The causal impact of children?

Rethinking the identification and estimation of child penalties

Katharina Kaeppel*

January 30, 2026

Abstract

A fast-growing literature estimates child penalties — the impact of parenthood on the labor market outcomes of mothers relative to fathers — using event study design. This paper clarifies the identification assumption underlying the most commonly used specification with event time, year, and age fixed effects, and shows that this approach is prone to bias in the presence of heterogeneity by age and calendar year, even under valid identification. I then propose an unbiased estimator, using the same identification framework. Applied to French administrative data, my estimates for mothers are 27% smaller than those using the standard approach. The paper also discusses identification under an alternative parallel trend assumption, that is, when controlling for individual unobserved heterogeneity, and highlights key trade-offs in defining counterfactuals, as one cannot simultaneously absorb individual heterogeneity, age, age at birth, and event time. I provide a framework to build an estimator step-by-step, which is easily adaptable depending on which comparisons researchers assess most appropriate. Finally, I document substantial heterogeneity by age at childbirth for both mothers and fathers. By eight years after the birth of their first child, mothers aged 20–25 at birth lose 50% compared to their counterfactual earnings, compared to 12–20% for older mothers. Fathers aged 20–25 at first birth earn on average 26% less eight years later, while effects for older fathers are close to zero. These findings have important implications for both policy and applied research using event study designs to estimate the causal impact of children.

Keywords: child penalty, female labor supply, heterogeneous treatment effects, event studies

JEL: C51, J13, J16, J22, J31

Acknowledgements: I am grateful to my supervisors, Thomas Breda and Marc Gurgand, for their guidance and support throughout this project. I also thank Philipp Ketz, David Margolis, Maxime Tô, and participants at the Labor Chair seminar and the internal Labor and Public Economics seminar at PSE for helpful discussions. Earlier conversations with Clément de Chaisemartin on related work were influential in shaping the ideas developed in this paper. All remaining errors are my own.

*Paris School of Economics (PSE).

1. INTRODUCTION

Despite substantial progress over the last century, gender gaps persist in the labor market, and progress has stalled in recent decades. Recent research suggests that, in high-income countries, child penalties — the relative impact of parenthood on women's versus men's labor market outcomes — account for most of the remaining gender inequality (Kleven et al., 2019; Cortés and Pan, 2023; Kleven et al., 2025). These penalties are estimated using event study designs, which allow researchers to trace the dynamic effects of parenthood for the full population of parents (Angelov et al., 2016; Kleven et al., 2019; Andresen and Nix, 2022; Cortés and Pan, 2023), offering advantages over earlier identification strategies such as twins and same-sex instruments (Rosenzweig and Wolpin, 1980; Angrist and Evans, 1998) or IVF-based designs (Lundborg et al., 2017).

However, credible causal identification in this setting is complicated. Decisions of whether and when to have a child are not exogenous: parents differ from non-parents, and among parents, fertility timing, that is, age at birth, is endogenous and correlated with factors—education, career trajectories, unobserved ability—that shape potential labor market outcomes. At the same time, earnings trends differ with age and macroeconomics conditions. As a result, constructing a credible counterfactual in event study settings requires navigating non-trivial identification trade-offs, since one cannot at the same time compare individuals at the same age who also have their child at the same moment in life. Yet the literature offers limited guidance on how different empirical strategies navigate these challenges, their underlying identification framework, and how to assess their credibility. This paper seeks to fill that gap by providing guidance on identification and estimation of child penalties under alternative empirical strategies.

The most commonly used specification[1], pioneered by Kleven et al. (2019), estimates gender-specific impacts of children by regressing earnings on a set of event time dummies, along with calendar year and age fixed effects (FE). Event time refers to the years relative to the event (zero indicating the year of the childbirth), and age and year FE are added to control flexibly for macroeconomic shocks and life-cycle trends. A second set of approaches instead adopts a difference-in-differences (DiD) design[2], controlling for individual time-invariant unobserved heterogeneity. These strategies rely on different identification assumptions that give rise to distinct estimands and estimation approaches.

The paper begins by formalizing the identification assumptions underlying the standard approach

[1] See, for instance, Kleven et al. (2019); Meurs and Pora (2019); Kleven et al. (2021); Kleven (2022); Andresen and Nix (2022); Cortés and Pan (2023); Lim and Duletzki (2023); Kleven et al. (2024, 2025); Andresen and Nix (2025).

[2] Examples include Casarico and Lattanzio (2023); Coudin et al. (2025); Bonnet et al. (2025); Meurs and Pora (2025).

with age and year fixed effects. Identification of gender-specific impacts of children relies on a conditional independence assumption (CIA): fertility timing must be mean independent of untreated potential outcomes after conditioning flexibly on age and calendar year. Applying insights from Goldsmith-Pinkham et al. (2024), I show that even if this assumption holds, event-time coefficients are prone to bias due to contamination from other periods. I discuss the conditions under which this bias is small or large in child penalty estimations and provide descriptive evidence that it is likely relevant in practice. In particular, older parents differ substantially from younger parents[3], suggesting that treatment effects may plausibly vary with age, while age also predicts selection into parenthood. I illustrate the resulting bias under varying degrees of treatment effect heterogeneity using a numerical example.

In light of these insights, I next propose an estimator that avoids contamination bias while relying on the same identification framework. Under CIA, counterfactual outcomes can be estimated from untreated observations. This motivates the use of an imputation approach à la Borusyak et al. (2024, *BJS* hereafter). Following the standard specification, untreated potential outcomes are modeled as a flexible function of age and calendar year and estimated using untreated observations only. Counterfactual earnings for treated individuals are then obtained by plugging their age and year values into this model. Individual treatment effects are defined as the difference between observed and predicted outcomes and can be aggregated for any population of interest. This estimator is easy to implement[4] and flexible, for instance allowing the inclusion of additional covariates[5] or restrictions on the group used to estimate counterfactual earnings. Whereas CIA appears a strong assumption in the context of child penalties, it allows to estimate child penalties without the need for high-quality admin data using cross-sectional data, and hence for a larger set of countries (see also Kleven et al., 2025). I separately discuss two strategies to formally test the plausibility of CIA in this setting.

Applying the imputation approach to a sample of about 130,000 French parents observed from 2010 to 2020, I find that for women, estimated earnings losses eight years after childbirth are 27.4% smaller in absolute terms than those obtained using the standard approach, consistent with the presence of contamination bias. Mothers by that time lose on average €5,400, corresponding to 20% of counterfactual earnings. Earnings losses are highly heterogeneous by age at birth and decline sharply with age at childbirth: they are largest for mothers in their early twenties and

[3]Differences in labor market outcomes for parents with different ages at birth have previously been documented (Wilde et al., 2010; Adda et al., 2017; Goldin et al., 2022). Melentyeva and Riedel (2023) discuss the implications for child penalty estimations using TWFE models, as does contemporaneous work by Leventer (2025) in Israel.

[4]It can be implemented directly using the *did_imputation* command provided by BJS, which includes built-in tools for inference.

[5]For example, Adams et al. (2024) argue that time-since-graduation and education should be accounted for, as both predict earnings and fertility timing.

substantially smaller for those giving birth in their early thirties, with little additional change at later ages. For men, results differ across approaches, but for them CIA, that is, as-good-as-random timing of birth conditional on age and year, is clearly rejected in the data using the provided tests, suggesting alternative strategies.

Motivated by this evidence, I next discuss identification strategies that control for individual time-invariant unobserved heterogeneity. In this case, identification relies not on conditional independence but on a weaker parallel trends (PT) assumption: absent children, parents' earnings would have evolved in parallel to those of an appropriate control group.[6] When rich panel data are available, PT-based strategies provide a natural alternative to CIA. To formalize identification under PT and derive a heterogeneity-robust estimator aligned with this framework, I adopt the forward-engineering approach of Baker et al. (2025). This approach begins with a clearly defined target parameter, specifies the identifying assumptions required to recover it, and then derives an appropriate estimator. In practice, credibility hinges on how the control group is defined and on which dimensions parallel trends are assumed to hold. My preferred strategy compares parents' earnings trajectories around childbirth to individuals of the same age, observed over the same calendar period, who have not yet had children but will have them later, yielding age- and cohort-specific effects that can be transparently aggregated. I also relate this approach to the stacked estimator proposed by Melentyeva and Riedel (2023), which represents an alternative heterogeneity-robust strategy within a parallel trends framework, offering gains in precision at the cost of stronger restrictions on the time-series evolution of untreated potential outcomes.

The forward-engineering perspective makes explicit a key identification trade-off when using not-yet-treated parents as controls. Researchers must choose whether to align individuals by age—ensuring comparisons within the same life-cycle stage but across different fertility timing—or by age at childbirth, which aligns fertility timing but compares individuals at different ages. While challenges when comparing individuals who become parents at a different age have been discussed in related papers (Melentyeva and Riedel, 2023; Adams et al., 2024; Leventer, 2025), I discuss the alternative strategy of matching on age at childbirth, highlighting this key-trade off. I also evaluate both strategies empirically, and my results suggest that fixing age provides the more credible comparisons.

In terms of aggregate effects using my preferred estimator, I find that by eight years after childbirth, mothers earn approximately €7,100 less than their counterfactual earnings, corresponding to a 23% decline. These estimates are similar to those obtained using the standard event study approach. For fathers, results differ markedly across estimation strategies. The PT-based strategy

[6] I use "parallel trends" to refer to the identifying assumption in canonical TWFE models with unit and time fixed effects, following standard usage.

yields flat pre-trends and therefore appears more credible in this setting, revealing non-negligible costs for fathers: seven years after childbirth, fathers earn about €1,900 less than their counterfactual earnings, corresponding to a 4.7% loss.

Beyond aggregate effects, I uncover substantial heterogeneity by age at childbirth for both women and men under the PT framework. While all women experience a sudden drop, earnings losses are largest among those who become mothers at younger ages, both in the short and medium term. After eight years, women aged 20–25 at first birth earn around €14,000 less than their counterfactual earnings, corresponding to a 46.4% reduction. Losses are smaller for older mothers, amounting to 19.9% for mothers aged 26–30, 13.1% for those aged 31–35, and 15.3% for those aged 36–40.

Striking heterogeneity also emerges for fathers. Younger fathers (aged 20–25 at birth) experience substantial earnings losses of approximately €9,000 by eight years after childbirth, corresponding to a 25.5% decline. Fathers aged 26–30 also face a significant loss of 8.5%, while I cannot reject the null of no effect for men who become fathers between ages 31 and 40, consistent with the broader finding in the literature of no average fatherhood penalty. All results are robust to alternative estimators.

These results are of clear policy relevance and highlight the importance of accounting for heterogeneity in the timing of parenthood.

Related literature This paper contributes to the fast-growing applied literature on child penalties (Angelov et al., 2016; Kleven et al., 2019, 2021; Goldin et al., 2022; Andresen and Nix, 2022; Kleven, 2022; Kleven et al., 2025; Coudin et al., 2025). This work documents large and persistent labor market costs of motherhood, and close-to-zero effects for fathers. I add to this empirical literature by providing new evidence on substantial heterogeneity in the impact of children, including sizeable penalties for young fathers. Age at birth remains a largely underexplored dimension of heterogeneity, with the exception of Melentyeva and Riedel (2023), who document larger effects for younger mothers in Germany but do not observe fathers, and Adams et al. (2024), who study heterogeneous effects by parents' skill level—which is highly correlated with timing of birth—in Denmark.

Within the child penalty literature, a more recent strand examines identification and estimation of the cost of parenthood (Melentyeva and Riedel, 2023; Thakral and Tô, 2025; Leventer, 2025). My contribution differs in scope: I provide a synthesis of the empirical strategies used in practice, formalize the distinct identification assumptions underlying each approach, and provide unbiased estimators aligned with these frameworks.

In addition to providing a unified framework, I complement recent methodological contributions. Parallel work by Thakral and Tô (2025) emphasizes the collinearity that prevents age, event time, age-at-birth, and individual fixed effects from being jointly included in an OLS regression; I instead focus on the related underlying identification trade-off. Whereas they discuss OVB when falsely omitting dummies, I also show that bias may arise even under valid identification. Contemporaneous work by Leventer (2025) studies a different estimand—differences in gender-specific estimates—and establishes identification under a triple-difference design; in contrast, I focus on the gender-specific effects that constitute the building blocks of child penalties and discuss their identification under both CIA and PT. The closest paper is Melentyeva and Riedel (2023) who discuss the heterogeneity across older and younger mothers and document bias in child penalty estimations using canonical TWFE models with individual FE. I complement their insight by characterizing contamination bias in the more commonly used specification with age and year fixed effects, formalizing the identification framework underlying this approach, and extending the heterogeneity analysis to fathers.

The identification trade-off I highlight is not specific to child penalties. It arises in designs where individuals are treated at different ages, age is correlated with selection into treatment, and earnings follow highly non-linear life-cycle profiles. This applies to a range of other life events affecting earnings, such as cohabitation or marriage, as well as health-related shocks. In these settings, researchers face the same choice between imposing restrictions on untreated earnings trajectories and selecting appropriate comparison groups.

The rest of this paper is organized as follows. Section 2 presents the standard approach to estimating child penalties, formalizes the underlying identification assumptions, and discusses the sources of potential bias. Section 3 introduces an alternative imputation estimator, describes the data and sample, compares it to existing estimators, and documents heterogeneity by age at birth. Section 4 studies identification under parallel trends, compares estimates across alternative specifications, and documents heterogeneity in the costs of parenthood for both mothers and fathers. Section 5 concludes.

2. BIAS IN STANDARD CHILD PENALTY ESTIMATIONS

This section outlines the standard event study approach used to estimate child penalties. I clarify the identification assumptions underlying this approach and show that, even if they hold, the resulting estimates are prone to bias when treatment effects are heterogeneous across age or year. I discuss conditions when this bias is relevant in child penalty estimations and show how it varies with the extent of TE heterogeneity using a numerical example.

2.1. The standard approach

The most commonly used approach to estimate child penalties, pioneered by Kleven et al. (2019), estimates gender-specific dynamic effects of children by regressing individual earnings on a set of event time dummies, age FE, and calendar year FE:

$$Y_{it}^g = \kappa + \sum_{\ell \neq -1} \alpha_\ell^g \cdot \underbrace{\mathbb{I}[\text{et} = \ell]}_{\text{event time FE}} + \sum_a \beta_a^g \cdot \underbrace{\mathbb{I}[\text{age}_{it} = a]}_{\text{age FE}} + \sum_t \gamma_t^g \cdot \underbrace{\mathbb{I}[y = t]}_{\text{year FE}} + \nu_{it}^g \tag{1}$$

Y_{it}^g denotes earnings (in levels) of individual i of gender g in calendar year t and event time ℓ. Individuals who do not work are included with zero earnings. Event time $\ell = -1$ is omitted, and standard error are clustered at the individual level.

In a second step, the coefficients α_ℓ^g are scaled by counterfactual earnings to express treatment effects in percentage terms. Counterfactual earnings $\tilde{Y}_{it}^g$ are predicted using the same regression (1) but setting the event-time dummies to zero, that is,

$$\tilde{Y}_{it}^g \equiv \kappa + \sum_a \hat{\beta}_a^g \cdot \mathbb{I}[\text{age}_{it} = a] + \sum_t \hat{\gamma}_y^g \cdot \mathbb{I}[y = t]. \tag{2}$$

The relative coefficients are, then, for gender g at event time t

$$P_\ell^g \equiv \frac{\hat{\alpha}_\ell^g}{E[\tilde{Y}_{it}^g \mid et = \ell]}. \tag{3}$$

Finally, the child penalty is defined as the difference between male and female coefficients (at event time 10, or averaged over several post-birth event times), normalized by counterfactual female earnings[7]:

$$P_\ell \equiv \frac{\hat{\alpha}_\ell^m - \hat{\alpha}_\ell^w}{E[\tilde{Y}_{it}^w \mid et = \ell]}. \tag{4}$$

There are several challenges with this approach. In this paper, I will mainly focus on the first step, the identification and estimation of α_ℓ^g, which are used to estimate child penalties. I will show that α_ℓ^g does not map to a clear causal estimand of interest and is biased under heterogeneity, even under valid identification. Another problem of the standard approach is that counterfactual earnings $\tilde{Y}_{it}$ are predicted from a model estimated on both untreated and treated observations, which may

[7] Kleven et al. (2025) instead define the child penalty as $P_\ell^w - P_\ell^m$, that is the difference in the scaled effects. The distinction between these quantities is irrelevant for this paper.

underestimate true untreated potential outcomes, and hence overstate the cost of children (as first pointed out by Melentyeva and Riedel, 2023). My alternative estimation approach will instead include a clean way to estimate counterfactuals from untreated observations only.[8]

2.2. Parameter of interest and identification

I formalize the target parameter of the child penalty literature using a potential outcomes framework (Rubin, 1974), closely following the notation in Baker et al. (2025). Let G_i denote the calendar year in which individual i has their first child. I refer to G_i as the treatment date or cohort. For any g, let $Y_{it}(g)$ denote the potential outcome for individual i at time t if the first child arrives in year g, and let $Y_{it}(0)$ denote the potential outcome if i never has a child. The individual treatment effect at time t for an individual whose first child arrives in g is

$$\tau_{it} = Y_{it}(g) - Y_{it}(0).$$

In practice, we never observe $Y_{it}(0)$ and $Y_{it}(g)$ for the same individual and time. Instead, we observe

$$Y_{it} = (1 - D_{it})\, Y_{it}(0) + D_{it}\, Y_{it}(g),$$

where D_{it} equals one if i has (ever) had a child by time t. The object τ_{it} allows for unrestricted heterogeneity across individuals and over time, but this level of heterogeneity is not point-identified without imposing strong additional structure. Instead, empirical DiD designs typically focus on estimating averages of heterogeneous treatment effects (Borusyak et al., 2024; Baker et al., 2025).

In standard child penalty estimations, the target parameter is the average treatment effect on the treated (parents), the ATT, for gender s[9] at a given event time ℓ:

$$ATT^s_\omega(\ell) \;=\; \sum_{it \in \Omega} \omega_{it}\, \tau_{it},$$

The weights ω_{it} determine the population Ω the parameter refers to. They may reflect the sample composition at event time ℓ or be chosen to recover the ATT for the population of parents through reweighting (see Baker et al., 2025). In addition to the aggregated $ATT_\omega(\ell)$, we may also be interested in subgroup-specific effects $ATT(\ell \mid Z_i = z)$ if we expect the effect of children to vary along dimensions that seem policy relevant, such age as at birth or education level (e.g. Melentyeva and Riedel, 2023; Adams et al., 2024).

[8] For a detailed discussion of the identification of the quantity in Equation (4), see Leventer (2025), who show the quantities targeted in the literature are not identified and suggest an alternative estimand.

[9] I change notation here slightly to denote s as gender, since g stands for cohorts in the notation of Baker et al. (2025) which I follow.

