
ECONtribute
Discussion Paper No. 266

**Child Penalty Estimation and Mothers' Age at
First Birth**

Valentina Melentyeva

Lukas Riedel

December 2023

www.econtribute.de



Child Penalty Estimation and Mothers' Age at First Birth

Valentina Melentyeva (JMP) * Lukas Riedel[†]

December 5, 2023[‡]

Motherhood continues to pose significant challenges to women's careers, and a correct assessment of its effects is crucial for understanding the persistent gender inequality in the labor market. We show that the prevalent approach to estimate post-birth earnings losses – so called “child penalties” – is prone to yield substantially biased results. We demonstrate that the biases stem from conventional event studies pooling together first-time mothers of all ages, without considering their distinct characteristics and the varying impact of motherhood. To address the biases, we propose a novel approach that accounts for the heterogeneity by building upon recent advancements in the econometric literature on difference-in-differences models. Applying it to administrative data from Germany, we demonstrate that considering heterogeneity by maternal age at birth is crucial for both methodological correctness and a deeper understanding of gender inequality. Our approach yields substantially larger estimates of earnings losses after childbirth (by 20 percent), indicating that the costs of motherhood and related gender gaps in Germany are even larger than previously thought. Moreover, we demonstrate that effects and their interpretation differ significantly depending on maternal age at birth. We show that younger first-time mothers experience larger career costs of motherhood, as they miss out on the phase of the most rapid career progression.

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

JEL: J13, J16, J31, C23

* University of Cologne and ECONtribute (vmelentyeva@wiso.uni-koeln.de)

[†] ZEW Mannheim and University of Heidelberg (lukas.riedel@zew.de)

[‡] We are very grateful to Jérôme Adda, James Banks, Anna Bindler, Richard Blundell, Monica Costa Dias, Clément de Chaisemartin, Scott Cunningham, Thomas Dohmen, Christian Dustmann, Bernd Fitzenberger, Sergei Guriev, Lena Janys, Xavier Jaravel, Hyejin Ku, Barbara Petrongolo, Pia Pinger, Uta Schönberg, Sebastian Siegloch, Tom Zimmermann, and seminar and conference participants at the University College London, University of Cologne, University of Bonn, University of Manchester, ZEW – Leibniz Centre for European Economic Research, EEA Annual Meeting 2023, AFSE Annual Congress 2023, EALE Annual Meeting 2023 and CReAM-RFB Workshop 2023 for helpful discussions and suggestions. Melentyeva is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866.

1 Introduction

The past century has been marked by major successes of the women's rights movement. In developed countries, women gained – among other achievements – widespread and unrestricted access to education and employment. Nevertheless, gender inequality remains prevalent in the labor markets around the world and one particular reason for that persists to this day: motherhood is still costly for women's careers. Multiple studies have identified gender differences in parenthood costs as the major driver of the remaining gender inequality in labor market outcomes.¹ Therefore, it is crucial to correctly track the dynamics of the career costs of motherhood and study the underlying mechanisms in order to fully understand gender inequality and give informed policy advice. In this paper, we show that the most popular approach to estimate the labor market impacts of motherhood is likely to produce biased results. We illustrate how the biases emerge due to unaccounted differences among mothers of different ages, propose a solution that mitigates the biases and use it to provide a deeper understanding of labor market costs of motherhood.

The career costs of motherhood have been both subject of research and a recurring topic in public debates for a long time. Recently, the approach based on event studies around first childbirth – estimating so called “child penalties” – received widespread attention as it provides a straightforward and intuitive way to visualize the career impact of childbirth. The paper by Kleven, Landais, and Sogaard (2019) popularized the method and gained more than 1,200 citations over the span of four years. Kleven, Landais, and Sogaard crucially contributed to an understanding of the challenges of combining motherhood and career among a wide range of audiences. Researchers have been actively using the method to estimate gender inequality in the labor market, within and across countries, and to evaluate policies.

These common event studies pool together younger and older first-time mothers, implicitly assuming uniform effects of childbirth across women of different ages and comparing mothers to women who have not yet had a child. However, mothers are very different depending on their age at birth and the effects of motherhood vary across women who give birth at different stages of life and career development. As existing literature shows (see, for instance, Adda, Dustmann, and Stevens 2017; Goldin, Kerr, and Olivetti 2022) and as we confirm, the age at which a woman gives birth to her first child is highly correlated with both pre- and post-birth outcomes as well as human capital levels and other characteristics that are relevant in the labor market. In this paper, we demonstrate that the heterogeneity in the effects of childbirth and the characteristics of mothers introduces biases in event-study-based estimations and undermines the validity of the commonly employed control group consisting of older first-time mothers.

¹ See, for example, the review by Andrew et al. (2021), the book by Goldin (2022) and the papers by Adda, Dustmann, and Stevens (2017), Goldin, Kerr, and Olivetti (2022), and Blundell et al. (2021).

Building on the emerging literature on difference-in-differences (DiD) models with staggered roll-out and heterogeneous treatment effects (see the summaries by Roth et al. 2022; de Chaisemartin and D’Haultfoeuille 2022b), we show that event studies with childbirth as treatment are prone to yield estimates which do not necessarily reflect the actual impact of childbirth. The underlying problems are typically coined as “forbidden comparisons” and “contamination”. “Forbidden comparisons” mean that observations from already-treated units are included in the control group to proxy for a state without treatment. Applied to the childbirth context, “forbidden comparisons” happen when the outcomes of mothers who have already given birth end up as part of the counterfactual trend. Even though they are experiencing the costs of motherhood, they are used as counterfactual as if they had no child. The second problem of “contamination” only applies to multi-period settings such as event studies. It means that the estimate for each relative time period around the event can contain not only the treatment effect at this specific period but also be “contaminated” by the treatment effects from all other periods.

The consequence of both problems for conventional child penalty event studies is that estimates are likely to be biased and pre-trends might be not informative about the plausibility of the parallel trend assumption. Using German administrative data, we illustrate how both issues materialize in the case of child penalty estimation and lead to a substantial underestimation of earnings losses, as the comparisons with already-mothers receive very large weights. We also discuss how in other settings, these problems can lead both to an over- or underestimation of child penalties, depending on the degree and pattern of heterogeneity and on the composition of a given sample.

These problems of DiD models with staggered roll-out have been addressed by newly developed estimators that are robust to effect heterogeneity (see, among others, Sun and Abraham 2021; Callaway and Sant’Anna 2021; Borusyak, Jaravel, and Spiess 2022; de Chaisemartin and D’Haultfoeuille 2020). Applying them, however, requires to employ specific control groups: either units that are the last to receive treatment, that have not yet received treatment or that are never treated. Given the differences among mothers by their age at first birth and the selection into having children, the validity of these control groups hinges on assumptions that are unlikely to hold.

Therefore, we introduce a new approach to estimate the labor market costs of motherhood that overcomes the issues induced by heterogeneity and ensures clean and valid control groups. We propose to use a “stacked” DiD design that estimates the effects of childbirth separately for each cohort of mothers (i.e. for each age at first childbirth), and thus allows for effect heterogeneity and avoids “forbidden” comparisons and “contamination”. To ensure comparability of treated and control units, we impose an additional restriction on the control groups that are specific for each cohort. For each group of mothers who give birth at a given age, we construct a control group exclusively from pre-birth observations of not-yet-treated mothers who give birth at slightly older ages. With this approach, we exploit the strong correlation between age at first birth and labor

market outcomes and bring together the most comparable mothers in the treated and control groups. Our approach builds on work by Cengiz et al. (2019) who propose “stacking” and Callaway and Sant’Anna (2021) who provide a solution to use not-yet-treated units as a control group and avoid forbidden comparisons (though, without making further restrictions that are necessary for the setting of motherhood effects). This combination of a stacked DiD with a rolling window of control cohorts enables us to estimate the unbiased cohort-specific effects of childbirth on post-birth labor market outcomes.

We apply our approach to administrative data from Germany to assess the magnitude of the bias and provide new insights about career costs of motherhood with respect to maternal age at birth. First, we document that the average earnings losses of mothers are substantially larger compared to the conventional approach. The difference in absolute earnings losses is around Euro 7 K or 38 percent; the difference in relative earnings losses amounts to 15 percentage points or 20 percent. We show that the underestimation in conventional event studies primarily stems from their inability to accurately capture the unrealized growth in earnings that would have happened absent children. Since a control group in such event studies includes already-treated mothers, who experience income losses or a slowdown in career progression, the counterfactual trend is biased downwards.

Second, we provide new evidence on how the costs of motherhood and their interpretation differ in maternal age at birth. When estimating cohort-specific effects, we find that absolute earnings losses after childbirth almost linearly increase in age at first birth. We further investigate the sources of this heterogeneity pattern and identify two major opposing factors at play. On the one hand, the pre-birth levels of earnings almost linearly increase until the age of 30 and flatten out thereafter, following the well-documented wage path over the life cycle (Bagger et al. 2014). This implies that leaving the labor market or reducing working hours becomes increasingly costly for older first-time mothers. On the other hand, younger first-time mothers have steeper counterfactual trends due to their control groups of not-yet-mothers actively progressing with their careers and skill development at the early career stage. When we consider occupational rank, task complexity, and educational attainment as outcomes of interest, we observe significant negative effects for younger first-time mothers, primarily manifesting as foregone career progression and skill development. Older first-time mothers experience much smaller negative effects relative to their pre-birth levels, and their losses are composed to a large extent from downgrading after birth rather than foregone growth. Overall, our analysis underscores that younger first-time mothers, particularly those giving birth before the age of 26, face larger career costs of motherhood as they miss out on a phase marked by the most rapid career progression and human capital accumulation.

Our paper makes several contributions. First, we contribute to the large literature assessing the effects of motherhood on women’s careers using event studies and related methods (Angelov, Johansson, and Lindahl 2016; Kuziemko et al. 2018; Bütikhofer, Jensen, and Salvanes 2018; Fitzenberger,

Sommerfeld, and Steffes 2013; Kleven, Landais, Posch, et al. 2019; Bruns 2019; Andresen and Nix 2022; Fitzenberger and Seidlitz 2023). We underscore the importance of considering maternal age at first birth for both methodological correctness and a deeper understanding of the labor market costs associated with motherhood. We start by explaining how biases emerge when heterogeneity by age at birth is not addressed. We then develop a new approach to estimate unbiased effects of childbirth and document that it yields substantially different results than those obtained through conventional event studies. Furthermore, we use this new approach to show that the post-birth losses in labor market outcomes and their interpretation differ substantially by the age at which women give birth. Our new approach can be used to analyze impacts of childbirth, gender earnings inequality as well as the effects of policy reforms. For the latter, it can be particularly useful to assess effects separately for women becoming mothers at different stages of life, since family and labor market policies can have distinct effects by maternal age at first birth, rendering average effects uninformative.

Secondly, we make a contribution to the recently emerging body of literature concerning DiD models within the context of staggered roll-out and heterogeneous treatment effects. Estimation of career costs of children turns out to be a case that is common in the empirical literature and of high policy relevance but where substantial issues materialize from heterogeneous treatment effects as discussed by Goodman-Bacon (2021), de Chaisemartin and D’Haultfoeuille (2020), Sun and Abraham (2021), and Callaway and Sant’Anna (2021) (see also the summaries by de Chaisemartin and D’Haultfoeuille 2022a; Roth et al. 2022). We provide a detailed illustration of how the biases manifest due to the heterogeneous nature of motherhood effects, following the decomposition of effects and weights as proposed by Goodman-Bacon (2021) and Sun and Abraham (2021). Building on the existing solutions to account for heterogeneity, we propose a new approach to estimate the labor market costs of motherhood that addresses the peculiarities of the childbirth setting.

The paper is organized as follows. Section 2 gives a brief overview of the datasets we use, Section 3 provides an overview of the heterogeneity of outcomes and characteristics among different cohorts of mothers, Section 4 discusses the issues with the conventional approach to child penalty estimation. In Section 5, we suggest a new solution to estimate child penalties and apply it in Section 6, Section 7 concludes.

2 Data

This paper uses survey and administrative data from Germany as both types of data have their strengths that complement each other. The survey data from the German Socio-Economic Panel (SOEP) provide the greater level of detail and more characteristics while the Sample of Integrated Labor Market Biographies (SIAB) provides the larger sample size along with the precision of administrative data. This section describes both datasets.

The German Socio-Economic Panel (SOEP) is a well-established panel study that started in 1984 (Goebel et al. 2019) and surveys around 12,000 households and their members each year. Along with detailed socio-demographic information it provides data on labor force status, labor earnings, working hours, occupations as well as on the household context of mothers. Importantly, it also records full birth histories that allow to identify mothers and when they have given birth. We use the SOEP data from the period 1984 until 2020 to provide a systematic overview of the heterogeneity in outcomes and characteristics for mothers by age at childbirth in section 3 and for some complementary analyses.

The Sample of Integrated Labor Market Biographies (SIAB, Frodermann et al. 2021) is provided by the Institute for Employment Research (IAB). It is a two percent sample drawn from the universe of German workers who are subject to social security contributions (i.e. individuals in self-employment and civil servants are not covered). It includes administrative records of individual labor market biographies of nearly 120,000 mothers for the period 1975 to 2019. It further provides information on employers, occupations and wages. From the latter we construct annual earnings. Since the data are taken from employers' reports to the social security system they have some shortcomings. The main two of them are that, first, wages are only recorded up to the threshold for social security contributions. For wages above that ceiling we apply an imputation method that follows Dauth and Eppelsheimer (2020) who build on work by Dustmann, Ludsteck, and Schönberg (2009) and Card, Heining, and Kline (2013). Second, births cannot be observed directly but have to be imputed following Müller, Filser, and Frodermann (2022). This imputation utilizes the maternity protection period around childbirth that mandates an employment break of at least 14 weeks. Müller, Filser, and Frodermann show that their method identifies around 60 percent of all births in Germany. Since it is applied to smaller subset of births by women in employment who are subject to social security contributions the share of identified births in our sample will be larger. We use this dataset to illustrate "forbidden" comparisons in section 4 and in the application of the new approach to estimate child penalties which we propose in section 5.

