PARENTHOOD TIMING AND GENDER INEQUALITY

JULIUS ILCIUKAS*

The Chinese University of Hong Kong, Department of Economics

October 7, 2025

Abstract

I develop a new methodology to estimate how parenthood and its timing affect labor market outcomes, leveraging variation from intrauterine insemination (IUI). The method exploits quasi-experiments where individuals not initially assigned to treatment may undergo repeated assignments, as when failed IUI attempts lead to subsequent procedures. By leveraging entire assignment sequences, it separates treatment effects (parenthood versus childlessness) from timing effects (earlier versus later childbearing). Using Dutch administrative data, I find that motherhood persistently reduces earnings (10–28%) and work hours (10–22%), causing up to half of post-childbirth gender inequality. Delayed childbearing—even when unintended—mitigates women's losses.

JEL codes: C21, C22, J13, J16,

Keywords: parenthood, gender inequality, treatment effects

*I thank Jérôme Adda, Francesco Agostinelli, Douglas Almond, Monique de Haan, Christian Dustmann, Phillip Heiler, Christine Ho, Artūras Juodis, Jura Liaukonyte, Petter Lundborg, 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, ESWC, 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 October 7, 2025.

1 Introduction

The differential impact of parenthood on women's and men's careers is widely viewed as a key driver of gender inequality in the labor market (Goldin, 2014; Bertrand, 2020; Cortés & Pan, 2023). Assessing this impact is challenging because of selection: individuals who become parents may differ systematically from those who remain childless. Many studies address selection by exploiting quasi-experimental fertility variation, such as in vitro fertilization success (Lundborg et al., 2017) or intrauterine device failure (Gallen et al., 2024). However, most women who do not conceive in the initial quasi-experiment (e.g., after a first IVF cycle or due to IUD failure) eventually become mothers. Existing work either estimates a weighted average of the effects of parenthood and its timing or identifies parenthood's impact by assuming away timing effects. Both approaches provide a limited understanding: the first conflates two forces that may run in opposite directions and call for distinct policies, while the second overlooks timing and yields estimates sensitive to assumptions about its role. The lack of evidence on the effect of timing itself presents a key gap in the literature—both as an independent driver of gender inequality and as a prerequisite for understanding the overall impact of parenthood.

This paper provides new evidence on the independent effects of parenthood and its timing. To do so, I develop a methodology to disentangle these effects and apply it to a common but previously unstudied source of fertility variation: the success of intrauterine insemination (IUI). The approach exploits quasi-experiments in which individuals not initially assigned to treatment may undergo repeated assignments, such as when women whose first IUI fails attempt additional procedures. By leveraging variation across entire assignment sequences, it separates treatment and timing effects. Using Dutch administrative data, I show that IUI success is uncorrelated with prior labor market outcomes, conditional on age, providing plausibly random fertility variation. Using this variation, I estimate (i) the effect of having children relative to remaining childless, and (ii) the effect of conceiving later than intended due to failed previous attempts. I find that parenthood substantially and persistently reduces women's earnings and work hours, and that later childbearing—even when unintended—results in smaller losses. My findings imply that both the incidence and timing of parenthood play a central role in shaping gender inequality. They indicate that women face a double trade-off: not only between parenthood and career, but also between delaying childbearing and permanent career setbacks.

The core idea behind the method developed in this paper is to leverage repeated random treatment assignment (e.g., a series of IUI procedures) to help identify both average treatment and timing effects. Ideally, such variation could be used to recover average outcomes in scenarios such as receiving treatment at the first assignment, receiving it later, or never receiving it. This would allow treatment and timing effects to be assessed separately without imposing restrictions on how they vary. The challenge is that, even with random assignment realizations, individuals may pursue different numbers of assignments, and some may obtain treatment regardless of assignment (e.g., conceive naturally). Both behaviors may correlate with outcomes independent of treatment, making it difficult to use such sequential variation without introducing selection.

To make progress, I introduce a treatment-effects framework that classifies individuals into latent types along two dimensions: how many assignments they would pursue absent prior assignment and whether they would eventually obtain treatment without assignment. I first use this framework to isolate the effect of treatment received at the first assignment relative to remaining untreated. I show that the outcomes of individuals who remained untreated after a given number of assignments recover the average untreated outcome for the corresponding type. I also show that continuation rates after each unsuccessful assignment, together with uptake rates conditional on prior failures, reveal the distribution of latent types. These components identify the average untreated outcomes for individuals whose treatment depends on assignment within the sequence. I then argue that the most effective way forward without restrictive assumptions is to bound the average treated outcome for this group. I show that this is possible using the observed outcomes of those treated initially together with the group's population share, yielding bounds on their average treatment effect.

Isolating treatment effects without restrictions on timing effects provides the foundation for analyzing them directly. I first leverage sequential assignment to identify a generalized analogue of the standard instrumental variable (IV) estimand, which instruments eventual treatment with initial assignment. When outcomes do not depend on timing, this estimand coincides with the average treatment effect captured by the bounds, that is, treatment at the first assignment for individuals whose uptake depends on the entire assignment sequence. Otherwise, it is biased in proportion to timing effects, just like conventional IV. Combining the baseline bounds with this estimand then yields bounds on timing effects, showing how outcomes differ between treatment at the initial assignment and treatment obtained after the assignment sequence.

The method relies on minimal assumptions, requiring only that each assignment realization be as good as random, conditional on observables. Crucially, no restrictions are imposed on heterogeneity of treatment effects across individuals or over time. The bounds are sharp in the sense that they cannot be narrowed without additional assumptions. Narrower bounds require assumptions about which initially treated individuals would obtain treatment absent assignment. For instance, one might assume that couples conceiving a first child through IUI but a second naturally would still have had at least one child if IUI had failed, reflecting the higher priority placed on having a first child. Such assumptions also yield testable implications, making them empirically assessable.

Finally, I discuss estimation. Inference for the bounds is challenging because they involve selecting observations at the tails of the outcome distribution, requiring nonparametric estimation of a quantile function. To address this, I propose orthogonal moment functions that eliminate the first-order sensitivity of the bounds to estimation error in the quantile function, justifying asymptotic inference as if the true distribution were known.

My analysis focuses on opposite-sex couples who undergo IUI for their first child. The procedure is typically used as a first-line treatment for couples experiencing difficulties conceiving naturally, particularly when fertility issues are male-factor-related or unexplained. By placing sperm directly into the uterus using a flexible catheter, it closely mimics natural conception while remaining minimally invasive, quick, and generally painless.

About 5% of Dutch women who became mothers in 2017 had undergone IUI. However, many more would likely have turned to the procedure had they not conceived naturally earlier. Because IUI is typically initiated only after at least a year of unsuccessful natural attempts, fecundity rates from the medical literature can be used to obtain a lower bound on the size of this group. These calculations suggest that IUI users can be viewed as an as-good-as-random sample representing at least half of all mothers (see Section 4.3). Hence, my estimates may be informative for the broader population of parents.

I present four sets of empirical results. First, I apply my method to estimate how parenthood affects women's and men's labor market outcomes and contributes to gender inequality. I find that when parenthood begins at the first IUI, women's annual work hours persistently decline by 10%–22% and their earnings by 10%–28%. For men, the bounds are of comparable width but centered near zero, ruling out small negative effects. Taken together, parenthood causes 32%–53% of the gender gap in work hours and 10%–45% of the gap in earnings among parents.

Second, I examine timing effects. I find that when women become mothers after IUI failure, they continue to work more and earn more than if they had children earlier. This is despite the fact that these women have become mothers more recently, when impacts are often thought to be the largest (see Lundborg et al., 2017, for discussion). It is likewise despite the fact that these delays are unintended—arising from IUI failure and resulting in conception later than planned—which is also thought to result in larger losses if women time fertility to minimize career costs (see Bensnes et al., 2025, for discussion). These results suggest that early childbearing systematically impedes career progression, whereas later childbearing helps to mitigate these losses.

Third, I use my estimates to assess selection into fertility timing. Using IUI timing as a proxy, I find that women who attempt but do not conceive underperform similarly aged women who will attempt later (but also not conceive). Strikingly, men in this group perform better than their counterparts. These results are consistent with two explanations. First, selection on characteristics that influence labor market outcomes independently of parenthood, such as human capital, with earlier mothers negatively selected and earlier fathers positively selected (see Adams et al., 2024, for discussion). Second, selection on fertility preferences that matter in the absence of children, whereby unmet childbearing goals depress women's outcomes and boost men's outcomes (see Bögl et al., 2024, for discussion). Regardless of the source, these differences explain a share of gender inequality similar in size to the effect of parenthood itself.

Fourth, I examine potential mechanisms. I focus on relationship breakdowns and mental health problems. These issues may arise when couples struggle to achieve pregnancy, raising questions about whether the estimated effects of parenthood generalize to voluntary childlessness. A direct extension of the method allows me to bound effects for women who, in the event of procedure failure, would remain childless without experiencing such problems. Comparing these bounds to the baseline quantifies their contribution. The results suggest that the most severe manifestations of these issues—separation and depression, proxied by antidepressant use—are not major drivers of overall labor market impacts.

I conclude by discussing the broader implications of my results for the role of parenthood in the labor market, especially for women deciding early in their careers whether and when to have children. Most importantly, I argue that the gains due to delayed childbearing that I estimate likely represent a lower bound on the gains from planned delays, as unplanned delays may depress outcomes through unsuccessful conception attempts and shift births to less optimal career moments. I end with a suggestive calculation of the extent to which postponing childbearing can mitigate the career costs of motherhood.

My work builds on research that uses various quasi-experiments to estimate how parenthood shapes labor market outcomes (Hotz et al., 2005; Agüero & Marks, 2008; Cristia, 2008; Miller, 2011). These studies compare women who give birth at a particular moment for plausibly random reasons with women who do not give birth then (e.g., due to miscarriage), many of whom later become mothers. Typically, they either focus on reduced-form estimates, which capture a weighted average of parenthood and timing effects, or, to isolate parenthood effects, they assume away timing effects and use the initial birth as an instrument for eventual parenthood. Yet such estimates can be difficult to interpret, since delayed parenthood, for example at the peak of one's career, may not be comparable to either childlessness or earlier childbearing. My key contribution is to provide the first evidence that disentangles the effects of parenthood from those of its timing.

Existing literature offers little insight on the effects of parenthood timing. A notable exception is Gallen et al. (2024), who find that Swedish women experiencing early unplanned births due to intrauterine device (IUD) failure face larger earnings losses than women who conceive via in vitro fertilization (IVF). What distinguishes my analysis, building on this work, is that I provide direct evidence on the causal effect of timing, independent of heterogeneity in effects across women with different fertility intentions. Furthermore, by leveraging IUI failures, I shed light on the impacts of unintended postponement. Complementing Gallen et al. (2024), my findings show that the benefits of delayed childbearing are not limited to avoiding unplanned births but also arise when parenthood occurs later than intended. These results suggest that delayed childbearing itself, rather than planning alone, improves women's labor market outcomes. They also underscore that policies can reduce gender inequality by enabling later childbearing, for example through reproductive technologies.

Within the literature on the effects of parenthood, the closest studies use IVF success (Lundborg et al., 2017, 2024; Gallen et al., 2024; Bensnes et al., 2025) in Scandinavia. I contribute along three dimensions. First, I provide estimates free from assumptions about timing effects. Second, I leverage a common and minimally invasive treatment (IUI), often used in cases of male-factor infertility, broadening the population covered by existing evidence. Third, I provide evidence from the Netherlands, where parental leave and childcare usage are close to OECD averages, in contrast to the generous family policies in Scandinavia. This strengthens the external relevance of my results. Compared to IVF-based estimates, which suggest relatively small earnings reductions that fade over time, my results indicate larger and more persistent declines. Instead, they are closer to the estimates of Gallen et al. (2024) for women who experience early unplanned births due to

IUD failure. Since IUI births are planned and typically occur later, this suggests that outside Scandinavia, where family policies are less generous, women may experience larger losses.

My work also contributes to the literature on dynamic treatments. In biostatistics, numerous methods have been developed to evaluate sequential experiments under full compliance (see Hernán & Robins, 2020, for an overview). In economics, Van den Berg & Vikström (2022) introduce an approach for settings where treatment is assigned among eligible individuals dynamically and individuals selectively exit eligibility. Heckman et al. (2016) and Han (2021) develop methods for settings where treatments may also be obtained outside assignment, provided instruments are available for each treatment margin. I contribute a method for settings where previously unassigned individuals selectively undergo additional assignments and may obtain treatment outside assignment, potentially without valid instruments. Beyond parenthood, such settings include education systems with admission lotteries; job-training programs that randomize access among applicants; legal settings with random assignment to sanctioning authorities (e.g., judges or police officers), and repeated sanctioning upon reoffending; and sequential clinical trials with imperfect compliance.

My work also relates to the literature on bounds for treatment effects, beginning with Manski (1989, 1990), and in particular the methods of Zhang & Rubin (2003) and Lee (2009). While bounding methods can in principle be applied to virtually any treatment-effect setting, their main limitation is that the resulting bounds are often too wide to be economically informative. I contribute by showing how bounding techniques can be used together with sequential quasi-experimental treatment variation to obtain bounds that are strictly narrower and valid for a broader population, without additional assumptions. This improvement proves crucial in application: existing methods adapted to separate treatment and timing effects yield bounds for my main outcomes that are at least 4.5 to 8 times wider and fail to rule out large positive or negative effects (Appendix SA1).

My work further relates to the literature on dynamic non-compliance with one-time instruments (Cellini et al., 2010; Ferman & Tecchio, 2023; Angrist et al., 2024; Gallen et al., 2024; Bensnes et al., 2025). These methods exploit initial treatment assignment and repeated outcome measures to identify treatment effects when the duration of treatment is the key factor. I contribute a complementary approach that yields sharp bounds when effects may depend on both duration and timing. My method also allows the importance of each dimension to be quantified and can be implemented using cross-sectional data. I discuss this connection in more detail in Section 3.

Finally, I contribute to the broader literature on parenthood and gender inequality in the labor market (see Bertrand, 2011; Blau & Kahn, 2017; Olivetti et al., 2024). This work includes studies that compare mothers to fathers or to women who do not yet have children (Fitzenberger et al., 2013; Angelov et al., 2016; Chung et al., 2017; Bütikofer et al., 2018; Kleven, Landais, & Søgaard, 2019; Kleven, Landais, Posch, et al., 2019; Eichmeyer & Kent, 2022; Kleven et al., 2024; Melentyeva & Riedel, 2023), as well as studies employing structural methods (Adda et al., 2017; Jakobsen et al., 2022). By separately quantifying the effects of parenthood, its timing, and other outcome differences systematically correlated with these factors, my results clarify the mechanisms underlying estimates that do not explicitly distinguish them.