Throughout, I focus on gender-specific parameters $ATT^s_\omega(\ell)$, defined separately for women and men. To keep notation light, I omit the superscript s in what follows; all parameters and estimators are understood to be gender-specific. As with the gender superscript, I omit the subscript ω going forward for notational simplicity, while keeping in mind that all ATT parameters are defined relative to an underlying weighting scheme and target population.

Identification under conditional independence Different assumptions can identify the unobserved counterfactual earnings at a given event time. For instance, assuming $Y_{it}(0)$ to be constant over time, the causal effect of children on parents is identified by the simple before-after comparison of their earnings. If fertility was as good as randomly assigned, potential earnings would be mean independent of the treatment by design, and the differences in earnings at a given time between parents and non-parents would identify the ATT (which in this case would equal the ATE, the average treatment effect).

Instead, the child penalty estimation in Equation (1) relies on a conditional independence assumption (**CIA**). Let $\mathbf{W}_{it}$ be the vector of controls, here age and year fixed effects, and denote by ET_{it} individual i's event time in year t. CIA requires that, conditional on age and year, the timing of childbirth is as good as random for untreated potential outcomes:

$$\mathbb{E}[Y_{it}(0) \mid ET_{it} = \ell, \mathbf{W}_{it}] = \mathbb{E}[Y_{it}(0) \mid \mathbf{W}_{it}] \quad \forall \ell. \tag{5}$$

Once we condition on age and calendar year, differences in timing of first birth (how many years before or after birth an individual is) do not predict the counterfactual earnings they would have in the absence of children. I will discuss tests for the plausibility of this assumption that use only untreated observations in Section 3.2.

In addition, we also need to impose a no-anticipation (**NA**) assumption, namely that children do not affect earnings *before* childbirth, so that for all $t < G_i$, we have

$$Y_{it} = Y_{it}(0) \tag{6}$$

NA may be violated if, for instance, pregnancy or mandated leave already affects earnings in the calendar year before birth. To address this concern, researchers have redefined the pre-treatment period $ET = -2$. To the extent there is anticipation for mothers in the form of a negative effect on earnings, estimates for the cost of motherhood will be underestimated. In contrast, violations of NA can also lead to overstatement of the cost of parenthood if earnings rise in anticipation of childbirth.

Under CIA and NA, untreated potential outcomes are identified from observed outcomes of untreated individuals:

$$\mathbb{E}[Y_{it}(0) \mid \mathbf{W}_{it}] = \mathbb{E}[Y_{it}(0) \mid ET_{it} < 0, \mathbf{W}_{it}] = \mathbb{E}[Y_{it} \mid ET_{it} < 0, \mathbf{W}_{it}] \,.$$

The first equality follows from **CIA**, and the second equality follows from **NA**.

Finally, the standard child-penalty specification in Equation (1) imposes an additional functional-form restriction on $Y(0)$. Let

$$\mu_0(\mathbf{W}_{it}) = \mathbb{E}[Y_{it}(0) \mid \mathbf{W}_{it}] \,.$$

The standard approach assumes that

$$\mu_0(\mathbf{W}_{it}) = \kappa + \mathbf{D}_{it}^{age} + \mathbf{D}_{it}^{year}, \tag{7}$$

that is, untreated earnings are modeled as the sum of an age effect and a year effect. In this approach, $\mu_0(\cdot)$ is not estimated in a separate step to estimate treatment effects, but interpreting the coefficients α_ℓ as causal effects implicitly requires this specification to recover the correct conditional mean of untreated potential outcomes. This assumption of the potential outcome model can also be seen in Equation (2).

To summarize, causal interpretation of the event-time coefficients α_ℓ requires two conditions. First, CIA and NA must hold so that $\mu_0(\mathbf{W}_{it})$ coincides with the conditional mean of observed earnings among untreated individuals. Second, the parametric restriction in Equation (7) must be correct so that age and year fixed effects recover this conditional mean. In the next subsection, I show that even if all these assumptions are satisfied, OLS estimates for $\hat{\alpha}_\ell$ are biased for the causal effect of children, unless much more restrictive assumptions on treatment effect homogeneity hold.

2.3. "Contamination" bias in child penalty estimations

The previous section states the causal estimand of interest and clarifies the identification assumptions underlying the empirical approach outlined in Section 2. This section shows that the event study estimates from Equation (1) do not generally map to this causal estimand of interest, even under valid identification assumptions.

2.3.1 Decomposition of event study coefficient (Goldsmith-Pinkham et al., 2024)

Goldsmith-Pinkham et al. (2024) show that an OLS regression of the sort

$$Y_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + g(\mathbf{W}_{it}) + \eta_{it}, \tag{8}$$

where $\mathbf{X}_{it}$ is a vector of mutually exclusive treatment indicators and $g(\mathbf{W}_{it})$ is a function of control variables produces biased TE estimates in the presence of heterogeneity. This setup nests many commonly-used approaches in applied research, including TWFE models or RCT with multiple treatments and controls. It also nests the standard child penalty estimation in Equation (1), with $\mathbf{X}_{it}$ being the vector of mutually exclusive treatment dummies $g(\mathbf{W_t})$ being a function of the controls, that is,

$$g(\mathbf{W}_{it}) = \kappa + \mathbf{D}^{\text{Age}}_{\mathbf{it}} + \mathbf{D}^{\text{Year}}_{\mathbf{it}} \tag{9}$$

Goldsmith-Pinkham et al. (2024)'s result thus shows that α_ℓ from Equation (1) is prone to bias if treatment effects vary by age or year. Event time coefficients at $et = \ell$ are a sum of two terms:

$$\alpha^g_\ell = \underbrace{\mathbb{E}[\lambda_{\ell\ell}(\mathbf{W_{it}})\,\tau_\ell(\mathbf{W_{it}})]}_{\text{own}} + \underbrace{\sum_{k\neq\ell}\mathbb{E}[\lambda_{\ell k}(\mathbf{W_{it}})\,\tau_k(\mathbf{W_{it}})]}_{\text{contamination}} \tag{10}$$

The first term on the right hand side of Equation (10) is a weighted average of conditional treatment effects $\tau_\ell(\mathbf{W_{it}})$ at own event time ℓ, with weights $\lambda_{\ell\ell}$ that average to 1, $\mathbb{E}[\lambda_{\ell\ell}(\mathbf{W_{it}})] = 1$. The second term is a weighted average of conditional treatment effects at *other* event times k, with weights averaging to 0, $\lambda_{\ell k}$, and $\mathbb{E}[\lambda_{\ell k}(\mathbf{W_{it}})] = 0$. This second term is the "contamination bias", as it contaminates the estimate of TE at event time ℓ.[10] This bias is likely to be present in the estimation of child penalty estimations, as I discuss in the next part.

2.3.2 Relevance of contamination bias in child penalty estimations

When is contamination bias relevant? In this part I discuss under what conditions contamination bias appears in the context of the child penalty. I also provide descriptive evidence that it is likely relevant.

The contamination term in Equation (10) can be written as follows:

[10] Another type of bias arises if $\lambda_{\ell\ell}$ is negative for some ℓ. In this paper, I focus on contamination bias. As the authors note, both biases stem from a misspecification of the implicit linear probability model $\mathbb{E}\left[X^{et=\ell}_{it} \mid \mathbf{X^{et\neq\ell}_{it}},\ \mathbf{W_{it}}\right]$: a given treatment dummy (event time dummy) is not mean independent of the controls *and* other treatment indicators, due to the mutually exclusive nature of the treatments.

$$\begin{aligned}\mathbb{E}[\lambda_{\ell k}(\mathbf{W}_{it})\,\tau_k(\mathbf{W}_{it})] \\ &= \operatorname{cov}[\lambda_{\ell k}(\mathbf{W}_{it}),\,\tau_k(\mathbf{W}_{it})] + \underbrace{\mathbb{E}[\lambda_{\ell k}(\mathbf{W_{it}})]}_{=0}\mathbb{E}[\tau_k(\mathbf{W_{it}})] \\ &= \operatorname{corr}[\lambda_{\ell k}(\mathbf{W}_{it}),\,\tau_k(\mathbf{W}_{it})]\,\operatorname{std}[\lambda_{\ell k}(\mathbf{W}_{it})]\,\operatorname{std}[\tau_k(\mathbf{W}_{it})]\,.\end{aligned} \tag{11}$$

Contamination bias is thus small if either i) TE heterogeneity, that is, $\operatorname{std}[\tau_k(\mathbf{W}_{it})]$, is small, ii) there is little variability in the contamination weights, that is, small $\operatorname{std}[\lambda_{\ell k}(\mathbf{W}_{it})]$, or iii) there is little correlation between the weights and conditional TE (small $\operatorname{corr}[\lambda_{\ell k}(\mathbf{W}_{it}),\,\tau_k(\mathbf{W}_{it})]$) (Goldsmith-Pinkham et al., 2024).

Small TE heterogeneity. Contamination bias vanishes if there is no variability in age-year specific treatment effects. This is unlikely to be true. Individuals who become parents at different ages differ in ways that likely predict their labor market trajectories before and after having children. Figure 1 plots annual earnings in the year before birth by age at childbirth, separately for mothers and fathers. Pre-birth earnings increase steeply between 20 and 30 for both mothers and fathers, before flattening. This pattern reflects differences in education (see Figure D.8 in the appendix). Similar evidence on differences between younger and older parents have been documented by Melentyeva and Riedel (2023) for Germany, Adams et al. (2024) for Denmark, and Leventer (2025) for Israel. This suggests that treatment effects in earnings, especially when measured in levels, are likely heterogeneous by age. In line with this, Melentyeva and Riedel (2023) find heterogenous earning losses by age at childbirth for mothers in Germany, and Adams et al. (2024) document heterogeneity by education level-which is highly correlated with age at birth-in Denmark. Beyond age, TE heterogeneity can be driven by macroeconomic events.

Small variability in contamination weights. Contamination weights are identified:

$$\lambda_{\ell k}(\mathbf{W}_{it}) = \frac{\mathbb{E}\left[\tilde{\tilde{X}}_{it}^{et=\ell}\,X_{it}^{et=k} \mid \mathbf{W}_{it}\right]}{\mathbb{E}\left[\tilde{\tilde{X}}_{it}^{2,et=\ell}\right]} = -\frac{p_k(\mathbf{W}_{it})\,E^*\left[X_{it}^{et=\ell} \mid X_{it}^{et=k}=1,\,\mathbf{W}_{it}\right]}{\mathbb{E}\left[\tilde{\tilde{X}}_{it}^{2,et=\ell}\right]}, \tag{12}$$

where $\tilde{\tilde{X}}_{it}^{et=\ell}$ are the residuals from regressing $X_{it}^{et=\ell}$ on $\mathbf{X}_{it,}^{et\neq\ell}$ and $\mathbf{W_{it}}$ and $E^*[X_{it}^{et=\ell} \mid \mathbf{X}_{it,}^{et\neq\ell}, \mathbf{W_{it}}]$ is the projection of $X_{it,}^{et=\ell}$ onto the space of linear combinations of the other event-time dummies and of the age and year fixed effects (Goldsmith-Pinkham et al., 2024).

Contamination weights do not vary with $\mathbf{W}_{it}$ if treatment assignment does not depend on $\mathbf{W}_{it}$, that is, if the propensity score $p_k(\mathbf{W}_{it}) = \mathbb{E}[X_{it}^{et=k} \mid \mathbf{W}_{it} = \mathbf{w}]$ is constant. In this case $\lambda_{\ell k}(\mathbf{W}_{it})$

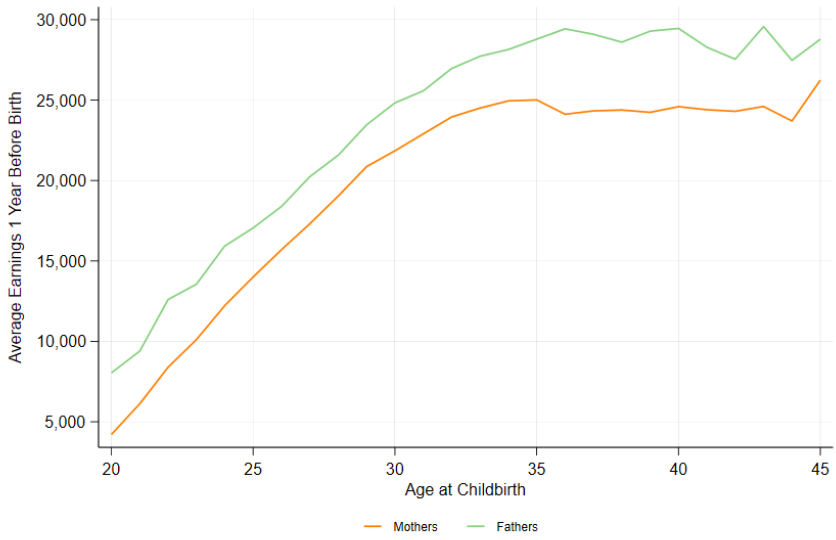


FIGURE 1. Earnings in ET=-1, by age at birth

Notes: This figure shows the average earnings (in 2010 euros) in the year before birth by age at childbirth, separate for women and men. Earnings are set to zero for individuals who do not work.

is a constant function of $\mathbf{W}_{it}$, and, since $\mathbb{E}[\lambda_{\ell k}(\mathbf{W}_{it})] = 0$, all contamination weights are zero and there is no contamination bias.

For child penalty estimations, contamination weights would therefore be constant—and the bias would vanish—if, across each age–year cell, individuals were equally likely to be at any event time. This is, of course, not the case. Figure 2 shows the distribution of age at birth for men and women. Most individuals have their first child between their late twenties and early thirties, they are hence much more likely to be at event time zero at these ages.

Small correlation between contamination weights and year-age specific TE. Finally, bias vanishes if contamination weights are uncorrelated with conditional treatment effects. This is the case when the factors influencing TE heterogeneity are unrelated to the factors predicting event time. However, as I argue above, age is related to both treatment timing and variability in conditional treatment effects; therefore, $\text{corr}[\lambda_{\ell k}(\mathbf{W}_i),\ \tau_k(\mathbf{W}_i)]$ is likely to be non-zero.

Taken together, event study coefficients from Equation (1) are likely biased due to heterogeneity in TE based on age, which also determines treatment timing.

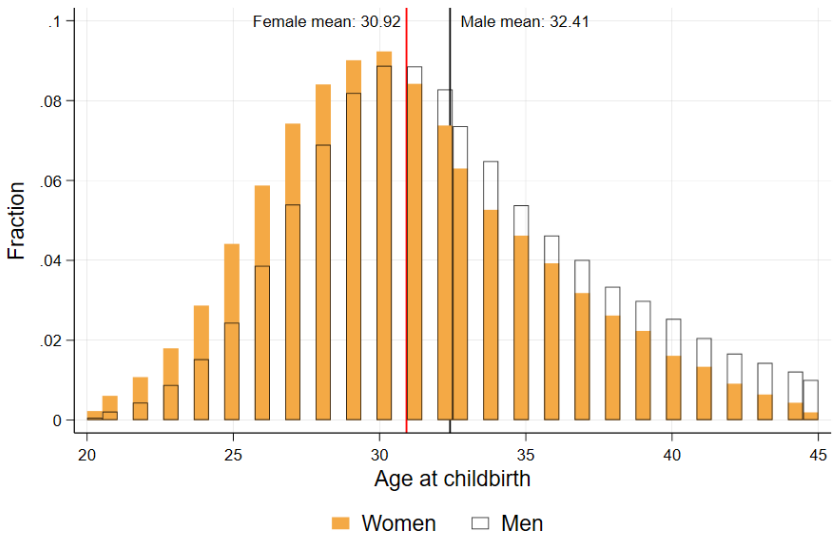


FIGURE 2. Distribution, age at birth

Notes: This figure shows the distribution of age at first childbirth for mothers and fathers in the estimation sample. Bars report the fraction of parents by age at birth. Vertical lines indicate mean age at childbirth for women and men, respectively.

2.3.3 Numerical example

To see how this bias materializes, consider an example with two birth cohorts, 2012 and 2014, observed from 2011 to 2015. The setup is illustrated in Figure 3. Each individual is randomly assigned a starting age between 25 and 31 in 2011, after which age evolves deterministically over time; age at birth is therefore the individual's age in their cohort's childbirth year. Potential outcomes without treatment are set to 20 plus some idiosyncratic noise, and treatment effects are then added according to the event time. Note that in this set up with constant $Y(0$, the before-after comparison identifies the causal impact of children.

Treatment effects are the same across cohorts: -10 in the year of childbirth, and -5 the years after. In the no-heterogeneity baseline, there is no heterogeneity by age or year. In the heterogeneity scenario, I introduce TE heterogeneity by age at $ET = 1$. One year after childbirth, only younger parents, defined by a below-median age at birth, experience earnings losses of -5. Older parents face zero losses. I then run specification (1) for both scenarios. Table 1 shows the results. In the absence of heterogeneity, the coefficients are accurate (column 1). However, they are off when TE are heterogeneous (column 2). For instance, we know that the true effect at $ET = 3$ is still -5. However, $\hat{\alpha}_3$ estimated via Equation (1) is -6.56, around 32% smaller than the true effect.

In this example, differences in treatment effects between younger and older parents are large. To

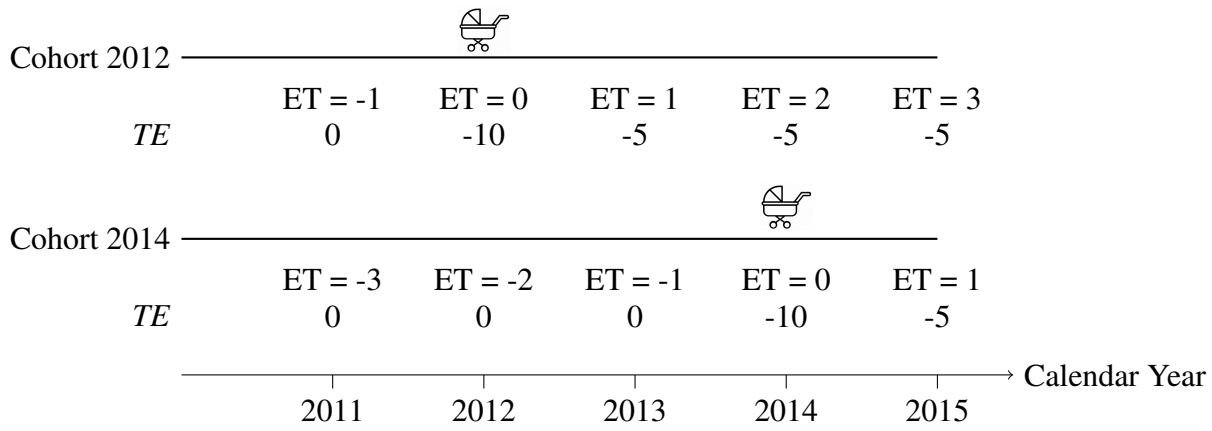


FIGURE 3. Illustration with two cohorts

show how contamination bias varies with varying levels of TE heterogeneity, I define $TE^{age}, age \in \{old, young\}$ as the TE for old and young parents, respectively, with $TE^{old} = TE^{young} + X$. TE for older parents one year after childbirth are smaller (in absolute terms) by X. I run Specification (1) for $X \in \{0, 1, 2, 3, 4, 5\}$ to estimate the parameters α_3 for these different scenarios, and I use Goldsmith-Pinkham et al. (2024)'s decomposition from Equation (10) to decompose these coefficients into both the *own* and *contamination* term.[11] I plot them together in Panel (a) of Figure 4. When $X = 0$, that is, with no treatment effect heterogeneity, $\hat{\alpha_3}$ (orange line) is close to the true effect, -5, and equals the first decomposition term in Equation (10), the weighted average of TE at $ET = 3$. Accordingly, the contamination bias term (red dashed line), is zero. The larger X, the higher the variability in treatment effect heterogeneity (gray dashed line), and thus the larger the contamination bias. In Panel (b), I introduce further TE heterogeneity at $ET = 3$. Now, the true effect is the average of $TE^{young} = -5$ and $TE^{old} = -5 + X$, since half of the sample are young, and half of the sample are old parents. This is reflected in the first term of the decomposition (green line). The coefficient continues to be biased downward — the contamination bias is non-zero — due to TE heterogeneity at other event times (here, at $ET = 1$).