Tables A.1 and A.2 in the Appendix provide summary statistics for both datasets. With respect to the key characteristic age at first childbirth the data from the SOEP (29 years) and the SIAB (28.6 years) give virtually identical values. Similarly, the values for earnings in the pre-birth year are close together (Euro 26.1 K for the SOEP and Euro 23.9 K² for the SIAB). The finding of larger earnings in the SOEP data is in line with expectations as the SOEP also includes women in self-employment and civil servants, i.e. two groups who are typically farther up in the earnings distribution.

We use both datasets according to their respective strengths. As the SOEP provides more detailed information on mothers' characteristics, we primarily use it to illustrate how educational levels, occupational rank or working hours of mothers differ by their age at first birth (in most applications

² All monetary values are in real terms for the base year 2015.

grouped into quartiles of age at first birth). Leveraging the larger number of observations in the SIAB allows us to precisely show how the effect of childbirth on maternal earnings changes with increasing age at first birth and to conduct a cohort-specific analysis of child penalties. The findings of this paper are not specific to either dataset. They are rather driven by the fact that mothers exhibit substantial differences in various characteristics depending on their age at the first childbirth (see next section). Our results, including those in section 5 where we present an alternative method to assess post-birth earnings losses, are qualitatively and quantitatively similar in both datasets.

3 The Source of Problems: Heterogeneity by Age at Birth

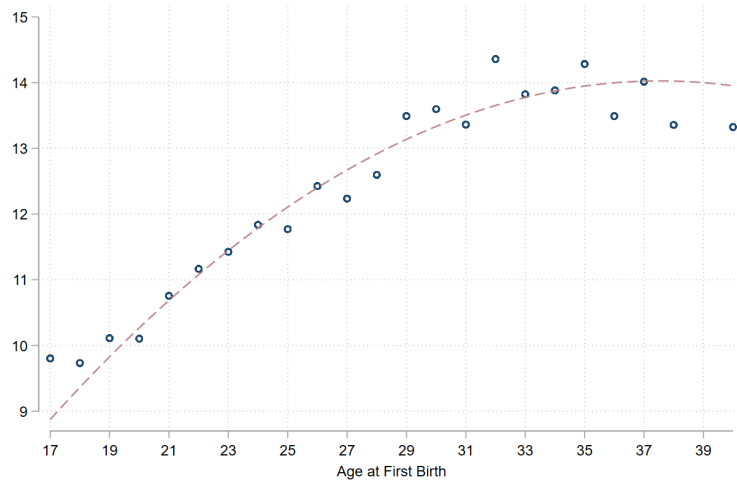
As our paper largely builds on the observation that the post-birth losses are heterogeneous with respect to timing of birth, this section provides detailed descriptive evidence on heterogeneity of mothers' outcomes depending on their age at birth and dynamics of losses over time since birth. Different types of heterogeneity between younger and older mothers and childless women have been mentioned in many economic papers (Goldin, Kerr, and Olivetti 2022; Wilde, Batchelder, and Ellwood 2010; Adda, Dustmann, and Stevens 2017). Nevertheless, we provide a systematic analysis of the across-cohorts differences in outcomes and relevant covariates to be able to fully support our reasoning on their consequences for the estimation of child penalties.

To start with, we show that older mothers tend to have higher levels of education (Figure 1a) – the outcome which is usually decided on during the early stages of career paths and is related to desired fertility (Adda, Dustmann, and Stevens 2017; Doepke et al. 2022). We observe that first-time mothers who are older than 30 on average have completed higher education, while women who become mothers before 25 are more likely to only hold a high school degree.

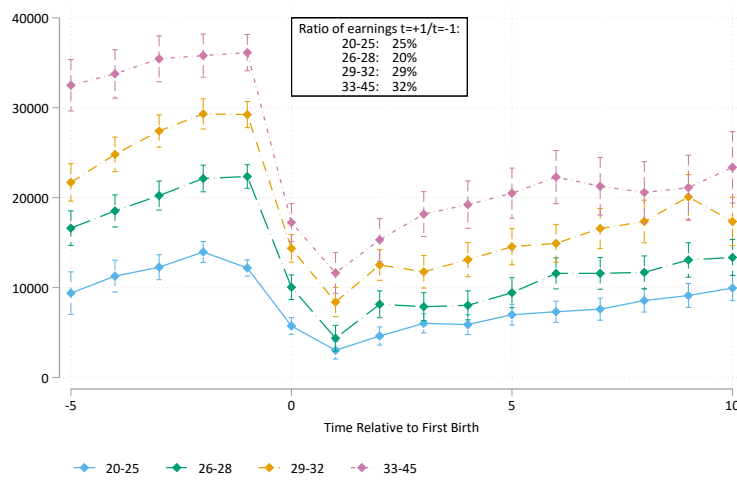
Furthermore, older mothers have on average a smaller number of children over life than early ones (Figure A.5 in the Appendix). First-time mothers younger than 25 tend to have more than two children, while those older than 35 are more likely to have one child in total. This also means that interpretation of effects estimated around the first childbirth differs for these groups of mothers as the estimates for the younger cohorts also capture the effect of having additional children.

With respect to labor market outcomes, we document in Figure A.3 in the Appendix that older first-time mothers work more hours and in positions with higher occupational rank, both before and after the childbirth, and are more likely to return to the labor market after becoming mothers. Women who delay the timing of their first birth also tend to have a shorter parental leave break, returning to work faster (see Figure A.1 in the Appendix).

The heterogeneity described above translates into mothers having different earnings trajectories depending on their age at first childbirth. As shown in Figure 1b and Figure A.2 in the Appendix, older mothers tend to have higher pre- and post-birth levels of earnings, larger magnitudes of drops



(A) Average total years of education of mothers by their age at first childbirth.



(B) Average earnings of mothers around birth for four quantiles of age at first birth.

FIGURE 1: Heterogeneity in education and earnings among mothers by their age at first birth.

Notes: Panel A shows years of education of mothers by their age at first childbirth as a binned scatter plot with an added quadratic fit. Panel B plots average annual labor earnings of mothers (including zero earnings) in time relative to their first birth for four quantiles of the distribution of age at first birth. The first quartile includes mothers aged 20–25 at first birth, the second those age 26–28, the third those aged 29–32 and the fourth those from 33–45.

Source: Own calculations based on the SOEP.

in both absolute and percentage terms, and faster recovery growth rates of their post-birth earnings. Noteworthy, the losses change over time since birth for every cohort. These differences also hold if we restrict the sample only to those women who continue to work after the childbirth (Figure A.3 in the Appendix). These results are in line with Wilde, Batchelder, and Ellwood (2010) who show

similar correlations between wages of mothers and childbirth timing. To sum up, we observe that, on average, older first-time mothers exhibit significantly better labor market outcomes and levels of human capital than younger ones, both before and after the childbirth.

Another important observation about the earnings trajectories discussed above is that post-birth losses change over time since birth for all groups of mothers. As can be seen in Figure 1b, all cohorts of mothers experience large drops in the year of birth, smaller continuing losses in the year just after birth, some recovery starting in second year, and steady growth thereafter. This dynamic is likely to be explained by maternity leave policies, with women returning to the labor market within the job protection period (which is currently 3 years for Germany). Since many European countries have maternity leave policies, such dynamics of post-birth earnings, with first realization of losses and then of some recovery, is prone to be universal. The magnitude of losses, speed of recovery and post-birth growth rates may vary depending on the setting.

Overall, these results provide a suggestive evidence that effects of motherhood are potentially heterogeneous depending on age at which women give birth to their first child and that the effects are likely to change over time since birth. As we show in the following section, these differences become the source of biases when estimating child penalties using event study regressions.

4 Child Penalty Estimation under Heterogeneity by Age at Birth

In this section, we explain how heterogeneity by age at birth and over time since birth poses a threat to estimating child penalties, since age is not just one of many characteristics of mothers but also a timing dimension. As Goodman-Bacon (2021) and Sun and Abraham (2021) have shown, if treatment effects differ across cohorts or if they just change over time, these differences enter the DiD estimates and bias the results. First, we show how these issues materialize in child penalty estimations and lead to significant biases. Second, we discuss how the heterogeneity of treatment effects invalidates the commonly-used control groups and limits the applicability of heterogeneity-robust estimators to the settings with childbirth as treatment.

4.1 The conventional approach to estimate child penalties

We start with presenting the common event study setup on which existing child penalty estimations are based. Child penalties aim to quantify and visualize the losses women experience with regard to some outcome – typically earnings – following the birth of their first child. Their estimation usually starts with an event study regressing the outcome on a set of event time dummies and additional control variables. A next step that re-scales the event study coefficients to get percentage changes is not strictly necessary but common.

The conventional specification proposed by Kleven, Landais, and Sogaard (2019) (henceforth KLS) looks as follows:

$$Y_{it} = \sum_{l, l \neq -1} \beta_l \times \mathbb{1}[t - t_i^0 = l] + \gamma_{it} + \lambda_t + \varepsilon_{it}. \quad (1)$$

Here, Y_{it} is the outcome of interest for unit i at calendar time t , t_i^0 is the time when unit i gives the first childbirth. Fixed effects for age (γ_{it}) and calendar time (λ_t) are added to control for life cycle as well as business cycle effects. The coefficients of the relative event time dummies ($\hat{\beta}_l$) then are intended to capture the effect of being l years away from the year of the first childbirth t_i^0 on the outcome of interest. This outcome of interest can be earnings – the one that gains most attention and that most of this paper focuses on – but other continuous (such as wages or working hours) or discrete (for instance, employment status) variables are used as well. In many studies, additional individual fixed effects are added to transform it to a DiD model and account for the heterogeneity among mothers that is constant over time.

4.2 Child Penalty Estimation as a Case of *Forbidden Comparisons and Contamination*

As discussed, among others, by Goodman-Bacon (2021), Sun and Abraham (2021) and de Chaisemartin and D’Haultfoeuille (2020), heterogeneity of effects for cohorts treated at different points in time can lead to several issues in the setting of event studies, on which child penalty estimations are based. According to the literature, there are two main sources of heterogeneity-induced biases. The first one are “forbidden” comparisons, in which already-treated units are used as controls, and their post-birth changes in earnings are used as counterfactual trends. Event studies do such comparisons and assign them the weights which depend on sample variance and composition, such that the resulting average can even lie outside of the range of actual effects. Second, in dynamic settings, where effects for multiple periods are estimated, estimates for one relative time period can be “contaminated” by effects from other periods. These two problems lead to uninformative pre-trends and biased estimates of treatment effects. We argue that both are likely to arise when using event studies to estimate child penalties.

“Forbidden” comparisons “Forbidden” comparisons are made when a DiD estimator compares units at non-matching points in time. To explain their origin and consequences, we take a step back and consider a static DiD model where only an average treatment effect is of interest. Goodman-Bacon (2021) has shown that a DiD estimator is a variance-weighted average of all possible 2x2 DiD estimators that compare cohorts to each other, i.e. compares each treated group to all other already-treated and non-treated groups. If effects are homogeneous – exactly the same across cohorts and constant over time – the differences between the trends of a treated and the already-treated groups

will be zero in each period. Then, the estimate will be unbiased. However, if effects change over time or if they are different across cohorts, the differential trends of the already-treated cohorts will be used as counterfactuals and introduce biases.

What does it mean for child penalty estimation? To answer this question, we use a stylized example and illustrate “forbidden” comparisons building on the Goodman-Bacon (2021) decomposition. We take annual earnings as an outcome and the SIAB data (see Section 2) for this illustration. As simplification, we focus on one treated cohort that gives birth at age 29 and are interested in estimating the effect of childbirth for this cohort for the year 0 (at age 29) relative to the pre-birth year (at age 28). For illustration purposes, we restrict the sample to one much earlier-treated cohort (that gave birth at age 24), one cohort that gave birth in the previous year at age 28 and one next-treated cohort that will give birth at age 30. The average earnings for these four cohorts over age are plotted in Figure 2. It shows that all cohorts exhibit similar earnings trajectories around childbirth: steady growth before birth, large losses in period 0, additional smaller losses in period +1, some recovery in the year +2 (likely since mothers start re-entering the labor market after maternity leave), and slow growth thereafter.

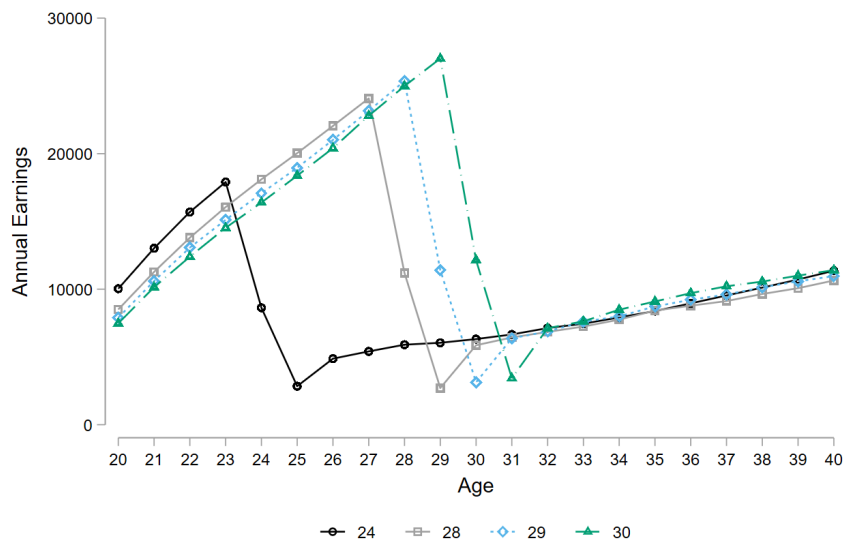


FIGURE 2: Average earnings of mothers around birth for four ages at first birth (24, 28, 29, 30)

Notes: The figure plots average annual labor earnings of mothers (including zero earnings for periods of non-participation) by age for four levels of age at first birth (24, 28, 29, 30). Source: Own calculations based on the SIAB.

