I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16966

Changing Fertility and Heterogeneous Motherhood Effects: Revisiting the Effects of a Parental Benefits Reform

Bernd Fitzenberger
Arnim Seidlitz

APRIL 2024

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16966

Changing Fertility and Heterogeneous Motherhood Effects: Revisiting the Effects of a Parental Benefits Reform

Bernd Fitzenberger
IAB, FAU Erlangen-Nürnberg, IFS, CESifo, IZA and ROA

Arnim Seidlitz
IAB

APRIL 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Changing Fertility and Heterogeneous Motherhood Effects: Revisiting the Effects of a Parental Benefits Reform*

Using a semiparametric event study approach with a control group, we estimate the effect of motherhood on labor market outcomes in Germany, the child penalty. We further investigate how the 2007 parental benefits reform changed the child penalty while accounting for fertility effects. A large novel data set linking data from two administrative sources provides information on all births. Our estimation approach accounts for motherhood being a staggered treatment. The reform has small positive medium-run effects employment outcomes. It changes the selection into fertility and shows heterogeneous effects. However, the reform did little to reduce the average child penalty.

JEL Classification: J08, J13, J16, J22

Keywords: parental benefit reform, child penalty, semiparametric event study approach

Corresponding author:
Bernd Fitzenberger
Institute for Employment Research (IAB)
Regensburger Str. 104
90478 Nürnberg
Germany
E-mail: bernd.fitzenberger@iab.de

* We gratefully acknowledge finacial support through the DFG project "Female Employment Patterns, Fertility, Labor Market Reforms, and Social Norms: A Dynamic Treatment Approach" (DFG project number: FI 692/14-1 and PA 2536/1-1). We thank Marie Paul, Anna Raute, Ralf Wilke and participants at various seminar presentations of the paper, including DTMC Workshop on Interactions between Labor and Marriage Markets in Aarhus 2022, SES conference 2022, IZA Summer School 2022, CReAM/RF Berlin workshop Topics in Labor Economics 2023, Committee on Population Economics in Berlin 2023 and a number of universities for their valuable input. All remaining errors are our own.

1 Introduction

The so-called "child penalty" for women due to motherhood is a major reason for the persistent gender gap in the labor market (Angelov et al., 2016; Goldin, 2014; Kleven et al., 2019b). Extensive parental leave policies involving job protection and parental benefits while being on leave have been implemented to support women to balance family and work as well as return to the pre-leave job (Olivetti and Petrongolo, 2017; Rossin-Slater, 2018). A large literature has estimated the effects of the introduction or the change of parental leave policies suggesting that entitlement periods up to about one year can be beneficial to keep mothers attached to the labor market while longer entitlement periods may reduce female labor supply or long-term career prospects (Olivetti and Petrongolo, 2017; Rossin-Slater, 2018). However, there is no consensus on these effects in the literature. The existing empirical literature focuses on the effect of parental leave policies for mothers who had been employed before birth (e.g. Dahl et al., 2016; Lalive et al., 2014). This literature typically does not explicitly estimate how the fertility effect of parental leavel policies affects the overall effect of these policies on labor market outcomes, see e.g. Kleven et al. (2020) as an example. However, Raute (2019) shows that the parental benefits reform examined here had an impact on fertility, although the effects varied depending on the subgroup of women.

Our study revisits the 2007 parental benefits reform in Germany whose effects on maternal employment has been estimated in a number of studies.[1] This reform introduced comprehensive paid parental leave up to 12 months after child birth for all mothers. It replaced lower and flat means-tested parental leave benefits for low-income mothers by benefits which increase in pre-birth earnings. In Germany, mothers are further entitled to job protection for three years after child birth. A key goal of the reform was to increase fertility among women, who are highly attached to the labor market.

Our paper makes five important contributions to the literature. First, to our knowledge, our study is the only one that estimates both the child penalty using a semiparametric event study approach with a control group and the reform effect. This allows us to systematically account for the heterogeneity both in the motherhood effect and in the reform effect.

