leq\mathbb{E}[1_{\{Y\leq q\}}/\Pi_{j=2}^{\overline{w}}(1-e_j^*(X_j^*))\mid X_1^*=x,D=0]/$ $\mathbb{E}[\Pi_{j=2}^{\overline{w}}(1-e_j^*(X_j^*))\mid X_1^*=x,D=0]\}$
$z_k^{L-}(x)$	$\mathbb{E}[Y/\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert Y\geq q^0\left(1-1/p(X_1^*),X_1^*\right),D=0,X_k^*=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert Y\geq q^0\left(1-1/p(X_1^*),X_1^*\right),D=0,X_k^*=x]$
$z_k^{U-}(x)$	$\mathbb{E}[Y/\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert Y\leq q^0\left(1/p(X_1^*),X_1^*\right),D=0,X_k^*=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert Y\leq q^0\left(1/p(X_1^*),X_1^*\right),D=0,X_k^*=x]$
$r_k^L(x)$	$\mathbb{E}[1_{Y>q^0(1-1/p(X_1^*),X_1^*)}/\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert D=0,X_k^*=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert Y\leq q^0\left(1/p(X_1^*),X_1^*\right),D=0,X_k^*=x]$
$r_k^U(x)$	$\mathbb{E}[1_{Y<q^0(1/p(X_1^*),X_1^*)}/\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert D=0,X_k^*=x]$ $\mathbb{E}[\Pi_{j=k+1}^{\overline{w}}(1-e_j^*(X_j^*)\vert Y\leq q^0\left(1/p(X_1^*),X_1^*\right),D=0,X_k^*=x]$

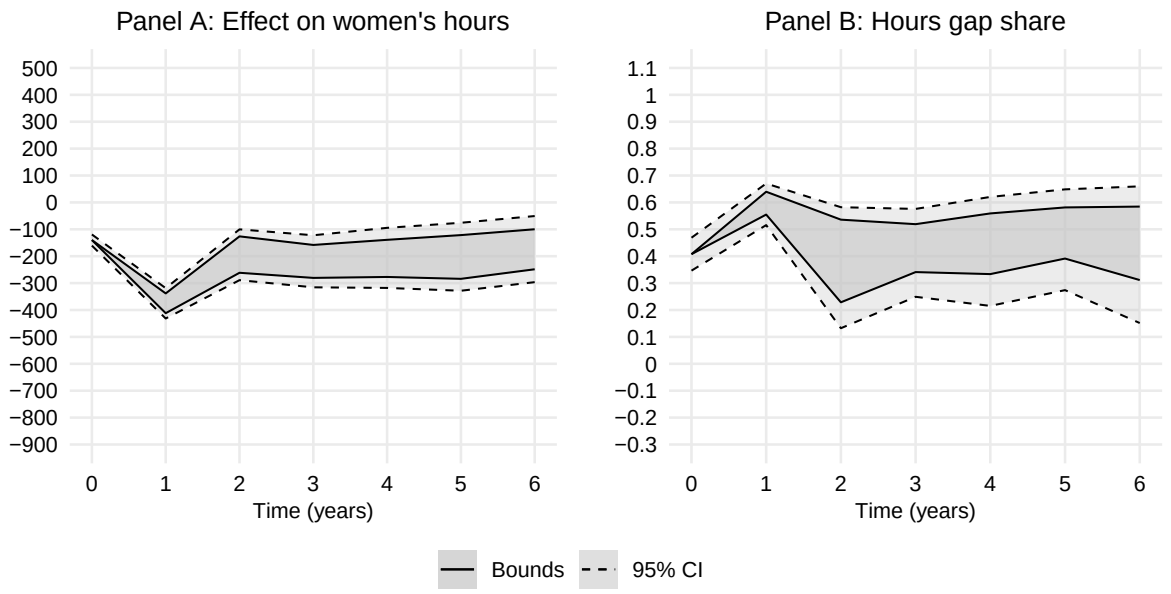


Figure SA5: Leave-Adjusted Estimates of Work Hours

Note: The effects of parenthood on women's annual work hours and the share of within-couple gender inequality in hours caused by parenthood, adjusting for up to 16 weeks of maternity leave; see Appendix SA1.4 for details. Time relative to the first intrauterine insemination (IUI). Sample includes all couples in which the woman underwent IUI for their first child between 2013 and 2016 and was cohabiting with a male partner in the year prior to the first procedure.

are measured at systematically older ages. If work hours and earnings rise with age, this could lead my estimates to understate the aggregate contribution of parenthood.