Formally, estimating the treatment effect in year 0 for the chosen treated cohort (giving birth at age 29) means estimating the following regression on a sample that is restricted to the cohorts that give birth at ages 24, 28, 29 and 30 and the age window from 28 to 29:

$$Y_{it} = \beta \times \text{treated}_{it} + \gamma_i + \lambda_t + \varepsilon_{it}. \quad (2)$$

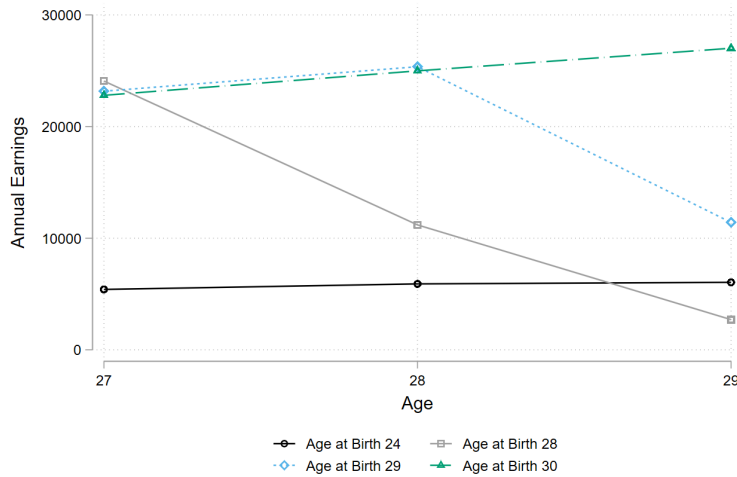
Here, Y_{it} indicates annual earnings for mother i at age t . Fixed effects for individual (γ_i) and age (λ_t) are included to implement the DiD design, accounting for pre-birth differences between treatment and control groups and the changes in outcomes of the control group over time. Treated_{it} is a treatment status indicator, which takes value of 0 if the individual is not-treated yet, switches to 1 when the treatment happens and stays 1 thereafter. The estimate $\hat{\beta}$ captures the change in the outcome for the treated group, compared to changes which the control group experiences. Estimating this regression using the SIAB sample and the restrictions described above yields the average estimate of Euro $-11,562$ (see Table 1, column “Average”).

However, decomposition of this average DiD estimate shows it is a weighted-average of three comparisons. First, the cohort giving birth at age 29 is compared to one not-yet-treated cohort that is going to give birth at age 30, which is a “clean” comparison. In addition, the cohort giving birth at age 29 is compared to the already-treated cohorts who gave birth at ages 24 and 28. As the latter two cohorts are already treated and experience effects of motherhood, these are “forbidden” comparisons.

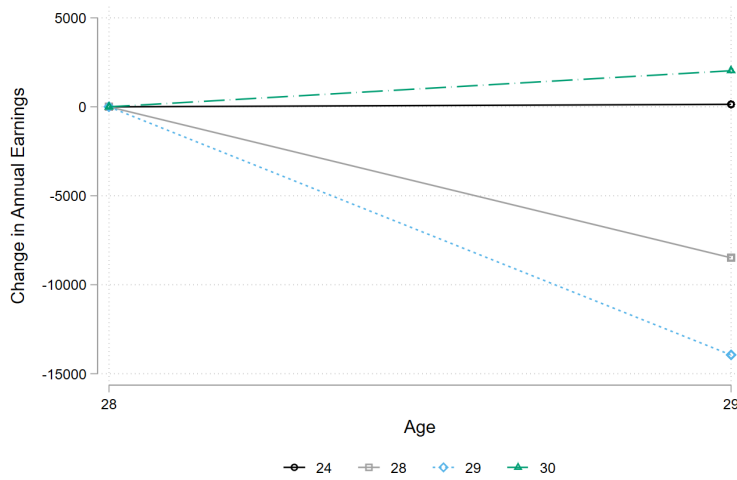
To see why this happens, we firstly look at the average annual earnings of these four cohorts over the estimation window in Figure 3a. We observe that the treated cohort (blue line) experiences growth until the age of 28 and then large losses in the year of childbirth (at the age of 29). The not-yet-treated cohort that gives birth at 30 (green line) experiences a steady increase in earnings. The much earlier-treated cohort that gave birth at age 24 (black line) exhibits only very slow earnings growth at age levels 28 and 29. The previously-treated cohort that gave birth at 28 (grey line) continues to face losses at the age of 29, although they become smaller in year after birth compared to the year of childbirth. The comparison that researchers usually intend to make is the one between the treated and the not-treated group – in the figure, between the blue and green lines. After the pre-birth differences in levels are taken into account, the trend of the not-yet-treated group is assumed to reflect a counterfactual development of earnings absent children³. However, this is not the only comparison that the regression makes.

The main issue is that the regression treats the cohorts that already gave birth at 24 and 28 as control groups, because the treatment dummy (treated_{it}) does not change for them and has the value of 1 in both periods. Inclusion of group (or individual) fixed effects is equivalent to demeaning

³ We discuss using not-yet-treated mothers as control group later in this section.



(A) Average earnings in levels for age range 27 to 29.



(B) Changes in average earnings for age range 28 to 29.

FIGURE 3: Average earnings in levels and changes for four ages at first childbirth (24, 28, 29, 30).

Notes: The figure plots average labor earnings for mothers who give birth at age 24, 28, 29 and 30. The upper panel plots earnings in levels, the lower panel plots the changes in earnings between age 28 and 29. Earnings include zero earnings in periods of non-participation. *Source:* Own calculations based on the SIAB.

of variables' values within the groups, which in our setting translates into subtracting the mean of treatment dummy for already-treated cohorts at age 28 and making them equal 0 in both periods, as if they are control units. The coefficient from the treatment dummy ($\hat{\beta}$) then captures the response in the outcome to the change in treatment status by 1 unit, which at the age of 29 happens only

for the cohort that gives birth at 29. All the other cohorts, for which the treatment status does not change, serve as control groups.

In our example, the estimator makes three comparisons that we illustrate in Figure 3b (Table 1 provides the according estimates). There, we plot the changes in earnings from age 28 to age 29, normalized to the pre-birth period (age 28). The “clean” comparison is the one where the changes in earnings of the treated cohort and the not-yet-treated one are compared (in our example the difference equals Euro $-15,976$). The forbidden comparisons are the other two that use already-treated mothers as control units. The cohort that gave birth in the previous year (at age 28) is still experiencing losses in its first post-birth year. However, the losses are not constant over time. At age 29, they are smaller than in the year before when the cohort gave birth. If the losses and thus the earnings trajectories after birth didn’t change over time and were equal to the losses of the group treated at age 29, this comparison would yield a difference of zero and, hence, would not introduce a bias. Instead, since the losses decrease from year 0 to year +1, this comparison yields an estimate of Euro $-5,430$, meaning that it underestimates the earnings losses for the treated cohort. The second “forbidden” comparison is the one with the cohort that gave birth already at age 24. This cohort exhibits a mostly flat trend in its earnings, highlighting the slow earnings growth of mothers from the second post-birth year onward (see Figure 2). Using this trend as counterfactual leads to an underestimation of the treatment effect as well (we estimate Euro $-14,158$).

The weights that each of these three comparisons receives in the average estimate depend on the group size and variance of the treatment dummy in each pair of cohorts, i.e. on parameters specific for a given sample (Goodman-Bacon 2021). In our example, the estimates from the two “forbidden” comparisons receive almost two thirds (66.2 percent) of the weight. It introduces a substantial upward bias into the average DiD estimate which equals Euro $-11,562$, whereas the estimate solely based on the clean comparison yields Euro $-15,975$. The bias is large and equals Euro $4,413$, or 38 percent of the average estimate.

In Figure 4, we extend this example to include all cohorts of mothers that give birth between ages of 23 and 34. In the figure, the blue line indicates the cohort treated at age 29. The orange line captures the losses of the previously treated cohort (at age 28), the set of almost flat trends comes from cohorts treated earlier (before age 27) and the set of upward trends depicts the earnings development of not-yet-treated mothers who give birth at age 30 and above. One additional upward trend – the pink upper line – is coming from the cohort that was treated two periods ago (at age 27) and experiences recovery due to returning from maternity leave (this dynamic can be also seen in Figure 2). Except for this one, all trends that come from the previous and the earlier cohorts lead to an underestimation of the treatment effect. Generally, the estimate of a given cohort in a given period is composed of comparisons of the treated cohort’s changes in earnings to mostly flat post-birth trends of much earlier already-treated cohorts, steep upward trend of the cohort treated

TABLE 1: Decomposition of average estimate: “Clean” and “forbidden” comparisons.

	Average	“Clean” (to 30)	All “Forbidden” (to 24 and 28)	“Forbidden” (to 24)	“Forbidden” (to 28)
Treatment status	-11,562*** (142)	-15,976*** (173)	-9,307*** (160)	-14,158*** (173)	-5,430*** (202)
Age FE (Age=29)	-2,384	2,030	-4,639	212	-8,516
Person FEs	YES	YES	YES	YES	YES
Included cohorts	24, 28, 29, 30	29, 30	24, 28, 29	24, 29	28, 29
Estimation window	28–29	28–29	28–29	28–29	28–29
Weight in average		33.8%	66.2%		
Weight in “forbidden”				44.4%	55.6%

Notes: The table reports the results from estimating the effect of childbirth on average annual labor earnings in the year after birth for the cohort that gives birth at age 29 following Equation (2). The first column shows results for the estimated average effect using the sample that includes four cohorts that give birth at ages 24, 28, 29 and 30. The second column shows the results from “clean” comparisons, when only the cohorts 29 and 30 – treated and not-yet-treated – are included in the sample. The third column shows the result from “forbidden” comparisons, when only the cohorts 24, 28, and 29 – already-treated and treated – are included in the sample. The fourth and fifth column show the coefficients separately for each of the two “forbidden” comparisons. Standard errors are reported in parentheses. *** indicates statistical significance at the 1 percent level. *Source:* Own estimation using the SIAB.

two periods before, comparably less steep downward trend from the cohort treated just before and to upward sloping trends of not-yet treated mothers. Repeating the decomposition of weights by Goodman-Bacon (2021) for this example reveals that the “forbidden” comparisons receive 60 percent of weights. Noteworthy, for different outcomes and samples the bias can go in any direction and weights can change, since they depend on patterns in pre- and post-birth trends and characteristics of the sample at hand.

“Contamination” In common applications, dynamic treatment effects rather than an average one are of interest, which, as we discuss next, introduces additional issues. The general finding by Goodman-Bacon (2021) is that in the case of a static DiD with average treatment effect, the estimates can be biased as they are a weighted average of “clean” and “forbidden” comparisons if multiple groups are treated at different points in time and treatment effects are heterogeneous among these groups. As shown by a growing number of recent papers, the same as well as additional problems apply to dynamic DiD models, that move from the static case to having dynamic treatment effects (reviewed by de Chaisemartin and D’Haultfoeuille 2022b; Roth et al. 2022).

To apply the insights from the TWFE literature to the case of child penalties, we, first, have to note which one is the relevant dimension of treatment timing. The literature on TWFE models assumes treatment effect heterogeneity among cohorts defined by *calendar time*. When looking

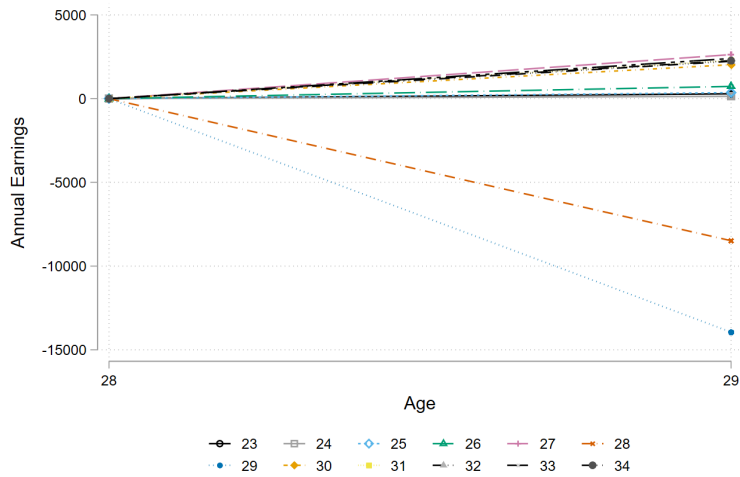


FIGURE 4: Changes in earnings of mothers from age 28 to 29 for all ages at birth (23-34)

Notes: The figure plots average annual labor earnings of mothers (including zero earnings) adjusted to pre-birth (at age 28) differences between cohorts, over ages 28-29 for all ages at birth between 23 and 34. Source: Own calculations based on the SIAB.

at childbirth as treatment, however, *age at first birth* is the more important dimension. As shown in the previous section, age at first birth is highly correlated with several characteristics relevant for the labor market as well as with the absolute and relative earnings losses after childbirth. In the following, we therefore use age at first birth as the variable to define cohorts of mothers. This means in practice, that we often observe relatively small differences in the effects of childbirth when comparing, for instance, mothers of the same age giving birth in the years 2000 and 2015, but we find much larger differences between mothers who give birth at ages 20 and 35 within the same calendar year. In the Figure A.6 we plot average absolute changes in earnings in the post-birth year relative to the pre-birth one over calendar time and age at birth. We observe that average losses in post-birth year remain quite stable over the years 1995-2010, while they vary much more over age at first birth. The standard deviation for the losses over age at birth is almost 8 times larger than one over calendar time. Overall, heterogeneity of losses is much more pronounced across ages at birth rather than across calendar years.

With this change in the cohort definition, we can apply the analysis by Sun and Abraham (2021) to the case of child penalties. Their setting considers event study regressions of a single, binary and absorbing treatment (i.e. where treatment status can only change from 0 to 1 and stays 1 thereafter) with a staggered roll-out. This treatment description exactly matches the birth of the first child as

an event that appears once and has consequences for the entire time thereafter.⁴ The set of treated units can be divided into cohorts based on when they receive treatment. This setup can be treated as equivalent to KLS’s specification in (i) with added individual fixed effects. As pointed out above, age at first birth now serves as the variable grouping individuals into cohorts based on when they receive treatment.⁵

Sun and Abraham (2021) decompose the point estimates $\hat{\beta}_l$ from TWFE regressions (similar to the one in Equation (i) with added person fixed effects) and show that for each relative time period l they consist of two parts. The first one is a weighted average of the cohort specific treatment effects in period l , which is similar to the decomposition of the static DiD from the previous illustration and can be biased if there is heterogeneity of effects across cohorts. The second part is again a weighted average, but of the treatment effects from the *other periods* l' . This part is only in place in the dynamic DiD setting and commonly referred to as “contamination”. When the single treatment dummy is replaced with a set of dummies for each relative time period around treatment interacted with treatment status, the treatment effect for a given period is estimated *conditioning on effects from other periods*. Intuitively, it means making an extra adjustment to the trends in Figure 3b. In addition to shifting the trends by pre-birth differences, estimation in dynamic setting includes subtracting effects from outcomes of already-treated cohorts specific to their relative time period at a given age. If the treatment effects are identical across cohorts, the average will reflect an unbiased estimate, since the trends of already-treated cohorts will be identical to the “clean” ones. If, however, treatment effects vary across cohorts, subtracting an average effect will not take this heterogeneity into account and, hence, introduce an additional bias.