Second, our approach addresses a number of concerns raised in the recent very active literature on difference-in-differences (DiD) estimates for staggered treatment (Roth et al., 2023; de Chaisemartin and d'Haultfoeuille, 2023; Borusyak et al., 2021). Most importantly, we estimate the treatment effects for each treatment cohort for a specific

[1] Frodermann et al. (2023), Kluve and Schmitz (2018), and Bergemann and Riphahn (2020) consider employment and earnings in the medium run. Earlier studies show the effects in the short run (Bergemann and Riphahn, 2011; Kluve and Tamm, 2013; Geyer et al., 2015).

treatment time separately using a control groups of only not-yet-mothers, similar to the methodological approach of Fitzenberger et al. (2013) and Cengiz et al. (2019). This way we avoid any forbidden comparisons with earlier mothers which are a major concern pointed to by the aforementioned studies when applying two-way-fixed-effects regressions (TWFE) to implement DiD estimates of treatment effects. The widely cited event-study approach of Kleven et al. (2019b, 2020) uses such an approach to estimate the child penalty.[2] However our semiparametric approach is no DiD estimator, thus we do not have to rely on a parallel trends assumption.

Third, we have access to rich novel monthly administrative data linking precise information on fertility from the pension insurance to administrative labor market data. In contrast to most of the existing literature, we use all births, thus not restricting attention to mothers who had been employed before motherhood.[3] Our data allow us to estimate changes in the selection of mothers due to the reform and to distinguish the reform effect from selection effects. In particular, we account for the differential reform effect on fertility (Raute, 2019) and for the possible heterogeneity of the reform effect, which is possible because our semiparametric approach allows us to estimate the causal motherhood effect and to account for its heterogeneity. Based on a sample of births three years before and after the reform date, we extrapolate beyond locally valid regression-discontinuity-design (RDD) estimates of the reform effect, which explicitly relies on the assumption that the selection of mothers did not change shortly around the reform date. Our rich administrative data allow for a comprehensive analysis of various outcomes involving employment, full-time employment, earnings, and second order fertility. Thus, we provide a more comprehensive assessment than the existing literature on the German parental leave reform. In particular, we assess whether the reform reduced the strong child penalty in full-time employment.

Fourth, we carefully distinguish between treatment time and event time which is important due to the gestation period. The decision to have a child is taken almost a year before birth, therefore outcomes during the year preceding are likely to be affected by the motherhood. In contrast to most of the literature using data at a yearly frequency, we have a monthly panel on earnings, employment, full-time status, and birth. This way, we can trace more precisely the evolution of outcomes relative to the event time, i.e. the month of birth. In fact, we do find strong anticipation effects which would be masked by an analysis at an annual frequency.

Fifth, our analysis accounts for possible entitlement effects caused be the reform.

[2]For estimating the child penalty in Germany, Melentyeva and Riedel (2023) illustrate that the Kleven et al. (2019b) approach suffers from the methodological issues raised for TWFE regressions.

[3]Existing studies for Germany using administrative labor market data (e.g. Schönberg and Ludsteck, 2014; Frodermann et al., 2023; Melentyeva and Riedel, 2023) only consider births for mothers who were employed shortly before birth. Most studies in the literature also restrict attention to mothers who were employed shortly before birth (e.g. Lalive et al., 2014; Stearns, 2018).

Eligibility for parental leave benefits hinges upon being employed, thus creating a work incentive before giving birth.

Our econometric method uses an approach for staggered treatment adoption as in Fitzenberger et al. (2013) and Sianesi (2008) which combines a semiparametric event study approach with a control group of non-mothers to estimate the causal motherhood effect. Stacking data for each possible calendar month of birth allows to compare mothers with not-yet-mothers avoiding comparisons with mothers who had given birth before. Similar to Fitzenberger et al. (2013) and Cengiz et al. (2019), we then aggregate the effect over all calendar months of birth weighting by the number of mothers in each calendar month, separately before and after the reform. We control in a flexible way for a rich set of covariates, the monthly labor market history aligned twelve months before birth, and general time trends. The child penalty estimates are then used to determine the reform effect and to account for various selection effects. These methodological differences to all other papers on the 2007-reform and to most papers on child penalties (for example Angelov et al. (2016) and Kleven et al. (2020)) allow us to assess the change in the selection of mothers due to the reform and to distinguish this composition effect from the causal reform effect.