Ideally, cumulative lifetime outcomes would address this concern, but such data is unavailable. Parametric age corrections would opaque assumptions. Instead, I adopt a simple adjustment: I lag men's outcomes to match the woman's age (e.g., by two years if she is two years younger), ensuring that gender gaps reflect comparable life-cycle stages. For some men, this implies using pre-parenthood outcomes, which is appropriate if fatherhood effects are small. The adjustment reduces the sample by 19%, yielding 10,310 observations.

Figure SA6 presents the results, showing that the adjustment has little impact on the estimates. The upper bound on the share of within-couple gender inequality in work hours decreases to at most 50% in each year, while the bound for earnings increases by no more than 10 percentage points per year.

SA1.6 Effects on Mental Health

To maximize precision in estimating the impact on antidepressant uptake, I use the sequential IV approach where outcome is uptaking antidepressants in a given year. In the absence of timing-dependent effects, it identifies τ_{ATR}. Figure SA7 presents the results, the effects are precisely estimated and indistinguishable from zero.

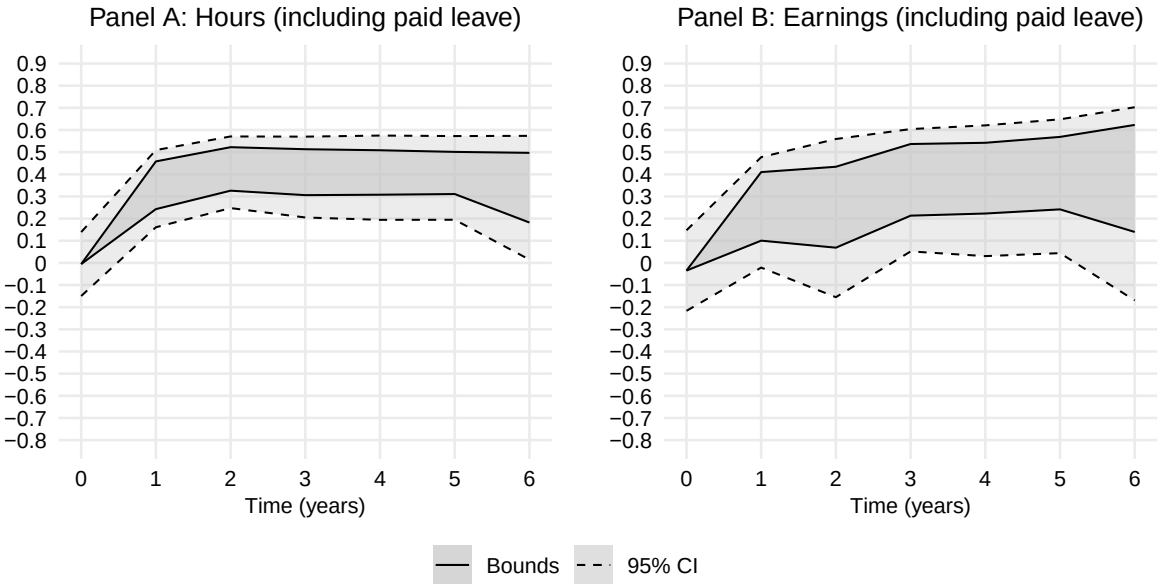


Figure SA6: Age-adjusted Share of Within-couple Gender Inequality Caused by Parenthood

Note: Share of within-couple gender inequality caused by parenthood. Calculated as $1 - a/b$, where a is the average gap in the control outcome and b is the lower or upper bound for the average treated outcome, both estimated using orthogonal moments from the baseline specification. Confidence intervals are based on the Delta method. Time relative to the woman's first intrauterine insemination, and outcomes for both men and women are measured at the specific age that the woman is at that moment. Sample includes all couples in which the woman underwent intrauterine insemination for their first child between 2013 and 2016, was cohabiting with a male partner in the year prior to the first procedure, and for whom partner outcomes in the respective year are observed (10,310 observations).