The remainder of the paper is structured as follows. Section 2 introduces a framework for sequential quasi-experiments. Section 3 discusses existing methods, presents intuition for my econometric approach, states the formal results, and outlines estimation. Section 4 describes the institutions, intrauterine insemination, and the data, and presents support for the assumptions. Section 5 presents the results. Section 6 discusses generalizability and life-cycle implications. Section 7 concludes.

2 A Framework for Sequential Quasi-experiments

Section 2.1 presents the framework. Section 2.2 discusses interpretation details.

2.1 Setup

My framework generalizes the local average treatment effect (LATE) framework (Angrist & Imbens, 1995) by making explicit how a treatment—parenthood—depends not only on initial treatment assignment, such as the success of the first IUI procedure, but also on subsequent decisions to pursue additional assignments and their outcomes. Although the method can also be applied in various contexts, I focus on IUI and parenthood as the leading example.

I consider a moment in time after a woman's first IUI procedure. I characterize each woman by two latent variables. First, $W \in \{1, \ldots, \overline{w}\}$ is the number of IUIs she would have undergone for her first child if all prior IUIs failed, with an upper bound $\overline{w}$; I refer to W as *willingness* to undergo IUI. Second, $R \in \{0, 1\}$ indicates whether she would have remained childless if all W IUIs failed; I refer to R as *reliance* on IUIs. I refer to women with $R = 1$ as *reliers*, who rely on IUI to have children, and to women with $R = 0$ as *non-reliers*, who would have conceived even if all IUIs failed.

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

The first parameter of interest is the *effect of parenthood*: the difference in potential outcomes between conceiving at the first IUI and remaining childless after its failure, $\tau = Y_1(1) - Y_0(0)$. The second is the *effect of parenthood timing*: the difference between conceiving at the first IUI and conceiving later, $\delta = Y_1(1) - Y_0(1)$. Note that both treated outcomes are measured after childbirth. For example, annual earnings three years after the first IUI, in a case where conception occurred immediately or one year later. This convention aids exposition, as timing effects then reflect the contemporaneous impact of having become a mother earlier rather than more recently.[1]

The two main effects I focus on are the average treatment effect among reliers, $\tau_{ATR} = \mathbb{E}[\tau \mid R = 1]$, and the average timing effect among non-reliers, $\delta_{ANR} = \mathbb{E}[\delta \mid R = 0]$. I discuss why focusing on these effects is particularly useful when outlining the intuition behind the method in Section 3.

The observed indicator for the success of the jth IUI for the first child is Z_j. It equals 1 if

[1]One may also consider the cumulative effect, which combines the periods before and after delayed childbirth. However, this is conceptually similar to the reduced-form estimand, as it mixes childlessness and later parenthood; I cover it in the main analysis.

the procedure succeeded and 0 if it failed or was not undertaken. The realized number of IUIs is $A = \min(\{j : Z_j = 1\} \cup \{W\})$, where a woman undergoes IUIs until one succeeds or until reaching her maximum willingness. The last IUI outcome is Z_A. The parenthood (*treatment*) indicator is $D = Z_A + (1 - Z_A)(1 - R)$, where a woman is a mother if an IUI has succeeded or if she is a non-relier who conceives independently of IUI success. A woman's realized labor market outcome is $Y = Y_1(1)Z_1 + (1-Z_1)DY_0(1) + (1-Z_1)(1-D)Y_0(0)$, and it depends on whether she conceived at the first IUI, after the first IUI, or not at all.

After introducing the core method, I also leverage information on non-IUI births among women whose first IUI succeeds. $R^+ \in \{0, 1\}$ is a latent indicator of whether a woman is reliant on IUI for all additional children after conceiving her first child via IUI. Women with $R^+ = 1$ (*subsequent reliers*) would have only IUI-conceived children after conceiving their first via IUI, whereas those with $R^+ = 0$ would also have one or more non-IUI children. The indicator for having at least one non-IUI child is $D^+ = Z_A(1 - R^+) + (1 - Z_A)D$, where a woman has at least one such child if an IUI succeeded and she is not a subsequent relier, or if all IUIs failed and she had a child regardless.[2]

2.2 Framework Discussion

In this subsection, I discuss the interpretation of potential outcomes, reliance, and willingness, and connect my framework to the LATE framework.

Reliers, the focus of my framework, are closely related to compliers. Compliers ($C = 1$) conceive only if their first IUI succeeds, while always-takers ($C = 0$) conceive regardless of the outcome. Reliers include all compliers and those always-takers who would conceive through later IUIs if the first failed but remain childless if all failed.[3] Focusing on reliers not only delivers results for a broader population but, crucially, also helps obtain substantially more informative estimates than would otherwise be possible, as I will discuss in Section 3.

As 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 IUIs a researcher would observe a woman undergo following repeated failures, while reliance describes her fertility behavior if those IUIs fail. Both may be random from the woman's perspective or under her control, correlated with potential outcomes and with each other. Their sole purpose is to formalize what can be learned by leveraging randomness in IUI success, without assumptions about how women determine the number of IUIs, how non-IUI fertility occurs, or how the two interact.

Like compliance, reliance and willingness are defined cross-sectionally: at a given moment, reliers are those who would remain childless up to that point if all IUIs failed. Appendix SA1 shows how the method can also be applied to estimate treatment effects for a consistent population across periods, namely women who would remain childless until the end of the sample period if all IUIs failed. This extension also addresses a further concern with conventional methods related to anticipation of parenthood, where estimates may be biased because control outcomes are identified

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

[3]Formalizing this involves heavy notation but can be summarized by $C(1-Z_1) = R\Pi_{j=1}^{A}(1-Z_j)$.

using women who eventually become mothers. Results remain similar to the baseline estimates.

Although women may conceive at different moments after the first IUI fails, for ease of exposition I group these scenarios into a single potential outcome. Fully characterizing them would require substantial additional notation without adding conceptual insight. The difference in treatment timing between outcomes is also left implicit, though observable in the data. It is determined by when a woman chooses to attempt conception after IUI failure, paralleling how treated outcomes are characterized by when she chooses to undergo her first IUI.

While parenthood may involve multiple children, this study follows the literature in focusing on the extensive margin: treated and later-treated outcomes reflect the consequences of having whatever number of children follow the first, and this number can differ between the treated outcomes. Since all women whose first IUI succeeds have a child, I do not define outcomes for remaining childless in such cases, though extending the method to include them is straightforward.

In practice, not all IUIs may be observed by the researcher. As will become clear below, classifying births from unobserved IUIs as non-IUI births will not introduce bias, as these cases will be 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 IUIs—either due to more complete data or because few births occur without IUIs—the fewer cases require bounding, resulting in tighter bounds.

3 Econometric Approach

Section 3.1 describes the limitations of conventional methods. Section 3.2 presents the intuition behind my approach. Section 3.3 formalizes the procedure. Section 3.4 outlines estimation.

To demonstrate the intuition, I introduce the local sequential unconfoundedness assumption:

Assumption 1 (Local 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 IUI j, the outcome of IUI 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.

The local 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 IUIs among women who actually undergo the procedure.

The assumption does not restrict how success rates may vary across procedures. To simplify the exposition, I abstract from covariates. The main method in Section 3.3 accounts for procedure-type-specific success rates and other observed factors that influence success, such as age at the time of the procedure. I provide empirical support for the assumption in Section 4.4.

3.1 Limitations of Standard Methods

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

$$\lambda_{RF} = \mathbb{E}[Y_1(1) - Y_0(0) \mid C = 1] \Pr(C = 1) + \mathbb{E}[Y_1(1) - Y_0(1) \mid C = 0] \Pr(C = 0) \tag{1}$$

$$= \mathbb{E}[\tau \mid C = 1] \Pr(C = 1) + \mathbb{E}[\delta \mid C = 0] \Pr(C = 0). \tag{2}$$

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.

The bias direction is ambiguous: career costs may be underestimated if younger children require more care or if delayed parenthood occurs at a less optimal career moment ($Y_0(1) < Y_1(1)$), and overestimated if early motherhood causes lasting career setbacks ($Y_0(1) > Y_1(1)$). The extent of the bias also depends on the scaling factor $\Pr(C = 0)/\Pr(C = 1)$. In the context of IUI (and IVF; Lundborg et al., 2017), 75% of women whose first procedure fails eventually have children, meaning $\Pr(C = 0) = 0.75$ and $\Pr(C = 0)/\Pr(C = 1) = 3$. This implies that even small timing effects in Equation (3) are amplified and can introduce non-negligible bias.[4]

One potential source of bias is that women who conceive earlier spend more time as mothers, and effects may depend on the duration of parenthood. To address this, Gallen et al. (2024) and Bensnes et al. (2025) propose a recursive IV approach, which assumes that effects are driven primarily by duration. Under this assumption, short-run effects are first estimated using standard IV and then used to adjust long-run estimates for differences in treatment duration between the treatment and control groups.

The key difference between conventional IV and recursive IV is that the former allows effects to vary with life-cycle stage (e.g., the woman's current age), while the latter allows them to vary with treatment duration (e.g., the age of the first child). Both approaches, however, face challenges when the timing of treatment matters (e.g., age at first birth). For example, if motherhood before age

[4] The bias can also be described in terms of negative weights; see Bensnes et al. (2025).

30 persistently reduces annual work hours by 50, while motherhood after 30 reduces them by 100, both methods would suggest reductions of less than 50 hours. More generally, depending on how timing shapes the effect, estimates from the two approaches may also point in opposite directions.

Bensnes et al. (2025) show that recursive IV estimates differ markedly from conventional IV, illustrating that existing evidence hinges on whether effects vary with duration, life-cycle stage, or timing. In Appendix SA1, I replicate this divergence and compare my estimates to both approaches. I find that the timing of motherhood plays an important role. Next, I present my method, which avoids restrictions on timing effects while also allowing treatment effects to vary freely with treatment duration and across the life cycle.

3.2 Intuition for the Econometric Approach

To illustrate the idea behind the method, note that the average treated outcome is straightforward to identify using women who conceive at the first IUI. The challenge is identifying the average control outcome: some women have children independent of IUI success, their outcomes are never observed, and they may differ systematically from the rest. Instead, the first step of the method aims to identify the average control outcome for as much of the population as possible. The idea is that if we could identify it for, say, 90% of women who initiate IUI, we might be able to impose highly conservative assumptions on the outcomes of the remaining 10% and, because this share is small, reach similar conclusions about the average effect regardless of the assumptions made.

The next question, then, is what is the largest group for which the average control outcome can be identified. As is well known, when leveraging only the first IUI attempt, this group consists of those whose parenthood status depends on its success—the compliers. To learn about average effects in this case, one must make assumptions about all individuals who would conceive if the first IUI failed. This is often a large group, resulting in wide bounds (see Appendix SA1).

One of the central results of this paper is that by exploiting variation across subsequent attempts, I can identify average outcomes for those whose parenthood status depends on the entire IUI sequence—the reliers. Existing methods cannot do this, as they cannot account for selective participation in subsequent IUIs and non-IUI conceptions. Once this is achieved, however, assumptions are required only about those who would conceive if all attempts failed. Since this group is smaller, the resulting range of possible average effects is narrower. Therefore, focusing on reliers is appealing not only because they are a more general group than compliers, but also because it helps obtain more informative results.

Next, I explain how I identify the relier average control outcome and bound their average treated outcome, and how I use additional information to tighten these bounds. Finally, I show how average treatment effects for reliers can be identified under the assumption of no timing effects, which in turn yields bounds on timing effects for non-reliers.

3.2.1 Relier Average Control Outcome

To demonstrate how the relier average control outcome is identified, I first express it as a weighted average of childless outcomes among reliers with different willingness to undergo IUIs, 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 IUIs and remained childless form an as good as random sample of reliers willing to undergo w IUIs, 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 IUIs, since non-reliers would have children, and women willing to pursue more than w IUIs would have done so. Second, for such reliers, experiencing w failed IUIs 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.

I identify the shares of different types following a similar argument. Women who experience at least w failed IUIs form an as good as random sample of those willing to undergo at least w IUIs. Thus, the share of these women initiating an additional IUIs identifies the share willing to undergo at least $w + 1$ IUIs among those willing to undergo at least w IUIs:

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

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

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

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

3.2.2 Relier Average Treated Outcome

Having identified the average relier control outcome, one can proceed in two ways. First, when outcomes have bounded support, the average control outcome of non-reliers can be bounded by assigning them the highest or lowest possible values, thereby bounding the average treatment effect. However, this approach may yield very wide bounds. Instead, I focus on bounding the average treated outcome for reliers to obtain bounds on the average effect for this group. This strategy will yield tighter bounds by exploiting the data rather than the limits of outcome support.

I start from the observation that, since IUI outcomes are as good as random, the outcome distribution among women whose first IUI succeeded represents the full distribution of treated outcomes in the IUI sample. Using the relier share identified in the previous step, I construct worst-case bounds by assuming that reliers are those with the highest or lowest treated outcomes. For instance, suppose 100 women had a successful first IUI and the relier share is 80%. It is not known which 80 are reliers, but I can obtain the upper (lower) bound on their average outcome by taking the average of the top (bottom) 80 outcomes[5]

I then refine the bounds by incorporating pre-IUI covariates. Suppose that after splitting the sample by pre-IUI earnings, each group is estimated to have an 80% relier share. Selecting the bottom 80% of outcomes among women whose first IUI succeeded, without accounting for pre-IUI earnings, may yield a set of potential reliers whose pre-IUI earnings are inconsistent with the group-specific shares. Because this selection produces the most conservative lower bound, any alternative selection can only raise it. To construct the refined bounds, I identify the relier share within each pre-IUI earnings group and select the corresponding share of the lowest and highest treated outcomes in that group.

I can further narrow the bounds only by 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 IUI after having their first through IUI would have had at least one child even if all IUIs had failed—or, equivalently, that women who are reliant on IUI for their first child are also reliant for subsequent children. To see how this helps narrow the bounds, consider the previous example. If 10 of the 100 women whose first IUI succeeded subsequently have a non-IUI child, they can be excluded from the pool of potential reliers, as they are certainly not reliant on IUI. 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.

Assumption 2 (Subsequent reliance monotonicity). *$R^+ \geq R$.*

One way to interpret the monotonicity condition is 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 seems reasonable, as couples who initiate IUI intend to have at least one child and are likely aware that they may not be able to have multiple.