Sun and Abraham show that if there is heterogeneity in treatment effects across cohorts, then other periods will enter the estimation of a coefficient for a given period with non-zero weights and “contaminate” it. In Figure A.7 in the Appendix, we provide the evidence that “contamination” is likely to be in place when estimating child penalties. We plot the weights obtained from the decomposition Sun and Abraham propose. Both for a pre-birth (−4) as well as a post-birth (+5) time period we find that estimates from other relative time periods (on the x-axis) will receive non-zero weights (on the y-axis), if effects are heterogeneous. It is especially of note that the estimate for year +5 after childbirth can include effects from pre-birth time periods. This decomposition confirms that under effect heterogeneity, “contamination” is likely to bias child penalty estimates.

Similar to the static DiD example, there are two main consequences. First, estimated pre-trends are uninformative. If they are flat, this is no clear indication of parallel trends before treatment.

⁴ Note, that we only consider the birth of the *first* child which is both for child penalty estimations and in the TWFE literature the common case. Considering higher order births can introduce additional complications. de Chaisemartin and D’Haultfoeuille (2022a) provide an econometric assessment of the case of multiple treatments.

⁵ For the illustrations in Figures 1b, A.2, A.3 and A.4 we restrict to four cohorts to improve readability. In practice, the number of cohorts is determined by the number of different levels of age at first birth in the data at hand.

Second, especially in the presence of treatment heterogeneity, estimated treatment effects can be biased. It implies that heterogeneity in the effects of childbirth on maternal outcomes can lead to estimates that are neither a numerically correct representation of the actual treatment effect nor an interpretable weighted average of treatment effects of multiple cohorts.

4.3 Validity of Control Groups and Applicability of Heterogeneity-Robust Estimators

A growing number of papers propose new estimators that avoid “contamination” and “forbidden” comparisons and – under certain conditions – obtain estimates that are robust to heterogeneous treatment effects. In the following we discuss the most common of these estimators and the assumptions one has to make to apply them.

As the underlying problem of static and dynamic DiD estimations is the inclusion of treated units in a counterfactual, researchers developed various new estimators that clearly specify their control groups. Generally, three types of control groups are suggested by the DiD literature. Units that are the last ones to receive treatment before they become treated (Sun and Abraham 2021), units that never receive treatment (Sun and Abraham 2021; Callaway and Sant’Anna 2021), and units that are not-yet treated (de Chaisemartin and D’Haultfoeuille 2020; Callaway and Sant’Anna 2021). Estimating child penalties using any of these control groups requires to make strong assumptions that they are suitable counterfactuals for each cohort of mothers. Moreover, heterogeneity among these control groups in the setting of childbirth leads to substantial differences regarding the interpretation of the estimates.

Last-treated units as control group As we show in Section 3, the timing of childbirth is correlated with some basic characteristics of mothers and their labor market outcomes, both before and after birth. It directly follows that those mothers who are the oldest when they have their first child are a selected group. On average, they have a higher socio-economic status, reflected in higher earnings, education levels, occupational ranks and other labor market outcomes. Additionally, they have the shortest post-birth labor market breaks (see Figure A.1 in the Appendix). Therefore, the outcomes of last-treated mothers cannot serve as a counterfactual for earlier-treated cohorts.

Not-yet-treated units as control group The case of using not-yet-treated mothers as a control group is related to the idea of using outcomes of those mothers who are the oldest when they have children as a counterfactual. Some share of the control group consists of later and the last-treated mothers, i.e. older ones. For them, the above reasoning that later childbirth is associated with fundamental differences in pre- and post-birth characteristics and outcomes continues to hold. In their entirety not-yet-treated units are therefore unsuitable as control group.

However, a control group of not-yet-treated units as well consists of mothers who have their child rather early. For instance, for mothers who have their child at some age a , mothers who give birth at age $a + 1$ are part of the control group. In order to use them as a control group, the necessary assumption is that age at childbirth is quasi-randomly allocated within a relatively small bandwidth of age. Even though we show that age at childbirth and a number of characteristics that are relevant for labor market outcomes are correlated, it seems plausible to assume that even women who put effort in planning when to have children cannot precisely manipulate the exact date. Then, childbirth has a random component such that within the span of a few years mothers are comparable. We build on this feature when constructing the control group in Section 5.⁶

Never-treated units as control group Using those who never receive treatment, i.e. never give birth to a child, as the control group implies using childless women or men to obtain the effect of childbirth on mothers.

If one uses men, the estimate will capture the effect of having a child on the outcomes of mothers under the assumptions that men are unaffected by childbirth and that, absent children, women would have the same outcome trajectories as men. In practice, these assumptions are unlikely to hold. First, there are gender differences in career paths because of various kinds of discrimination and different experiences that can have an effect already prior to birth, for instance when choosing a college major. After childbirth, being a father can have both positive and negative effects on outcomes of men. Negative ones can arise in societies that are more gender-equal where the burden of raising children is shared more equally between both parents. Positive effects for fathers are documented as well. Goldin, Kerr, and Olivetti (2022), for instance, show the existence of a “fatherhood premium” that is larger for more time-intensive occupations pointing towards an increase in productivity due to focus on market work (similar to the findings on male marital wage premiums; documented, among others, by Antonovics and Town 2004).⁷ Using the SOEP data, we also document that there is a positive correlation between labor market outcomes of fathers and the timing of parenthood, further showing that they are not unaffected by childbirth (see Figure A.4 in the Appendix).

To use childless women as control units one needs to make the assumption that not having children is a trait that is allocated to women in a quasi-random fashion and is unrelated to their labor market outcomes, such that outcome trajectories for mothers and childless women are similar. This assumption is restrictive and difficult to test.⁸ Further, it does not allow for the voluntary

⁶ This idea is also used by Fitzenberger, Sommerfeld, and Steffes (2013).

⁷ There is the related special case where the control group is constructed from the male partners of mothers (as for instance in Angelov, Johansson, and Lindahl 2016; Andresen and Nix 2022). Here, the within-couple distribution of roles and tasks could introduce an additional source of effect heterogeneity.

⁸ Lundborg, Plug, and Rasmussen (2017) utilize the special case of IVF treatment success to estimated career effects of childbirth conditional on receiving IVF treatment.

decision of being childless for career reasons.⁹ In practice, childless women represent a very heterogeneous group of those who decided to have no children and those who wanted, but could not have them for different reasons.

Both for men and those childless women who know they will stay childless (for instance due to medical reasons) or explicitly plan to do so, there remains the issue of anticipation. Including them in an event study leads to a situation where the control group knows it will never receive treatment while the treatment group at least anticipates it is very likely to be treated at some point in the future (and possibly plans to influence when). Both groups are thus able to make according decisions such that level-differences and trends differ even more between them.

4.4 Re-scaling

The second common step when estimating child penalties is to re-scale the event study estimates from levels to changes in percentages. Percentage changes are easier to interpret and allow straight-forward comparisons of child penalties across different settings such as countries, points in time or policy regimes. KLS's canonical approach re-scales by calculating

$$P_l = \frac{\hat{\beta}_l}{E[\tilde{Y}_{it}|l]}, \quad (3)$$

where \tilde{Y}_{it} is the prediction from the previous regression (as in Equation 1) when omitting the relative event-time dummies $\hat{\beta}_l$. $E[\tilde{Y}_{it}|l]$, the earnings levels that only depend on fixed effects for age and calendar year, are intended to proxy for earnings in a counterfactual state of the world in which a woman does not have children. The child penalty P_l then gives the percentage difference between the earnings of mothers and counterfactual earnings of women without children. Note, that while the event study leads and lags make a comparison relative to the omitted pre-birth time period (typically one year before birth) the denominator changes in relative time around birth, thus introduces a second comparison that is specific for each relative period l .

The re-scaling step introduces an additional source of potential biases as it still contains treatment effects and makes comparisons between units that have been treated at different points in time. When predicting counterfactual earnings based on age and year fixed effects alone, the resulting \tilde{Y}_{it} is not restricted to only use specific observations and therefore consists of observations *both* from the pre- and the post-birth period. Since the share of women who already had their first child increases in age, the composition of the counterfactual in terms of including pre- or post-birth observations changes in age at childbirth as well. Figure 5 illustrates this situation by plotting standard child penalty estimates over -5 to $+10$ years around birth based on Equations (1) and (3) for the four

⁹ Steinhauer (2018) documents interrelations between gender roles, childlessness and choices in the labor market.

quartiles of the distribution of age at first birth. Along with the child penalty, we plot the share of pre-birth observations that is used at each point in relative time to construct the counterfactual earnings.

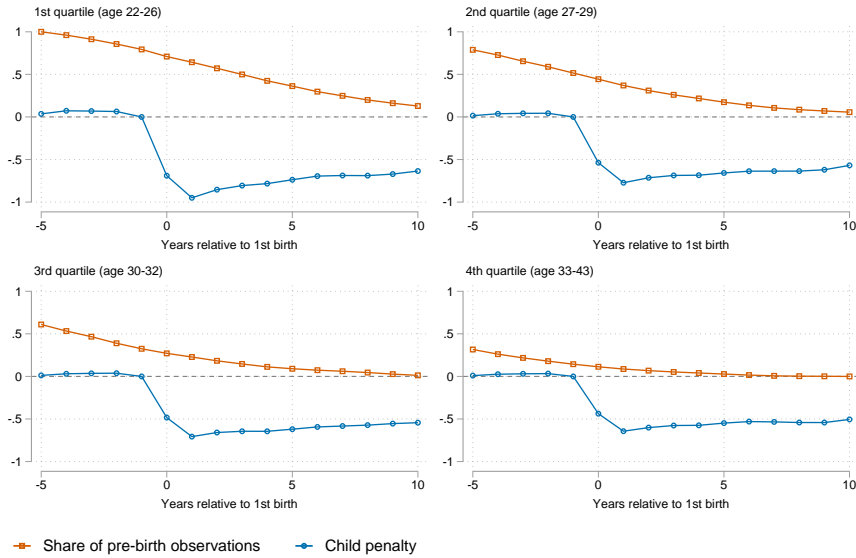


FIGURE 5: Composition of counterfactual earnings $E[\tilde{Y}_{ist}|t]$.

Notes: Child penalties (blue) and share of pre-birth observations (orange) in the counterfactual by time around first childbirth, plotted for quartiles of age at first birth.

Source: Own calculations based on the SOEP.

For younger mothers, the counterfactual earnings are primarily informed by observations of not-yet-mothers while the counterfactual for older mothers mostly consists of observations from women who are already mothers. For the youngest mothers in the first quartile – for whom the largest penalty is estimated – counterfactual earnings in the post-birth period consist to a large share of observations of other mothers *before* they give birth (for the two years directly after birth above 50 percent). For mothers in the fourth quartile, the counterfactual for their post-birth earnings primarily contains observations of other mothers *after* they have given birth (at the time of birth, the share of pre-birth observations is already below 25 percent and further decreases over time to virtually zero in year seven after birth).¹⁰ In other words, for the majority of the sample the counterfactual for post-birth earnings is constructed from post-birth earnings which do not give an adequate depiction for a situation without treatment. This way of constructing counterfactual earnings leads to smaller counterfactuals for older mothers. Along with their earnings generally being higher, this mechanically generates smaller penalty estimates with increasing age at first birth.

¹⁰ Table A.3 in the Appendix lists the shares of available pre-birth observations for five levels of age at first childbirth.

Abstracting from our example in Figure 5 that is estimated using SOEP data, it is straightforward to formalize the relationship between age at first birth and the composition of counterfactual earnings. Assume, age at first birth A is distributed on some interval $[A, \bar{A}]$. The – dataset specific – distribution function is $F_A(a) = P(A \leq a)$ and gives the share of observations that have given birth at or before age a . Therefore, for any age at first birth a and any post-birth year t , $\sigma_{\text{pre}} = 1 - F_A(a + t)$ gives the share of remaining pre-birth observations in the sample that is not influenced by effects of previous childbirth. By definition of a distribution function, σ_{pre} is decreasing in age, highlighting that fewer and fewer suitable observations to construct counterfactual earnings are available when we look at older mothers.

5 The Solution: Stacked DiD with Rolling-Window Control Groups

Having discussed the issues with the common event-study-based approaches to estimate child penalties, we develop a new solution within the DiD framework, which takes into account heterogeneity of treatment effects by age at first childbirth, employs a valid control group and – under certain assumptions – is able to bring the results closer to an estimate of the causal effects of motherhood on labor market outcomes. In this section, we present the new approach along with the required assumptions.

We suggest to combine the concepts of the “stacked” DiD design (Cengiz et al. 2019) and the Callaway and Sant’Anna (2021) estimator with an additional restriction imposed on the control group. The main idea behind any DiD design is that the counterfactual trend is borrowed from a control group and that level differences in outcomes between treated and control groups are taken into account. Identification within a DiD design then builds on the general comparability of treatment and control group, which translates into similar outcome trajectories absent treatment. To ensure comparability of both groups in the setting of a child penalty estimation, we propose a “stacked” DiD design combined with a rolling window of cohort-specific control groups over age at birth. Specifically, for each treated cohort a control group will consist of the pre-birth observations from only the closest (in terms of age at birth) not-yet-treated cohorts. Due to the strong correlation between age at birth and labor market outcomes (see Section 3), this approach brings the most comparable treated and control mothers together. By creating such valid and “clean” control groups for each treated cohort and “stacking” the cohorts with their control groups by event time, we are able to estimate unbiased cohort-specific treatment effects and avoid making “forbidden” comparisons.