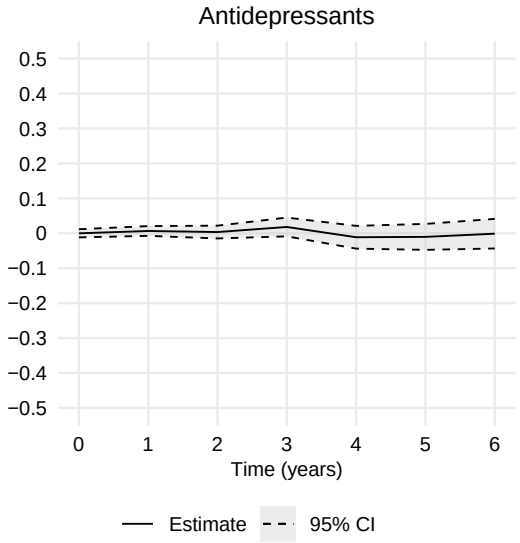


Figure SA7: Effects on Antidepressant Uptake

Note: Effects of motherhood on antidepressant uptake estimated using the sequential instrumental variable approach; see Appendix SA2. Time is measured relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

SA2 Auxiliary Estimation Details

Here I discuss implementation details for the conventional IV, sequential IV, recursive IV, and sequential recursive IV methods.

I implement the conventional IV estimator following Lundborg et al. (2017). The first-stage regression specification is:

$$D = Z_1\beta^{FS} + X_1\chi^{FS} + \varepsilon^{FS}, \tag{39}$$

and the second-stage regression specification is:

$$Y = \widehat{D}\beta^{IV} + X_1\chi^{IV} + \varepsilon^{IV}, \tag{40}$$

where ε^{FS} and ε^{IV} are individual-level error terms, $\widehat{D}$ denotes the fitted values from the first stage, and the coefficient β^{IV} captures the effect of parenthood.

For sequential IV, recursive IV, and sequential recursive IV methods, I first introduced auxiliary functions that will enable estimation in a double-robust manner to maximize precision:

$$g_a^{0+}(G) = \gamma_a^{0,1+}(X_1^*) + (a - \gamma^{0,1+}(X_1^*))\Pi_{j=1}^{\overline{w}}\frac{(1-Z_j)}{(1-e_j^*(X_j^*))} \tag{41}$$

$$+ \Sigma_{k=1}^{\overline{w}}\Big[1_{\{A\geq k\}}\Pi_{j=1}^{k-1}\frac{(1-Z_j)}{1-e_j^*(X_j^*)}\frac{(e_k^*(X_k^*)-Z_k)}{1-e_k^*(X_k^*)}[\gamma_a^{0,1+}(X_1^*) - \gamma_a^{0,k+}(X_k^*)]\Big] \tag{42}$$

$$g_a^0(G) = \gamma_a^0(X_1^*) + (a - \gamma_a^0(X_1^*))\frac{Z_1}{e_1^*(X_1^*)} \tag{43}$$

$$g_a^1(G) = \gamma_a^1(X_1^*) + (a - \gamma_a^1(X_1^*))\frac{1-Z_1}{1-e_1^*(X_1^*)}, \tag{44}$$

where $\gamma_a^1(X_1^*)$, $\gamma_a^0(X_1^*)$, $\gamma_a^{0,k+}(X_k^*)$ are differentiable functions.