However, fertility may not be entirely determined by choice. For example, having 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-IUI births that would not have occurred had the first IUI failed, thereby violating monotonicity. I relax the assumption and address such concerns in Section 5.4.

The monotonicity assumption also yields testable implications, namely that the subsequent relier share at each covariate value exceeds the relier share. I present the testing procedure and results in Appendix SA1. There, I also report estimates based on a relaxed specification that allows the direction of monotonicity to vary with covariates. Monotonicity is not rejected, and the estimates under the alternative assumption remain similar.

[5]Trimming outcome distribution tails to bound subgroup averages follows the logic of Zhang & Rubin (2003) and Lee (2009) for addressing sample selection. The innovation lies in identifying and overcoming the challenges involved in appropriately leveraging sequential assignment with bounding, yielding narrower bounds that apply to a broader subpopulation (see Appendix SA1).

The remaining results assume monotonicity. In settings without 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 everyone as subsequent reliers.

3.2.3 Interpreting the Uncertainty in the Relier Average Treatment Effect Bounds

The bounds on the relier average treatment effect are obtained by subtracting the identified relier average control outcome from the bounds on their average treated outcome. Before proceeding, it is useful to clarify what uncertainty these bounds do and do not reflect.

The bounds capture uncertainty about selection into treatment absent assignment. In the IUI context, this refers to who would conceive naturally if all procedures failed (conditional on pre-IUI covariates and subsequent relier status). If natural conception among IUI couples is partly random, the true effect lies within a narrower range than the conservative bounds; if entirely random, it is near the midpoint. More generally, the bounds shrink exactly in proportion to the share of random versus selective births. This share may be substantial in the IUI setting, as all couples are trying to conceive and natural conception after failed procedures is largely a matter of luck.

Importantly, the treatment effect bounds contain no information about the magnitude of timing effects, as they rely only on treated outcomes for those assigned treatment at baseline and on control outcomes. With this distinction, I now turn to timing effects.

3.2.4 Timing Effects and Treatment Effect Identification Without Timing

As a starting point for assessing timing effects, I first show how to point-identify τ_{ATR} under assumptions equivalent to the IV approach. Let $Y_0^* = Y_0(0)R + Y_0(1)(1 - R)$ denote the outcome when all IUIs 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 IUIs: $\mathbb{E}[Y \mid A = w, Z_w = 0] = \mathbb{E}[Y_0^* \mid W = w]$. Averaging over W using identified type shares gives $\mathbb{E}[Y_0^*]$. Subtracting this from the average treated outcome, and scaling by the relier share yields:

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

I refer to this as the *sequential IV* estimand, since it uses a sequence of IUI attempts to identify an estimand similar to the IV in Equation (3) of Section 3.1, but targeting reliers instead of compliers. Under the IV exclusion restriction ($\delta = 0$), sequential IV identifies τ_{ATR}. When $\delta \neq 0$, it no longer identifies τ_{ATR}, but the bias is attenuated relative to the IV because the timing term is weighted less: non-reliers are a subset of always-takers, so $\Pr(R = 0)/\Pr(R = 1) \leq \Pr(C = 0)/\Pr(C = 1)$.

Crucially, the sequential IV estimand need not be contained within the bounds. This is more likely to occur when timing effects are large and/or when variation in treated outcomes among sequential reliers is largely explained by past covariates (e.g., past income), resulting in narrow bounds. Comparing the estimand to the bounds thus offers a test of the exclusion restriction: under the restriction, it should lie within the bounds; otherwise, $\delta \neq 0$. Furthermore, since the relier share is identified, bounds on τ_{ATR} can be directly translated into bounds on δ_{ANR} using Equation (8).

Finally, it is worth noting that the same reasoning can be applied to bound timing effects across subpopulations and treatment moments. For example, one could bound the average timing effect for individuals willing to undergo at least two procedures by first identifying the later-treated outcomes using those who conceive at their second attempt, and then bounding their average treated outcome using the same logic as before. However, focusing on conception at the first versus the second attempt may reveal little about timing, since these events often occur only a month apart. The advantage of focusing on non-reliers is that it captures the largest possible timing shift induced by variation in IUI success: conceiving at the first IUI versus only after all procedures fail.

3.3 Identification

In this section, I establish the formal identification results. I begin by introducing the conditional local sequential unconfoundedness assumption:

Assumption 3 (Conditional local 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 assignment j, such as age at the time of the procedure. In words, the success of IUI j is independent of potential outcomes and type, conditional on undergoing at least j IUIs 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 IUI 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 IUI succeeding, having only IUI 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 IUIs 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 IUIs to account for the fact that women willing to undergo more IUIs are less likely to experience IUI failure, making them underrepresented in this group. Next, the covariate-conditional share of subsequent reliers is identified from the share of women having only IUI children among those whose first IUI 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 IUI succeeded and who have only IUI 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 IUI succeeded, who have only IUI 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 IUIs to account for the fact that reliers willing to undergo more IUIs are less likely to not experience IUI 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 1 (Bounds on the Treatment Effect)**.** *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.

The next moment is used to identify the sequential IV estimand, which equals the relier average treatment effect under the assumption of no timing effect:

$$m^{S}(G, \eta^{0}) = Y\left(\frac{Z_1}{e_1(X_1)} - \Pi_{j=1}^{A} \frac{(1 - Z_j)}{(1 - e_j(X_j))}\right) \tag{14}$$

Theorem 2 (Identification in the Absence of Timing Effect)**.** *Under Assumptions 3 and 4, and if $\delta = 0$, then $\tau_{ATR} = \mathbb{E}[m^{S}(G, \eta^{0})]/\mathbb{E}[r(X_1)]$.*

Proof. See Appendix A1.

Finally, the moments for bounding the average timing effect for non-reliers are constructed by combining the moment used to identify the sequential IV estimand with the moments used to bound the relier average treatment effect:

Theorem 3 (Bounds on the Timing Effect)**.** *Under Assumptions 2, 3, and 4, sharp lower and upper bounds on δ_{ANR} are given by $\gamma_L = \mathbb{E}[m^S(G,\eta^0) - m^U(G,\eta^0)]/\mathbb{E}[1-r(X_1)]$ and $\gamma_U = \mathbb{E}[m^S(G,\eta^0) - m^L(G,\eta^0)]/\mathbb{E}[1-r(X_1)]$.*

Proof. See Appendix A1.

3.4 Estimation

The bounds can be estimated by replacing the expectations and nuisance functions in Theorem 1 with their empirical counterparts. However, inference for this estimator is complicated by its sensitivity to first-stage estimation error in the conditional quantile function. If overly conservative—i.e., non-sharp—bounds are acceptable, one can overcome this challenge by restricting attention to discrete covariates or by partitioning continuous covariates into bins. In that case, covariate-conditional bounds can be estimated separately within each covariate cell and then aggregated. Asymptotic normality and analytical standard errors then follow from conventional GMM results.

Ensuring sharp bounds, however, requires using all available covariates, including continuous ones, making such an approach infeasible. To address this challenge, I draw on insights from Semenova (2025), who develops an estimator for the Lee (2009) bounds, which also rely on conditional quantiles. The method combines orthogonalization with sample splitting. I adapt this approach by constructing new orthogonal moment functions for my bounds, presented in Appendix A2. These moments generalize those in Semenova (2025) by incorporating the sequential step of my procedure, which shifts the focus from compliers to reliers and requires new correction terms for the sequential propensity scores. They satisfy the key conditions under which Semenova (2025) establishes the validity of standard GMM inference, as if the nuisance functions were known.

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 functions 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 procedure. While the scores could also be estimated nonparametrically using age fixed effects, procedure success rates are well documented to evolve smoothly with age, and leveraging this smoothness helps avoid sensitivity to sparsely populated age bins in later attempts. 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 procedure, along with pre-procedure earnings and work hours for both partners. Since past outcomes are highly predictive of future outcomes, these two covariates are the most important for obtaining narrow bounds. Confidence intervals are constructed following Stoye (2020). I use data on the first ten procedures, which makes the bounds only negligibly wider than if additional procedures 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 procedures failed. Varying the number of procedures used between 8 and 12 has little impact on the estimates.

4 Institutions, Intrauterine Insemination, and Data

Section 4.1 describes Dutch family policies and the labor market context. Section 4.2 discusses IUI. Section 4.3 overviews the data and compares the IUI sample to a representative sample. Section 4.4 provides empirical support for the unconfoundedness assumption and documents procedure profiles and birth timing.

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 at 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 (OECD, 2023d). Among two-parent families, only 14% had both parents working full-time, 52%

[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/.

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

My analysis focuses on Dutch couples who undergo IUI for their first child. As in most countries, the procedure is usually a first-line treatment for couples who are unable to conceive naturally within a year, especially in cases of male-factor or unexplained infertility. Prior to the procedure, women typically undergo cycle tracking and may receive hormonal stimulation to enhance egg production. During IUI, sperm is injected directly into the uterus via a catheter, mimicking natural conception by facilitating fertilization within the body. With a success rate of about 10% per attempt, the procedure is minimally invasive, typically lasting five minutes and causing little or no discomfort.

Some couples who do not succeed with IUI eventually proceed to IVF, which has previously been used to study the career impact of parenthood in Denmark, Norway, and Sweden (Lundborg et al., 2017, 2024; Gallen et al., 2024; Bensnes et al., 2025). IVF is a surgical procedure in which eggs are retrieved through the vaginal wall, fertilized in the laboratory, and transferred as embryos into the uterus. It is more invasive than IUI, performed under sedation or anesthesia, and in most countries it is used for women with severe fertility problems, such as tubal damage or advanced endometriosis. It also has a higher success rate than IUI, of about 25% per embryo transfer. Because IVF also generates as-good-as-random fertility variation, I include eventual IVF attempts in the analysis. To account for differences in success rates between IUI and IVF, all terms in each propensity score are interacted with a procedure-type indicator.

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 cost approximately 4,000 euros, but because multiple embryos can be frozen per cycle, subsequent transfers could cost 1,000 euros or less.

4.3 Data and Sample Characteristics

I use administrative data from Statistics Netherlands, covering all residents. Medical 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 as-good-as-random variation in fertility outcomes. Procedure 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

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

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 procedures, and were cohabiting with a male Dutch-born partner at the time of the procedure. Throughout the analysis, I refer to these men as the partners. I address potential complications related to separation in Section 5.4. 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 assisted conception procedures 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 women whose first IUI succeeded to a representative sample of mothers (column 6). A year before pregnancy, 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%).

Beyond observable characteristics, an important question is whether IUI mothers differ systematically in unobservables, which could limit the generalizability of the estimates. They may, for instance, have a stronger desire or preparation for raising children, leading to different effects than most women. Since only about 5% of Dutch mothers undergo IUI before their first child, such selection could be substantial. Yet fertility problems are largely unpredictable, meaning that IUI mothers can be seen as drawn from a broader group of women with similar preferences who would have initiated treatment but conceived naturally earlier. The size of this group can be assessed by considering medical evidence on fecundity and the institutional context. IUI is typically initiated after 1.5 years of unsuccessful natural attempts, meaning that with a mean initiation age of 32, most women had been trying since about age 30.5. At this age, the probability of conceiving naturally within 1.5 years is about 90% (Leridon, 2004). This implies that for each IUI mother, nine

Table 1: First IUI 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 IUI succeeded; *Fail* – average among women whose first IUI 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 assisted conception procedures; *Suc. vs rep* – difference between *Success* and *Rep.*. Reference year: year of first IUI (IUI sample); 9 months before first birth (representative sample). IUI 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 assisted conception procedures 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.

similar women conceived naturally before becoming eligible. Calculations based on full age profiles yield comparable figures. These results suggest that the 5% who undergo IUI before motherhood may be comparable in unobservables to at least half of all mothers, making estimates from this group relevant for a broad population of parents.

4.4 Unconfoundedness, Procedure Profiles, and Career Trajectories

Since women have limited direct control over IUI outcomes, the main threat to 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-IUI outcomes, unconfoundedness can be assessed by comparing pre-IUI characteristics between women whose procedure succeeded and those whose procedure failed.

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 reports average covariate differences after adjusting for age and education (following Lundborg et al., 2017) using inverse probability weights from the main specification (column 3). After the adjustment, the remaining gaps are negligible. Excluding education has no effect. Appendix SA3 presents equivalent results for subsequent procedures. These findings support conditional sequential unconfoundedness. All remaining analyses account for differences in success rates by age at the time of each procedure, education, and procedure type.

Another reason local sequential unconfoundedness may fail is if women pursue additional procedures 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 procedure 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 of procedure attempts and natural births. Panel B plots complier and relier shares over time, estimated using the baseline specification. Two 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. The gradual decline in the relier share implies that women who become mothers within the sample period after failed procedures do so, on average, 3.1 years later than those who conceive after the first procedure, meaning that procedure success substantially influences parenthood timing.