The literature mostly ignores possible fertility effects of parental leavel policies changing the selection of mothers when such a policy is in place. The parental leave reform analyzed in this paper replaced the old means-tested parental leave benefits lasting for up to two years which generates both reform winners and losers. Raute (2019) shows that the reform increases fertility among women with higher pre-birth earnings ("reform winners"). Many studies use a RDD-approach in time to estimate the effect of a change in parental leave policy in order to obtain a selection-free estimate (e.g. Lalive et al. (2014), Kluve and Schmitz (2018)). This is justified as concrete timings or manipulating of birth dates are hardly possible.[4] Thus, around the reform date there is no change in the selection of mothers - except for possible seasonality effects (Lalive et al., 2014). Hence, the RDD approach by definition misses changes in the selection of mothers due to fertility effects of the reform.[5] Our estimation approach, however, accounts for the fertilty effect of the reform.

The literature on the effects of parental leave policies presumes that the negative effects of longer employment interruptions after birth on labor market outcomes are likely to be higher for women with higher earnings (Olivetti and Petrongolo, 2017). Thus, in addition to a possible change in overall fertility, a second selection effect may involve a change in the composition of mothers with regard to the size of the child

[4] In many cases like the German reform in 2007 the announcement of the reform was less than 10 months before implementation which makes strategic behavior virtually impossible.

[5] Lalive et al. (2014) and Stearns (2018) report that there is no change in fertility in response to a change in parental leave policies without providing detailed results.

penalty. If the child penalty is higher among high-earning women and a policy change increases their fertility rate compared to other women, then the average child penalty would increase in size.

Another important effect on pre-birth employment and earnings arises when the introduction of parental benefits results in an entitlement effect such that the labor market attachment before and during pregnancy is increased, an effect which to our knowledge has rarely been analyzed in the literature (Olivetti and Petrongolo, 2017; Rossin-Slater, 2018).[6] Because being employed may entitle mothers to (higher) parental leave benefits after the reform, women have an incentive to find or maintain a job, which by itself may increase their labor market prospects after birth, thus involving a potential positive effect of paid parental leave on labor market outcomes.

The child penalty for mothers often entails a shift from full-time work to part-time work, which may be fostered by a legal right to part-time work Olivetti and Petrongolo (2017). Such a right was introduced in Germany in 2001 (Fitzenberger et al., 2013), which is likely to contribute to the high share of part-time work among mothers. Part-time work can have ambiguous effects on long-term labor market outcomes for mothers. On the one hand, part-time work may reduce human capital accumulation and career progression. On the other hand, part-time work may allow mothers of young children to remain attached to the labor market.

The existing literature almost exclusively focuses on the birth of the first child and the subsequent labor market outcomes, often neglecting second order fertility (Olivetti and Petrongolo, 2017; Rossin-Slater, 2018). We look into the latter because more than 50% of first-time mothers in Germany have a second child within five years after first birth (Fitzenberger et al., 2013). Thus, second order fertility adds to the child penalty and could be affected itself by the parental leave reform.

Our main findings are as follows: The 2007 parental benefit reform in Germany strongly reduces employment and earnings during the first year after birth, while it has small positive effects afterwards. The positive reform effects decrease when job protection ends after three years. There is no significantly positive effect on full-time employment and on second order fertility. The reform shows a positive effect on first-time births among high-earning women and among non-employed women. There is a noticeable entitlement effect with higher earnings and employment rates before birth. Furthermore, our analysis reveals some interesting and plausible heterogeneity of the reform effect. Finally, the methodological differences between our approach and other common approaches prove important. In contrast to our control group approach, RDD and event study estimates show no medium-run reform effects on earnings and employ-

[6]This entitlement effect is mentioned in the study by Olivetti and Petrongolo (2017, page 214) and Girsberger et al. (2023) provide supportive evidence in the case of Switzerland.

ment.