Applying the lemma yields:

$$\frac{\mathbb{E}[g_Y^1(G) - g_Y^0(G)]}{\mathbb{E}[g_D^1(G) - g_D^{0+}(G)]} = \mathbb{E}[\tau|C=1] + \mathbb{E}[\delta|C=0]\frac{\Pr(C=0)}{\Pr(C=1)} \tag{45}$$

and:

$$\frac{\mathbb{E}[g_Y^1(G) - g_Y^{0+}(G)]}{\mathbb{E}[g_D^1(G) - g_D^{0+}(G)]} = \mathbb{E}[\tau|R=1] + \mathbb{E}[\delta|R=0]\frac{\Pr(R=0)}{\Pr(R=1)}, \tag{46}$$

which correspond to the IV and sequential IV estimands, respectively. I construct the corresponding estimators using empirical counterparts of the expectations on the left-hand side of equations (45) and (46). $\gamma_a^1(X_1^*)$ is the OLS prediction of a given X_1^*, estimated using observations with $Z_1 = 1$ and weights $1/\widehat{e}_1^*(X_1^*)$; $\gamma_a^0(X_1^*)$ is the analogous prediction using observations with $Z_1 = 0$ and weights $1/(1-\widehat{e}_1^*(X_1^*))$; and $\gamma_a^{0,k+}(X_1^*)$ is the OLS prediction of a at X_k^*, estimated using observations with $Z_1 = 0$ and $A \geq k$, and weights $1/(\Pi_{j=1}^{\overline{w}}(1-\widehat{e}_j^*(X_j^*)))$, where $\widehat{e^*}$ denotes the estimated propensity scores.

For the recursive IV approach, assume that a woman's outcome in period k since the first ACP

Table SA2: Balance in Later ACPs

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Work (W)	0.003	-0.015	0.010	-0.003	0.017	0.002	-0.007	0.016	0.022
	(0.009)	(0.010)	(0.010)	(0.011)	(0.012)	(0.016)	(0.015)	(0.015)	(0.019)
Work (P)	0.007	0.018	0.009	0.016	-0.009	-0.003	-0.017	0.005	0.040
	(0.010)	(0.010)	(0.012)	(0.012)	(0.016)	(0.015)	(0.020)	(0.021)	(0.024)
Hours (W)	11.828	-17.024	30.313	16.523	43.473	23.825	-27.325	63.826	67.821
	(17.745)	(19.650)	(19.857)	(22.052)	(24.966)	(30.057)	(31.365)	(34.966)	(41.903)
Hours (P)	16.880	16.806	27.356	29.854	-8.164	-7.812	-43.177	4.368	29.210
	(21.447)	(21.469)	(23.492)	(25.973)	(31.884)	(32.247)	(38.712)	(45.942)	(50.318)
Earnings 1000s EUR (W)	1.130	-0.343	0.908	0.938	1.369	-0.090	-0.068	0.683	1.679
	(0.661)	(0.649)	(0.746)	(0.858)	(0.952)	(0.953)	(1.102)	(1.215)	(1.732)
Earnings 1000s EUR (P)	-0.377	0.120	2.262	1.384	-0.224	-0.873	0.136	-0.322	4.169
	(0.991)	(0.917)	(1.073)	(1.217)	(1.412)	(1.381)	(1.624)	(1.781)	(3.537)
Observations	10,744	8,960	7,357	5,922	4,642	3,412	2,386	1,632	1,041
Joint p-val.	0.665	0.335	0.548	0.775	0.786	0.647	0.326	0.853	0.045

Note: Each column reports the difference in average characteristics between women whose respective ACP succeeded and those for whom it failed, among those who underwent the procedure, adjusted for age and education using inverse probability weights from the baseline specification. The sample consists of women who underwent intrauterine insemination for their first child between 2013 and 2016, with no prior assisted conception procedures, and who were cohabiting with a male partner in the year before the first procedure. Labor market outcomes measured in the year before first procedure. Earn. – earnings, (W) – woman, (P) – partner. Standard errors in parentheses.

is given by

$$Y_k = Y_{0k}(0) + \sum_{j \leq k} 1_{\{K_k = j\}} \tau_j + \varepsilon_k, \tag{47}$$

where $Y_{zk}(d)$ denotes the potential outcome in period k when the first ACP outcome is z and parenthood status is d; K_k is the number of years since first birth (taking an arbitrary negative value if the woman never has a child); and $\tau_j = \mathbb{E}[Y_{1j}(1) - Y_{0j}(0)]$ denotes the average effect of having been a parent for j years. This specification assumes that parenthood effects depend on motherhood duration but not on the timing of becoming a parent and are otherwise homogeneous between women.