Panel C shows the cumulative distribution of the number of procedures women undergo after the 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 failed 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

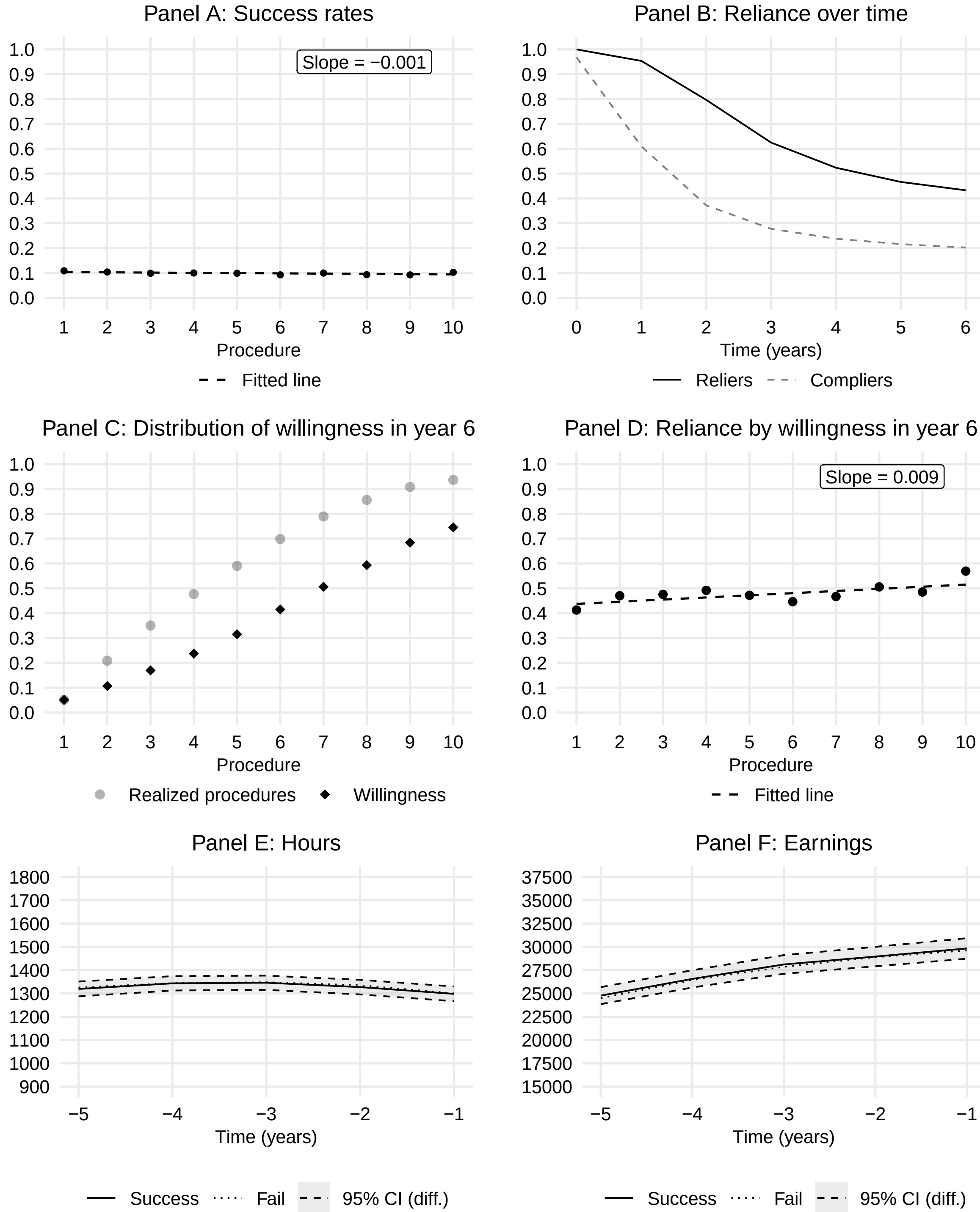

Figure 1: Procedure Descriptives and Pre-trends

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 procedures to illustrate the share of women exceeding 10 procedures; all other panels use only the first 10 procedures. Time is measured relative to the first procedure.

increasing. Differences between the two groups are small. 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

Section 5.1 quantifies the effects of parenthood. Section 5.2 examines the impacts of parenthood timing. Section 5.3 decomposes gender inequality into components explained by parenthood effects and by differences associated with parenthood timing. Section 5.4 assesses the role of mental health and relationship breakdowns.

5.1 Effects of Parenthood

Panels A and B in Figure 2 present effects on women's annual work hours and earnings. 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 (see Appendix SA1). For earnings, the average upper and lower bounds indicate reductions of 3,400 and 9,300 euros, or 10% to 28%. The midpoint of the bounds—which corresponds to the scenario with limited selection into natural births after failed IUIs, conditional on observables—suggest anual losses of about 190 hours and 6,400 euros.

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 present the share of within-couple gender gap among parents caused by parenthood—that is, the average effect on the gender gap relative to the average gap after parenthood.[8] Between the second and sixth years of parenthood, the effects account for 21% to 58% of the gender gap in annual work hours and up to 48% of the gap in annual earnings. Over the first six years, parenthood causes 32% to 53% and 10% to 45% of the gaps in work hours and earnings, respectively.[9] Assuming limited selection into natural births, the midpoints correspond to 42% for work hours and 28% for earnings.

I present supplementary analyses in Appendix SA1. First, I address that measuring effects by time since birth may result in men's and women's outcomes being observed at systematically different ages, which may limit the relevance of the estimates for lifetime inequality. Results remain similar when outcomes are measured at the same age for both partners. Second, I report heterogeneity analyses by women's pre-IUI age, work hours, and earnings, following Heiler (2024).

[8] It is calculated as $1-a/b$, where a is the relier average within-couple gap in the control outcome and b is the lower or upper bound for the relier average within-couple gap in the treated outcome. Both are estimated using orthogonal moments from the baseline specification.

[9] Summing estimates across years yields non-sharp bounds on cumulative impacts, since period-specific estimates do not account for within-couple correlations over time (see Semenova, 2025, for discussion). To estimate the total effect, I use cumulative outcomes in the last period.

The bounds do not reject homogeneous effects across these dimensions. Finally, I analyze women who would remain childless until the end of the sample period if all IUIs failed. This ensures that estimates over time are based on a consistent group and are not influenced by anticipated motherhood. The results are essentially unchanged.

5.2 Effects of Parenthood Timing

Next, I turn to the effects of parenthood timing. I focus on outcomes starting three years after the first IUI, when the non-relier share stabilizes at 50% (Figure 1); earlier bounds are wide because this share is too low. At that point, women who conceived after failed IUIs did so, on average, about three years later than if they had conceived at the first IUI (as implied by Figure 1). Hence, the estimates reflect the effect of conceiving three years later rather than immediately, on average.

Panels A and B in Figure 3 present the effects of delayed childbirth ($-\delta$) on women's annual work hours and earnings. Between the third and sixth years, the bounds imply average annual gains in work hours ranging from 10 to 160, with a midpoint of about 85. Although the lower limit of the 95% confidence interval lies near zero, this corresponds to the most extreme form of selection into natural conception. Under even minimal randomness in natural conception, the estimates therefore imply that the true effect lies above zero. For earnings, the bounds range from a loss of 1,900 euros to a gain of 4,500 euros, with a midpoint of about 1,300 euros. Hence, the evidence on earnings is less conclusive thus far.

To shed further light on these effects, I assess the cumulative impacts since the first IUI, which yield more informative estimates by accounting for within-woman correlations over time. To clarify the interpretation, suppose all natural conceptions after IUI failure occur three years after the first procedure. The year-three estimates then primarily capture the total effect of remaining childless up to that point. The key to assessing timing effects is the change from that point onward: if the cumulative impacts stabilize, timing effects are limited; if they continue to rise, it indicates that even after childbirth, women who conceive later maintain better outcomes. Panels C and D present the estimates. Because the cumulative bounds rise steadily, the results suggest persistent benefits of delaying parenthood for both work hours and earnings.

It is worth emphasizing that these estimates compare contemporaneous losses: after later childbirth, women experience smaller reductions than if they had conceived earlier. Notably, this occurs even though they have become mothers more recently and impacts are often thought to be largest at the onset of motherhood (see Lundborg et al., 2017, for discussion). It also occurs even though these delays are unintended, arising from IUI failure, which should lead to larger losses if fertility is timed to minimize career costs (see Bensnes et al., 2025, for discussion). Overall, the evidence points to early childbearing as an important driver of career setbacks. It suggests that postponement helps mitigate the losses not only by eliminating them before the later birth but also by reducing them afterward. I present indicative calculations of how much postponement may offset the career cost of motherhood in Section 6.

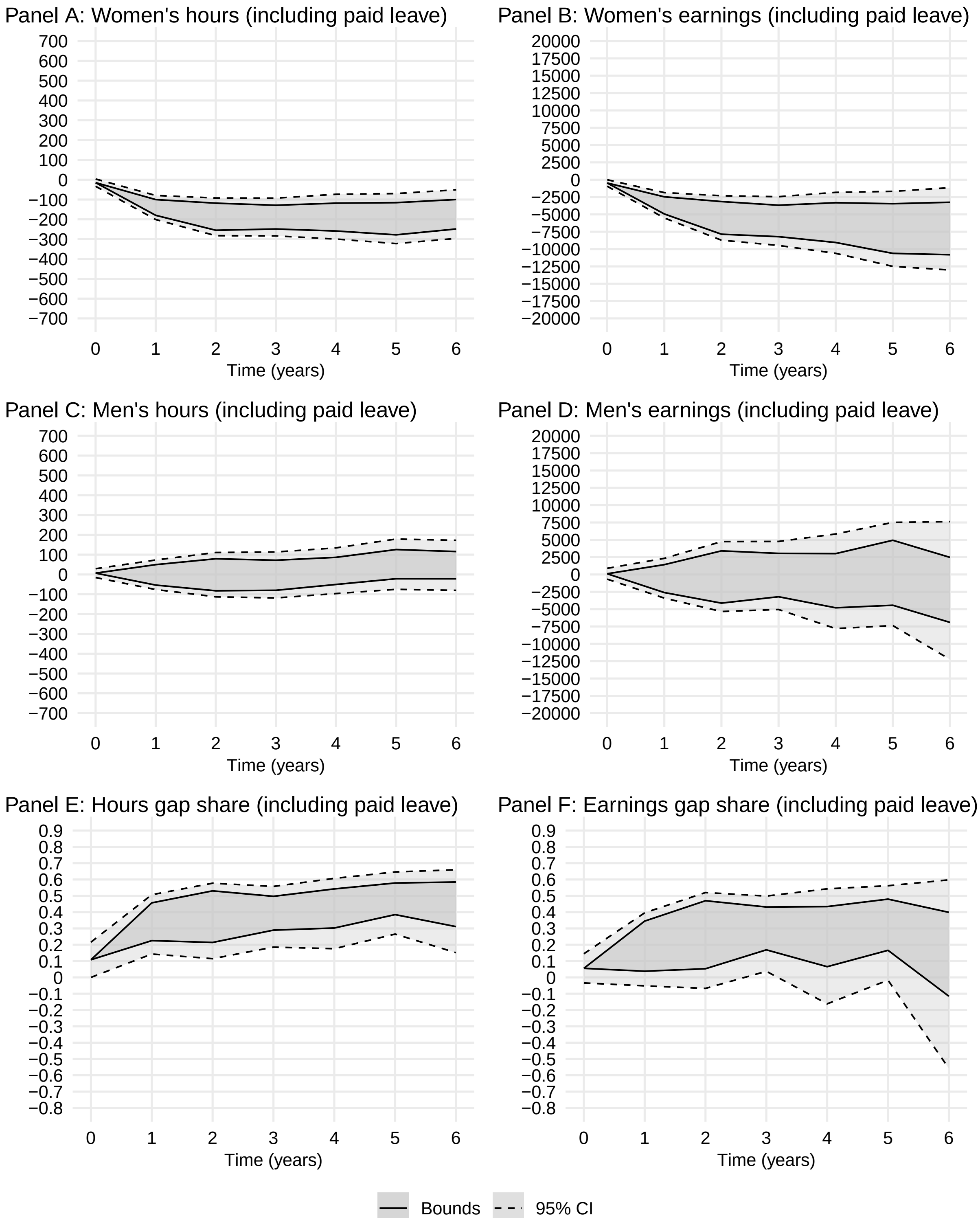

Figure 2: Effects of Parenthood

Note: Panels A–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 within-couple gap in the control outcome and b is the lower or upper bound for the average within-couple gap in treated outcome, both estimated using orthogonal moments from the baseline specification. Confidence intervals are 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.

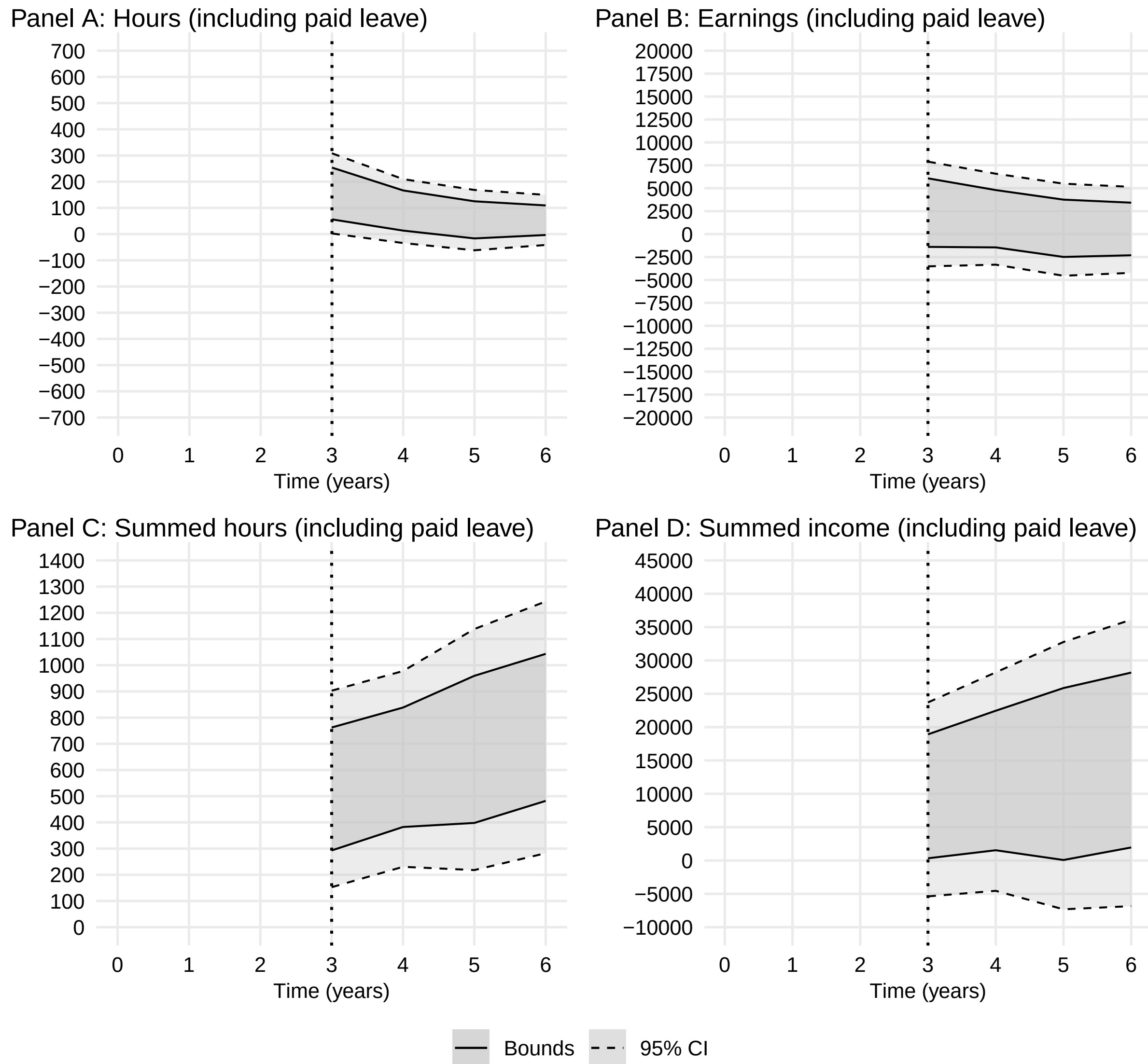

Figure 3: Effects of Delayed Parenthood on Women's Outcomes

Note: Panels A and B: effects of delayed childbearing on annual work hours and earnings (in EUR). Panels C and D: effects of conceiving later on cumulative work hours and earnings (in EUR). See Appendix SA2 for implementation details. 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.