I conclude this section with a final remark on sample composition. Whereas I use a panel that is balanced in calendar time, many papers, including Kleven et al. (2019), use instead a sample that in balanced in event time, that is, where individuals are observed over a fixed window around childbirth. The potential for contamination bias remains in this setup (see Section A in the appendix).

[11] This is possible because I know the true conditional treatment effects and weights are identified (see Goldsmith-Pinkham et al., 2024).

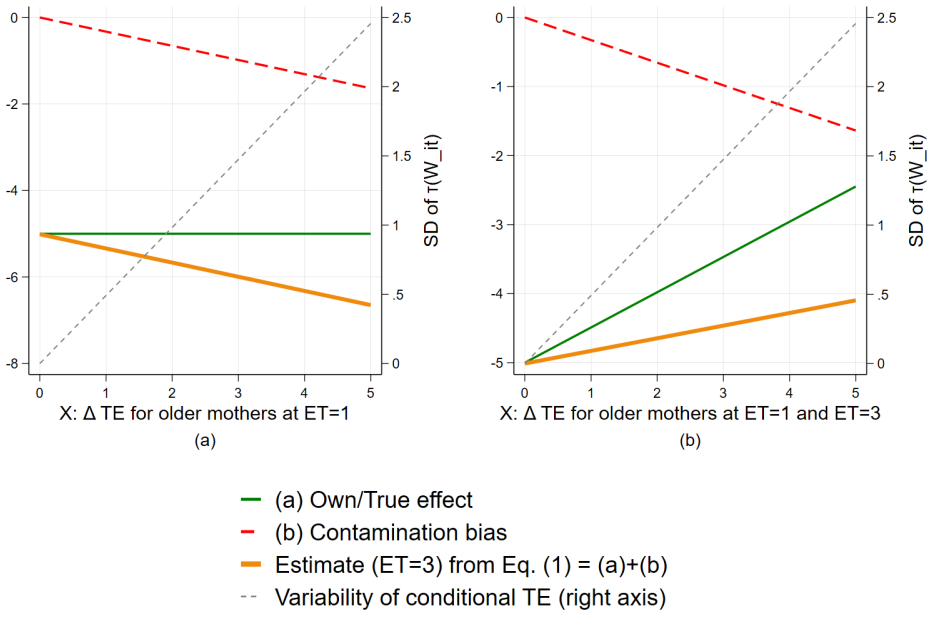


FIGURE 4. Contamination bias with varying TE heterogeneity

Notes: The figure illustrates the decomposition of $\hat{\alpha}_3$ from Specification (1) into (a) the own/true component and (b) the contamination component due to treatment effect heterogeneity at other event times, as implied by Equation (10). Panel (a) varies heterogeneity by increasing the gap X between treatment effects for younger and older parents at $ET = 1$. Panel (b) additionally varies treatment effects at $ET = 3$. The grey dashed line plots the variability of conditional treatment effects $\tau(W_{it})$ (right axis). As X increases, heterogeneity grows and so does contamination bias.

	Scenario	
	Homogenous TE	Heterogeneous TE
ET=-3	-0.021	-0.020
	(0.020)	(0.027)
ET=-2	-0.020	-0.662
	(0.030)	(0.040)
ET=0	-10.049	-10.675
	(0.022)	(0.029)
ET=1	-5.000	-3.306
	(0.020)	(0.027)
ET=2	-5.047	-5.693
	(0.022)	(0.029)
ET=3	-5.012	-6.658
	(0.028)	(0.038)
N	50,000	50,000

TABLE 1. Estimates using Specification (1)

Notes: This table reports estimates from Specification (1) in a simulated panel with two birth cohorts (2012 and 2014) observed from 2011–2015. Individuals are assigned baseline ages and age deterministically; untreated potential outcomes are constant over time. True treatment effects are -10 in the year of birth and -5 thereafter. Column (1) presents estimates under homogeneous treatment effects. Column (2) imposes heterogeneity at $ET = 1$, where only younger parents (below-median age at birth) experience a -5 loss. Standard errors in parentheses.

3. UNBIASED ESTIMATION UNDER CONDITIONAL INDEPENDENCE

The insights in the previous section have two important implications for child penalty estimations. First, the gender-specific event study coefficients from the standard approach are prone to contamination bias when treatment effects vary across age–year cells. Second, coefficients on pre-treatment event times, since they, too, are prone to contamination bias, do not provide a valid test of the identification assumption. This section presents an imputation estimator that is unbiased under CIA and NA, discusses tests for the plausibility of the identification assumptions separately, and compares the resulting estimates with those obtained from Equation (1) for a sample of French parents.

3.1. Imputation estimator

As shown in Section 2.2, under NA and CIA untreated potential outcomes are identified from untreated observations. An intuitive and easy-to-implement way to estimate treatment effects in this identification framework is an imputation estimator as in Borusyak et al. (2024). The procedure has three steps:

1. Estimate age and year fixed effects in the sample of untreated observations only:

$$Y_{it} = \kappa + \beta_a \mathbf{D}_{it}^{age} + \gamma_y \mathbf{D}_{it}^{year} + \epsilon_{it}. \tag{13}$$

2. For treated observations, predict untreated potential outcomes using the estimated fixed effects:

$$\hat{Y}_{it}(0) = \hat{\kappa} + \hat{\beta}_a D_{it}^{age} + \hat{\gamma}_y D_{it}^{year},$$

and compute individual treatment effects

$$\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0).$$

3. Estimate the target parameter $ATT(\ell)$ as a weighted sum $\tau_\omega = \sum_{it\in\Omega} \omega_{it}\hat{\tau}_{it}$ for the relevant target population (Borusyak et al., 2024).[12]

Under CIA and NA, this estimator consistently recovers $ATT(\ell)$ for each event time ℓ, as well as subgroup-specific $ATTs$. Unlike the standard event study regression, this estimator does not jointly regress earnings on mutually exclusive event time indicators and therefore avoids contamination bias. Moreover, counterfactual earnings are obtained solely from untreated observations, in line with the identification assumptions.

The estimator is straightforward to implement following the three steps above or using Borusyak et al. (2024)'s *did_imputation* command, with built-in tools for inference.

Overlap Imputation in Step 2 is feasible only for the age–year cells in which the age and year fixed effects can be estimated in Step 1. In particular, for each treated observation (i, t) with event time $\ell = t - G_i$, there must exist untreated observations in the same age–year bin. Otherwise the counterfactual $Y_{it}(0)$ cannot be imputed. For example, when restricting the sample to parents only, all units are treated by the final observation year T, so treatment effects for the last cohort $G_i = T$ are not identified. Likewise, restricting the sample to individuals who have their first child by age A implies that no untreated observations exist at age A, and treatment effects for this age-at-birth group cannot be identified. More generally, at a given event time ℓ, treatment effects are identified only for individuals belonging to cohorts $g \leq T - 1 - \ell$ and whose first birth occurs no later than age $c \leq A - 1 - \ell$. As a result, $ATT(\ell)$ averages treatment effects over the subset of treated individuals for whom the counterfactual is identified at event time ℓ, and this subset varies with

[12] Borusyak et al. (2024)'s main application is a TWFE, that is, $Y(0)$ is modeled as the sum of unit and time effects, but their insights explicitly also apply to cases where the potential outcome model includes other fixed effects or continuous covariates.

ℓ.[13]

These support restrictions are important for interpretation: overlap determines the set of treated observations for which treatment effects are identified, and therefore the implicit population of interest underlying the $ATT(\ell)$ estimates.

Flexibility The potential outcome model in Equation (13) can accommodate additional controls without introducing bias, including continuous covariates or additional fixed effects. For example, Adams et al. (2024) argue that time-since-graduation and education should be included, as both predict earnings and fertility timing. While adding these controls introduces further sources of bias in the standard approach, they can be included safely in (13).

The imputation framework also allows the control group to be restricted. For instance, following Melentyeva and Riedel (2023), one may limit the sample of untreated observations used to estimate counterfactuals to those who give birth not too much later in life. One could also estimate Equation (13) on a sample of non- or not-yet-treated parents that are matched to parents based on key demographic characteristics (similar to the matching strategy underlying the pseudo-panel approach in Kleven, 2022).

The imputation approach clearly separates the estimation procedure discussed above from the test of the underlying identification assumptions. I will discuss such tests in the next section.

3.2. Test of identification assumption

Borusyak et al. (2024) suggest an OLS-based testing procedure using only untreated observations that I can adapt to my case.

Identification in the standard child penalty estimation relies on the assumption that timing is as good as random for untreated earnings once we control for age and year. The plausibility of this assumption can be tested by regressing, for not-yet-treated observations, the outcome on non-parametric controls for age and year *and* leading event time dummies:

$$Y_{it} = \alpha_\ell \cdot \mathbf{D}^{\text{Event}}_{\mathbf{it},\ell<\mathbf{0}} + \beta_{\mathbf{a}} \cdot \mathbf{D}^{\text{Age}}_{\mathbf{it}} + \gamma_{\mathbf{y}} \cdot \mathbf{D}^{\text{Year}}_{\mathbf{it}} + \nu_{it}, \tag{14}$$

Under CIA, that is, if the timing of childbirth is mean independent of untreated potential outcomes conditional on age and year, the coefficients on leading event-time indicators, α^g_ℓ, are zero, which

[13] This is unless one restricts the horizon over which to estimate treatment effects and fixes the population of interest to those for whom individual tau_{it} are identified at all these event times.

can be tested using a Wald test (for more details, see Section 4.4 of Borusyak et al., 2024). When implementing this test below, I will refer to is as “Test 1".

An alternative test relies on a placebo strategy instead (e.g. De Chaisemartin and d'Haultfoeuille, 2020). It involves pretending that treatment happened at some event time $et = -k, k > 0$, and estimating treatment effects for $et \in (-k, .., -1)$ via imputation using as control group untreated observations at $ET < -k$. The null hypothesis here is that these placebo treatment effects are zero. I will refer to this test as “Test 2".

Crucially, both tests rely exclusively on untreated observations and therefore do not suffer from contamination bias induced by treatment effect heterogeneity.

3.3. Data and sample

I use data from the *Echantillon démographique permanent (EDP)*, a longitudinal version of the French census covering a representative sample of approximately 4% of the population. The EDP includes birth, death, and marriage records linked to sampled individuals, as well as their socio-economic characteristics since 1968. It also contains partial information on education. I focus on the 2010–2020 period, for which I can link socio-fiscal data at the individual and household level. This enables me to precisely measure income.

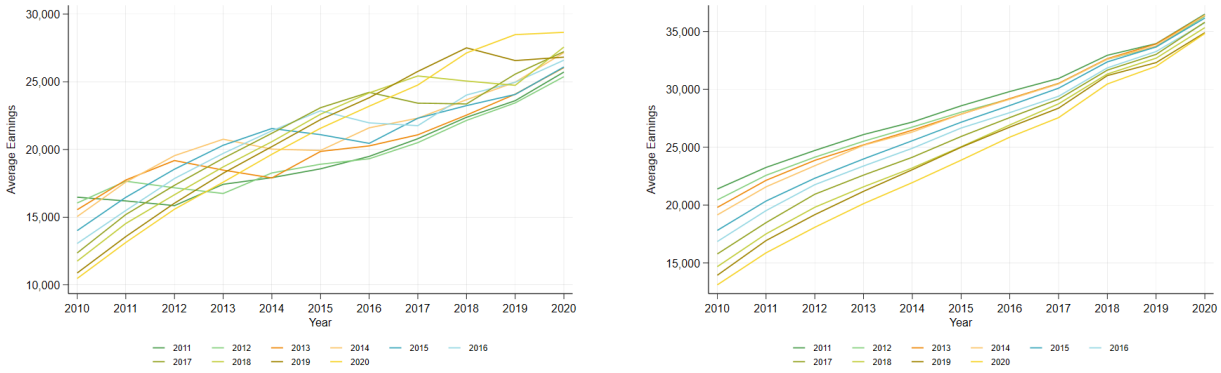
I restrict the sample to a balanced panel of individuals for whom detailed earnings data is available for all eleven years. I further limit the sample to individuals who were aged 18–55 during the observation window and had their first child between ages 20 and 45. I keep individuals who have their first child between 2011 and 2020.[14] The final sample consists of 131,393 mothers and fathers from these ten cohorts, defined by the year they had their first child. Descriptive statistics are shown in Table C.1 in the appendix.

My main outcome variable is gross annual earnings, which include earnings-replacement parental leave[15] benefits but exclude pensions and unemployment benefits. Mothers are entitled to 16 weeks of maternity leave (six weeks before and ten after birth) for the first and second child, and 26 weeks (eight before and 18 after) from the third child onward. Paternity leave was introduced in 2002, initially granting 14 days, and was extended in 2021 to 21 days, of which seven are mandatory (for more details on the institutional background, see Section B.1 in the appendix). All monetary values are deflated to 2010 euros. Following Kleven et al. (2019), individuals who do not work are assigned zero earnings. Figure 5 shows average earnings by cohort, for women and men,

[14] As further discussed below, the 2010 cohort is always-treated and cannot be used in the estimation of child penalties in an event study framework, as pre-birth earnings are not observed.

[15] Including these benefits avoids mechanically overstating short-run motherhood costs; see Adams et al. (2024) for a detailed discussion on the treatment of parental leave benefits in child penalty estimations.

respectively. Later cohorts enter the sample at a younger age in 2010, which explains the earning differences in that year. Figure 6 pools these by cohort and event time. From these descriptive statistics, the presence of a child penalty is already visible. For women, earnings fall in the year of the birth and the year after, and the trend appears flatter at later event times. In contrast, for men, not much is visible in terms of the level or trend of earnings in the year they become fathers.



(a) Women **(b)** Men

FIGURE 5. Raw earnings, by cohort

Notes: This figures shows raw earnings separately for women and men and for different cohorts of individuals. A cohort is defined by the calendar year in which individuals have their first child. My sample includes individuals aged 18 to 55 who have their child between the age 20 and 45 between calendar years 2011 and 2020. Earnings include zeros for individuals who do not work and are deflated to 2010 euros.

3.4. Results

Mothers Figure 7 shows event study estimates for mothers obtained by Specification (1) (black solid line) and the imputation approach (orange solid line), respectively. The dashed lines show the result from the two tests described in Section 3.2. Estimates for $ATT(\ell)$ from the imputation estimator come from the subset individuals for whom treatment effects are identified (see the discussion on overlap in the previous section).[16]

Independent of the estimator, we recognize the familiar pattern of the child penalty literature: earnings of mothers drop suddenly and significantly in the year of childbirth, compared to the counterfactual. The year after the first of their first child, my estimator shows mothers earn €4,187 less compared to their counterfactual, corresponding to 18.6% of counterfactual earnings (see Figure D.4 in the appendix). These short-term estimates do not differ from the standard approach: At $ET = 1$, the OLS estimate for α_ℓ from Equation (1) is €-4,313, corresponding to 18.8%. From

[16]Figure D.3 in the appendix shows, for each cohort g, the range of event times for which $ATT(g, g+\ell)$ is observed and indicates which $ATT(\ell)$ each cohort contributes to. Table C.2 shows descriptive statistic by cohort.

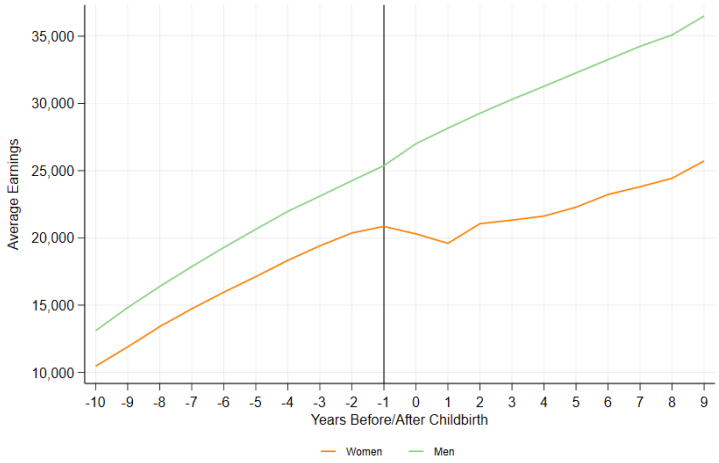


FIGURE 6. Average earnings by event time

Notes: This figure shows earnings pooled across cohorts and by event time, separately for mothers and fathers. A cohort is defined by the calendar year in which individuals have their first child. Event time zero corresponds to the year of the first child. My sample includes individuals aged 18 to 55 who have their child between the age 20 and 45 between calendar years 2011 and 2020. Earnings include zeros for individuals who do not work and are deflated to 2010 euros.

two years after childbirth onward, the estimates diverge. The imputation estimator yields consistently smaller estimates for the earnings losses related to children. Eight years after childbirth, the estimate is € -5,416, 27.4% smaller than the estimate of € -7,460 in the standard approach. Scaled by counterfactual earnings, the estimate using the imputation approach correspond to a loss of 20% of counterfactual earnings eight years after childbirth, whereas the scaled estimate from the standard approach is 25%, a 5 percentage point (pp) difference Figure D.4.

Heterogeneity by age at birth The difference between the two estimates is consistent with heterogeneity by age or year, which induces bias in the standard event–study specification. Figure 8 shows estimates by mothers' age at birth, obtained from aggregating individual τ_{it} by age at birth. The top row plots absolute and scaled estimates for selected ages at birth and shows substantial variation by birth timing. Children-related losses are largest for young mothers: women giving birth in their early twenties experience nearly € 15,000 lower earnings eight years after childbirth, around 60% relative to their counterfactual. Losses are smallest for women who have their first child in their early thirties; mothers giving birth at 30 earn about € 1,600 less eight years later, or 5.8%.