Practically, we apply the following procedure. First, we define an estimation window of relative time around childbirth that is the same for each cohort. The number of post-treatment periods is determined by the number of control cohorts (N_c) that are included in each control group. This

number is based on an assumption of how long mothers in the control groups who give birth at an older age remain comparable. We treat childbirth for each cohort as separate sub-events and construct unique control groups consisting of only pre-birth observations from cohorts that give birth at age $[a + 1, a + N_{cc}]$ and fall in the same estimation window. We then estimate the dynamic DiD model in the form of a TWFE regression:

$$Y_{ias} = \sum_{l=-L_{min}, l \neq -1}^{L_{max}} \beta_l^s \times \mathbb{1}[a - a_i^0 = l] \times \mathbb{1}[a_i^0 = s] + \gamma_{as} + \lambda_{is} + \varepsilon_{ias}. \quad (4)$$

In the equation, Y_{ias} indicates the outcome of mother i who belongs to cohort s at age a . Contrary to the conventional approach that focuses on calendar time, we treat age as the relevant time dimension.¹¹ Accordingly, the model includes a set of event-time indicators β_l^s that identify when a mother is l years away from her first childbirth at age a_i^0 . The indicator function $\mathbb{1}[a_i^0 = s]$ identifies each cohort of mothers along with the assigned control units, i.e. it allows the event-time indicators to vary by sub-event. The fixed effects for age and individual, γ_{ts} and λ_{is} , are allowed to vary by sub-event as well (we omit additional indicator functions in the equation for readability). This model then estimates the cohort-specific effects β_l^s on the outcome of interest over the estimation window from L_{min} to L_{max} . Since it allows treatment effects and fixed effects to differ across sub-events (i.e. childbirth for each cohort), it is equivalent to running separate TWFE regressions for each sub-event. We cluster the standard errors at the level of the individual to account for correlation of the error term over time and across observations of the same mothers that are used as controls in multiple sub-events.

Most commonly, the outcome Y is used in absolute terms which allows to keep zero earnings during times of non-participation in the sample. To show percentage changes that are often more informative, an additional transformation is required. As we discuss in Section 4.4, re-scaling by a counterfactual composed of average age and year fixed effects is problematic. More generally, any transformation involves a choice made by the researcher and different transformations can lead to different results. Depending on the outcome, we suggest one of the following. The first transformations explicitly calculates counterfactual outcomes. Separately for each cohort and event-time, we take the average outcomes of the control group (i.e. women who thus-far have not given birth) and adjust them to the differences in the pre-birth levels between the treated and control groups. This removes potential level-differences between both groups and assumes that mothers, in the counterfactual state without childbirth, follow the trajectory of the control group not-yet but soon-to-be treated women. Dividing the estimates $\hat{\beta}_l^s$ by the counterfactual gives the penalty of having a child relative to not having a child l years after childbirth. This approach stays close to the established notion of the child penalty as it adds an event-time-specific comparison to the

¹¹ See also Section 4.2.

comparison relative to a baseline that the event study estimates. A second possibility is to divide the estimates $\hat{\beta}_l^s$ by the cohort-specific pre-birth level of the outcome. Sticking to the pre-birth period as reference point is more suitable to highlight the interruption of a trend (i.e. a mothers' career) by childbirth. We apply and discuss both transformations in the following section.

Even though the cohort-specific estimates discussed so far provide additional information, it is often of interest to calculate an average effect over all cohorts. To this end, we follow Sun and Abraham (2021) and weight the cohort-specific estimates by the sample shares of each cohort:

$$\hat{\beta}_l = \sum_{s=S_{min}}^{S_{max}} \frac{N_s}{N} \times \hat{\beta}_l^s, \quad (5)$$

where N_s indicates the number of observations per cohort and N the total number of observations. The result is a weighted average estimate over all cohorts.

The main assumption behind this “stacked” DiD design is that absent treatment, the trends of the treatment and control groups are parallel for each cohort (i.e. sub-event). The similarity of outcome trajectories before treatment is testable by plotting the cohort-specific pre-trends. The comparability of the post-treatment trajectories absent children has to be assumed. Intuitively, this means for our setting to assume that women, who give birth at a certain age, absent children, would have followed the same outcome trajectory as those women, who give birth a few years later, have at the same age. The strong correlation of age at birth with both outcomes and relevant covariates provides the foundation for this assumption, because we bring together treated and control units from a narrow window of ages at birth. Age at first birth is arguably a better proxy for career trajectories than, for instance, education as it combines information on the levels of education and the pre-birth realized career.

Since we do not employ all not-yet-treated cohorts, but a smaller subset of them as a unique control group for each cohort, it is now more straightforward to assume that timing of childbirth is quasi-random within the span of the few years covered by the control cohorts. Thus, the restricted control group allows to interpret our estimates as the causal effects of becoming a mother compared to delaying childbirth. The intuitive reasoning behind this assumption is that, even though mothers can select into a time range during which they give birth, it is substantially less likely that they can precisely choose at which age they give birth. Having a general plan to have a child rather early or late in life can be a common trait among groups of women, while ensuring that a child is born in a specific year hinges on a number of factors that are not entirely within a woman's control.

A particular limitation of our approach is that it allows to estimate the costs of having a child only within a medium-run time horizon. The number of post-birth time periods (L_{max}) depends on an assumption on how long the cohorts in the control group remain comparable to the treated cohort.

Since labor market outcomes almost monotonically increase in age at birth (see Section 3), assuming comparability between mothers who give birth at ages 26 and 28 is easier than for ages 26 and 38. The number of cohorts for which the researcher is willing to make such an assumption determines how many not-yet-treated cohorts can be included in the control group and, consequently, the time horizon the estimation can cover. The width of the window in which cohorts are comparable can be formally tested (currently work in progress) and may differ across cohorts, such that for some cohorts effects for more post-birth periods can be estimated than for others. Noteworthy, the conventional approach implicitly assumes the comparability across all ages at birth by employing all older first-time mothers as control units.

A further advantage of the presented approach is the option to study effect heterogeneity across different cohorts of mothers by age at birth, which is of particular importance in policy evaluation settings. For example, given the positive correlation of earnings and age at first birth, an evaluation of a policy separately for younger and older first-time mothers would be informative about the distributional effects. In the following section we provide evidence on the importance of this heterogeneity dimension for getting a complete picture of how motherhood affects labor market outcomes of women.

6 New Evidence on the Career Costs of Motherhood

We exploit the advantages of the new approach and apply it to estimate child penalties in the German labor market and study their heterogeneity by age at first birth. For this task, we rely on the *SIAB* data (see Section 2), which offers a large sample size and, thus, allows to conduct analyses of effect heterogeneity with respect to the age at which women become mothers¹². We restrict the sample to women of age 20 to 45 who become mothers and who live in West Germany. We observe women who give their first birth between ages of 20 and 40. We choose to include the next five cohorts in terms of age at birth in the cohort-specific control groups, which allows us to estimate the effects up to the year +4 after childbirth (the formal test for window width selection is currently work in progress). To ensure a balanced panel over the estimation window (from -3 to $+4$) and enough observations in the control group, we include cohorts of mothers that give birth between ages of 23 and 32.

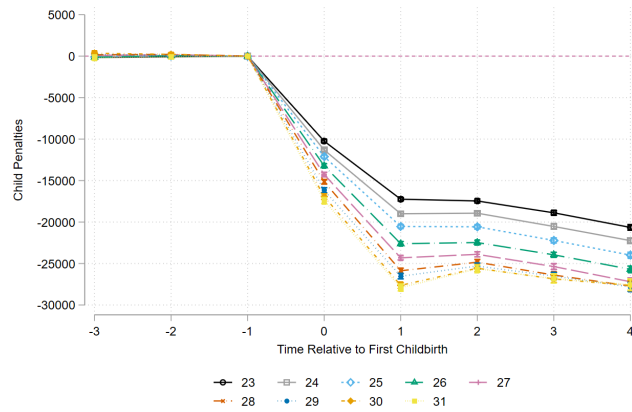
¹² We prefer the *SIAB* data for this exercise as the large sample size ensures precise and clearly distinguishable estimates for the different cohorts. Using the smaller but easier available *SOEP* data yields average estimates that are qualitatively and quantitatively similar.

6.1 Main Results: Annual and Cumulative Labor Earnings

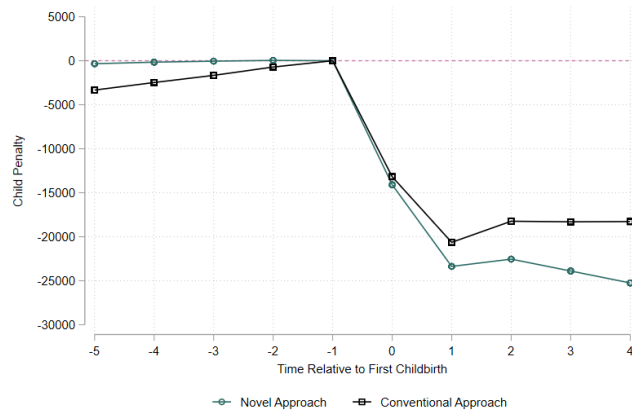
We start by exploring the effects of motherhood on annual labor earnings, since this outcome contains information on both the intensive and extensive margins of labor supply. We estimate Equation (4) with annual pre-tax earnings which include zeros for years of non-participation as an outcome. Results for cohort-specific and average effects in absolute terms are plotted in Figure 6. First, we document that losses in earnings are heterogeneous by age at first birth. As shown in Figure 6a, child penalties in absolute terms differ significantly across cohorts and almost linearly increase in age at first birth. Second, in Figure 6b, we compare the average estimates from our approach with a conventional event study as in Equation (1). The conventional approach substantially underestimates the child penalty, in year four after birth by around Euro 7 K (or 38 percent). These results are consistent with our discussion in Section 4.2 that “forbidden” comparisons with already-treated units underestimate the earnings growth that would have happened without childbirth such that the post-birth trends of earlier-treated mothers bias the counterfactual downwards. This finding also confirms that “forbidden” comparisons receive a substantial weight in common event studies, which leads to a large underestimation of earnings losses. Instead, our approach avoids these “forbidden” comparisons and correctly captures both the losses in levels as well as the unrealized earnings growth, as it compares mothers to the control group of the soon-to-be mothers. The downward sloping trend of the estimates that we find for each cohort and for the average effect reflects that women do not only lose earnings in levels, but also experience a slowdown of their earnings progression after the first childbirth compared to the control group, which leads to an increase of the child penalty over time since birth. Importantly, the pre-trends are flat and insignificant for all cohorts. This provides evidence that the parallel trends assumption is likely to hold in our application and further, that there is no anticipation of treatment that differs over time before birth or across cohorts.

Using annual labor earnings as an outcome allows us to assess losses within a specific year but does not give insights into the broader perspective of the costs associated with motherhood over the life cycle. To provide a more comprehensive understanding, we use cumulative earnings as an outcome within our stacked DiD approach. The results are depicted in Figure 7, where we observe a similar pattern as for annual earnings: older first-time mothers incur greater absolute losses compared to younger ones. For instance, women who give birth at the age of 24 accumulate losses amounting to Euro 80 K over the course of four years following childbirth, while those who give birth at the age of 31 accumulate losses of Euro 100 K over the same period.

For both annual and cumulative earnings we observe that losses almost linearly increase in age at birth. Ex ante, this pattern may appear ambiguous. Older first time mothers are in the later stages of their careers and have higher earnings prior to giving birth, so they have more to lose if they exit the labor force or reduce their working hours. At the same time, they have stronger incentives to



(A) Post-birth earnings losses by age at first childbirth.



(B) Average post-birth earnings loss.

FIGURE 6: Earnings losses of mothers after the first childbirth.

Notes: The figure plots the estimates of absolute losses in annual labor earnings after the birth of the first child following the approach described in section 5. Panel A reports the estimates for cohort-specific losses as in Equation (4). Panel B reports the estimates of the weighted average across-cohorts as in Equation (5) and the conventional approach as in Equation (1). The estimates are reported for the periods from -3 to $+4$, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015.

Source: Own calculations based on the STAB.

return to the labor market and continue their careers due to the larger investments in human capital they have already made and the larger returns on their employment. Therefore, in the following section we decompose this heterogeneity pattern to identify what drives such increase in absolute losses with age at birth.

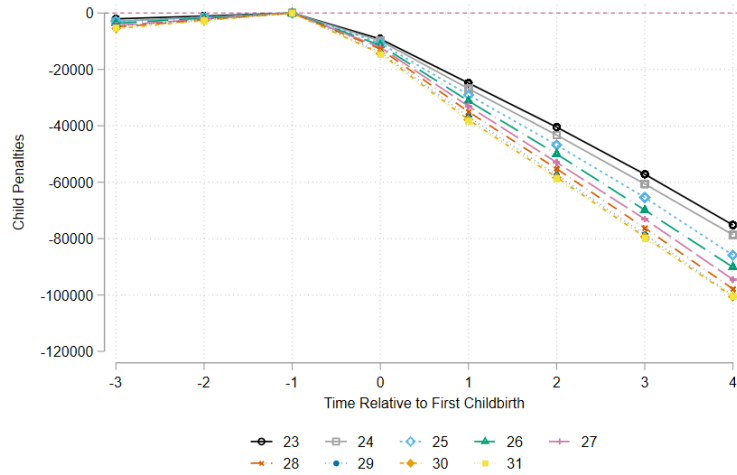


FIGURE 7: Cumulative earnings losses after the first childbirth by age at birth.

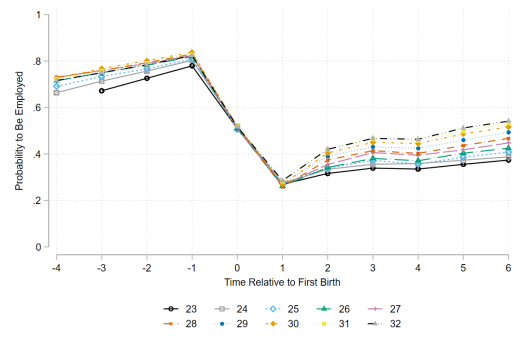
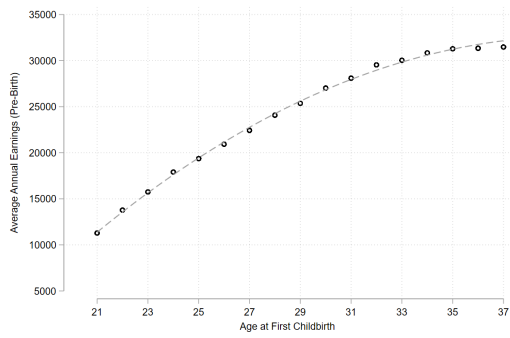
Notes: The figure plots the estimates of losses in cumulative annual labor earnings after the birth of the first child following the approach described in section 5. The figure reports the estimates for cohort-specific losses in absolute terms as in Equation (4). The estimates are reported for the periods from -3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015.