5.3 Selection into Parenthood Timing

Next, I assess selection into parenthood timing. More specifically, I ask how labor market outcomes would differ between individuals who chose to have children at different moments in the absence of children. One way to assess this is to use a representative sample (or all IUI mothers) and compare average outcomes between individuals who already have children and those who will have them later. My estimates of the effects of parenthood could then be used to net out the part of this difference caused by parenthood, leaving only differences that would arise in the absence of children. Yet this approach is problematic, because the effects for women reliant on IUI may differ from those in the broader population (or even the full IUI sample), and such a comparison would erroneously attribute these differences to selection into timing.

I introduce an alternative approach. I use the timing of the first IUI as a proxy for chosen fertility timing. Then, leveraging IUI success, I quantify differences in childless career trajectories among reliers who opt for childbearing at different ages. Finally, I combine these estimates with bounds on the effect of parenthood in this group to quantify the share of the gender gap that can be explained by each of the two factors.

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

$$Y_{it}^{g} = \sum_{k \neq -1} \alpha_k^g 1_{\{T_{it}=k\}} + \sum_{a} \beta_a^g 1_{\{\text{Age}_{it}=a\}} + \sum_{s} \gamma_s^g 1_{\{t=s\}} + \nu_{it}^{g}, \tag{15}$$

where Y_{it}^{g} is the labor market outcome of individual i of gender g in period t. The first term reflects 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 (15) 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 average 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 4 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.

These patterns suggest positive selection into early fatherhood and negative selection into early motherhood. This selection may reflect differences in traits affecting labor market outcomes independent of parenthood, such as human capital differences (see Adams et al., 2024, for discussion). It may also reflect factors that shape outcomes specifically in the absence of children, such as unmet early-childbearing goals that depress women's outcomes and enhance men's (see Bögl et al., 2024, for discussion). Regardless of the underlying cause, these results indicate that couples who delay childbearing would exhibit smaller gender gaps even in the absence of children.

I next ask how much of the gender inequality among parents reliant on IUI can be jointly explained by the effect of parenthood and by differences in outcomes between late and early parents in the absence of children. More specifically, how much of the gender inequality among parents

[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.

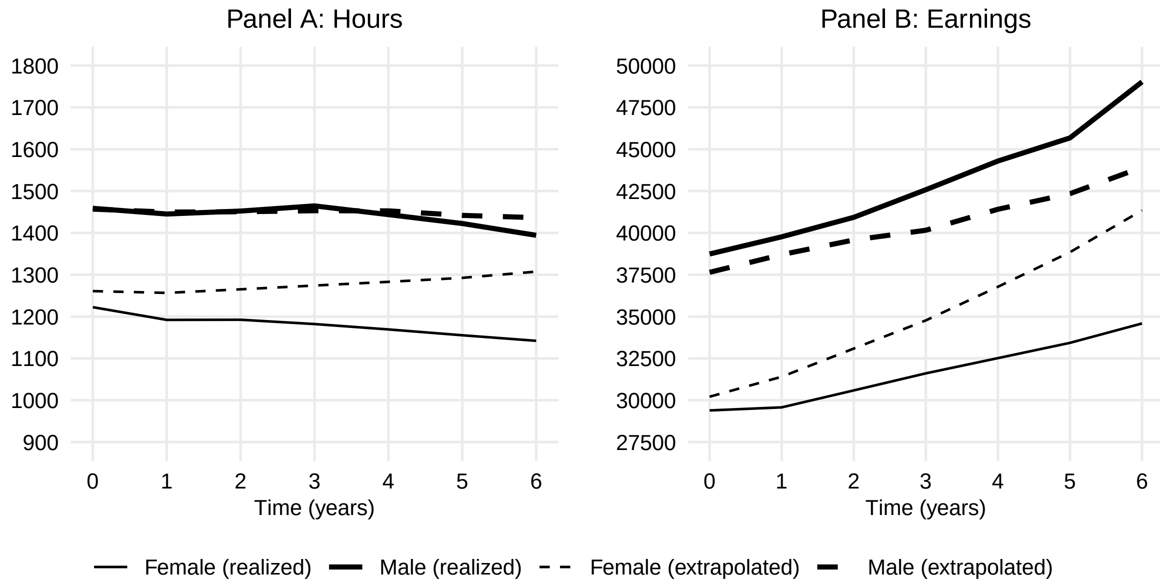

Figure 4: Career Progression in the Absence of Children

Note: Estimated annual work hours and earnings (in EUR) for relier couples in the absence of children. *Realized* – estimated using couples who remain childless at the end of the sample period; *Extrapolated* – constructed using similarly aged individuals in couples one year away from their first IUI; see Section 5.3 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.

would disappear if they had no children and if early parents performed as well as late parents in the absence of children?[11] Figure 5 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 parenthood alone accounts for up to 58% and 40% of the respective gaps. These results suggest that both factors play a substantial role in explaining the difference in gender gaps between early parents and couples who delay childbearing.

5.4 Mental Health and Relationship Stability

Women who remain childless after trying to achieve pregnancy may experience mental health deterioration or relationship breakdowns, potentially affecting labor market outcomes. On the one hand, these consequences may be an integral part of the effect of not having children, making it important to understand the extent to which these mechanisms drive the labor market impacts. On the other hand, such issues may also stem not from childlessness itself but from fertility treatments or from the experience of trying and failing to conceive, which could limit the generalizability of the estimates to settings where women voluntarily remain childless.

[11] I start with a point in the bounds on the effect of parenthood on the gender gap for reliers: $\mathbb{E}[Y^{gap}_{1k}(1) - Y^{gap}_{0k}(0) \mid R_k = 1]$, where $Y^{gap}_{zk}(d)$ 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^{male}_k, and subtract the corresponding difference for relier mothers, α^{female}_k. I divide the result by the gender gap in the case of parenthood, $\mathbb{E}[Y^{gap}_{1k}(1) \mid R_k = 1]$.

Figure 5: 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 5.3 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.

Before assessing these factors, it is useful to clarify when the side effects of failing to conceive matter for generalizability. They are arguably less concerning for my estimates of timing effects: I find that women who experience delays perform better, and doing so despite potential side effects reinforces that delaying childbearing mitigates losses. By contrast, they are more important for my estimates of the effects of parenthood and of selection into timing, where I find that large gender gaps persist even when couples fail to conceive, and women who fail to conceive underperform those who have not yet tried. If these patterns reflect such side effects, the estimates may understate the role of parenthood and overstate selection beyond my sample.

To provide suggestive evidence on mental health effects, one might estimate the impact of conceiving at the first IUI on antidepressant uptake. The estimates, reported in Appendix SA1, are precise and indistinguishable from zero. However, this does not directly address the concern, since both conceiving and failing to conceive may worsen mental health relative to not attempting.

To directly assess the role of mental health and relationship impacts, I modify the baseline approach to bound the labor market effects for women who, in the event of procedure failure, would (i) remain childless, (ii) be cohabiting with or registered to a partner, and (iii) not initiate antidepressant use. If the adjusted bounds remain close to the baseline, it indicates that these issues, independent of the source, have limited empirical relevance for the labor market impacts.

The formal procedure classifies women who meet condition (i) but not (ii) or (iii) as having conceived naturally, thereby removing them from the group used to construct the bounds. This

Figure 6: Effects for Women Less Affected by Failed Conception

Note: Effects of motherhood on women's annual work hours and earnings (in EUR). *Baseline* refers to effects for the full sample; *Less affected women* refers to effects for women who would continue cohabiting with a partner and would not initiate antidepressant use after failing to conceive. See Section 5.4 for procedural 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.

adjustment also addresses potential monotonicity violations by excluding women who would have conceived naturally had IUI succeeded but did not after failure because of subsequent mental health problems or relationship breakdowns.[12]

The procedure is conservative because it also excludes women who would have experienced such issues irrespective of parenthood, as well as those whose labor market outcomes would have remained unaffected even if the issues arose due to childlessness or procedure failure. This occurs because it is not possible to identify which women who conceived at the first IUI fall into these categories, so the trimming step must discard more observations from the tails, which in turn produces wider bounds. If such women are primarily the ones experiencing these issues, the bounds will widen without shifting systematically in either direction, much as if some childless women had been excluded at random.

Figure 6 presents the results. Overall, the estimates change only modestly: the bounds remain close to the baseline in the early years and widen slightly over time, with no systematic pattern in either direction. In the most extreme case, the share of within-couple gender inequality caused by parenthood rises to 63% for work hours and 58% for earnings (not reported). Thus, while mild mental health or relationship difficulties cannot be ruled out, the analysis indicates that severe mental health problems and separation play at most a modest role in driving the estimates.

[12]The formal argument amounts to adapting Theorem 1 by redefining R and D.

6 Discussion

In this section, I discuss how my estimates may generalize to scenarios in which women make different fertility and career plans, and I present suggestive calculations on how much delaying motherhood may mitigate its career costs.

The focal point of my analysis is the moment of the first IUI procedure, whose timing is chosen by the woman. My estimates for the effect of parenthood therefore reflect the consequences of failing to conceive when intended, while my estimates for the effect of timing reflect the consequences of conceiving later than intended. To understand the broader role of parenthood in the labor market, two other counterfactuals are especially important: planned childlessness and deliberate postponement of childbearing.

How my estimates differ from the impacts of parenthood relative to planned childlessness depends on whether women would invest more or less in their careers under that scenario. On the one hand, women may invest more if they plan to remain childless, since the absence of children may allow them to reap greater returns. In that case, my estimates would serve as a lower bound on the career cost of motherhood. Alternatively, women may invest more if they anticipate motherhood—both to accumulate resources for raising children and to develop careers compatible with family life. In that case, my estimates would instead provide an upper bound on the career cost.

The impacts of planned delayed childbearing are arguably easier to anticipate. One reason why planned and unplanned delays may differ concerns the experience of infertility itself: women's outcomes may be depressed by the failure to achieve pregnancy. Another reason is that women may plan fertility and their careers, at least in part, to minimize career costs. Both imply that planned delays would result in better labor market outcomes than unplanned delays. This suggests that the results may be seen as a lower bound on the gains from postponement.

The final question I consider is the extent to which delayed childbearing can mitigate the career cost of motherhood. This analysis is necessarily suggestive, as my estimates for the effects of parenthood and its timing cover distinct groups and capture only short-run impacts. I focus on the midpoints of the bounds. Before the delayed birth, the estimates imply stable gains from the absence of children of about 17% in annual hours and 20% in earnings. After the delayed birth, these gains converge to 5% in annual hours and 2% in earnings.[13] Taking 33 years as the horizon from conception at the first IUI (mean age 32) to retirement, these figures imply that a one-year delay would raise women's hours and earnings from the age of 32 onward by roughly 6% and 2.5%, respectively. A five-year delay would raise hours by 7% and earnings by 4.5%. Since the estimates of the effect of parenthood converge to annual losses of 15% in hours and 20% in earnings, these calculations suggest that a five-year delay would reduce average annual losses to about 9% for hours

[13]The gains before delayed childbearing are computed by comparing estimates of the effect of parenthood for reliers in Figure 2 to the corresponding relier average treated outcomes. The gains after delayed childbearing are computed by comparing estimates of the average effect of delayed childbearing for non-reliers in Figure 3 to the corresponding relier average treated outcomes.

and 16% for earnings.[14]

7 Conclusion

This paper develops a new methodology to estimate treatment and treatment-timing effects in quasi-experiments where individuals not initially assigned to treatment may obtain it later through repeated assignments. I use this method together with Dutch administrative data to assess the career impacts of parenthood. I find that motherhood persistently reduces women's work hours and earnings, whereas fatherhood has little effect. Taken together, parenthood causes up to half of the gender inequality among parents. Finally, my results show that women's losses are substantially mitigated when childbearing occurs later in their careers.

My results suggest that there is considerable scope for policies that reduce the direct effects of parenthood, such as ensuring that work and family can be more easily combined, as well as for policies that support delayed childbearing, for example through broader access to reproductive technologies or contraceptives. Yet such interventions may be insufficient to eliminate gender inequality, as substantial gaps persist even in the absence of children. My findings point to the importance of better understanding how expectations and anticipatory behavior before childbearing influence women's careers, and which policies may help mitigate these effects.

Appendix

A1 Proofs

To simplify 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:*

$$(Y_1(1), Y_0(0), R^+, R, W, Z_1, \ldots, Z_{j-1}, X_1^*, \ldots, X_{j-1}^*) \perp\!\!\!\perp Z_j^* | X_j^* \text{ for all } j > 1. \tag{16}$$

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

[14] This is computed by adding the estimates of the effect of delayed childbearing for non-reliers in Figure 3 to the estimates of the effect of parenthood for reliers in Figure 2 and comparing the result to the average relier control outcome.