The bottom row (Panel (c) and (d)) illustrates this steep gradient more clearly by showing the average earnings loss from $ET = 0$ to $ET = 6$ by age at birth. In absolute terms, annual losses range from € 7,000 to € 8,000 for women who become mothers between ages 20 and 25—up to

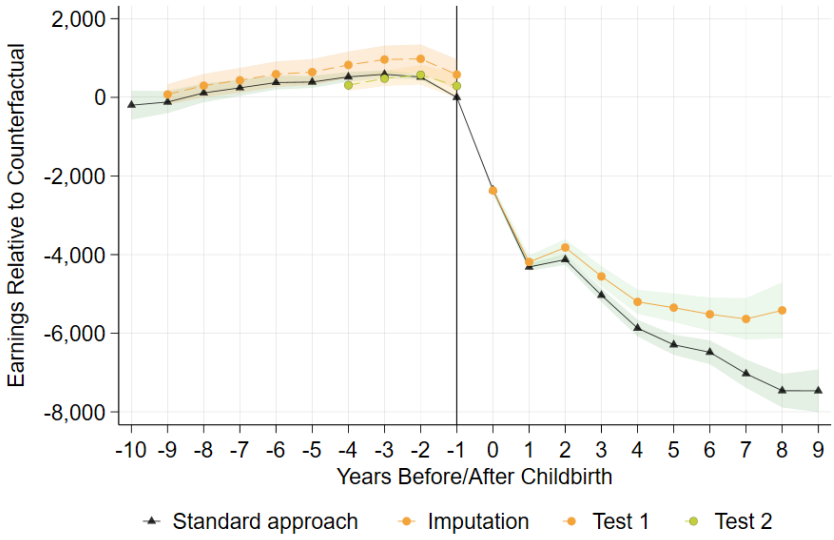


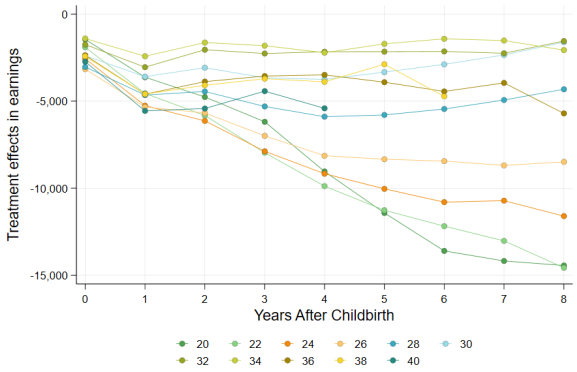
FIGURE 7. Standard approach vs. imputation estimator, women

Notes: The figure plots estimated event–study effects on mothers' earnings relative to their counterfactual using the standard specification in Equation (1) (black line) and the imputation estimator (orange line). Dashed lines show the two specification tests described in Section 3.2. All estimates are relative to predicted counterfactual earnings. The sample includes individuals aged 18–55 who have their first child between ages 20 and 45 in calendar years 2011–2020. Earnings include zeros for non-employed individuals and are deflated to 2010 euros.

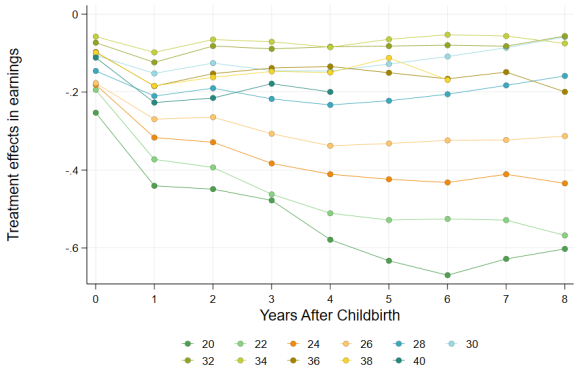
50% of counterfactual earnings for the youngest mothers (Panel (d)). In contrast, average losses fall to about €2,000 for those who give birth in their early thirties, corresponding to 5–10%.

The cost of motherhood is thus considerably larger for young mothers in both absolute and relative terms. This result partly aligns with Melentyeva and Riedel (2023) who find larger relative (but smaller absolute) losses for young mothers in Germany.

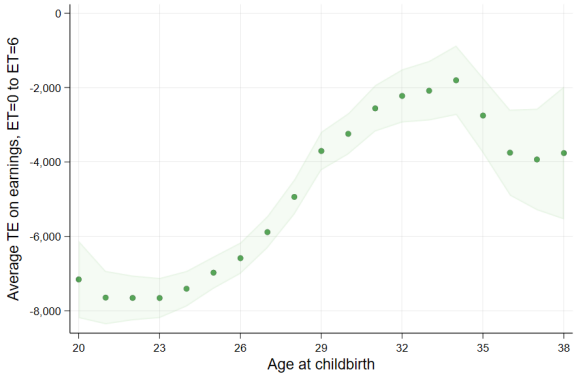
Restricted control group In the baseline specification, I use all untreated observations to estimate parents' counterfactual earnings. Instead, akin to the approach Melentyeva and Riedel (2023) take in their stacked DiD estimator, I now instead restrict the sample on which I estimate counterfactual outcomes to not-yet-mothers who have their child not too much later, compared to the treatment group, as these might be more comparable. For each age at birth c, I keep only observations who become parents by age $c + 5$, estimate the model in Equation (13) on untreated observations within that subsample only, and impute counterfactuals for groups c. This limits the horizon for which treatment effects can be estimated to four years after childbirth. In Figure 9, I plot the average estimated treatment effect between $ET = 0$ and $ET = 4$ by age at childbirth when using the full control group (in green) and the restricted control group (in orange). Estimates are not distinguishable across these approaches in my sample.



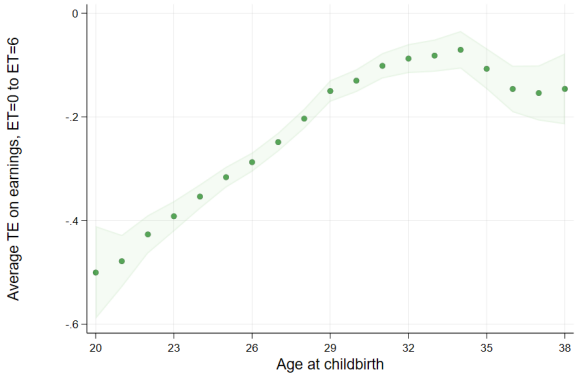
(a) TE estimates, by age at birth



(b) Scaled TE estimates, by age at birth



(c) Average TE, $ET = 0$ to $ET = 6$



(d) Average TE, scaled, $ET = 0$ to $ET = 6$

FIGURE 8. Child penalties by mothers' age at first birth

Notes: The figure shows estimated child penalties by mothers' age at first birth. The top row plots event study treatment effects on annual earnings at selected ages at birth, in levels (Panel (a)) and relative to counterfactual earnings (Panel (b)). The bottom row reports the average treatment effect on earnings from event time $ET = 0$ to $ET = 6$, in levels (Panel (c)) and relative terms (Panel (d)). The sample includes individuals aged 18 to 55 who have their first child between ages 20 and 45 in calendar years 2011–2020. Earnings include zeros for non-employed individuals and are deflated to 2010 euros.

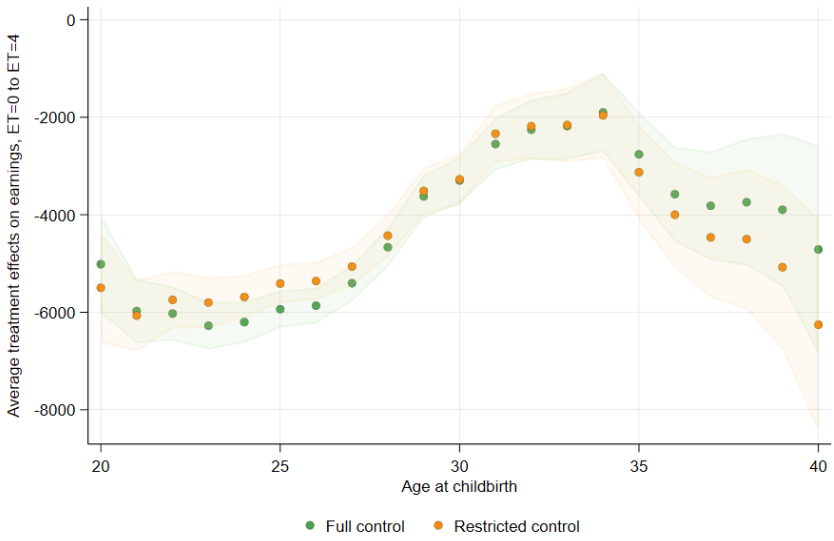


FIGURE 9. Average treatment effect, ET=0 to ET=4, by mothers' age at birth

Notes: The figure shows the average estimated treatment effect on mothers' earnings from event time $ET = 0$ to $ET = 4$ by age at childbirth. The green line uses the full set of untreated observations to estimate counterfactual earnings, while the orange line restricts the control group to not-yet-mothers who give birth within five years of the treated group's age at birth. For each age-at-birth group c, counterfactual outcomes are estimated from untreated observations only, following Equation (13). The sample includes individuals aged 18–55 who have their first child between ages 20 and 45 in calendar years 2011–2020. Earnings include zeros for non-employed individuals and are deflated to 2010 euros.

A final comment relates to estimates of the standard approach at $ET = 9$. I restrict the sample to individuals who have their child between 2011 and 2020 and are observed between 2010 and 2020. Accordingly, everyone is treated by year 2020, which corresponds to $ET = 9$ for the earliest cohort, $g = 2011$. As there are no untreated observations to estimate the fixed effect for that year to impute counterfactual earnings, treatment effects are not identified. Despite, the standard approach produces an estimate by extrapolation, an issue also pointed out by Borusyak et al. (2024).

Fathers Results for fathers are shown in Figure 10. Estimates from the standard approach suggest a small medium-term impact of children on fathers' earnings. However, the figure displays sizeable pre-trends, raising concerns about the plausibility of the identifying assumptions. Similar pre-trends for men, at least in the five years before birth, have been documented in other settings, including Kleven et al. (2019) and Cortés and Pan (2023).

Consistent with this finding, both diagnostic tests described in Section 3.2 reject the null of no relationship between (negative) event time and earnings (dashed lines). The positive coefficients on pre-treatment dummies indicate positive selection into fatherhood: men who are about to become fathers earn more than comparable men who remain childless for longer, conditional on age and year. Using men who are further away from childbirth as controls therefore risks understating counterfactual earnings and, as a result, understating the cost, or overstating the implied "fatherhood premium." In light of these test results, the imputation estimator's positive post-birth estimates for fathers cannot be interpreted as causal.

The rejection of CIA for fathers suggests that, beyond age and calendar year, there are additional determinants of earnings that correlate with fertility timing. Even for mothers, despite smaller estimated effects, the same tests reject the null of no relationship between event time and earnings when conditioning on age and year. This pattern is consistent with omitted factors—such as education or unobserved factors—that jointly influence earnings and the timing of parenthood (Adams et al., 2024). In this setting, conditional independence may therefore be an overly strong assumption. Motivated by this evidence, the next section considers an alternative identification strategy commonly used in the child penalty literature, which relies instead on parallel trends, controlling for individual unobserved heterogeneity.

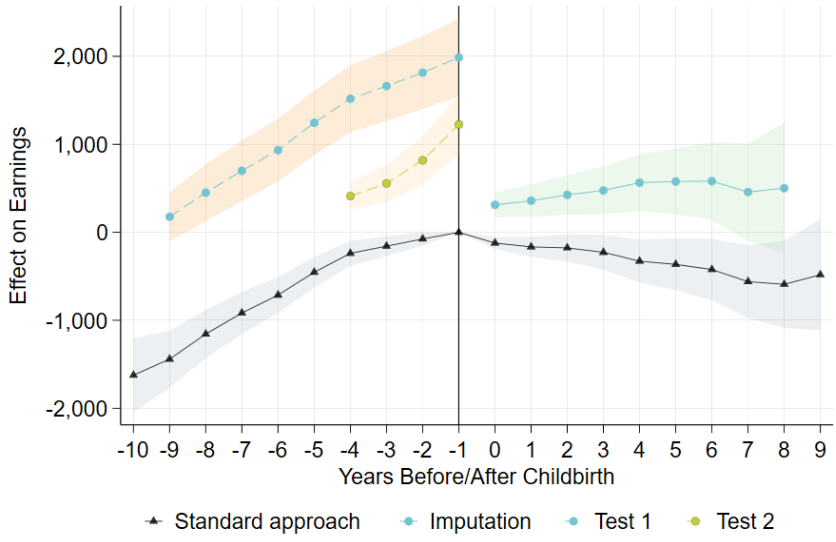


FIGURE 10. Standard approach vs. imputation estimator, men

Notes: The figure compares event–study estimates of fathers' earnings around childbirth using the standard specification in Equation (1) (black line) and the imputation estimator (blue line). Dashed lines correspond to the two diagnostic tests described in Section 3.2. Estimates are expressed relative to predicted counterfactual earnings. The sample includes men aged 18–55 who have their first child between ages 20 and 45 in calendar years 2011–2020. Earnings include zeros for non-employed individuals and are deflated to 2010 euros.

4. ALTERNATIVE IDENTIFICATION UNDER PARALLEL TRENDS

A set of studies deviate from the approach outlined in Section 2 and instead employ a DiD framework to estimate gender-specific effects of children. By absorbing individual unobserved heterogeneity that is constant over time, this approach may address some concerns with respect to the endogeneity of fertility timing. Yet causal identification still hinges on the choice of a credible control group and assumptions on how (knowledge of) $Y(0)$ affects fertility timing. When it comes to estimation, building on the DiD literature, Melentyeva and Riedel (2023) highlight the potential for bias when adding individual fixed effects in an OLS regression as Equation (1).

To address both identification and estimation challenges, this section adopts a forward-engineering approach, as proposed by Baker et al. (2025). Rather than starting with a regression specification such as Eq. (1) and inferring the underlying causal estimand post hoc, this framework begins with a clearly defined target parameter, specifies the assumptions required for identification, and then derives an appropriate estimator. This framework highlights key trade-offs in identifying child penalties, makes transparent the exact comparisons being made, and allows researchers to choose those they view as most credible.

4.1. Identification under parallel trends

I have previously defined the target parameter of interest, $ATT(\ell)$, as a weighted average of individual treatment effects at event time ℓ (see Section 2.2). In a DiD framework, it is useful to decompose $ATT(\ell)$ into its *cohort–time* building blocks, $ATT(g,t)$, defined as:

$$ATT(g,t) = \mathbb{E}[Y_{it}(g) - Y_{it}(0) \mid G_i = g], \quad \forall t \geq g.$$

Since $\ell = t - g$, the parameter of interest $ATT(\ell)$ can be obtained by aggregating cohort-specific effects $ATT(g, g+\ell)$ by event time, using weights proportional to the number of observations at each ℓ (Baker et al., 2025).

Rather than conditional independence, identification in DiD designs relies on the parallel trends (**PT**) assumption: absent treatment, outcome *trends* would have evolved similarly for the treated cohort and the comparison group. As is standard in the literature, I use the year before childbirth, $t = g - 1$, as the baseline. Formally, for treated cohort g and a comparison group defined by $C_i = 1$, the PT assumption is:

$$\mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i = g] = \mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid C_i = 1]. \tag{15}$$

I further maintain the no-anticipation assumption (Equation (6)).

Under **NA** and **PT**, counterfactual outcomes $Y_{i,t}(0)$ are identified, and so is $ATT(g,t)$, based on differences in outcome changes between the treated cohort and the control group, indicated by dummy C, between $t = g - 1$ and $t = g + \ell$:

$$ATT(g,t) = \mathbb{E}[Y_{it} - Y_{i,g-1} \mid G_i = g] - \mathbb{E}[Y_{it} - Y_{i,g-1} \mid C_i = 1]. \tag{16}$$

The key step that remains is the choice of the control group. Several groups could, in principle, serve this role, and the plausibility of the parallel trends assumption—and thus identification—depends directly on this choice.

4.2. The choice of the control group in staggered designs: Non-treated vs. not-yet-treated

The plausibility of the parallel trends (PT) assumption depends on context and, crucially, on the choice of the control group. In this section, I discuss two control groups commonly used in the child penalty literature—non-parents and not-yet-parents—and assess the credibility of PT for each.

All parameters are defined, identified, and estimated separately by gender. Accordingly, when discussing control groups, comparisons should be understood as mothers versus non-mothers and fathers versus non-fathers, and analogously for not-yet-parents.

Non-parents. Using non-parents as a control group entails using the earnings trends of individuals observed without children as the counterfactual for each treated cohort g. The main advantage of this approach is that the control group remains fixed over time, avoiding compositional changes. However, individuals who never have children may differ systematically from parents in ways that predict untreated earnings trends, $\Delta Y_{it}(0)$. As a result, non-parents may be too dissimilar to serve as a credible control group.[17]

Not-yet-parents. The staggered timing of childbirth creates a second natural control group: individuals who will be treated in the future. Using not-yet-parents as controls means comparing earlier cohorts to later cohorts that have not yet had children. Because these individuals eventually select into parenthood, they may be more comparable to parents than never-parents. This approach, however, implies time-varying composition of the control group, as cohorts enter treatment and exit the comparison group. Not-yet-parents are implicitly used as controls in the standard approach to estimating child penalties, which restricts the sample to individuals who eventually become parents.

In the context of childbirth, I favor using not-yet-parents as the control group. Alternative approaches are possible. For instance, Coudin et al. (2025) use matching techniques to construct a control group among comparable non-parents, and Kleven et al. (2019) implement a similar procedure as a robustness check. I do not pursue these strategies here.

Using not-yet-parents as controls, where the comparison group consists of cohorts g' such that $g' > \max\{g, t\}$, the PT assumption becomes:

$$\mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i = g] = \mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i = g']. \tag{17}$$

The corresponding estimand in Equation (16) is:

$$\begin{aligned} ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1} \mid G_i = g] \\ - \mathbb{E}[Y_{i,t} - Y_{i,g-1} \mid G_i = g']. \end{aligned} \tag{18}$$

In post-treatment periods ($t > g$), the control group consists of all cohorts not yet treated by time

[17] Fertility is right-censored. "Non-parents" are therefore individuals who have not had children by time T (2020 in my case). Those observed childless at age 45—the upper fertility bound in my sample—can be considered never-treated, while younger individuals may still become parents in the future.

t, that is, $g' > t$. In pre-treatment periods, under **NA**, the estimand in Equation (18) compares *untreated* earnings trends between cohort g and cohorts not yet treated at $t = g - 1$ ($g' > g$). These pre-treatment estimates serve as a plausibility check for the PT assumption. While small and precise pre-trends lend credibility to the design, they cannot validate post-treatment PT, which is inherently untestable.