Source: Own calculations based on the SIAB.

6.2 Sources of Heterogeneity in Earnings Losses

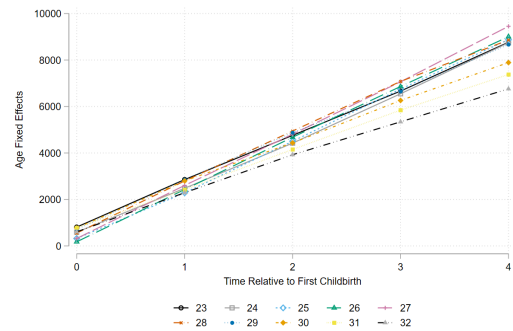
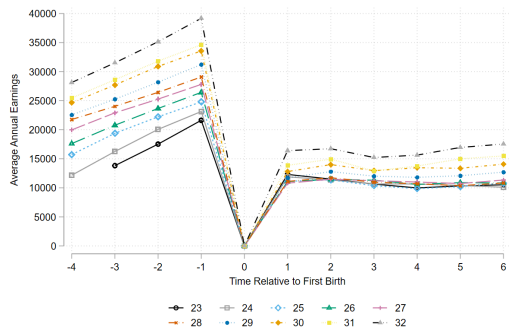
In Figures 6a and 7 we observe that losses almost linearly increase in maternal age at birth. Does it mean that women who delay fertility face larger career costs when giving birth? To answer that question we decompose the heterogeneity in earnings losses into four parts: differences in pre-birth levels, in employment rates (rates of return to the labor market), in post-birth earnings growth and in counterfactual trends of the control groups. We plot these four components in Figure 8. First of all, in Figure 8a we document that pre-birth levels in earnings almost linearly increase in age at birth, driving the pattern that we observe in Figures 6a and 7. However, two other factors affect earnings losses in the opposite direction and turn out to be hidden by the differences in pre-birth levels. Specifically, younger first-time mothers tend to return to the labor market at lower rates (Figure 8b) and their control groups have higher earnings growth rates, i.e. steeper trends (Figure 8d). In terms of post-birth earnings growth (Figure 8c), all mothers exhibit similar flat post-birth trends, even when conditioning on those who have returned after a labor market break of at most one year (thus ensuring that earnings growth is not biased by different lengths of maternal leave).

Therefore, if we account for the pre-birth differences in levels, the heterogeneity pattern reverses. When we divide the effects by the cohort-specific pre-birth levels (as plotted in Figure 9), we observe that earnings losses in relative terms instead decrease in age at birth, with younger first-time mothers



(A) Pre-birth levels of annual labor earnings by age at first birth

(B) Employment rates by age at first birth around the event of childbirth



(C) Annual labor earnings by age at first birth around the event of childbirth (conditional on returning to work after 1 year break)

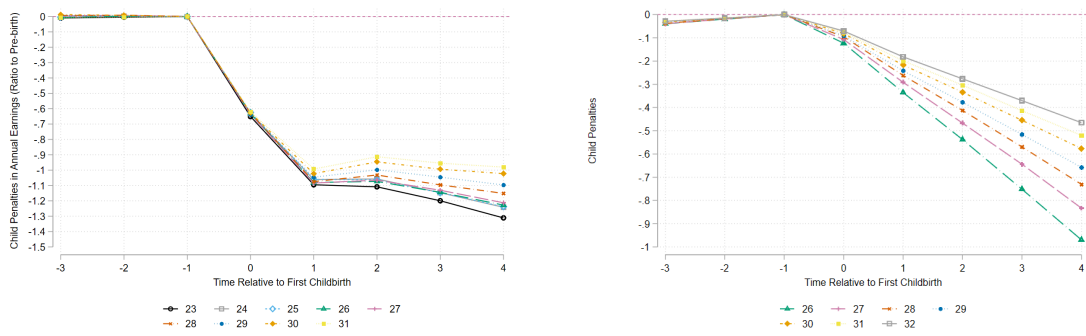
(D) Annual labor earnings growth of control groups by age at first birth around the event of childbirth

FIGURE 8: Decomposition of Heterogeneity in Annual Earnings by Age at Birth

Notes: The figure plots the following four outcomes by age at first birth: pre-birth levels of annual earnings, employment rates, annual earnings conditional on working and annual earnings growth experienced by the cohort-specific control groups. Pre-birth levels of annual earnings include zero earnings. Employment rates are calculated under the possibilities to be employed, unemployed and out of labor force. Panel 8c plots average annual earnings conditional on taking a maternity leave for one year and returning to the labor market after the leave. Panel 8d plots the average growth rates of earnings in absolute terms of the control groups defined in section 5 (equivalent to age fixed effects from the equation Equation (4)).

Source: Own calculations based on the STAB.

experiencing considerably greater losses than older ones. By the fourth year following childbirth, for those who become mothers at the age of 31, the foregone earnings accumulate to 50 percent of the total earnings they had amassed prior to childbirth. In contrast, for mothers who give birth at the age of 26, the losses in income levels and unrealized earnings progression accumulate to nearly 100 percent of what they had cumulatively earned before childbirth. These findings reflect that younger mothers tend to have steeper counterfactual trends due to their control groups being at the early and most important stage of career development (as argued by Bagger et al. 2014).



(A) Annual labor earnings losses after the first childbirth (relative to pre-birth levels). (B) Cumulative labor earnings losses after the first childbirth (relative to pre-birth levels).

FIGURE 9: Relative losses in annual and cumulative labor earnings after the first childbirth

Notes: The figure plots the estimates of relative losses in annual and cumulative labor earnings after the birth of the first child following the approach described in section 5. The figure reports the estimates for cohort-specific losses as in Equation (4). The estimates are reported for the periods from -3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015.

Source: Own calculations based on the SIAB.

This phenomenon becomes more apparent when we look at occupational rank¹³, the probability of engaging in complex tasks, and the attainment of higher education degrees as outcomes within our stacked DiD approach. For the first two outcomes we restrict the sample to those mothers who return to the labor market within 3 years following childbirth, i.e. to exclude potentially confounding effects of compositional changes in the sample from the pre- to the post-birth years. For the same purposes, we adjust the outcomes during the maternity leave to pre-birth levels.¹⁴ In Figure 10, we present the averages of the outcomes by age at first childbirth in the left column, and in the right column, we plot the cohort-specific estimates of the effect of motherhood on these outcomes. Focusing first on the left column, we notice that the major part of career progression in terms of occupational advancement, educational attainment and the development of skills to conduct more complex tasks at work happens between the ages of 25 and 30. This phase is likely to be missed by younger first-time mothers, since we observe almost no growth in labor market outcomes after the first childbirth. Subsequent to the first childbirth, not only do younger cohorts of mothers experience a plateau in their occupational ranks, education levels, and task complexity, but their control groups who have not yet become mothers rapidly progress in their careers. In summary, younger first-time mothers miss out on the phase of most rapid career advancement and human capital accumulation.

¹³ We measure occupational rank as the median earnings by occupation at the 3-digit level (defined in the KldB 2010 by Bundesagentur für Arbeit 2021).

¹⁴ This is motivated by the German parental leave legislation that grants mothers the right to return to their pre-birth employer in the same or a similar job.

For younger mothers, the effects of motherhood on these three outcomes – occupational rank, complexity of task and education levels – in the right column of Figure 10 should be then interpreted as unrealized progression rather than losses in levels. We document that younger mothers lose relatively much more than older ones in terms of foregone growth of occupational rank, complexity of tasks and educational levels. For example, those women who give birth at 23 and 31 experience a negative effect of –4 percentage points on the probability to do complex tasks in the fourth year after childbirth (Figure 10d). However, for the first group of younger mothers, these 4 percentage points make up 80 percent of the pre-birth levels and are fully composed of the foregone growth, as the likelihood to do complex tasks stagnates at 5 percent over their life cycle. At the same time, for those who give birth at 31, these 4 percentage points make only 16 percent of their pre-birth levels and are composed of both downgrading to doing less complex tasks after childbirth and foregoing some progression in the complexity of work tasks. We observe a similar pattern if we look at the results for the occupation rank as an outcome of interest (Figure 10b). For older first time-mothers who give birth at 31 the coefficient in the fourth year after birth equals to Euro –500 (or only 2 percent of the pre-birth level) and is not significant. For the younger first time mothers, the coefficient is significant and equals Euro –900, which is 4.5 percent of their pre-birth level and reflects only unrealized growth. For the older first-time mothers we observe that in addition to foregone progression, these small losses also include some downgrading in occupational rank after birth, with post-birth levels still being substantially larger than those of younger first-time mothers. These results also indicate that the well-documented sorting into lower-paying but more flexible and family-friendly workplaces and occupations (Bolotnyy and Emanuel 2022; Goldin 2014; Bertrand, Goldin, and Katz 2010) is likely to be driven by older first-time mothers.

As we have already seen in Figure 8c, which plots annual earnings conditional on working, career progression seems to stop after childbirth for mothers of all ages – the post-birth trends are almost flat independent of age at birth. From Figure 10, we learn that occupational rank and complexity of tasks also stagnate after childbirth for younger mothers, with older first-time mothers even experiencing some downgrading. It means that after childbirth, the differences across cohorts by age at birth appear only in levels – mothers tend to return to the careers they have built prior to birth but face almost no growth thereafter. For younger mothers, the estimated effects of childbirth in the right column of Figure 10 then mainly represent the foregone growth, since they stop climbing the concave career ladder at its very beginning. For older first-time mothers such negative effects reflect both unrealized growth and career downgrading as an adjustment to motherhood. This differences in the interpretation of the effects indicate that policies (as, for instance, expanding public childcare) might have differential impacts depending on which channels they work through and that for mothers of different ages different policies might be needed in order to the reduce career costs of motherhood.

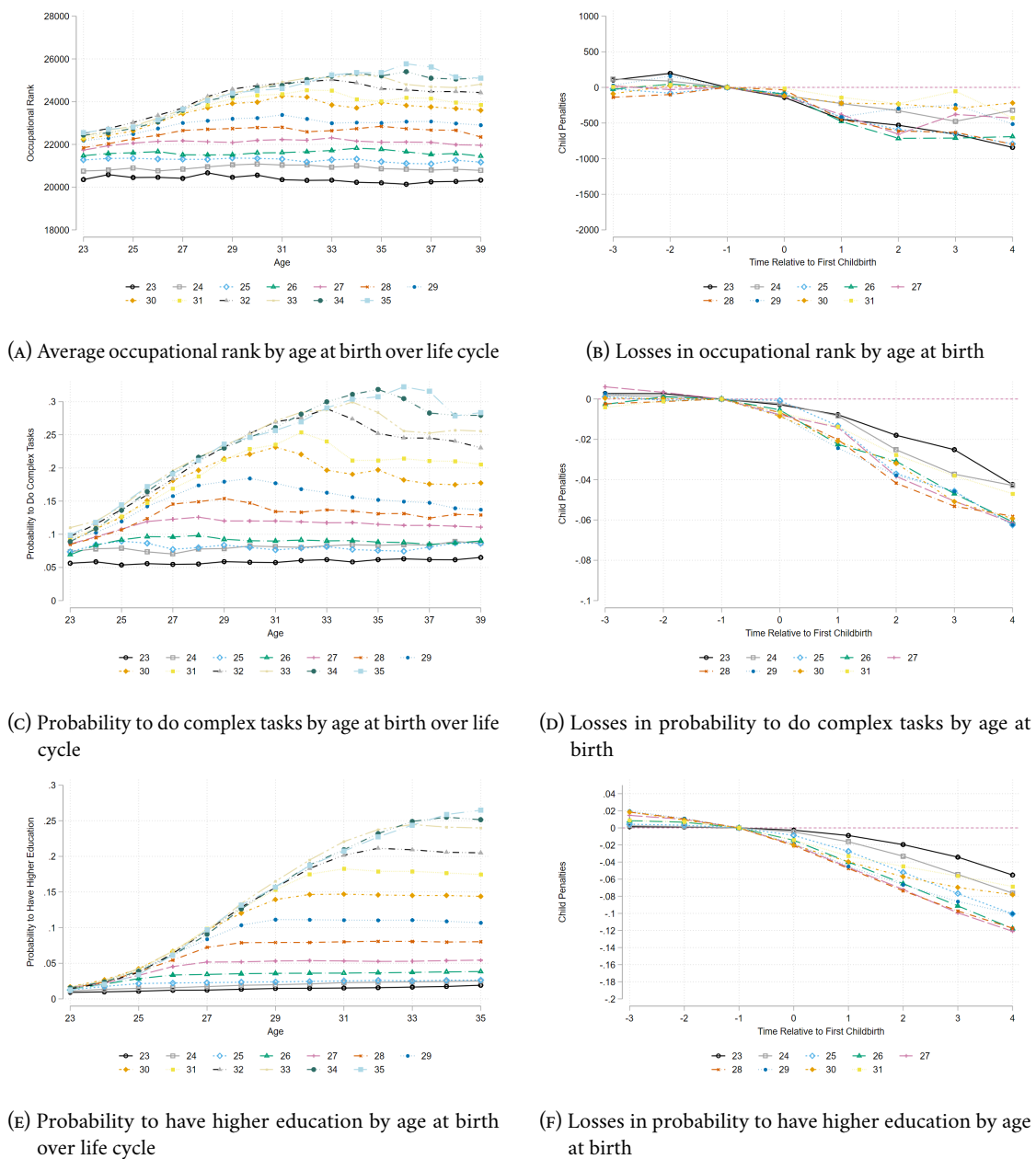
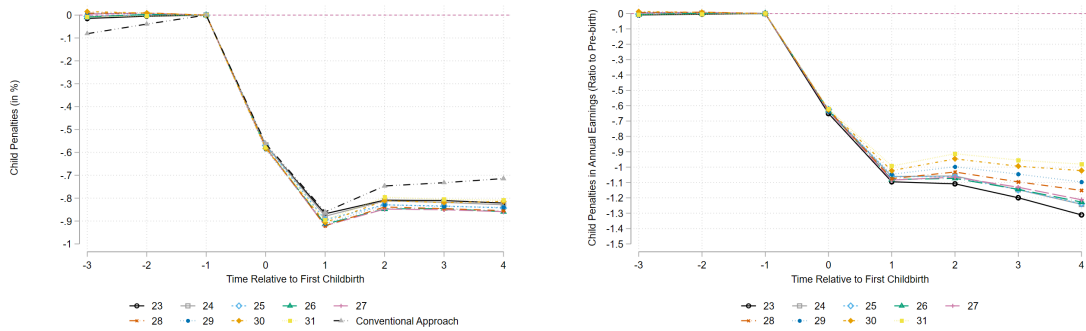


FIGURE 10: Career progression and human capital accumulation by age at birth over life cycle

Notes: The figure plots the average outcomes by age at birth (left column) and estimates of losses in outcomes (right column) after the birth of the first child following the approach described in section 5. The outcomes are occupational rank (median earnings within 3-digit occupations), probability to do complex tasks at work and probability to have higher education degree. The figures report the estimates for cohort-specific losses as in Equation (4). The estimates are reported for the periods from -3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015.