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

$$\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} Z_j^* \middle| X_l^*\right] = \mathbb{E}\left[\mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}} Z_j^* \middle| X_{\overline{w}}^*\right] \middle| X_l^*\right] \tag{17}$$
$$= \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}-1} Z_j^* \, \mathbb{E}\left[Z_{\overline{w}}^* | X_{\overline{w}}^*\right] \middle| X_l^*\right] \tag{18}$$
$$= \mathbb{E}\left[g(M_l)\Pi_{j=l+1}^{\overline{w}-1} Z_j^* \middle| X_l^*\right] \tag{19}$$
$$= \mathbb{E}\left[g(M_l)|X_l^*\right], \tag{20}$$

where (17) holds by law of iterated expectations and because X_j^* includes X_l^* for $j \geq l$, (18) holds by the Corollary, (19) holds because $\mathbb{E}[Z_l^*|X_l^*] = 1$ under Assumption 4, and where (20) follows from steps similar to (17) through (19) for X_j^* for j s.t. $l < j < \overline{w}$. □

Proof of Theorem 1. 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:

$$\mathbb{E}\left[Y(1 - D^+)\Pi_{j=1}^{\overline{w}} Z_j^*\right] = \mathbb{E}[Y_0(0)R\Pi_{j=1}^{\overline{w}} Z_j^*] \tag{21}$$
$$= \mathbb{E}\left[\mathbb{E}\left[Y_0(0)R\Pi_{j=1}^{\overline{w}} Z_j^* \middle| X_1^*\right]\right] \tag{22}$$
$$= \mathbb{E}\left[\mathbb{E}\left[Y_0(0)RZ_1^*|X_1^*\right]\right] \tag{23}$$
$$= \mathbb{E}\left[\mathbb{E}\left[Y_0(0)R|X_1^*\right]\mathbb{E}\left[Z_1^*|X_1^*\right]\right] \tag{24}$$
$$= \mathbb{E}\left[Y_0(0)R\right] \tag{25}$$
$$= \mathbb{E}\left[Y_0(0)|R = 1\right]\Pr(R = 1), \tag{26}$$

where (21) follows from the definitions, (22) holds by law of iterated expectations, (23) holds by Lemma, (24) holds by Assumption 3, and (25) holds because $\mathbb{E}[Z_l^*|X_l^*] = 1$ under Assumption 4. Next, note that:

$$\mathbb{E}\left[r(X_1)|X_1\right] = \mathbb{E}\left[R\Pi_{j=1}^{\overline{w}} Z_j^* \middle| X_1^*\right] \tag{27}$$
$$= \mathbb{E}\left[RZ_1^*|X_1^*\right] \tag{28}$$
$$= \Pr(R = 1|X_1^*), \tag{29}$$

where (27) follows from definitions, (28) holds by Lemma, and where (29) 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 $\mathbb{E}[Y(1 - D^+)1_{\{Y > q(1-p(X_1^*), X_1^*)\}} Z_1/e_1^*(X_1^*)]/\mathbb{E}[r(X_1^*)]$ 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 initially assigned treatment, 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:

$$\int_{\mathcal{X}_1^*} \gamma_x f_{x|R=1}(x)dx = \mathbb{E}[Y(1 - D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}} Z_1/e_1^*(X_1^*)]/\,\mathbb{E}[r(X_1^*)]. \tag{30}$$

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

$$\begin{aligned} &\mathbb{E}[Y(1 - D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}} Z_1/e_1^*(X_1^*)] &(31)\\ &= \mathbb{E}[\mathbb{E}[\gamma_{X_1^*}|X_1^*] \Pr(D^+ = 0, Z_1 = 1, Y > q(1 - p(X_1^*), X_1^*)|X_1^*)/e_1^*(X_1^*)]. &(32)\end{aligned}$$

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

$$\Pr(D^+ = 0, Z_1 = 1, Y > q(1 - p(X_1^*), X_1^*)) = r(X_1^*)e_1^*(X_1^*) \tag{33}$$

Thus:

$$\mathbb{E}[Y(1 - D^+)1_{\{Y>q(1-p(X_1^*),X_1^*)\}} Z_1/e_1^*(X_1^*)] = \mathbb{E}[\gamma_{X_1^*} r(X_1^*)]. \tag{34}$$

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.

□

Proof of Theorem 2. Following analogous reasoning as in the first step of Theorem 1 yields:

$$\mathbb{E}[Y(Z_1/e_1(X_1) - \Pi_{j=1}^{\overline{w}} Z_j^*)] = \tau_{ATR} \Pr(R = 1) + \delta_{ANR} \Pr(R = 0). \tag{35}$$

Similarly, as in Theorem 1, $\mathbb{E}[r(X_1)] = \Pr(R = 1)$. Since $\delta = 0$ implies $\delta_{ANR} = 0$, the statement holds. □

Proof of Theorem 3. I demonstrate the result for the upper bound, the result for the lower bound is symmetric. Following reasoning analogous to Theorem 2, yields:

$$\mathbb{E}[Y(Z_1/e_1(X_1) - \Pi_{j=1}^{\overline{w}} Z_j^*)] = \tau_{ATR} \Pr(R = 1) + \delta_{ANR} \Pr(R = 0), \tag{36}$$

$\mathbb{E}[r(X_1)] = \Pr(R = 1)$, and $\mathbb{E}[1-r(X_1)] = \Pr(R = 0)$. Theorem 1 further implies that $\mathbb{E}[m^U(G, \eta^0)]$

is a sharp upper bound on $\tau_{ATR}\Pr(R=1)$. Together, these results imply that $\mathbb{E}[m^S(G,\eta^0) - m^U(G,\eta^0)]$ is a sharp lower bound on $\delta_{ANR}\Pr(R=0)$. Thus, the statement holds. □

A2 Estimation

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

$$\mathbb{E}[\psi^{L+}(G,\xi^0)] = \mathbb{E}[m^L(G,\eta^0)],\ \mathbb{E}[\psi^{U+}(G,\xi^0)] = \mathbb{E}[m^U(G,\eta^0)]. \tag{37}$$

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

$$\xi_r = \{e_1^*(x_1^*),\ldots,e_l^*(x_l^*,r),\ldots,e_{\overline{w}}^*(x_{\overline{w}}^*),r_1(x_1^*),\ldots,r_{\overline{w}}(x_{\overline{w}}^*),r^+(x_1^*),q(p(x_1^*),x_1^*), \tag{38}$$

$$q(1-p(x_1^*),x_1^*),\beta_1(x_1^*),\ldots,\beta_{\overline{w}}(x_{\overline{w}}^*),\beta^+(x_1^*),z^{U+}(x_1^*),z^{L+}(x_1^*)\}, \tag{39}$$

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 (2025). A proof of orthogonality 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 $e_j^*(X_j^*)$.

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

$$\widehat{\theta_b} = \left(\Sigma_i\left(\psi^{b+}(G_i,\widehat{\xi_i})1_{\{p(X_1^*)\leq 1\}} + \psi^-(G_i,\widehat{\xi_i})1_{\{p(X_1^*)>1\}}\right)\right) / \left(\Sigma_i \psi^R(G_i,\widehat{\xi_i})\right) \tag{40}$$

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, which may occur in estimation when the true shares are close. ψ^- is constructed such that under the assumptions in Theorem 1 $\mathbb{E}[\psi^-(G,\xi)]/\mathbb{E}[r(X_1)] = \mathbb{E}[Y_1(1) \mid R^+ = 1] - \mathbb{E}[Y_0(0) \mid R=1]$. I address potential monotonicity violations further in Appendix SA1.

I estimate z_t^{U+} and z_t^{L+} by trimming the sample above or below estimated quantiles and estimating conditional expectations. While theoretical results for nonparametric estimation of truncated expectations exist (Olma, 2021), no implementation is currently available. The covariates in X_1^* include the woman's and her partner's earnings and work hours measured in the year before the woman's first IUI, and other covariates included in the first propensity score. The covariates in X_k^* additionally include those from the propensity scores at all procedures up to and including 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).

I discuss estimation of the sequential IV estimand in Appendix SA2. I translate the estimated

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

bounds on treatment effects and the sequential IV estimates into bounds on timing effects using Equation 8.

References

Adams, A., Jensen, M. F., & Petrongolo, B. (2024). Birth timing and spacing: Implications for parental leave dynamics and child penalties. *IZA Discussion Paper*.

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., 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. (2025). Reconciling estimates of the long-term earnings effect of fertility. *Unpublished working 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*.

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.

Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, *125*(1), 215–261.

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 working paper*.

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.

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. (2024). The labor market returns to delaying pregnancy. *Unpublished working paper*.

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

Han, S. (2021). Identification in nonparametric models for dynamic treatment effects. *Journal of Econometrics*, *225*(2), 132–147.

Heckman, J. J., Humphries, J. E., & Veramendi, G. (2016). Dynamic treatment effects. *Journal of econometrics*, *191*(2), 276–292.

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.

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.

Jakobsen, K. M., Jørgensen, T. H., & Low, H. (2022). Fertility and family labor supply.

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

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*, *109*, 122–126.

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

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

Leridon, H. (2004). Can assisted reproduction technology compensate for the natural decline in fertility with age? a model assessment. *Human reproduction*, *19*(7), 1548–1553.

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.

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*.

Semenova, V. (2025). Generalized lee bounds. *Journal of Econometrics*, *251*, 106055.

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.

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 SA1.2 compares my estimates with conventional methods. Appendix SA1.2 presents an extension to bound the effects over time for a stable group. Appendix SA1.3 tests the monotonicity assumption, while Appendix SA1.4 reports the main estimates allowing the direction of monotonicity to vary with covariates. Appendix SA1.5 presents estimates of work hours adjusted for potential parental leave. Appendix SA1.6 reports effect sizes relative to remaining childless. Appendix SA1.7 addresses age differences between partners. Appendix SA1.8 examines heterogeneity by pre-IUI covariates. Appendix SA1.9 presents estimates of the impact on antidepressant uptake.

SA1.1 Comparison with Conventional Methods

I first compare my bounds with those based solely on the first IUI, without using monotonicity, which are mathematically equivalent to the approaches in Zhang & Rubin (2003) and Lee (2009) (ZRL). Figure SA1 shows that, in every year following childbirth, ZRL bounds are 4.5 to 8 times wider and do not rule out either large positive or large negative effects on women's outcomes. Even when monotonicity is leveraged, these bounds remain 3.8 to 5.7 times wider than mine. Thus, exploiting sequential quasi-experimental assignment yields substantially more informative estimates. Notably, this improvement is not a data artifact but a theoretical guarantee.[15]

Next, I compare my estimates to those based on the IV and recursive IV approaches. Implementation details are provided in Appendix SA2. Figure SA2 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 are more negative. The lower bounds are attainable under strong negative selection of reliers with respect to treated labor market outcomes. One reason for this difference could be that the IV covers compliers whereas the bounds cover reliers. To explore this further, I estimate effects on reliers using the sequential IV approach described in Section 3.2.4. The results are similar to the conventional IV estimates. Since both the bounds and the sequential IV approach estimate effects for reliers, the comparison indicates that the timing of treatment may play an important role. Moreover, because the non-reliers who drive this gap are a subset of always-takers (see Section 3.2.4), this implication extends to conventional IV estimates as well.

Figure SA2 also presents recursive IV estimates for women's outcomes. In contrast to the conventional IV estimates, these estimates are consistent with the smallest negative effects implied by the bounds, which are attainable under strong positive selection of reliers with respect to treated

[15] ZRL bounds are wider because, with only the first IUI, the average control outcome can be identified for compliers but not for the broader group of reliers. The smaller the group, the more extreme the bounds for their average treated outcomes (if everyone is a complier, no trimming is needed and the effect bounds collapse to the average treatment effect).

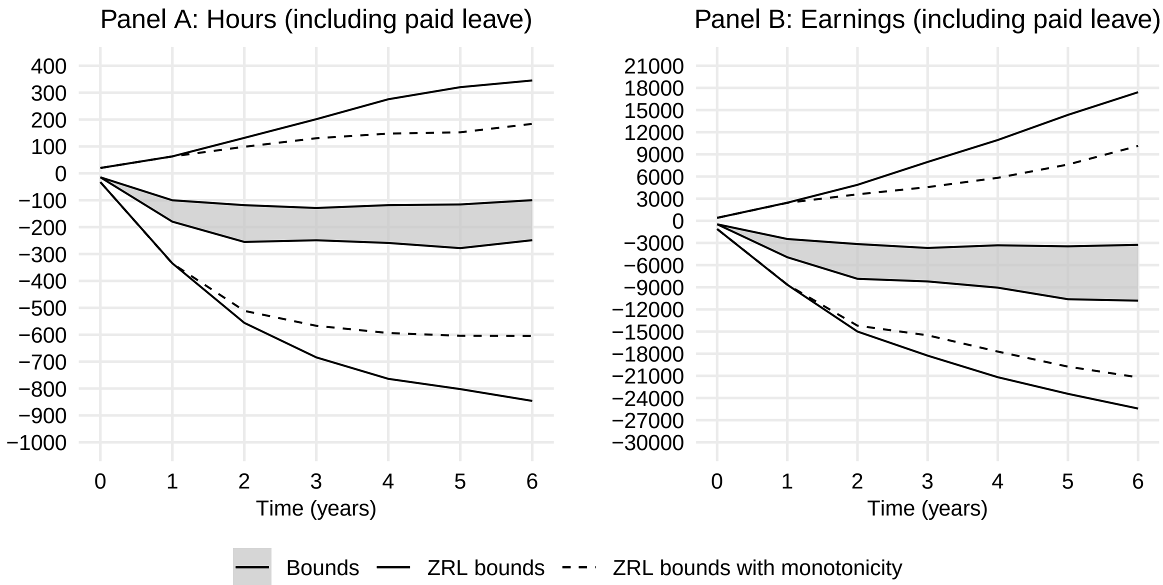

Figure SA1: Comparison with ZRL Bounds for Effects on Women's Outcomes

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 subsequent 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.

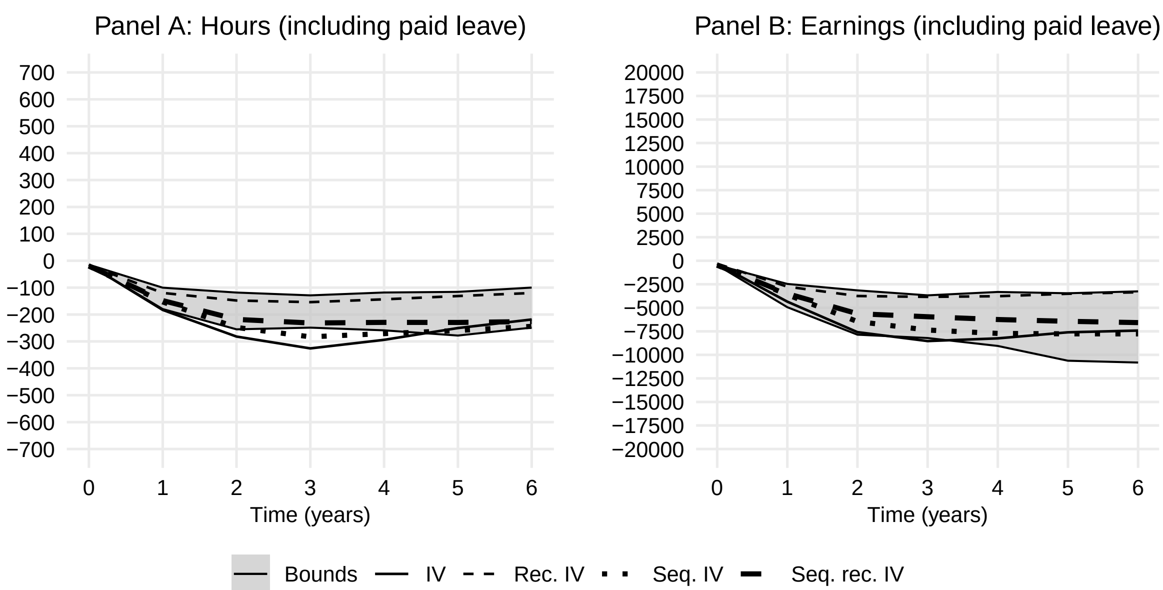

Figure SA2: Comparison of Different Methods for Effects on Women's Outcomes

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.

labor market outcomes. Figure SA2 additionally shows recursive estimates constructed using sequential IV (see Appendix SA2 for details). These estimates differ from the conventional recursive IV results, lying near the lower bound for hours and closer to the center of the bounds for earnings. Because these estimates would be expected to coincide if timing effects were absent, their divergence indicates that timing effects may play an important role.