The credibility of PT using not-yet-parents ultimately depends on the comparability of earlier and later cohorts. If these groups differ in characteristics that predict untreated earnings trends, PT is likely to fail. In contrast to the standard approach, age is not explicitly accounted for in this setup so far. Since age strongly predicts earnings dynamics, the parallel trends assumption may fail if it varies systematically across cohorts, even absent other sources of selection. The next section examines this issue in detail.

4.3. Accounting for age and conditional PT

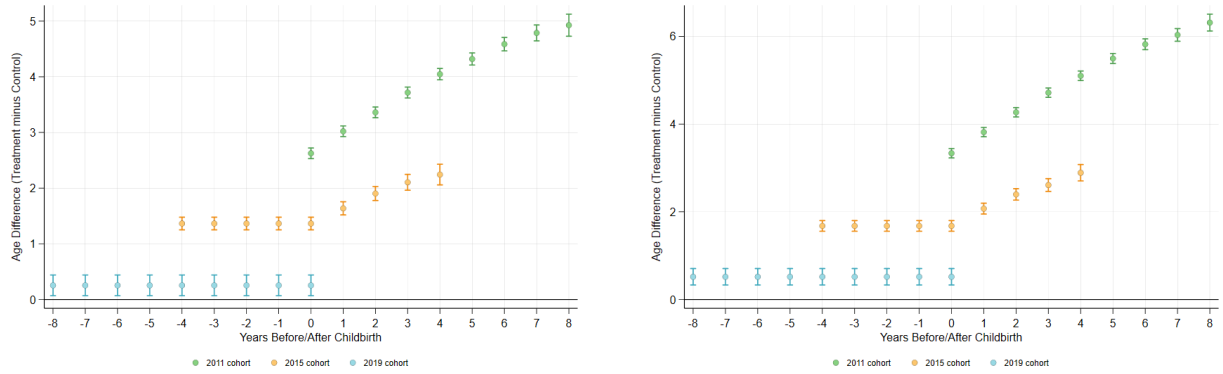
I begin by documenting systematic age imbalances between earlier and later cohorts when not-yet-parents are used as controls. I then show how these imbalances undermine unconditional parallel trends, motivating identification under a weaker, conditional PT assumption.While conditioning on age improves comparability in life stage, it necessarily implies comparisons to individuals who delay childbirth. I later contrast this approach with an alternative strategy that fixes age at childbirth instead, highlighting a key trade-off in identifying child penalties using event study designs.

Imbalance in age across cohorts Figure 11 shows imbalances in age for selected cohorts and different event times. At each event time, the comparison group consists of all cohorts not yet treated. For example, for the 2011 cohort, the control group at $et = 1$ includes all cohorts with $g' > 2012$, whereas at $et = 8$ it includes only the 2020 cohort. As a result, age imbalance between treated cohorts and the control group increases with event time. In my sample, women (men) in the 2020 cohort are on average around five (six) years younger than those in the earliest cohort ($g = 2011$). Whereas they only present a small part of the comparison group at $ET = 1$ (the 2020 is small), they are the only control cohort at $ET = 8$ where it is compared to cohort $g = 2011$.

These differences matter for earnings trends. Figure 12 plots earnings trajectories for the treated 2011 cohort alongside the disaggregated trends of each control cohort contributing to the counterfactual at different event times. Later cohorts exhibit steeper earnings growth, consistent with their younger age. This pattern suggests that untreated earnings for earlier-treated cohorts may be overstated, particularly at higher event times.

Taken together, these findings indicate that unconditional PT is likely violated when not-yet-parents are used as controls without accounting for age. In the next subsection, I formalize identi-

fication under a weaker PT assumption that conditions explicitly on age.



(a) Women **(b)** Men

FIGURE 11. Imbalance in age between treatment and comparison group, selected cohorts

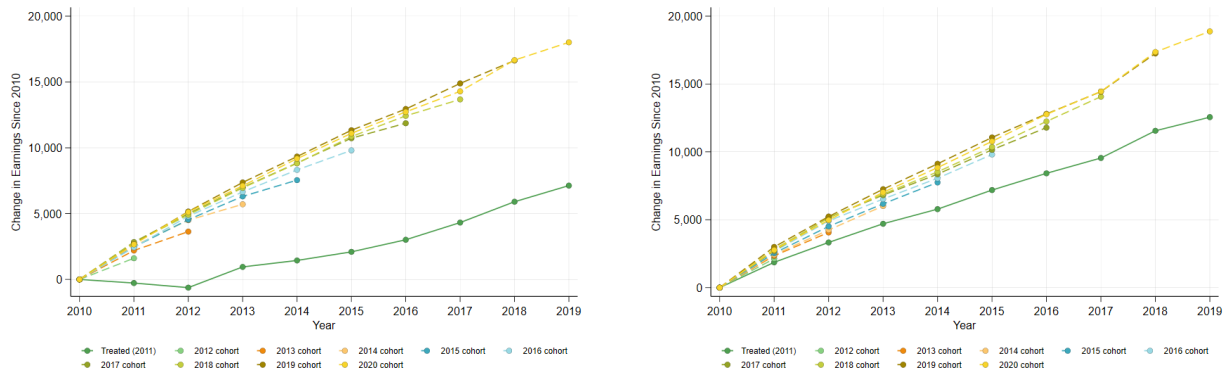
Notes: This figure shows the mean difference in age between selected treated cohorts and their comparison group at different event times, separate for mothers and fathers. A cohort g is defined by the calendar year in which individuals have their first child. Event time zero corresponds to the year of the first child. At each time, the comparison group contains all cohorts $g' > g$ and $g' > t$, that is, cohorts not yet treated at the time of comparison. My sample includes individuals aged 18 to 55 who have their child between the age 20 and 45 between calendar years 2011 and 2020. Vertical lines indicate 95% confidence intervals.

Identifying age–cohort–time specific ATTs One way to address the age imbalance documented above is to estimate cohort–time–specific treatment effects within age groups and aggregate these to obtain $ATT(\ell)$. This approach conditions on lifecycle stage while continuing to compare treated individuals to not-yet-parents.

Baker et al. (2025) show that $ATT(\ell)$ can be identified under a weaker, *conditional* parallel trends assumption that conditions on covariates predictive of untreated outcome trends. For discrete covariates such as age, a transparent strategy is to estimate covariate-specific treatment effects and then aggregate them. Using not-yet-parents as controls, age-specific treatment effects $ATT^a(g,t)$ are defined as:

$$\begin{aligned} ATT^a(g,t) &= \mathbb{E}[Y_{it}(g) - Y_{it}(0) \mid G_i = g,\, age_{it} = a] \\ &= \mathbb{E}[Y_{it}(g) - Y_{i,g-1}(0) \mid G_i = g,\, age_{it} = a] \\ &\quad - \mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i > \max\{g,t\},\, age_{it} = a]. \end{aligned} \tag{19}$$

Identification of $ATT^a(g,t)$ requires parallel trends *within* age groups and an additional overlap condition: at each age and time, there must exist both treated and untreated individuals. Baker et al. (2025) refer to this requirement as *strong overlap*. Formally, for some $\epsilon > 0$ and all cohorts



(a) Women **(b)** Men

FIGURE 12. Counterfactual trend, disaggregated

Notes: This figure shows the change in average earnings since 2010 for the treated 2011 cohort and for each cohort that serves as part of the comparison group at different event times. Earnings include zeros and are expressed in 2010 euros. For each event time, the counterfactual is constructed using not-yet-parents from all later cohorts $g' > t$. A cohort g is defined by the calendar year in which individuals have their first child. Steeper earnings growth among later cohorts raises concerns that this approach may overestimate untreated earnings for earlier treated cohorts.

$g \in \mathcal{G}$,

$$\epsilon < \mathbb{P}(G_i = g \mid age_{i,t} = a) < 1 - \epsilon.$$

The comparison underlying $ATT^a(g,t)$ is transparent: for individuals in cohort g observed at age a, earnings growth from $t = g - 1$ to $t = g + \ell$ is compared to the contemporaneous earnings growth of individuals of the same age who have not yet had children and will be treated later.

This is precisely the identifying variation that the standard approach claims to exploit,[18] but here it is formalized under a DiD framework and requires only conditional parallel trends, rather than the stronger conditional independence assumption.

The counterfactual outcome is well defined and identified under the same assumptions:

$$\begin{aligned}\mathbb{E}[Y_{it}(0) \mid G_i = g,\, age_{it} = a] = \mathbb{E}[Y_{i,g-1}(0) \mid G_i = g,\, age_{i,t} = a] \\ + \mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i > \max\{g,t\},\, age_{i,t} = a],\end{aligned} \tag{20}$$

which can be used to scale treatment effect estimates.

[18] See, for instance, the identification discussion in Kleven et al. (2019).

4.4. Estimation and inference under PT using heterogeneity-robust estimators

Identification of $ATT(g,t)$ under parallel trends suggests a natural estimator based on sample analogues of conditional expectations. When using not-yet-treated cohorts as controls, the cohort–time average treatment effect for $t > g$ can be estimated as:

$$\widehat{ATT}(g,t) = \frac{\sum_{i=1}^{n} \mathbf{1}\{G_i = g\}(Y_{i,t} - Y_{i,g-1})}{\sum_{i=1}^{n} \mathbf{1}\{G_i = g\}} - \frac{\sum_{i=1}^{n} \mathbf{1}\{G_i > t\}(Y_{i,t} - Y_{i,g-1})}{\sum_{i=1}^{n} \mathbf{1}\{G_i > t\}}. \tag{21}$$

This estimator corresponds to the cohort–time DiD estimator proposed by Callaway and Sant'Anna (2021). I refer to it as the CS estimator. For $t < g-1$, the same expression can be used to assess pre-treatment differences in trends. In staggered designs, this estimator recovers $ATT(g,t)$ without requiring the researcher to subset the data manually; throughout, I implement it using the `csdid2` command in Stata, clustering standard errors at the individual level.

Age-specific effects $\widehat{ATT}^{a}(g,t)$ are obtained by applying Equation (21) within age groups. Aggregate cohort–time effects $\widehat{ATT}(g,t)$ are then constructed by averaging age-specific estimates using weights proportional to the distribution of ages within cohort g at each event time. Finally, $\widehat{ATT}(\ell)$ is obtained by aggregating $\widehat{ATT}(g, g+\ell)$ across cohorts, weighting by cohort size at each ℓ.

Why not OLS with fixed effects? Event study specifications estimated by OLS with individual fixed effects are inappropriate in settings with staggered treatment timing and heterogeneous treatment effects. As shown in the DiD literature (e.g. Sun and Abraham, 2021; Goodman-Bacon, 2021) and discussed in the child-penalty context by Melentyeva and Riedel (2023), such specifications do not recover well-defined causal estimands under heterogeneity.

Relation to alternative estimators The forward-engineering approach of Baker et al. (2025), adopted here, makes explicit the link between the estimand, the identifying assumptions, and the estimator. Estimating age-specific $ATT^{a}(g,t)$ is the most transparent way to condition on age, but it comes at a cost in precision relative to approaches that partial out age parametrically or impose additional restrictions on untreated potential outcomes.

For instance, the CS framework allows the inclusion of covariates, which improves precision compared to subsetting the data. Yet it requires additional parametric assumptions, and does not deliver age-specific treatment effects, which are of direct substantive interest in this context (see Sections 4 and 5.1.4 of Baker et al., 2025).

Melentyeva and Riedel (2023) implement a stacked estimator that, similarly to my estimator, com-

pares individuals of the same age to not-yet-parents, with two key differences: First, they restrict the control group to individuals who have their child not too much later in life. This approach relies hence on a weaker assumption of PT between parents and not-yet-parents who are closer in age at childbirth, while restricting the horizon for which treatment effects can be estimated.[19] Second, they pool observations across calendar years and therefore abstract from aggregate shocks and macroeconomic conditions. This imposes additional structure on untreated earnings dynamics, which are implicitly assumed to evolve in parallel within a given life-cycle stage, independent of macroeconomic conditions. Depending on the period under study, this assumption may be restrictive. Either way, from a forward-engineering perspective, these restrictions should be made explicit, justified ex-ante, and, where possible, formally tested empirically.

The estimator derived here prioritizes unbiasedness under clearly stated identifying assumptions. By working directly with the 2×2 DiD building blocks underlying $ATT(g,t)$ and conditioning explicitly on age and calendar time, it makes both the estimand and the source of identifying variation transparent. This comes at a cost in precision relative to estimators that pool across calendar years or rely on stronger functional form restrictions. Comparing point estimates across such approaches can help assess the empirical relevance of these trade-offs. In my application, I will, therefore, compare estimates across approaches.

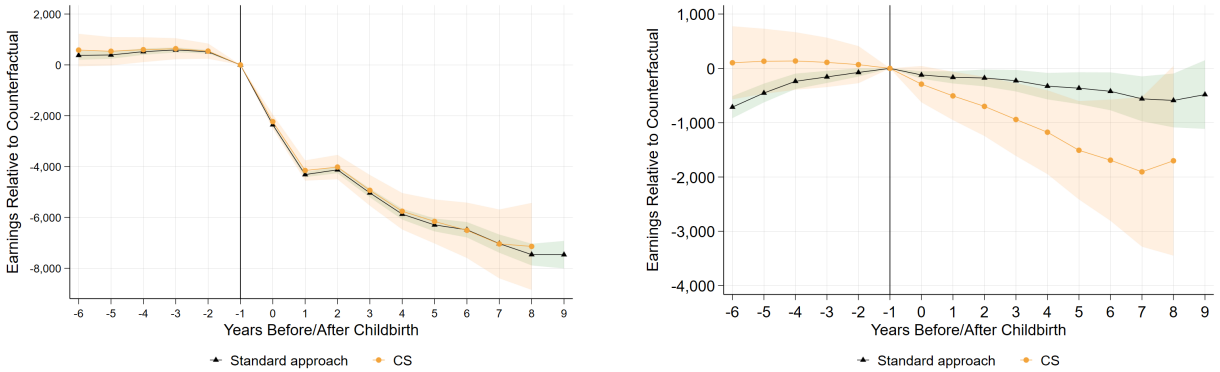
4.5. Difference-in-differences results

Figure 13 reports difference-in-differences estimates using my preferred estimator, relying on conditional parallel trends, for mothers (Panel a) and fathers (Panel b), together with estimates from the standard event study specification in Equation (1).

Because of limited power, I do not estimate $ATT^a(g,t)$ separately for each age at childbirth. Instead, I group parents into four age-at-birth bins (20–25, 26–30, 31–35, and 36–40). For each cohort g, I restrict treatment and control groups to individuals born within the corresponding birth-year range and estimate $ATT(g,t)$ within each bin using the CS estimator. I then aggregate age-specific estimates to obtain $\widehat{ATT}(\ell)$, as described above.

Mothers. For mothers, estimates obtained by aggregating age-specific effects are very similar in magnitude to those from the standard approach, albeit less precisely estimated. Eight years after childbirth, mothers earn on average €7,137 less than their counterfactual earnings, corresponding to a 23% loss (Figure D.5a).

[19] In their design, cohorts are defined by age at first birth, denoted c, rather than by calendar year, and treatment effects are estimated by comparing earnings changes between ages $c-1$ and $c+\ell$ for treated cohorts and not-yet-treated cohorts with $c' \leq c+5$. Treatment effects in this setup can be estimated until four years after childbirth.



(a) Mothers **(b)** Fathers

FIGURE 13. TE estimates under C-PT

Notes: This figure reports event-study estimates of earnings effects relative to counterfactual earnings for mothers (Panel a) and fathers (Panel b). The black line ("Standard approach") shows estimates from Equation (1), which relies on conditional independence of fertility timing given age and year. The orange line ("CS") shows estimates obtained using the age–cohort–time DiD estimator aggregated by event time (see Section 4.3), relying on conditional parallel trends and using not-yet-treated individuals as the control group. Shaded areas indicate 95% confidence intervals. The sample includes individuals aged 18–55 who have their first child between ages 20 and 45 in calendar years 2011–2016. Earnings include zeros for individuals who do not work and are deflated to 2010 euros.

Fathers. For fathers, results differ markedly from the standard approach (that relies on CIA, which for men was clearly rejected in the data). Aggregating age-specific $ATT^a(g,t)$ yields near-zero estimates in pre-treatment periods, lending credibility to the weaker conditional parallel trends assumption. Seven years after childbirth, fathers earn on average €1,904 less than their counterfactual earnings, corresponding to a 4.7% loss (Figure D.5b).

Taken together, these findings indicate that the standard approach may understate the cost of fatherhood in this sample. While limited precision prevents strong conclusions, the point estimates point to economically meaningful losses for fathers—larger than those typically reported in the literature, which often finds no effect of children on men's earnings.

Robustness to alternative age controls. Estimating $ATT(g,t)$ within age bins provides a transparent way to address age-related imbalance. Results are very similar if age is instead controlled for non-parametrically within the CS framework, or if cohorts are defined by age at childbirth following Melentyeva and Riedel (2023) (see Figure D.6).[20]

Heterogeneity by age at birth To examine heterogeneity in the impact of children by age at childbirth, Figure 14 plots age-specific $ATT^a(\ell)$ obtained by aggregating $ATT^a(g,t)$ within age-

[20] I replicate their stacking procedure but apply the CS estimator separately within each age-at-birth cohort, rather than estimating treatment effects via OLS.

at-birth bins across event time. Results scaled by counterfactual earnings are reported in Figure D.10 in the appendix.

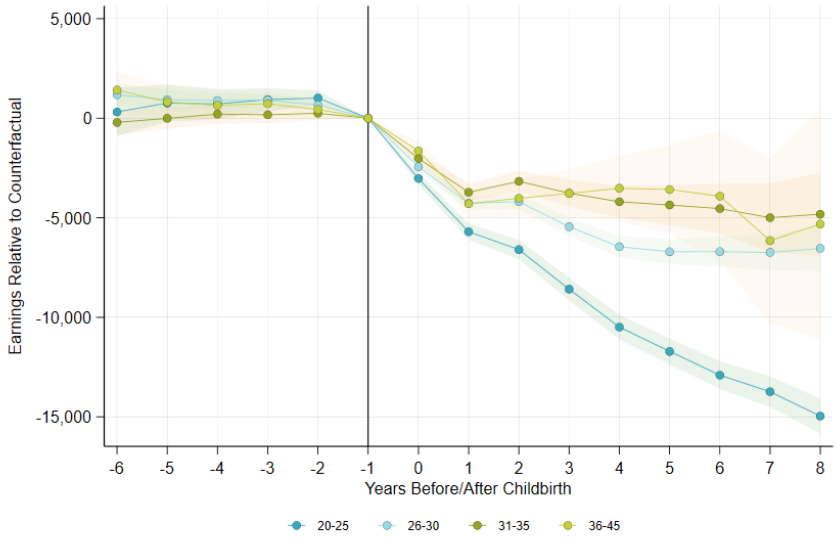
Mothers. All groups of mothers experience a sharp earnings decline following childbirth, but the magnitude of the penalty varies substantially by age at first birth. Women who become mothers between ages 20 and 25 face by far the largest losses, amounting to approximately €15,000 eight years after birth—nearly 50% of counterfactual earnings (Figure D.10a). These losses are significantly larger, both statistically and economically, than those experienced by older mothers. For women who have their first child between ages 26 and 30, earnings losses eight years after birth amount to €6,500 (21%). For those aged 31–35 and 36–40 at childbirth, losses are smaller, at €4,800 and €5,300, corresponding to 15% and 17.8%, respectively. Across all age groups, pre-treatment trends are flat, lending credibility to the conditional parallel trends assumption.