Source: Own calculations based on the SIAB.

To sum up, we have established that the major sources of heterogeneity in earnings losses by age at birth are the pre-birth levels and the growth rates of the control groups. Therefore, when we account for both pre-birth differences in levels and differences in counterfactual trends, as it is conventionally done in the literature through normalizing the estimated effects by counterfactual levels, we observe that the heterogeneity across maternal ages at birth mostly disappears (Figure 11a). For all cohorts of mothers, losses appear to be around -85 percent (which is by 15 percentage points, or 20 percent, larger than what the conventional method estimates using the same data). This finding confirms that pre-birth level differences and differential counterfactual trends are the major factors driving the heterogeneity in absolute earnings losses that we observe in Figure 6a. The results plotted in Figure 11 also indicate that depending on the choice of the re-scaling benchmark (e.g., pre-birth levels or counterfactual levels), the costs in relative terms will have different interpretations. Re-scaling by pre-birth levels conditions only on pre-birth differences in levels, while dividing by counterfactual levels of control groups additionally takes away differences in the trends of control groups across different ages at birth. Therefore, re-scaling the estimated effects by pre-birth levels is a more informative approach, because it allows to compare the costs in career progression across mothers of different ages.



(A) Annual labor earnings losses after the first childbirth (relative to **counterfactual** levels). (B) Annual labor earnings losses after the first childbirth (relative to **pre-birth** levels).

FIGURE 11: Re-scaled labor earnings losses after the first childbirth

Notes: The figure plots the estimates of relative losses in annual labor earnings after the birth of the first child following the approach described in section 5, for two approaches to re-scaling. The figure reports the estimates for cohort-specific losses as in Equation (4) and the conventional approach as in Equation (1). The estimates are reported for the periods from -3 to $+4$, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015.

Source: Own calculations based on the STAB.

7 Conclusion

Estimating motherhood effects based on event studies has quickly become a widely used and valuable tool to assess the labor market effects of childbirth for mothers. It comes, however, at the cost of making strong assumptions. In this paper, we show that the considerable amount of heterogeneity in both maternal outcomes and characteristics by age at first childbirth leads to a violation of these assumptions and introduces issues similar to those documented in the literature on staggered difference-in-differences models. By making “forbidden” comparisons to already-treated units as well as by introducing “contamination”, i.e. the possibility that the treatment effects for one period are confounded by effects from other periods, child penalty event studies are likely to produce biased estimates. In some settings such issues can be addressed by using heterogeneity-robust estimators. We discuss that they are not fully applicable with childbirth as treatment as they employ control groups that are not comparable.

Instead of relying on conventional event studies, we propose a novel approach to estimate child penalties. We use a “stacked” DiD design with an additional restriction of the control group. We construct cohort-specific control groups from observations of only the closest not-yet-treated mothers. This rolling window of included control cohorts ensures the comparability of the treatment and control group which is crucial to justify the plausibility of the parallel trends assumption. This approach allows us to estimate the unbiased effects of childbirth, both specific for each cohort and on average.

Our application, that revisits the estimation of child penalties in the German labor market, highlights that common event studies substantially underestimate the earnings losses following the first childbirth. Furthermore, we demonstrate that recognizing the heterogeneity based on the age at which mothers give birth is not only crucial in terms of methodological correctness but also essential to gain a precise understanding of the career costs associated with motherhood. Notably, our analysis reveals that younger first-time mothers experience larger career costs of children, which are primarily attributed to unrealized earnings growth and career advancement rather than occupational downgrading. Our results also suggest that taking this heterogeneity into account might be of particular importance for policy evaluations.

References

- Adda, Jérôme, Christian Dustmann, and Katrien Stevens (2017). “The career costs of children”. In: *Journal of Political Economy* 125.2, pp. 293–337 (cit. on pp. 2, 7).
- Andresen, Martin Eckhoff and Emily Nix (2022). “What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples”. In: *Journal of Labor Economics* 40.4, pp. 971–1004 (cit. on pp. 5, 19).
- Andrew, Alison, Oriana Bandiera, Monica Costa-Dias, and Camille Landais (2021). “Women and men at work”. In: *IFS Deaton Review of Inequalities* (cit. on p. 2).
- Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016). “Parenthood and the Gender Gap in Pay”. In: *Journal of Labor Economics* 34.3, pp. 545–579 (cit. on pp. 4, 19).
- Antonovics, Kate and Rober Town (2004). “Are All the Good Men Married? Uncovering the Sources of the Marital Wage Premium”. In: *American Economic Review* 94.2, pp. 317–321 (cit. on p. 19).
- Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin (2014). “Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics”. In: *American Economic Review* 104.6, pp. 1551–1596 (cit. on pp. 4, 29).
- Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz (2010). “Dynamics of the gender gap for young professionals in the financial and corporate sectors”. In: *American Economic Journal: Applied Economics* 2.3, pp. 228–255 (cit. on p. 31).
- Blundell, Richard, Monica Costa-Dias, David Goll, and Costas Meghir (2021). “Wages, experience, and training of women over the life cycle”. In: *Journal of Labor Economics* 39.S1, S275–S315 (cit. on p. 2).
- Bolotnyy, Valentin and Natalia Emanuel (2022). “Why do women earn less than men? Evidence from bus and train operators”. In: *Journal of Labor Economics* 40.2, pp. 283–323 (cit. on p. 31).
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2022). “Revisiting Event Study Designs: Robust and Efficient Estimation”. URL: <https://arxiv.org/abs/2108.12419> (visited on 04/03/2023) (cit. on p. 3).
- Bruns, Benjamin (2019). “Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap”. In: *American Economic Journal: Applied Economics* 11.2, pp. 74–113 (cit. on p. 5).
- Bundesagentur für Arbeit, ed. (2021). *Klassifikation der Berufe 2010 – überarbeitete Fassung 2020*. Nürnberg (cit. on p. 30).
- Bütikhofer, Aline, Sissel Jensen, and Kjell Salvanes (2018). “The role of parenthood on the gender gap among top earners”. In: *European Economic Review* 109, pp. 103–123 (cit. on p. 4).
- Callaway, Brantly and Pedro H. C. Sant’Anna (2021). “Difference-in-Differences with multiple time periods”. In: *Journal of Econometrics* 225, pp. 200–230 (cit. on pp. 3–5, 18, 22).

- Card, David, Jörg Heining, and Patrick Kline (2013). “Workplace Heterogeneity and the Rise of West German Wage Inequality”. In: *The Quarterly Journal of Economics* 128.3, pp. 967–1015 (cit. on p. 6).
- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). “The Effect of Minimum Wages on Low-Wage Jobs”. In: *The Quarterly Journal of Economics* 134.3, pp. 1405–1454 (cit. on pp. 4, 22).
- Dauth, Wolfgang and Johann Eppelsheimer (2020). “Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide”. In: *Journal for Labour Market Research* 54 (cit. on p. 6).
- de Chaisemartin, Clément and Xavier D’Haultfoeuille (2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects”. In: *American Economic Review* 110.9, pp. 2964–2996 (cit. on pp. 3, 5, 10, 18).
- (2022a). “Two-way Fixed Effects and Differences-in-Differences Estimators with Several Treatments”. URL: <https://sites.google.com/site/clementdechaisemartin/> (visited on 04/03/2023) (cit. on pp. 5, 17).
- (2022b). “Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey”. NBER Working Paper No. 29691 (cit. on pp. 3, 15).
- Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt (2022). *The economics of fertility: A new era*. Tech. rep. National Bureau of Economic Research (cit. on p. 7).
- Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). “Revisiting the German Wage Structure”. In: *The Quarterly Journal of Economics* 124.2, pp. 843–881 (cit. on p. 6).
- Fitzenberger, Bernd and Arnim Seidlitz (2023). “Changing Fertility and Heterogeneous Motherhood Effects: Revisiting the Effects of a Parental Benefits Reform”. URL: https://population-economics.committee.socialpolitik.de/sites/default/files/2023-02/Fitzenberger_Changing%20Fertility.pdf (visited on 10/30/2023) (cit. on p. 5).
- Fitzenberger, Bernd, Katrin Sommerfeld, and Susanne Steffes (2013). “Causal effects on employment after first birth – A dynamic treatment approach”. In: *Labor Economics* 25, pp. 49–62 (cit. on pp. 4, 19).
- Frodermann, Corinna, Alexandra Schmucker, Stefan Seth, and Philipp vom Berge (2021). “Sample of Integrated Labour Market Biographies (SIAB) 1975–2019”. In: *IAB FDZ Datenreport* 01|2021 (cit. on p. 6).
- Goebel, Jan, Markus M. Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp (2019). “The German Socio-Economic Panel (SOEP)”. In: *Jahrbücher für Nationalökonomie und Statistik* 239.2, pp. 345–360 (cit. on p. 6).
- Goldin, Claudia (2014). “A grand gender convergence: Its last chapter”. In: *American Economic Review* 104.4, pp. 1091–1119 (cit. on p. 31).

- Goldin, Claudia (2022). “Career & Family: Women’s Century-Long Journey Toward Equity. By Claudia Goldin. Princeton: Princeton University Press, 2021. Pp v, 325., hardcover.” In: *The Journal of Economic History* 82.3, pp. 915–916 (cit. on p. 2).
- Goldin, Claudia, Sari Pekkala Kerr, and Claudia Olivetti (2022). “When the Kids Grow Up: Women’s Employment and Earnings across the Family Cycle”. NBER Working Paper No. 30323 (cit. on pp. 2, 7, 19).
- Goodman-Bacon, Andrew (2021). “Difference-in-differences with variation in treatment timing”. In: *Journal of Econometrics* 225.2, pp. 254–277 (cit. on pp. 5, 9–11, 14, 15).
- Kleven, Henrik, Camille Landais, Johanna Posch, Adreas Steinhauer, and Josef Zweimüller (2019). “Child Penalties across Countries: Evidence and Explanations”. In: *American Economic Review: Papers & Proceedings* 109, pp. 122–126 (cit. on p. 5).
- Kleven, Henrik, Camille Landais, and Jakob Egholt Sogaard (2019). “Children and Gender Inequality: Evidence from Denmark”. In: *American Economic Journal: Applied Economics* 11.4, pp. 181–209 (cit. on pp. 2, 10).
- Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington (2018). “The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?” NBER Working Paper No. 24740 (cit. on p. 4).
- Lundborg, Petter, Erik Plug, and Astrid Würtz Rassmussen (2017). “Can Women Have Children and a Career? IV Evidence from IVF Treatments”. In: *The American Economic Review* 107.6, pp. 1611–1637 (cit. on p. 19).
- Müller, Dana, Andreas Filser, and Corinna Frodermann (2022). “Update: Identifying mothers in administrative data”. In: *IAB FDZ Methodenreport* 01/2022 (cit. on p. 6).
- Roth, Jonathan, Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe (2022). “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature”. URL: <https://psantanna.com/research/> (visited on 04/03/2023) (cit. on pp. 3, 5, 15).
- Steinhauer, Adreas (2018). “Working Moms, Childlessness, and Female Identity”. SciencesPo, LIEPP Working Paper No. 79 (cit. on p. 20).
- Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. In: *Journal of Econometrics* 225.2, pp. 175–199 (cit. on pp. 3, 5, 9, 10, 16–18, 24, 42).
- Wilde, Elizabeth Ty, Lily Batchelder, and David T. Ellwood (2010). “The mommy track divides: The impact of childbearing on wages of women of differing skill levels”. NBER Working Paper No. 16582 (cit. on pp. 7, 8).

A Additional Tables and Figures

TABLE A.1: Summary statistics of mothers one year prior to their first childbirth (SOEP).

	Mean	SD	Min.	p5	p50	p95	Max.
Labor earnings	26,079	17,730	0	0	26,072	53,412	182,917
Share fulltime	0.71	0.45	0	0	1	1	1
Share not working	0.10	0.30	0	0	0	1	1
Experience (fulltime)	5.41	4.33	0	0	5	13	24
Experience (parttime)	0.89	1.95	0	0	0	5	17
Total years of education	12.83	2.81	7	9	12	18	18
Age at first birth	29.02	4.59	21	22	29	37	45
<i>N</i>	1,992						

This table provides the summary statistics for the SOEP dataset.

TABLE A.2: Summary statistics of mothers one year prior to their first childbirth (SIAB).

	Mean	SD	Min.	p5	p50	p95	Max.
Labor earnings	23,908	17,702	0	0	24,367	51,349	343,804
Share fulltime	0.81	0.40	0	0	1	1	1
Share not working	0.15	0.35	0	0	0	1	1
Experience (fulltime)	5.01	4.08	0	0	4	13	20
Age at first birth	28.57	4.18	22	22	28	36	40
<i>N</i>	151,979						

This table provides the summary statistics for the SIAB dataset.