This comparison illustrates how the proposed method complements existing approaches. Even without restrictions on how effects vary with treatment timing, duration, or life-cycle stage, the bounds establish a relatively narrow range for the impacts. This shows that informative conclusions can be drawn without imposing strong assumptions about treatment effect heterogeneity.

SA1.2 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 IUI 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 SA3 presents estimates for women who remain reliers six years after their first IUI, which are comparable to the baseline results.

SA1.3 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 SA4 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 nonparametrically, formally testing whether their differences allow rejecting monotonicity is not trivial, but comparing them offers some insight. Panel A in Figure SA5 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 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.

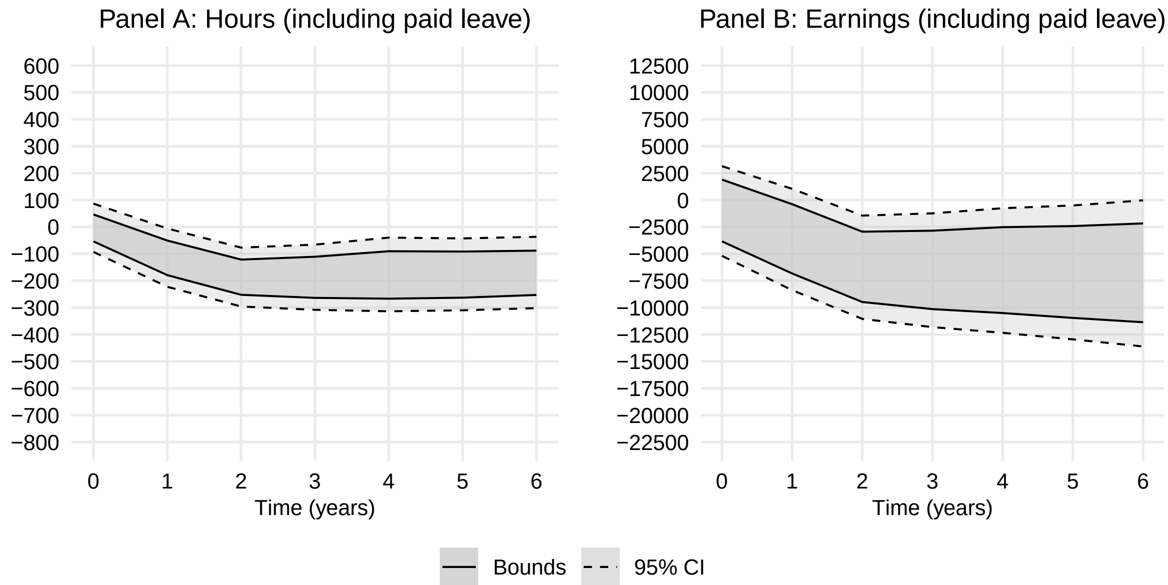

Figure SA3: Effects for Women who Remain Childless

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.

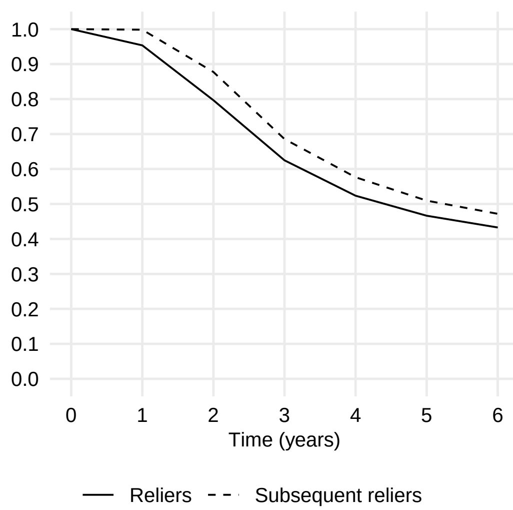

Figure SA4: Estimated Relier and Subsequent Relier Shares

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.

To formally test monotonicity using covariates, I adapt the approach of Semenova (2025). I partition the sample into $J = 25$ discrete cells $\mathcal{X}^*_{1,j}$ based on quintiles of women's work hours and age 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

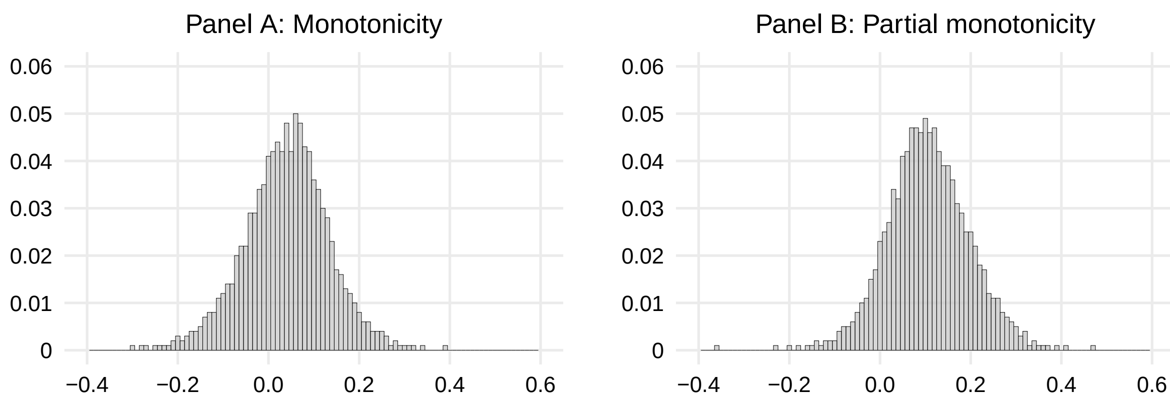

Figure SA5: Histogram of Difference in Subsequent Relier and Relier Shares

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* – estimated treating women who uptake antidepressants or separate from their partner after failing to conceive as if they had conceived naturally, and then using 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.

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 SA5, 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 the same conclusion.

The Panel B in Figure SA5 repeats the above when restricting focus to reliers who would remain cohabiting with their partners and not take up antidepressants after failing to conceive (and treating those who separate from their partners or take up antidepressants after procedure failure as if they had children, excluding them from the population covered by the bounds). The estimated difference between the subsequent relier and relier shares is below zero for only 11% of observations and below -0.1 for just 1%, providing stronger support for monotonicity in this group.

SA1.4 Relaxing Monotonicity

To address violations of monotonicity in the Lee (2009) setting, Semenova (2025) 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 yield nearly identical results.

Define $\mathcal{X}^*_{help} = \{x : r^{+}(x) \geq r(x)\}$ and $\mathcal{X}^*_{hurt} = \mathcal{X}^*_1 \backslash \mathcal{X}^*_{help}$. The relaxed monotonicity assumption

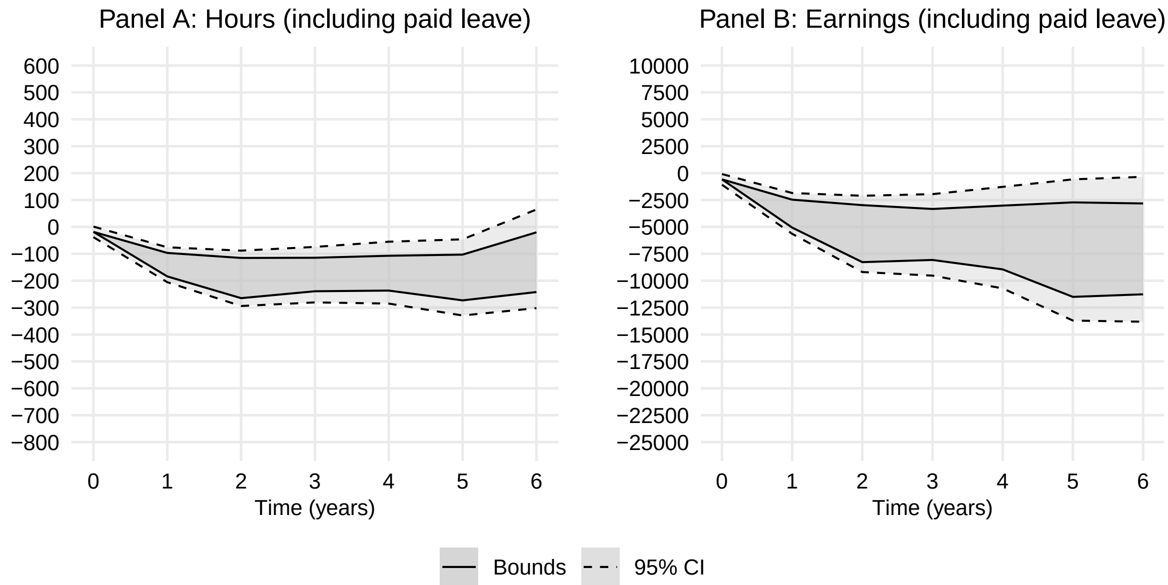

Figure SA6: Effects on Women Under Covariate-Conditional Monotonicity

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 SA1.4. 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.

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 orthogonal 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:

$$\widehat{\theta}_b = \frac{\sum_i \left(\psi^{b+}(G_i, \widehat{\zeta}_i)1_{\{p(X_1^*)\leq 1\}} + \psi^{b-}(G_i, \widehat{\zeta}_i)1_{\{p(X_1^*)>1\}}\right)}{\sum_i \left(\psi^{R}(G_i, \widehat{\zeta}_i)1_{\{p(X_1^*)\leq 1\}} + \psi^{R+}(G_i, \widehat{\zeta}_i)1_{\{p(X_1^*)>1\}}\right)}. \tag{41}$$

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 SA6 presents the estimates for women's outcomes, which closely resemble the baseline results.

SA1.5 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 effects on actual hours worked fall within the union of the bounds obtained using reported and adjusted values. Figure SA7 shows the results for female work hours and the corresponding gender gap. The estimates change little beyond the first year of parenthood.

SA1.6 Relative Effects

Figure SA8 reports the estimated effects of parenthood relative to remaining childless.

SA1.7 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

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

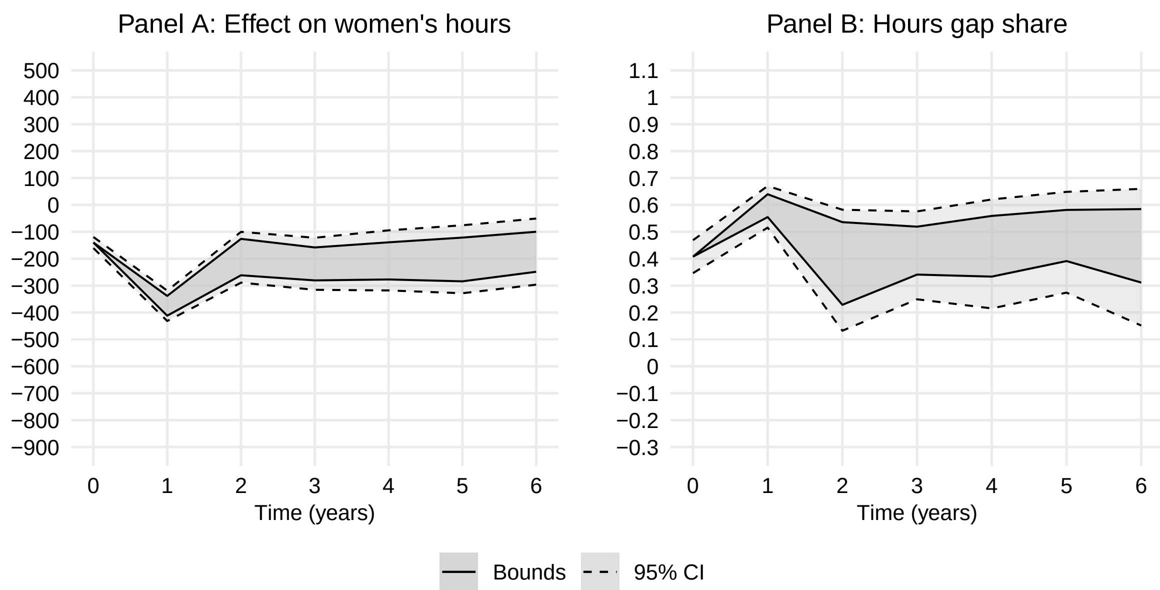

Figure SA7: Leave-Adjusted Estimates of Work Hours

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 SA1.5 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.

age differences, which may distort its relation to aggregate gender inequality, as men's outcomes 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. Instead, I lag men's outcomes to match the woman's age (e.g., by two years if she is younger), so gender gaps reflect comparable life-cycle stages. This means that in the first few years after the first IUI, some men's outcomes are measured before parenthood, which is appropriate if fatherhood effects are small. The adjustment reduces the sample by 19%, yielding 10,310 observations.

Figure SA9 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.

SA1.8 Heterogeneity by Covariates

To assess heterogeneity by pre-IUI covariates, I adapt the approach of Heiler (2024), which involves regressing the constructed moments on dummies for different heterogeneity dimensions (e.g., below- vs. above-median age). Figure SA10 presents the results for women's cumulative work hours and earnings over the first six years after the first IUI, split by median age, earnings, and work hours in the year before the first IUI. The results do not rule out homogeneous effects. Bounds are wider for groups with higher pre-IUI earnings and hours, reflecting greater dispersion in their outcome distributions.