Fathers. A similar but distinct pattern emerges for fathers. Early fatherhood is associated with substantial and persistent earnings losses. Fathers who become parents between ages 20 and 25 experience the largest declines, losing approximately €9,300 eight years after childbirth—about 26% of counterfactual earnings. Fathers aged 26–30 at first birth also experience meaningful losses (€3,200, or 8.8%). In contrast, men who become fathers at older ages—particularly those aged 31–35 and 36–40—do not experience statistically significant post-birth earnings losses. These findings suggest that the absence of a fatherhood penalty in aggregate estimates masks substantial heterogeneity: early fatherhood carries sizable earnings costs, whereas later fatherhood does not.

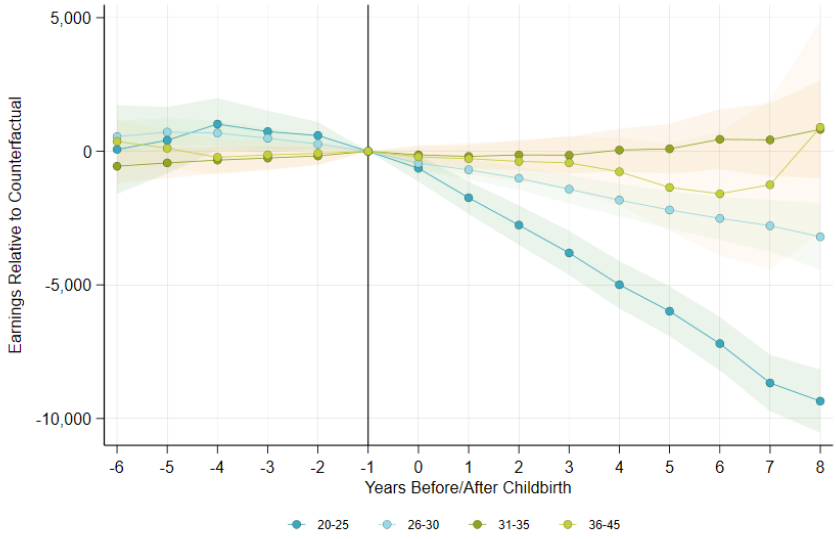
One potential explanation for larger penalties among younger parents is higher subsequent fertility. Indeed, younger first-time parents are more likely to have additional children during the sample window (Figure D.11). To assess whether this mechanism drives the observed heterogeneity, I repeat the analysis restricting the sample to individuals observed with only one child. The resulting pattern is very similar (Figure D.12), indicating that differential fertility alone does not account for the age gradient in earnings losses.

This analysis is descriptive: younger and older parents differ along many observed and unobserved dimensions.Nonetheless, the heterogeneity documented here is large and likely of policy relevance. Not only do young mothers and fathers experience especially large earnings losses, but the gender gap in the impact of parenthood is also widest among the youngest parents. Taken together, these findings underscore the importance of accounting explicitly for timing in the estimation and interpretation of child penalties and suggest that policies targeting young parents—particularly young mothers—may be critical for reducing remaining gender gaps.

The results above are identified by comparing parents to not-yet-parents of the same age, that is,



(a) Mothers



(b) Fathers

FIGURE 14. Event study results, by age group

Notes: This figure shows event study estimates of child penalties by bins of age at first birth, separately for women and men. The estimates are based on aggregating age-cohort-time-specific $ATT^a(g,t)$ as defined in Section 4.3 within four age-at-childbirth bins by event time. The vertical axis shows earnings relative to the counterfactual (in 2010 euros), and the horizontal axis denotes event time, where 0 is the year of childbirth. The outcome variable earnings is specified in levels and includes zeros for individuals who do not work. My sample includes individuals aged 18 to 55 who have their child between the age 20 and 45 between calendar years 2011 and 2020. The shaded area indicates 95% confidence intervals for the respective estimates.

to individuals who delay childbirth. The causal interpretation therefore hinges on the assumption that these delayers would have followed similar untreated earnings trajectories as earlier-treated parents. This assumption is not innocuous. As shown here and documented in other contexts (Melentyeva and Riedel, 2023; Adams et al., 2024; Leventer, 2025), individuals who have children at later ages differ systematically from younger parents, which may threaten identification. The next subsection examines an alternative alignment—conditioning on age at childbirth instead—and evaluates its credibility.

4.6. The age vs. timing trade-off in identifying child penalties

An alternative alignment is to fix age at childbirth rather than age. Conceptually, this amounts to considering treatment effects for individuals who become parents at the same age c:

$$ATT^c(g,t) = \mathbb{E}[Y_{it}(g) - Y_{it}(0) \mid G_i = g,\ age_i^{1ch} = c].$$

This strategy compares parents to individuals who will also become parents at age c, but who are younger at time t and therefore at earlier career stages.

Figure D.9 illustrates the implications of conditioning on age at childbirth. Aggregating treatment effects over ATT^c yields clear positive pre-trends, indicating that the comparison group—who are younger at time t—exhibits steeper earnings growth prior to childbirth. Based on these pre-trends, conditioning on age rather than age at childbirth appears more credible in this setting, with the caveats outlined above. Especially at longer event times, imbalances in age between treated and control groups become more pronounced, weakening the plausibility of parallel trends. Long-run estimates should therefore be interpreted with this limitation in mind.

More generally, this discussion highlights a fundamental challenge in identifying child penalties using event study designs: one cannot simultaneously fix age, age at childbirth, and calendar time, and any design must navigate the implied trade-offs.

5. CONCLUSION

Child-related gender inequality remains a key driver of persistent gender gaps in labor market outcomes across high-income countries. Event study designs have played a central role in documenting the dynamic impact of children on women's and men's labor market outcomes across a wide range of countries, offering clear advantages over earlier identification strategies. At the same time, credible causal identification in this setting is challenging. Fertility timing is endogenous, earnings trajectories vary strongly over the life cycle, and researchers face non-trivial trade-offs when constructing counterfactual outcomes. This paper revisits existing approaches to estimating child penalties with the aim of providing guidance on identification and estimation in this context.

I first formalize the identification assumptions underlying the standard event study approach with age and calendar year fixed effects and show that it relies on a conditional independence assumption. I show that even when conditional independence holds, the standard estimator is prone to contamination bias in the presence of treatment effect heterogeneity, akin to biases documented in the TWFE literature. An imputation approach provides a transparent and easy-to-implement alternative estimator under conditional independence, separating identification from estimation and avoiding such bias. While CIA appears a strong assumption in this context, it allows to estimate child penalties for a large set of countries without high-quality panel data. I provide tests to formally assess the credibility.

When rich panel data are available, identification strategies based on a weaker parallel trends assumption, controlling for individual time-invariant unobserved heterogeneity, provide a natural alternative. However, these approaches require careful consideration of the control group and explicit acknowledgment of underlying trade-offs. In particular, researchers cannot simultaneously compare individuals at the same age who also have their child at the same age. By adopting the forward-engineering framework of Baker et al. (2025), this paper derives heterogeneity-robust estimators step by step, making the identifying assumptions, comparisons, and implicit weighting schemes explicit. This framework provides guidance for researchers to choose and justify the empirical strategies they view as most credible.

Applying these methods to French administrative data, my aggregate results confirm a substantial and persistent cost of motherhood, in line with the existing literature and reveal a non-negligible cost of fatherhood. At the same time, the framework motivates and enables a transparent analysis of heterogeneity—an aspect of child penalties that has received comparatively less attention. I document pronounced heterogeneity by age at childbirth and uncover large and economically meaningful earnings losses for young fathers, an effect that is not visible in aggregate estimates. These insights carry clear policy implications: the large losses faced by young parents suggest

that policies aimed at improving the reconciliation of work and family life for this group may be particularly important, including for closing remaining gender gaps.

While substantial progress has been made in documenting child-related inequalities, further work is needed to understand the underlying mechanisms and policy-relevant heterogeneity. The framework developed in this paper aims to contribute to this agenda by clarifying identification trade-offs and providing transparent tools for applied research.

References

Adams, A., M. F. Jensen, and B. Petrongolo (2024). *Birth Timing and Spacing: Implications for Parental Leave Dynamics and Child Penalties*. IZA-Institute of Labor Economics.

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

Andresen, M. E. and E. Nix (2022). What causes the child penalty? evidence from adopting and same-sex couples. *Journal of labor economics 40*(4), 971–1004.

Andresen, M. E. and E. Nix (2025). You can't force me into caregiving: Paternity leave and the child penalty. *The Economic Journal*, ueaf057.

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

Angrist, J. D. and W. N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review 88*(3), 450–477.

Baker, A., B. Callaway, S. Cunningham, A. Goodman-Bacon, and P. H. C. Sant'Anna (2025). Difference-in-differences designs: A practitioner's guide. *Journal of Economic Literature*. forthcoming.

Bonnet, C., L. Dubreuil, B. Garbinti, and P. Pora (2025). Child penalty & the rise in within-couple income inequality. Technical report.

Borusyak, K., X. Jaravel, and J. Spiess (2024, 02). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies 91*(6), 3253–3285.

Breda, T., P. Dutronc-Postel, J.-S. Parraud, and M. Tô (2021). Gender pay gaps within companies. IPP Policy Briefs 68, Institut des Politiques Publiques (IPP).

Caenen, Y. and P. Virot (2023). La part des enfants de moins de 3 ans confiés principalement à une assistante maternelle ou une crèche a presque doublé entre 2002 et 2021. *Études et Résultats* (1257).

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics 225*(2), 200–230.

Casarico, A. and S. Lattanzio (2023). Behind the child penalty: understanding what contributes to the labour market costs of motherhood. *Journal of Population Economics 36*(3), 1489–1511.

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

Coudin, , S. Maillard, J. Tréguier, and T. Breda (2025). Sorting into lower-paying firms: The firm premium child penalty in the french labour market. Manuscript.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review 110*(9), 2964–2996.

European Institute for Gender Equality (2019). Gender equality index 2019 – france – domain: Work. Accessed: 2025-05-13.

Eurostat (2019). International women's day 2019 – women in the eu earned on average 16% less than men in 2017. Accessed: 2025-05-13.

Givord, P. and C. Marbot (2015). Does the cost of child care affect female labor market participation? an evaluation of a french reform of childcare subsidies. *Labour Economics 36*, 99–111.

Goldin, C., S. P. Kerr, and C. Olivetti (2022). When the kids grow up: women's employment and earnings across the family cycle. Technical report, National Bureau of Economic Research.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2024). Contamination bias in linear regressions. *American Economic Review 114*(12), 4015–4051.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics 225*(2), 254–277.

Kleven, H. (2022). The geography of child penalties and gender norms: A pseudo-event study approach. Working Paper 30176, National Bureau of Economic Research.

Kleven, H., C. Landais, and G. Leite-Mariante (2025). The child penalty atlas. *Review of Economic Studies 92*(5), 3174–3207.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019). Child penalties across countries: Evidence and explanations. In *AEA Papers and Proceedings*, Volume 109, pp. 122–126. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2024). Do family policies reduce gender inequality? evidence from 60 years of policy experimentation. *American Economic Journal: Economic Policy 16*(2), 110–149.

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

Kleven, H., C. Landais, and J. E. Søgaard (2021). Does biology drive child penalties? evidence from biological and adoptive families. *American Economic Review: Insights 3*(2), 183–198.

Leventer, D. (2025). Conditional triple difference-in-differences. *arXiv preprint arXiv:2502.16126*.

Lim, N. and L.-M. Duletzki (2023). The effects of public childcare expansion on child penalties: Evidence from west germany.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review 107*(6), 1611–1637.

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

Meurs, D. and P. Pora (2019). Gender equality on the labour market in france: A slow convergence hampered by motherhood. *Economie et Statistique 510*(1), 109–130.

Meurs, D. and P. Pora (2025). Traditional views, egalitarian views, and the child penalty: Insights from immigrant populations in france.

Pora, P. and L. Wilner (2024). Dissecting child penalties. *Industrial and Labor Relations Review*.

Rosenzweig, M. R. and K. I. Wolpin (1980). Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political economy 88*(2), 328–348.

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

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics 225*(2), 175–199.

Thakral, N. and L. T. Tô (2025). The child penalty and an age-old problem.

Wilde, E. T., L. Batchelder, and D. T. Ellwood (2010). The mommy track divides: The impact of childbearing on wages of women of differing skill levels. Technical report, National Bureau of Economic Research.

A. CONTAMINATION BIAS IN SAMPLE BALANCED IN EVENT TIME

Figure A.1 shows the setup with a sample that is balanced in event time, observed from $ET = -2$ to $ET = 3$. I introduce heterogeneity at $ET = 1$. Older mothers, that is, above or equal to the median, have smaller TE in absolute terms. As before, I estimate α_3 for varying levels of TE heterogeneity and use Goldsmith-Pinkham et al. (2024)'s composition to estimate both components. Figure A.2 shows the results, and confirms contamination bias also arises in this setup.

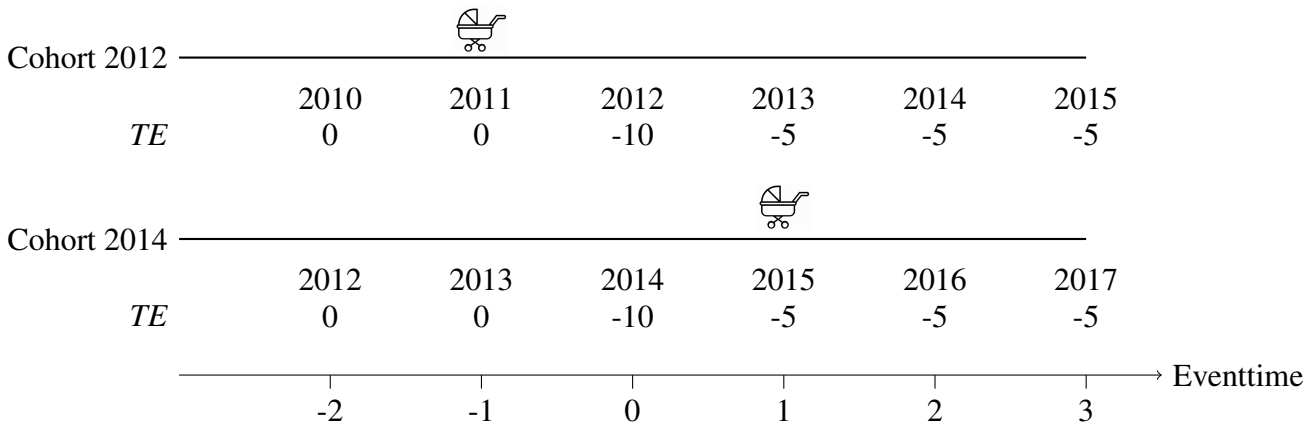


FIGURE A.1. Illustration with two cohorts, balanced in ET

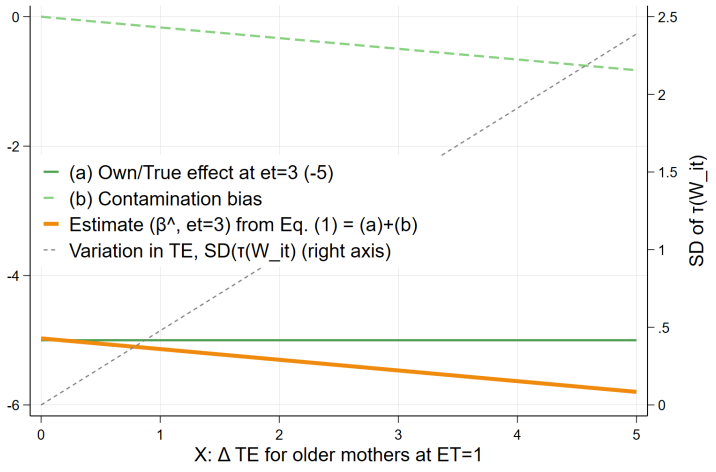


FIGURE A.2. Decomposition of $\hat{\alpha_3}$

B. INSTITUTIONAL CONTEXT

B.1. Parental leave, childcare policy, and taxation

France has a comprehensive, family-oriented social protection system. Mothers are entitled to 16 weeks of maternity leave (six weeks before and ten after birth) for the first and second child, and 26 weeks (eight before and 18 after) from the third child onward. Paternity leave was introduced in 2002, initially granting 14 days, and was extended in 2021 to 21 days, of which seven are mandatory. Since 2004, various additional parental allowances have been consolidated under the *Prestation d'Accueil du Jeune Enfant (PAJE)*. This includes a means-tested birth bonus, monthly family allowances, and benefits for parents who interrupt work after childbirth (formerly the *Complément de Libre Choix d'Activité, CLCA*). In 2015, the CLCA was replaced by the *Prestation Partagée d'Éducation de l'Enfant (PreParE)*, which grants each parent six months of leave and includes incentives to share entitlements across partners (see Pora and Wilner (2024) for details).

Parental leave policies are complemented by support for early childhood care. From the age of three, children attend preschool (*école maternelle*), which is free of charge and became legally compulsory in 2019. However, enrollment was already near-universal throughout the sample period (2010–2020). For children under three, care is provided either in formal *crèches* or, more commonly, by licensed childminders (*assistantes maternelles*). Over the past two decades, reliance on informal care has declined: in 2002, 74% of children under age three were cared for by parents or relatives, compared to 59% in 2021. In parallel, the share of children cared for by childminders rose from 13% to 20%, and formal crèche enrollment from 9% to 18% during the same period (Caenen and Virot, 2023). Families receive financial support through both direct subsidies and indirect tax credits, typically covering up to 50% of childcare expenses, subject to a ceiling. While these measures help reduce the effective cost of care, benefits are often tied to parental income, creating uneven incentives across the wage distribution. In particular, mothers at the lower end of the earnings spectrum who, therefore, have lower opportunity costs of staying at home, are more likely to interrupt or reduce their labor supply (Givord and Marbot, 2015).

In addition to leave and care policies, the tax system also plays a role in influencing parents' behavior. Household taxation takes the number of dependents into account and provides tax relief to families with children. Joint taxation between partners can create incentives for within-household specialization, further reinforcing gendered labor supply responses.

In sum, France's family policy framework offers an extensive set of tools that not only influence fertility behavior but also shape post-birth labor market outcomes—making them central to understanding post-birth labor market trajectories and the estimation of child penalties.