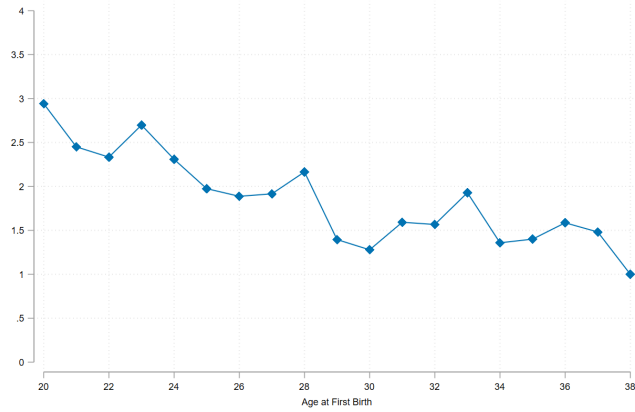


FIGURE A.1: Average length of employment break by age at childbirth (conditional on returning to work).

Notes: The figure plots average length of employment break for mothers by age at childbirth, conditional on returning to work.

Source: Own calculations based on the SOEP.

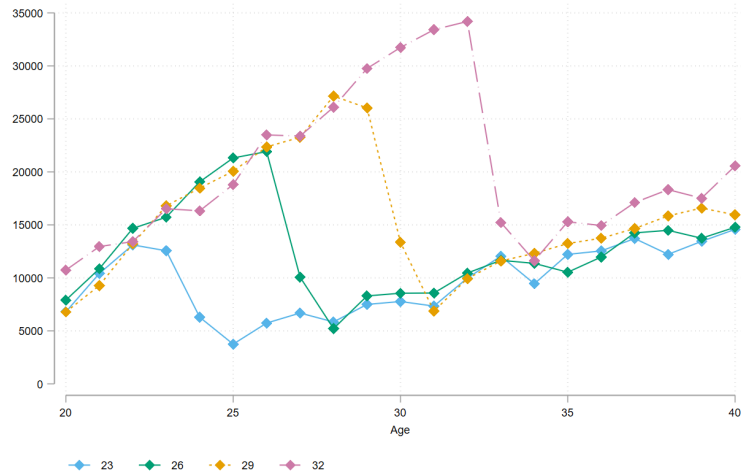
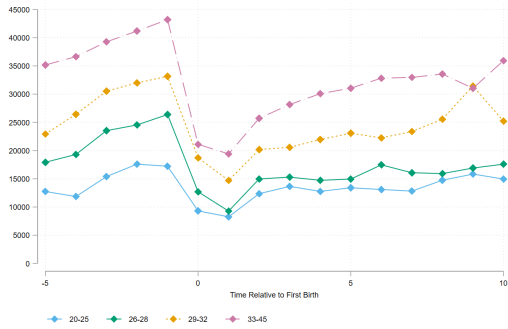


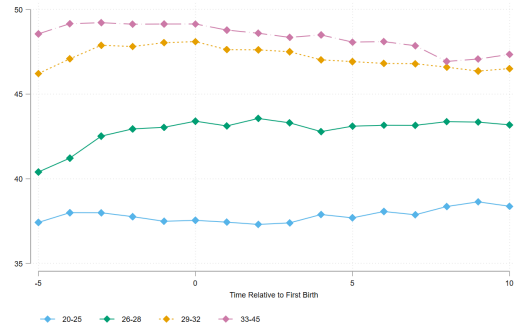
FIGURE A.2: Average earnings of mothers around birth for four ages at first birth, incl. zeros

Notes: The figure plots average annual labor earnings of mothers (including zero earnings) over life cycle for four ages at first birth (23/26/29/32).

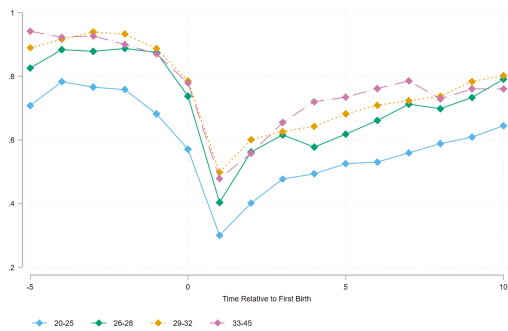
Source: Own calculations based on the SOEP.



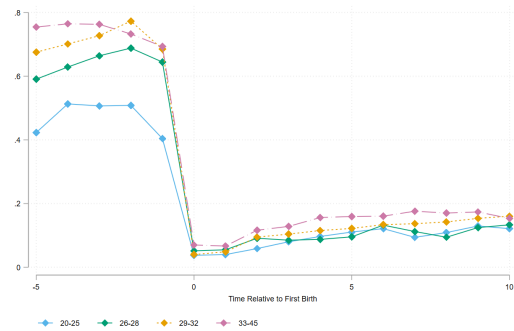
(A) Average annual earnings of *working* mothers.



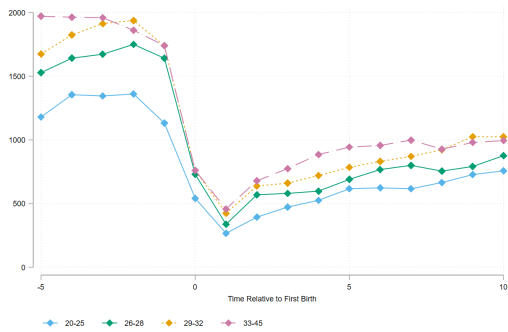
(B) Standard international occupational prestige scale (SIOPS).



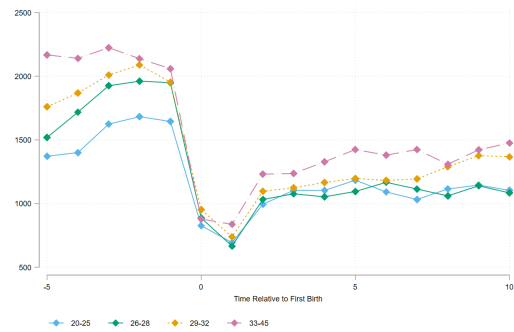
(C) Share working mothers.



(D) Share of mothers working full-time.



(E) Annual working hours for mothers (including zeros).

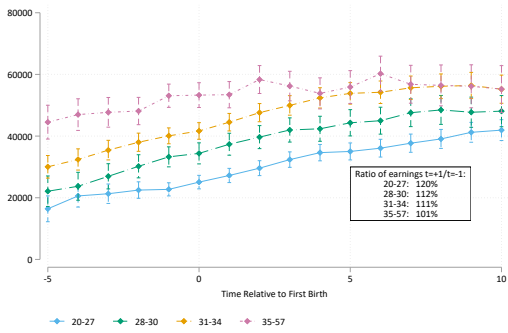


(F) Annual working hours for mothers conditional on working.

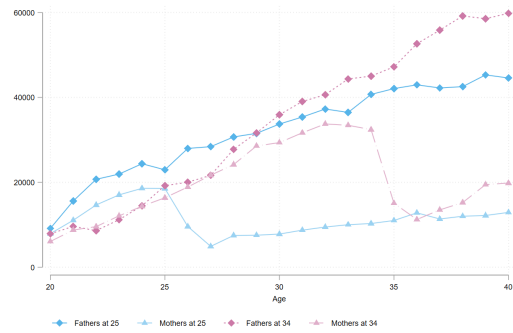
FIGURE A.3: Annual earnings and working hours, occupational rank, labor force and full-time status of mothers by their age at first childbirth.

Notes: The figure shows a measure of annual earnings and working hours, occupational rank, labor force and full-time statuses of mothers for four quantiles of the distribution of age at first birth. The first quantile includes mothers aged 20–27 at first birth, the second those age 28–30, the third those aged 31–34 and the fourth those from 35–57.

Source: Own calculations based on the SOEP.



(A) Annual earnings of fathers by event time.



(B) Annual earnings of fathers and mothers by age (child-birth at the age of 24 and 35).

FIGURE A.4: Average earnings of fathers and mothers around birth.

Notes: The left figure plots average annual labor earnings of fathers in time relative to their first birth for four quantiles of the distribution of age at first birth. The first quantile includes fathers aged 20–27 at first birth, the second those age 28–30, the third those aged 31–34 and the fourth those from 35–57. The right figure plots annual labor earnings of mothers and fathers who became parents at the age of 24 and 35.

Source: Own calculations based on the SOEP.

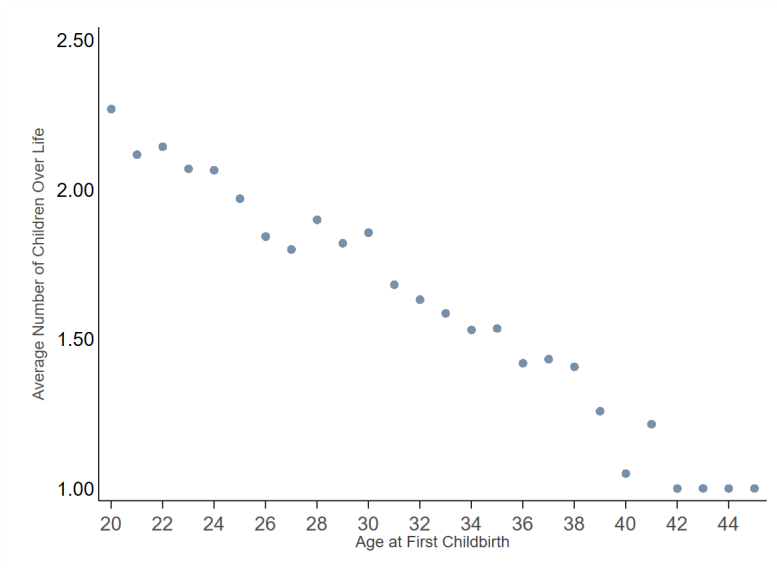


FIGURE A.5: Average number of children over life by age at childbirth.

Notes: The figure plots average number of children in total over life for mothers by age at first birth.

Source: Own calculations based on the SOEP.

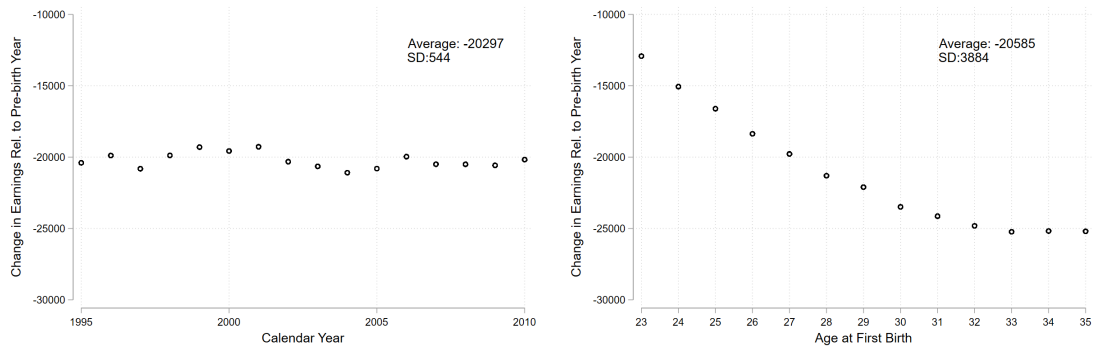
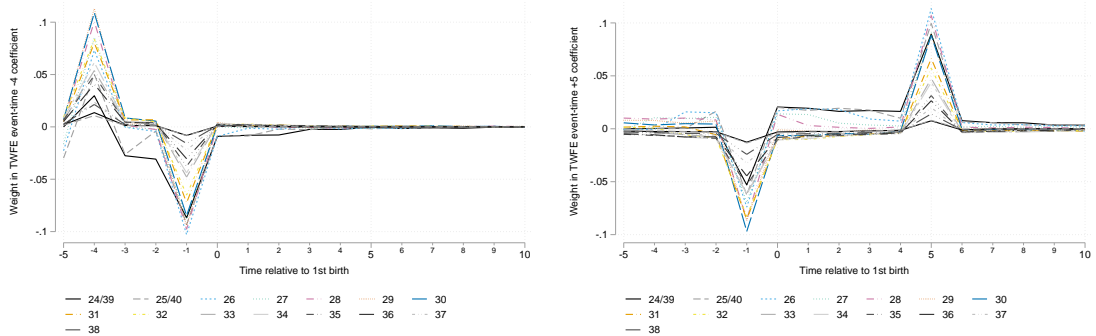


FIGURE A.6: Average Change in Annual Earnings after Birth over Calendar Time and Age at First Childbirth

Notes: The figure plots average changes in annual earnings of mothers in the year after first childbirth relative to the year before birth. The left panel plots the average losses over calendar time, while the right panel plots the average losses over age at first birth.

Source: Own calculations based on the SIAB.



(A) 4 years before childbirth.

(B) 5 years after childbirth.

FIGURE A.7: Sun and Abraham (2021) decomposition of weights: “contamination” from other periods.

Notes: The figure plots the weights obtained from the Sun and Abraham (2021) decomposition for the relative time periods -4 (left-hand panel) and $+5$ (right-hand panel). The weights are calculated with the `eventstudyweights` Stata module provided by Sun and Abraham (2021).

Source: Own calculations based on the SOEP.

TABLE A.3: Share of pre-birth observations to construct counterfactual earnings by relative time and age at first childbirth.

Time relative to 1st birth	Age at 1st birth				
	24	26	29	32	35
-5	1.000	1.000	0.733	0.537	0.310
-4	1.000	0.866	0.679	0.455	0.268
-3	1.000	0.794	0.582	0.373	0.216
-2	0.866	0.733	0.537	0.310	0.181
-1	0.794	0.679	0.455	0.268	0.136
0	0.733	0.582	0.373	0.216	0.106
1	0.679	0.537	0.310	0.181	0.083
2	0.582	0.455	0.268	0.136	0.071
3	0.537	0.373	0.216	0.106	0.060
4	0.455	0.310	0.181	0.083	0.042
5	0.373	0.268	0.136	0.071	0.020
6	0.310	0.216	0.106	0.060	0.006
7	0.268	0.181	0.083	0.042	0.008
8	0.216	0.136	0.071	0.020	0.000
9	0.181	0.106	0.060	0.006	0.000
10	0.136	0.083	0.042	0.008	0.000

The table reports the share of pre-birth observations that are available to construct counterfactual earnings as in Equation (3) by relative time around childbirth and for 5 levels of age at first childbirth. The age levels correspond to the 10th, 25th, 50th, 75th and 90th percentiles of age at first childbirth.

Source: Own calculations based on the SOEP.