To implement the recursive IV estimator, I first apply the IV estimator corresponding to equation (45) to data on Y_1 and D_1, yielding an estimate $\widehat{\tau}_1$ of τ_1. I then construct the one-year-of-parenthood-corrected outcome $\widehat{Y_k^1} = Y_k - 1_{\{K_k=1\}}\widehat{\tau}_1$. Applying the same estimator to $\widehat{Y_2^1}$ and D_1 yields an estimate $\widehat{\tau}_2$ of τ_2. I then construct the two-year-of-parenthood-corrected outcome $\widehat{Y_k^2} = \widehat{Y_k^1} - 1_{\{K_k=2\}}\widehat{\tau}_2$, allowing me to estimate τ_3, and so on. For the sequential recursive IV approach, I repeat these steps using the sequential IV estimator corresponding to equation (46).

Confidence intervals for comparing the two sets of estimates are obtained via a multiplier bootstrap with 100 draws and weights $w_i \sim \exp(1)$, applied at each step of the regressions used to estimate nuisance functions and construct the moment functions.

SA3 Additional Balance Results

Table SA2 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 IVF based on

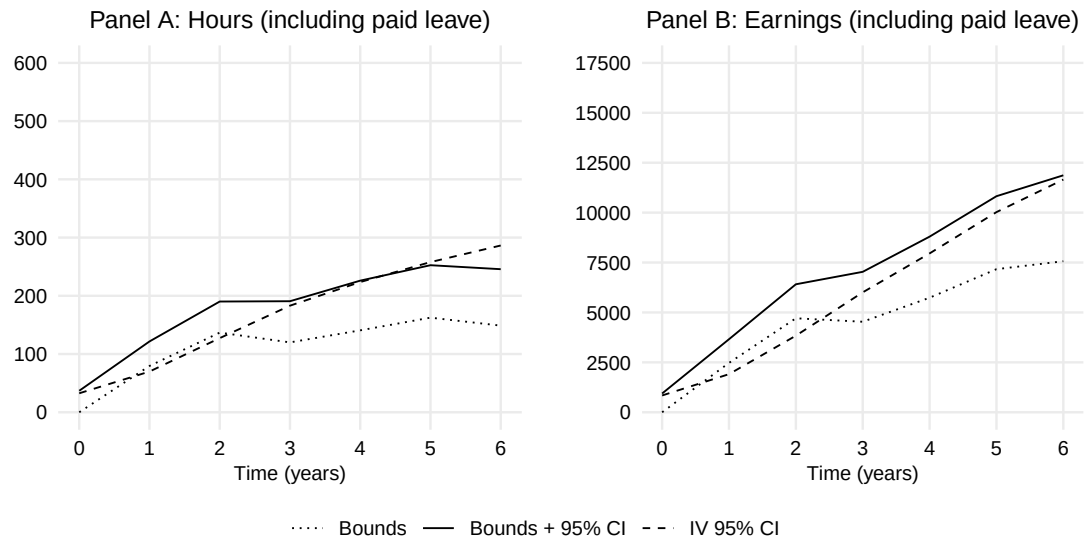


Figure SA8: 95% Confidence Interval for Effects on Women's Outcomes

Note: 95% confidence intervals for the effects of parenthood on women's labor market outcomes, estimated using different methods. *Bounds* – bounds estimated using the baseline specification, not including confidence intervals; *Bounds + 95% CI* – 95% confidence interval for the bounds estimated using the baseline specification; *IV 95% CI* – 95% confidence interval for conventional instrumental variable estimates. Time is measured relative to the first intrauterine insemination. Sample includes all women who underwent intrauterine insemination for their first child between 2013 and 2016 and were cohabiting with a male partner in the year before the first procedure.

women's types and potential outcomes. Overall, the results suggest no systematic differences in pre-IUI outcomes between those with successful and unsuccessful subsequent ACPs, supporting the conditional sequential unconfoundedness assumption.

SA4 Confidence Interval Comparison

While my method only partially identifies the effects, my estimates are substantially more precise. Figure SA8 compares the width of 95% confidence intervals for my bounds to IV estimates. The confidence intervals for the three methods are almost identical. 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.