B.2. Female labor market participation and gender gaps

In 2017, the employment rate for women aged 20–64 in France was 66%, close to the EU average of 66.5% (European Institute for Gender Equality, 2019; Eurostat, 2019). For men, the corresponding figure was 73%, resulting in an employment gender gap of 7 percentage points (p.p.), below the EU average of 11.5 p.p. In the same year, 30% of women worked part-time compared to only 8% of men (European Institute for Gender Equality, 2019). In full-time equivalent (FTE) terms, female employment amounted to 41%, versus 53% for men. Among couples with children, both parents' FTE employment is higher, but so is the gender gap: 66% for mothers and 85% for fathers, a gap of 19 p.p. In contrast, the gap is just 3 p.p. for couples without children (Eurostat, 2019).

The gender pay gap remains substantial despite progress in recent decades. In 2018, women earned on average 17.4% less per hour than men (Breda et al., 2021). Occupational segregation partly explains this disparity: in 2017, 35% of women worked in education, health, and social work, compared to just 10% of men. Conversely, only 7% of women were employed in Science, technology, engineering, and mathematics (STEM) occupations, against 33% of men (European Institute for Gender Equality, 2019). Yet, as shown by Breda et al. (2021), even after controlling for occupation and age, a residual wage gap of 11% remains.

A few recent studies have estimated child penalties for the French context using event study frameworks. Based on data from the French Labor Force Survey, **?** document a 26 p.p. employment penalty over ten years following childbirth, accounting for 89% of the remaining gender employment gap. Using a similar methodology, Meurs and Pora (2019) find a 30% relative decline in total wage income for mothers, along with a 17% premium for fathers—though this may be biased by substantial pre-trends. Pora and Wilner (2024) estimate wage losses of 33% five years after childbirth for mothers, stabilizing at 20% after ten years. Over a longer horizon, Coudin et al. (2025) find a 34% decline in income for mothers over the twenty years following the birth of their first child. In a European context, the child penalty in France is lower than in Germany or Italy (41% and 33% for employment, respectively), yet higher than in Nordic countries such as Denmark or Sweden (14% and 9%, respectively) (**?**). None of these papers documents a loss for fathers associated with parenthood. If anything, fatherhood is associated with a premium for men.

Together, these studies highlight substantial and persistent penalties to women's, but not men's, employment and earnings in France.

C. TABLES

TABLE C.1. Descriptive Statistics by Year

	Year											
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
N	131,393	131,393	131,393	131,393	131,393	131,393	131,393	131,393	131,393	131,393	131,393	1,445,323
Female (%)	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
# treated	.	21,337	19,328	17,282	15,409	13,664	11,904	10,264	8,630	7,382	6,193	14,963
# (women) treated	.	11,064	9,773	8,629	7,678	6,679	5,843	4,887	3,991	3,366	2,802	7,558
# (men) treated	.	10,273	9,555	8,653	7,731	6,985	6,061	5,377	4,639	4,016	3,391	7,424
Female age	26.59	27.59	28.59	29.59	30.59	31.59	32.59	33.59	34.59	35.59	36.59	31.59
Male age	27.88	28.88	29.88	30.88	31.88	32.88	33.88	34.88	35.88	36.88	37.88	32.88
Any higher education (%)	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Number of children	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70
Female age at childbirth (g=year)	.	29.77	29.92	30.14	30.69	31.09	31.43	32.01	32.50	33.10	33.85	30.92
Male age at childbirth (g=year)	.	31.70	31.73	31.88	32.19	32.28	32.73	32.93	33.22	33.90	34.39	32.41
Female income (incl 0)	14,402	16,310	17,556	18,640	19,513	20,514	21,175	22,131	23,471	24,563	26,414	20,426
Male income (incl 0)	18,214	20,637	22,538	24,128	25,514	27,087	28,508	29,922	32,216	33,432	36,034	27,112
Female share of couple (gross) income (excl 0)	0.41	0.43	0.42	0.43	0.43	0.43	0.43	0.43	0.42	0.43	0.43	0.43
Female working hours	1,437	1,418	1,430	1,451	1,454	1,457	1,460	1,446	1,454	1,464	.	1,447
Male working hours	1,564	1,600	1,621	1,637	1,657	1,670	1,703	1,706	1,687	1,695	.	1,654

Notes: This table reports descriptive statistics by calendar year for the full sample (2010–2020), as well as for the pooled sample (last column). A cohort is defined by the year in which individuals have their first child. "Female (%)" refers to the share of women in the sample. "Female age" and "male age" refer to age at the time of observation. "Female age at childbirth (g = year)" and "male age at childbirth" refer to age at the year of first childbirth. "Income (incl 0)" includes individuals with zero earnings; income is expressed in 2010 euros. "Working hours" are annual hours worked, conditional on working. "Female share of couple (gross) income" is computed within couples and then aggregated, excluding couples where one partner has zero earnings.

TABLE C.2. Descriptive Statistics by Cohort

	Cohort										
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
N	234,707	212,608	190,102	169,499	150,304	130,944	112,904	94,930	81,202	68,123	1,445,323
# treated	21,337	19,328	17,282	15,409	13,664	11,904	10,264	8,630	7,382	6,193	14,963
# (women) treated	11,064	9,773	8,629	7,678	6,679	5,843	4,887	3,991	3,366	2,802	7,558
# (men) treated	10,273	9,555	8,653	7,731	6,985	6,061	5,377	4,639	4,016	3,391	7,424
Female age	33.77	32.92	32.14	31.69	31.09	30.43	30.01	29.50	29.10	28.85	31.59
Male age	35.70	34.73	33.88	33.19	32.28	31.73	30.93	30.22	29.90	29.39	32.88
Any higher education (%)	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Number of children	1.92	1.89	1.84	1.78	1.73	1.64	1.55	1.42	1.20	1.06	1.70
Age at childbirth	30.70	30.81	31.01	31.44	31.70	32.09	32.49	32.89	33.54	34.14	31.68
age_1ch_fem	29.77	29.92	30.14	30.69	31.09	31.43	32.01	32.50	33.10	33.85	30.92
age_1ch_male	31.70	31.73	31.88	32.19	32.28	32.73	32.93	33.22	33.90	34.39	32.41
Female income (incl 0)	19,496	19,601	20,251	21,149	20,742	20,882	21,124	21,076	21,060	20,931	20,426
Male income (incl 0)	28,672	28,159	28,007	27,807	27,103	26,489	25,917	25,172	24,804	23,978	27,112
Female share of couple (gross) income (excl 0)	0.42	0.42	0.42	0.43	0.43	0.44	0.44	0.44	0.45	0.45	0.43
Female working hours	1,407	1,425	1,434	1,456	1,456	1,460	1,469	1,482	1,481	1,505	1,447
Male working hours	1,683	1,673	1,672	1,664	1,656	1,646	1,635	1,623	1,609	1,583	1,654

Notes: This table reports descriptive statistics by calendar year for the full sample (2010–2020), as well as for the pooled sample (last column). A cohort is defined by the year in which individuals have their first child. "Female (%)" refers to the share of women in the sample. "Female age" and "male age" refer to age at the time of observation. "Female age at childbirth (g = year)" and "male age at childbirth" refer to age at the year of first childbirth. "Income (incl 0)" includes individuals with zero earnings; income is expressed in 2010 euros. "Working hours" are annual hours worked, conditional on working. "Female share of couple (gross) income" is computed within couples and then aggregated, excluding couples where one partner has zero earnings.

D. FIGURES

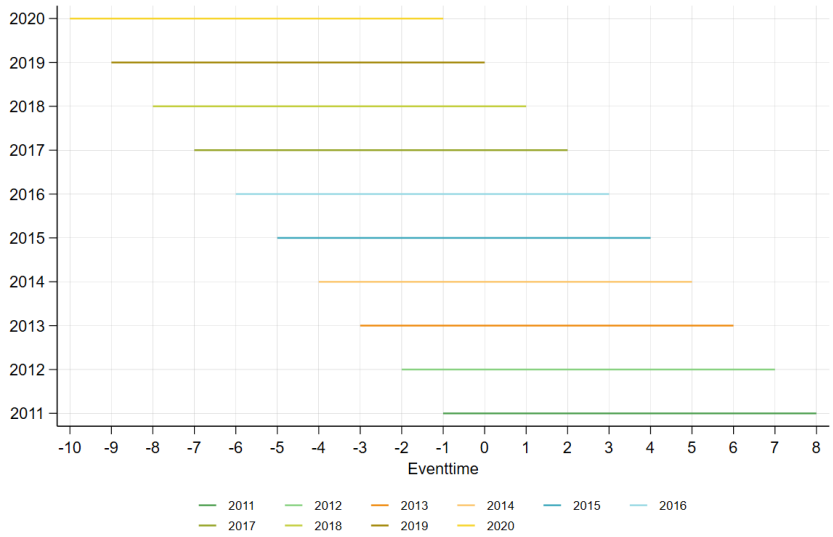


FIGURE D.3. Cohort contributions to event time ATT estimates

Notes: Each horizontal line shows the event-time values $\ell = t - G_i$ for which cohort G_i is observed with untreated counterparts. Treatment effects at event time ℓ are identified and computed only for the cohorts whose lines include that ℓ. The aggregated ATT_ℓ is therefore a vertical aggregation across these contributing cohorts at event time ℓ.

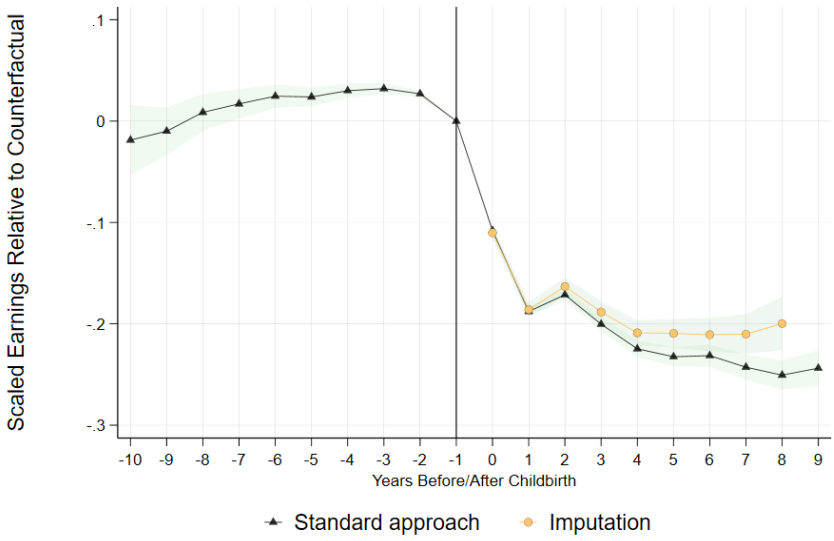
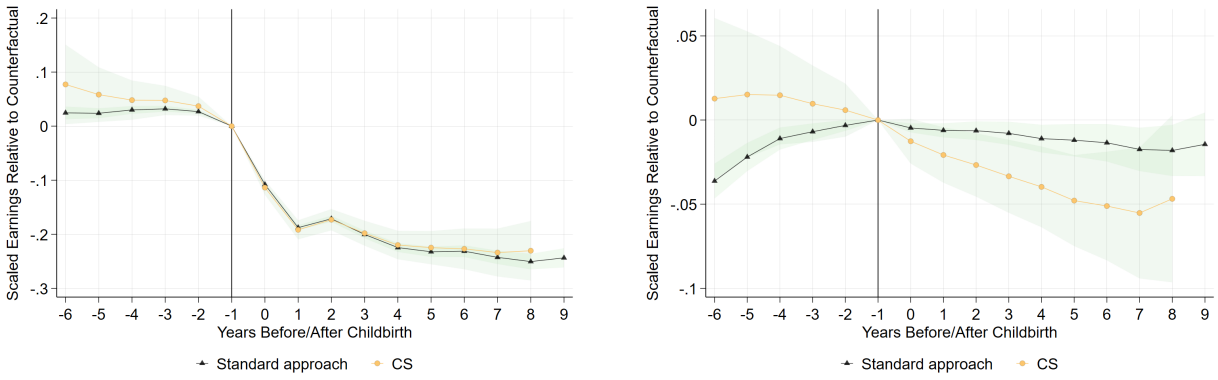


FIGURE D.4. Standard approach vs. imputation estimator, women

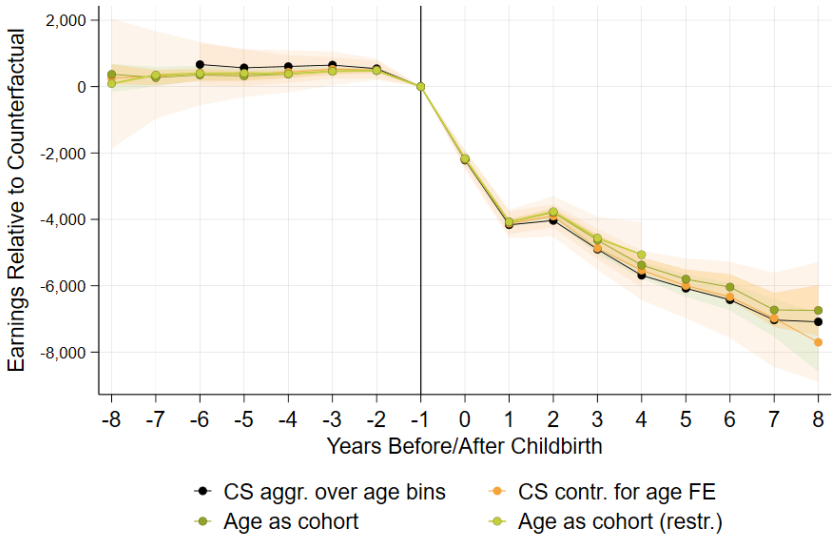
Notes: The figure plots estimated event–study effects on mothers' earnings relative to their counterfactual, scaled by counterfactual earnings, using the standard specification in Equation (1) (black line) and the imputation estimator (orange line). Counterfactual earnings in the standard approach are defined in (2). In the imputation approach, they are predicted for all treated observations based on age and calendar-year fixed effects estimated on untreated observations as explained in Section 3. The sample includes individuals aged 18–55 who have their first child between ages 20 and 45 in calendar years 2011–2020. Earnings include zeros for non-employed individuals and are deflated to 2010 euros.



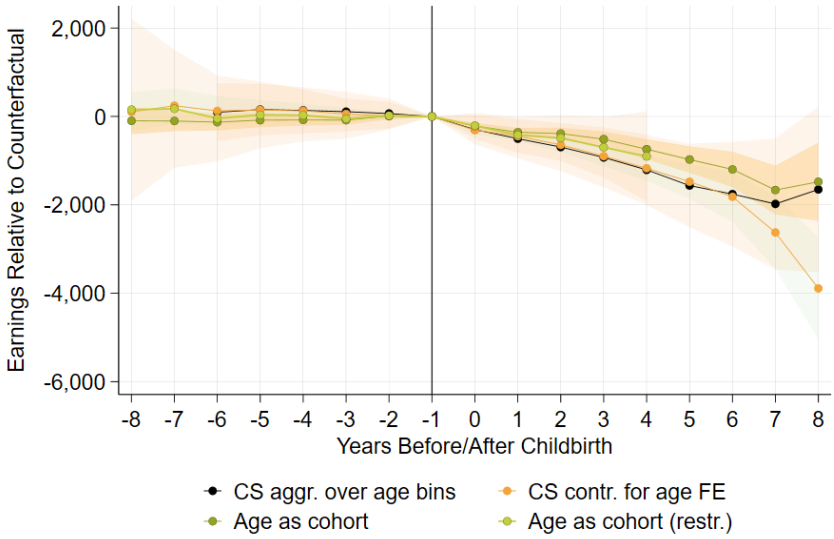
(a) Mothers **(b)** Fathers

FIGURE D.5. Scaled TE estimates under C-PT

Notes: This figure compares estimated event study coefficients across two specifications: (i) the standard approach from Eq. (1), and (ii) the CS estimator aggregated over age-specific $\widehat{ATT}^{a}(g,t)$. Estimates are scaled by the counterfactual. In the standard approach, counterfactual earnings are obtained using results from Eq. (1) and omitting the event time dummies. For CS, in time t for individuals of cohort g and age bin a, the counterfactual is the sum of pre-birth earnings (in $t = g-1$) plus the average trend in the comparison group between $t = g-1$ and t. The sample includes individuals aged 18 to 55 who have their child between the age 20 and 45 between calendar years 2011 and 2016. Earnings include zeros for individuals who do not work and are deflated to 2010 euros.



(a) Mothers



(b) Fathers

FIGURE D.6. Event study results, scaled by counterfactual earnings

Notes: This figure compares estimates of earnings effects relative to counterfactual earnings obtained using alternative ways to account for age, under parallel trends. The black line ("CS agg. over age bins") reports the baseline estimates, obtained by aggregating age-specific $ATT^a(g,t)$ over age bins as described in Section 4.3. The orange line ("CS contr. for age FE") reports estimates from a CS specification that conditions on age non-parametrically by including age fixed effects as covariates. The dark green line ("Age as cohort") redefines cohorts by age at first birth, following the alignment proposed by Melentyeva and Riedel (2023), but applies the CS estimator without restricting the control group. The light green line ("Age as cohort (restr.)") also defines cohorts by age at first birth, but restricts the control group to individuals who have their first child at most five years later ($c' \leq c + 5$), closely mirroring the stacked design of Melentyeva and Riedel (2023), while still using the CS estimator rather than OLS. Shaded areas indicate 95% confidence intervals. The sample includes individuals aged 18–55 who have their first child between ages 20 and 45 in calendar years 2011–2016.

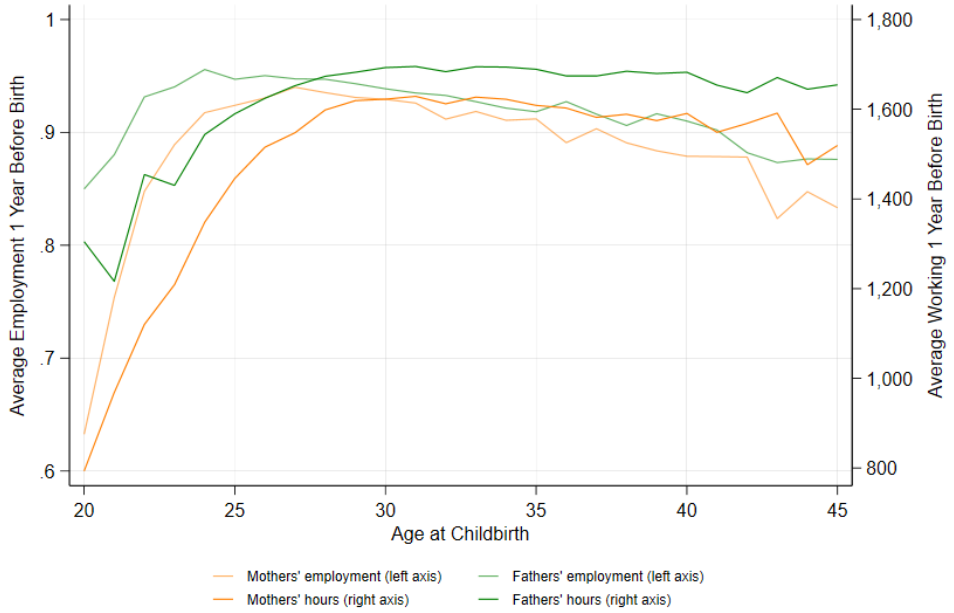


FIGURE D.7. Pre-birth employment rate and working hours by age at childbirth

Notes: This figure shows the average employment rate across cohorts in the year before birth, by age at childbirth and separate for women and men. Working hours are conditional on working.