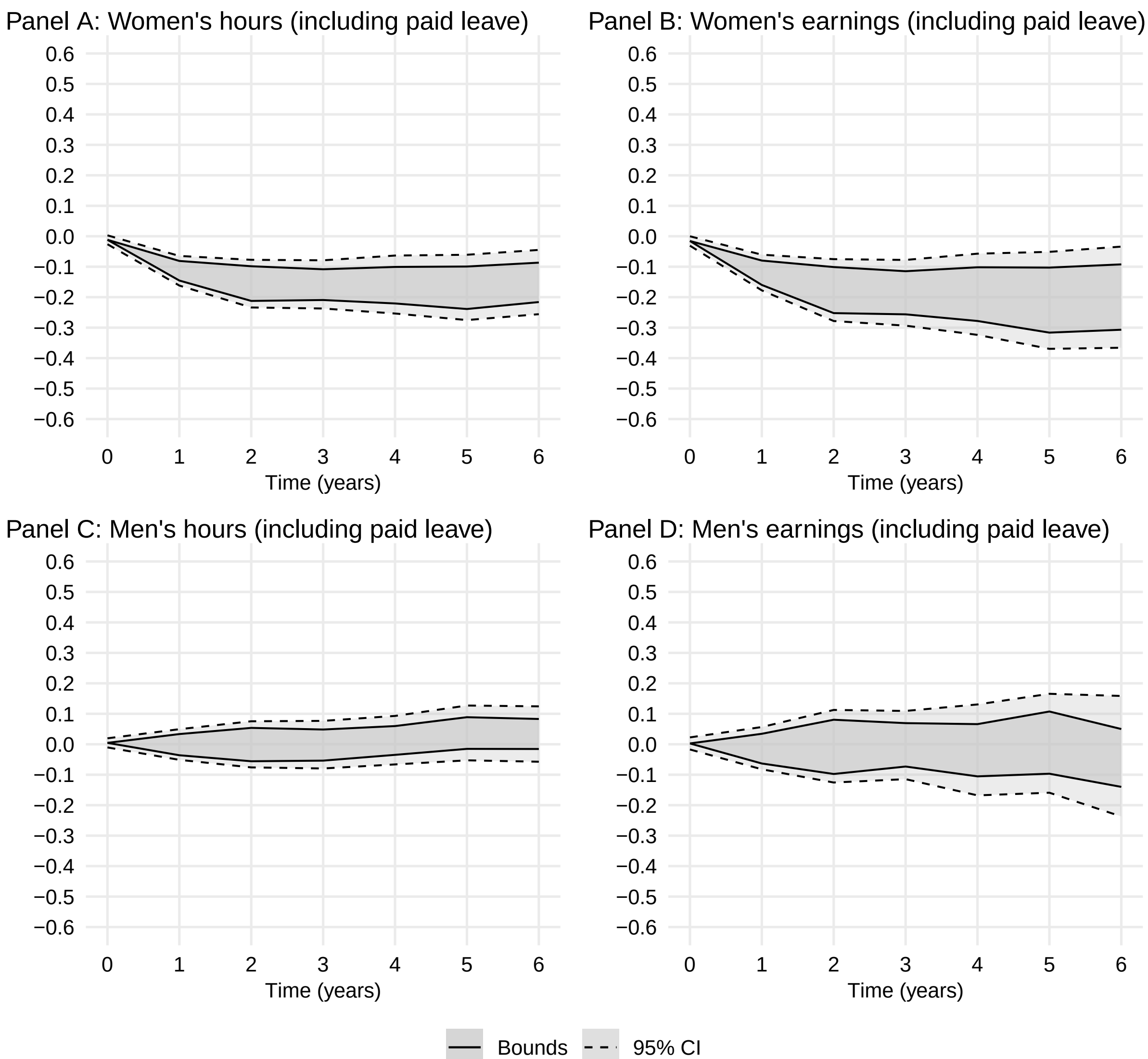

Figure SA8: Relative Effects of Parenthood

Note: Effects relative to childlessness are calculated as a/b, where a is the average effect for women reliant on procedure success (or their partners), and b is the average point-identified control outcome for the same group. Both are estimated using orthogonal moments from the baseline specification. Confidence intervals are 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.

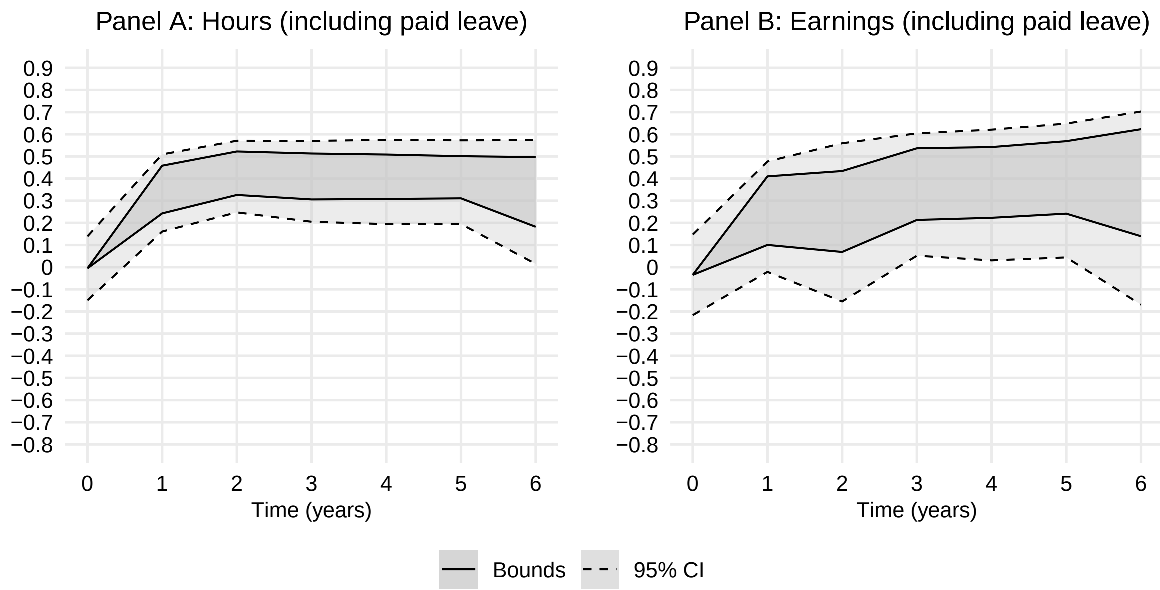

Figure SA9: Age-adjusted Share of Within-couple Gender Inequality Caused by Parenthood

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).

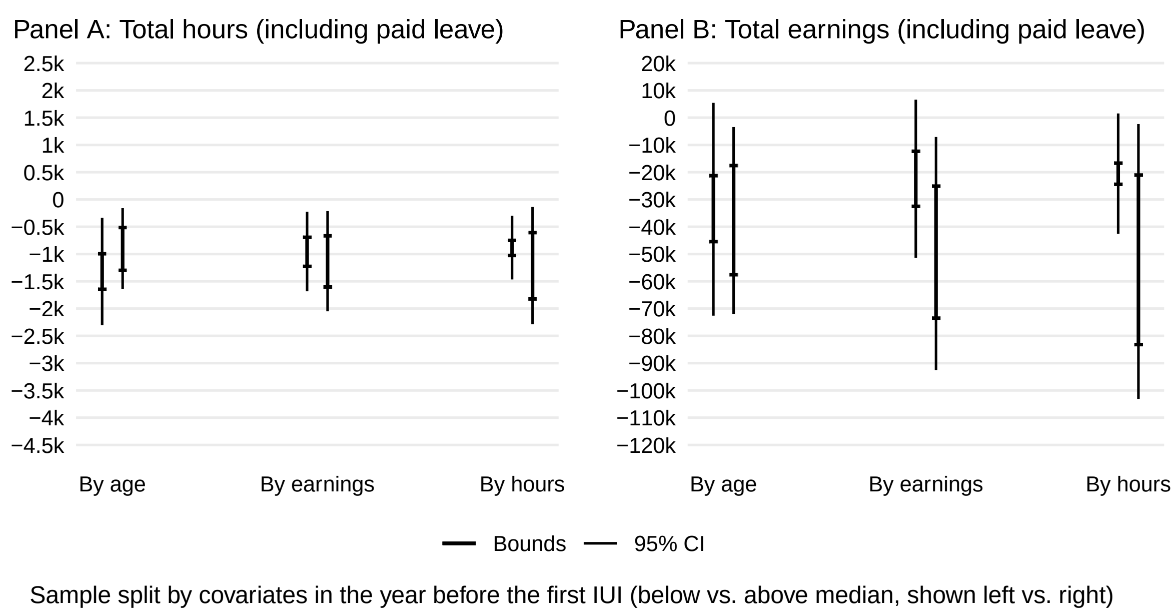

Figure SA10: Heterogeneity in Cumulative Effects for Women in the First Six Years

Note: Heterogeneity in cumulative effects on women's work hours and earnings (EUR) in the first six years of motherhood, estimated using the baseline specification. The sample is split by median age, earnings, and work hours in the year before 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 procedure.

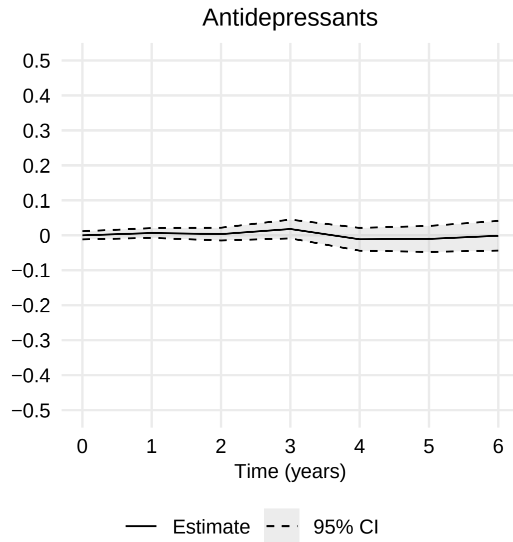

Figure SA11: Effects on Antidepressant Uptake

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.

SA1.9 Effects on Mental Health

Figure SA11 presents the estimates for the effect on antidepressant uptake based on the sequential IV approach. The effects are precisely estimated and indistinguishable from zero.

SA2 Auxiliary Estimation Details

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

$$D = Z_1\beta^{FS} + X_1\chi^{FS} + \varepsilon^{FS}, \tag{42}$$

and the second-stage specification is:

$$Y = \widehat{D}\beta^{IV} + X_1\chi^{IV} + \varepsilon^{IV}, \tag{43}$$

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 the sequential IV, recursive IV, and sequential recursive IV methods, I first introduce auxiliary functions that enable double-robust estimation. As with conventional augmented inverse probability weighting estimators, the key idea, in intuitive terms, is to residualize outcomes with respect to covariates, thereby improving precision when estimating average differences between treatment and control groups. For a comprehensive discussion of robust estimators for time-varying treatments, see Hernán & Robins (2020). The asymptotic properties of these estimators apply directly to the moment conditions introduced below.

$$g_a^{0+}(G) = \gamma_a^{0,1+}(X_1^*) + (a - \gamma^{0,1+}(X_1^*))\Pi_{j=1}^{\overline{w}} \frac{(1-Z_j)}{(1-e_j^*(X_j^*))} \tag{44}$$

$$+ \Sigma_{k=1}^{\overline{w}} \Big[1_{\{A \geq k\}} \Pi_{j=1}^{k-1} \frac{(1-Z_j)}{1-e_j^*(X_j^*)} \frac{(e_k^*(X_k^*) - Z_k)}{1-e_k^*(X_k^*)} [\gamma_a^{0,1+}(X_1^*) - \gamma_a^{0,k+}(X_k^*)] \Big] \tag{45}$$

$$g_a^0(G) = \gamma_a^0(X_1^*) + (a - \gamma_a^0(X_1^*)) \frac{Z_1}{e_1^*(X_1^*)} \tag{46}$$

$$g_a^1(G) = \gamma_a^1(X_1^*) + (a - \gamma_a^1(X_1^*)) \frac{1-Z_1}{1-e_1^*(X_1^*)}, \tag{47}$$

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

Applying the Lemma yields:

$$\frac{\mathbb{E}[g_Y^1(G) - g_Y^0(G)]}{\mathbb{E}[g_D^1(G) - g_D^{0+}(G)]} = \mathbb{E}[\tau|C=1] + \mathbb{E}[\delta|C=0] \frac{\Pr(C=0)}{\Pr(C=1)} \tag{48}$$

and:

$$\frac{\mathbb{E}[g_Y^1(G) - g_Y^{0+}(G)]}{\mathbb{E}[g_D^1(G) - g_D^{0+}(G)]} = \mathbb{E}[\tau|R=1] + \mathbb{E}[\delta|R=0] \frac{\Pr(R=0)}{\Pr(R=1)}, \tag{49}$$

which correspond to the IV and sequential IV estimands, respectively. I construct the corresponding estimators using empirical counterparts of the expectations in equations (48) and (49). $\gamma_a^1(X_1^*)$ 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_1^*, 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.

If the conditional expectation of Y is correctly specified, it is straightforward to show using the Delta method that orthogonality allows the nuisance functions to be treated as fixed, so standard errors are proportional to the standard deviation of the sample moment averages. To ensure robustness to misspecification, standard errors for the sequential IV estimator are obtained via a multiplier bootstrap with 100 draws, using weights $w_i \sim \exp(1)$. The weights are applied at each step of the regressions used to estimate the nuisance functions and in computing the averages of the moment functions. Standard errors for the bounds on δ_{ANR} are obtained by, for each estimate in the resulting distribution of the sequential IV estimator (and the corresponding relier share), applying the same observation weights to compute the averages of the moments in the bounds for τ_{ATR} (but not in the estimation of the nuisance functions for these bounds, since they are estimated nonparametrically and thus robust to misspecification), and using the resulting estimate pairs to construct the estimate distribution following the formula in Theorem 3.

For the recursive IV approach, assume that a woman's outcome in period k from first IUI is

$$Y_k = Y_{0k}(0) + \sum_{j \leq k} 1_{\{K_k = j\}} \tau_j + \varepsilon_k, \tag{50}$$

Table SA2: Balance in Later Procedures

	(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 procedure 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.

where $Y_{zk}(d)$ denotes the potential outcome in period k when the first IUI outcome is z and parenthood status is d; K_k is the number of years since first birth; and $\tau_j = \mathbb{E}[Y_{1j}(1) - Y_{0j}(0)]$ denotes the average effect of being in the jth year of parenthood. 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 (48) 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 (49).

SA3 Additional Balance Results

Table SA2 presents balance results for subsequent procedures up to the tenth. Since these procedures also include IVF, I additionally control for each partner's age interacted with treatment type. This ensures that procedure 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 women's types and potential outcomes. Overall, the results suggest no systematic differences in pre-IUI outcomes between those with successful and unsuccessful subsequent procedures, supporting the conditional local sequential unconfoundedness assumption.