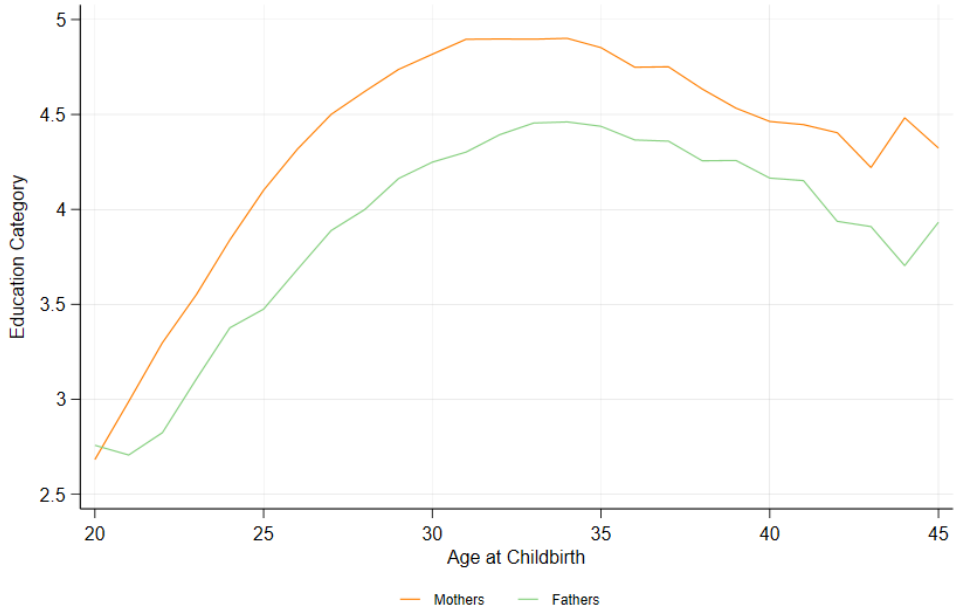
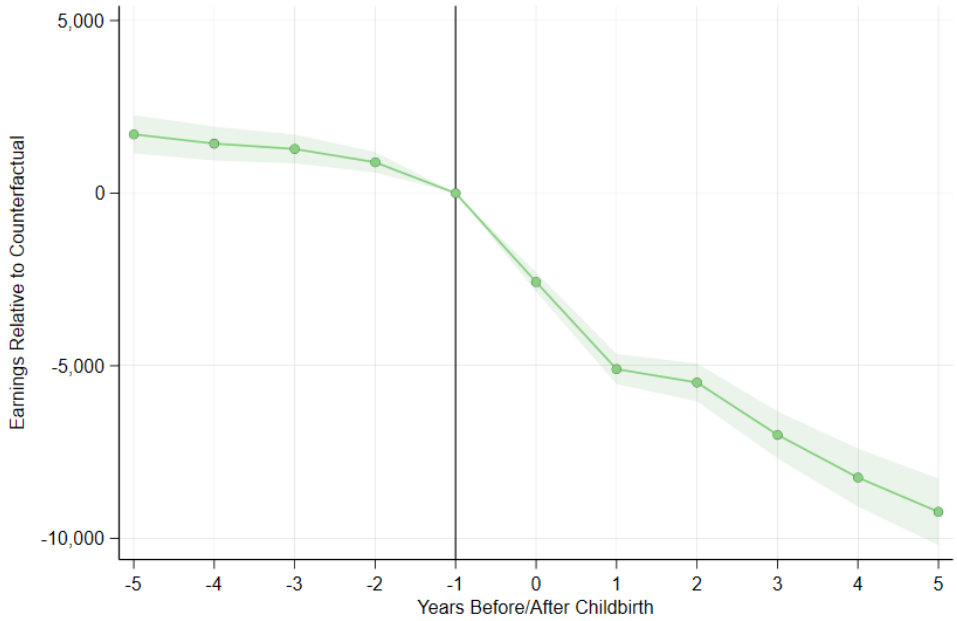
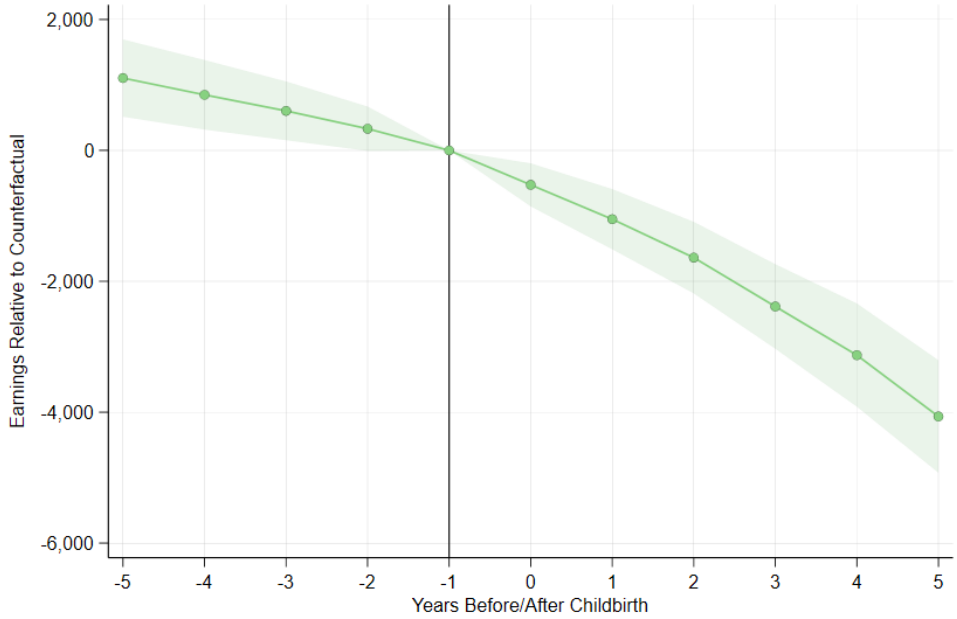


FIGURE D.8. Education category by age at childbirth

Notes: This figure shows the average education category across cohorts in the year before birth, by age at childbirth, separately for women and men. Level 4 corresponds to the equivalent of a technical or general high school diploma. Level 5 corresponds to undergraduate studies, whereas level 6 corresponds to master of PhD level studies.



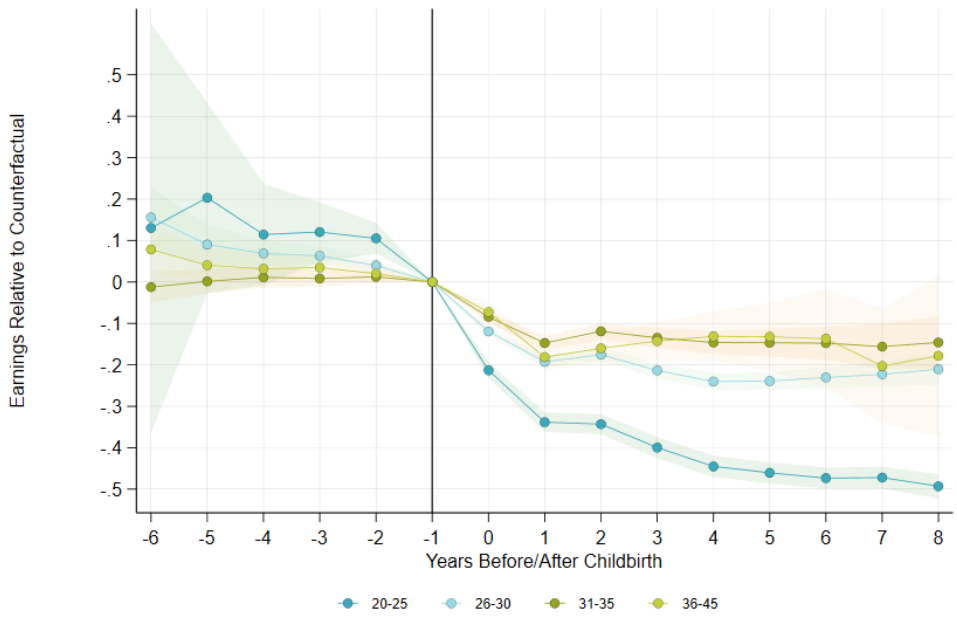
(a) Women



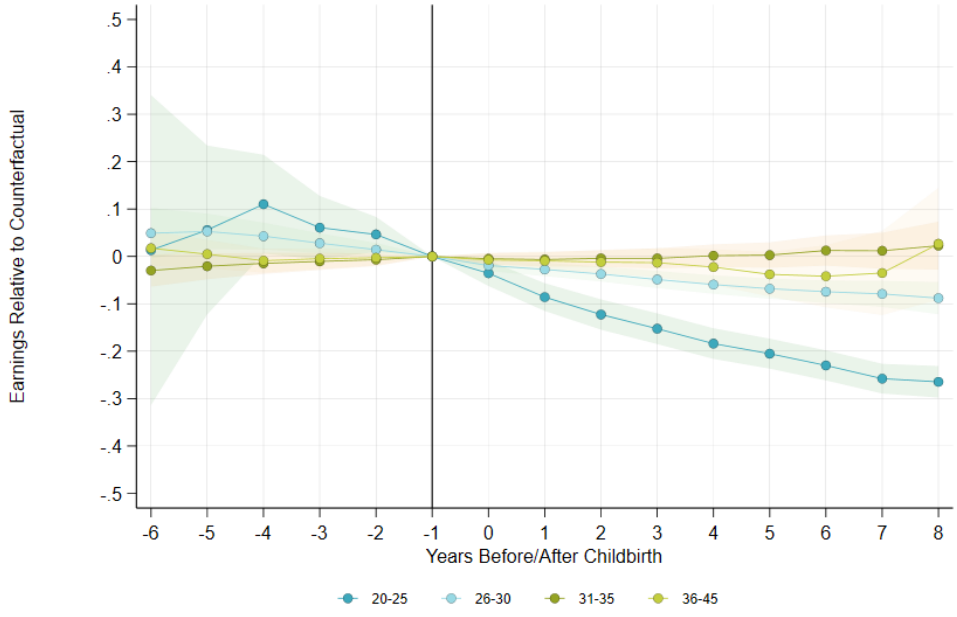
(b) Men

FIGURE D.9. Aggregated event study results matching on age at childbirth

Notes: This figure shows estimated event study coefficients obtained by aggregating age-at-childbirth specific $ATT^c(g,t)$ as defined in Section 4.6. $ATT^c(g,t)$ are estimated within four bins of age at childbirth: 20-25, 26-30, 31-35, 36-45. The vertical axis shows earnings relative to the counterfactual (in 2010 euros), and the horizontal axis denotes event time, where 0 is the year of childbirth. The outcome variable earnings is specified in levels and includes zeros for individuals who do not work. My sample includes individuals aged 18 to 55 who have their child between the age 20 and 45 between calendar years 2011 and 2020. The shaded area indicates 95% confidence intervals for the respective estimates.



(a) Women



(b) Men

FIGURE D.10. Event study results by age bin, scaled by counterfactual earnings

Notes: This figure shows event study estimates of child penalties by bins of age at first birth, separately for women and men. The estimates are based on aggregating age-cohort-time-specific $ATT^a(g,t)$ as defined in Section 4.3 within four age-at-childbirth bins by event time. The estimates are then scaled by counterfactual earnings, in time t for cohort g defined as the sum of pre-birth earnings (in $t = g-1$) plus the average trend in the comparison group between $t = g-1$ and t. My sample includes individuals aged 18 to 55 who have their child between the age 20 and 40 between calendar years 2011 and 2020. The shaded area indicates 95% confidence intervals for the respective estimates. Estimates at early event times are imprecise for some age groups due to limited overlap and smaller sample sizes within bins.

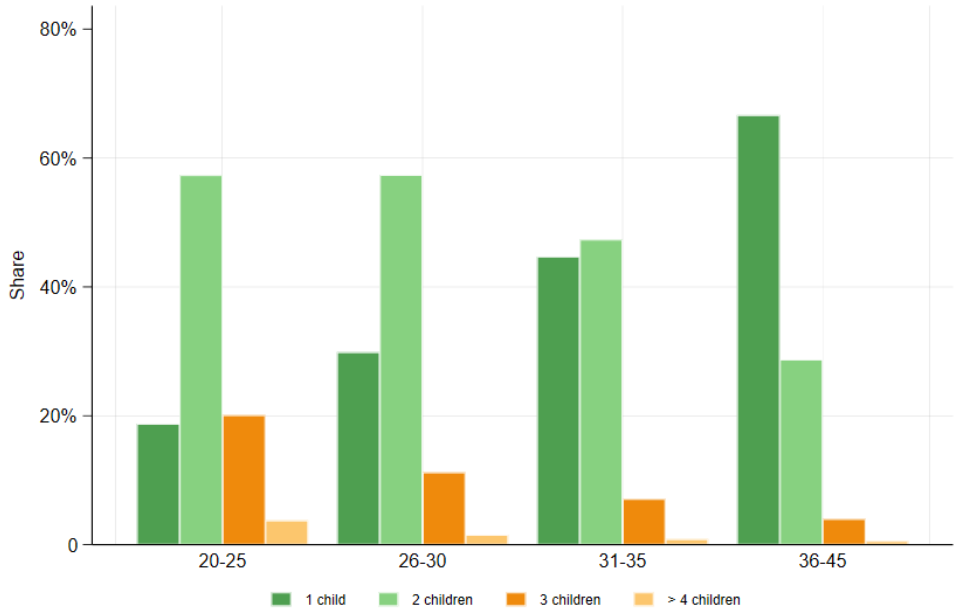
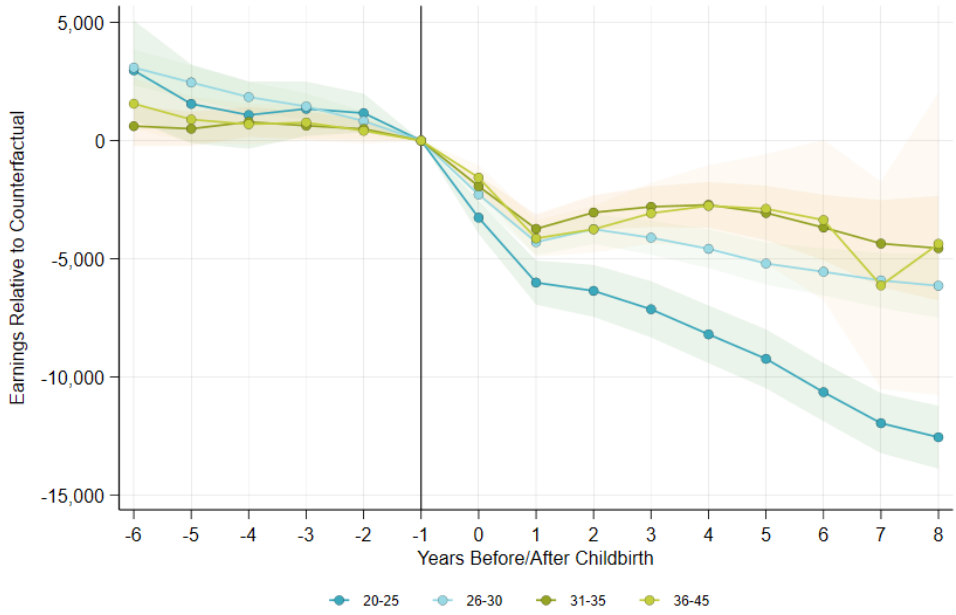
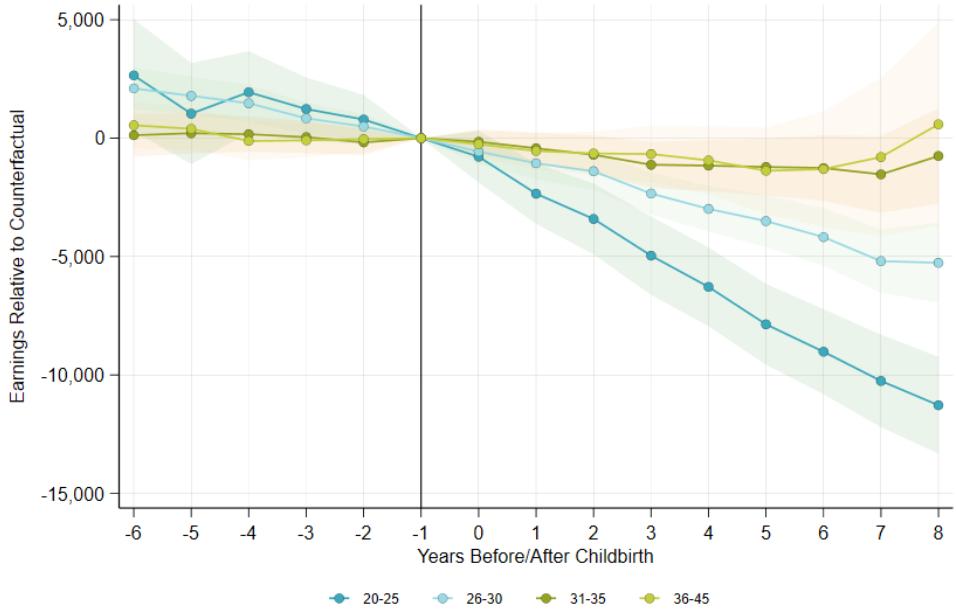


FIGURE D.11. Observed number of children, by age at childbirth

Notes: This figure shows the share of individuals with 1, 2, 3, or 4 or more children by four bins of age at childbirth. Younger first-time parents are more likely to be observed with 2 or 3 children.



(a) Women



(b) Men

FIGURE D.12. Event study results by age bin, sample only 1 child

Notes: This figure shows event study estimates of child penalties by bins of age at first birth, separately for women and men, in a restricted sample of individuals with only 1 child. The estimates are based on aggregating age-cohort-time-specific $ATT^a(g,t)$ as defined in Section 4.3 within four age-at-childbirth bins by event time. The vertical axis shows earnings relative to the counterfactual (in 2010 euros), and the horizontal axis denotes event time, where 0 is the year of childbirth. The outcome variable earnings is specified in levels and includes zeros for individuals who do not work. The shaded area indicates 95% confidence intervals for the respective estimates.