The remainder of this paper is organized as follows. Section 2 discusses the reform and the literature on it, Section 3 contains a description of the data and Section 4 shows the reform effects on fertility. Section 5 describes the estimation approach for identifying the child penalty and reform effects which are robust to changes in the composition of mothers. The empirical results are discussed in Section 6 and Section 7 compares our approach to alternative estimation strategies. Section 8 concludes and the appendix contains detailed findings.

2 The 2007 parental benefits reform

The 2007-reform in Germany introduced one of the most generous parental benefits systems worldwide. All mothers of children born from January 2007 onward – previously employed or non-employed – are entitled to the new parental benefits until the 12th month after giving birth. For mothers of children born until December 2006, eligibility for benefits depended upon the household's income and the amount was flat and in most cases substantially lower.

2.1 Institutional setting before and after the reform

From six weeks before the expected date of delivery until eight weeks after giving birth, employed mothers have to take a mandatory leave from work during the so-called "maternity protection" (Mutterschutz) period while receiving the "maternity allowance" (Mutterschaftsgeld) replacing 100 percent of the pre-birth earnings.[7] This remained unchanged by the 2007 reform (BMFSFJ, 2020b).

Under the old system, the eligibility for further benefits depends on the mother's and potentially her partner's income.[8] Mothers in high income households received no financial benefit from the third month after giving birth onward. Medium and low income households were paid the "child raising allowance" (Erziehungsgeld) of € 300 per month. Mothers in medium (low) income households were eligible until the sixth (24th) month after birth.[9]

Since 2007 all employed women are eligible for parental benefits after the maternity protection period. The level of benefits (Elterngeld) varies by the individual pre-birth

[7] To be precise, the pre-birth earnings are calculated as average of the three calender months before the start of the pregnancy. It is paid jointly by health insurances and the employers.

[8] The yearly net-income of the mother if she raised her child as single-mother and the combined income of her and her partner if they lived in one household was used as relevant quantity. For both groups of mothers different income thresholds were used.

[9] Mothers in low income households had the second option to receive higher monthly benefits (€ 450) only during the first year and zero benefits afterwards. However, this option was rarely used (BMFSFJ, 2004; Kluve and Tamm, 2013).

earnings. Mothers receive 65 percent of their previous monthly earnings during the 12 months before entering "maternity protection" but at least € 300 and at most € 1800. They may obtain benefits until the end of the first year after giving birth (Ehlert, 2008).[10] Before and after the reform, mothers lose eligibility for benefits, when working full-time, but working part-time for up to 30 hours weekly is possible. Furthermore, mothers enjoy job protection (Elternzeit) for three years after birth (BMFSFJ, 2020a), which remained unchanged by the reform.

Figure 1 shows in a stylized way how potential parental benefits for mothers depend upon household income in the pre- and post-reform period. It is easy to distinguish reform-winners from reform-losers. Medium and especially high income households receive unambiguously longer and higher benefits while low-income households may fare worse compared to the old system. In a nutshell, Germany changed from a system paying higher benefits to low-earners to one paying higher benefits to medium- and high-earners.

[Figure 1 about here.]

The system Germany introduced in 2007 is one of the most generous parental benefits systems worldwide. Figure 2 orders various OECD countries with parental benefits by the average replacement rate in terms of the previous earnings. The height of the bars gives the maximum duration of benefits. Most of the countries considered have both a smaller replacement rate and a shorter duration (like France and the UK). Some replace 100% but for a much shorter time period (like Mexico and Spain). Only Poland has a system similar to Germany. This makes Germany an interesting showcase for a country with generous parental benefits in combination with three year job protection.

[Figure 2 about here.]

2.2 Literature on the 2007-reform

The existing literature focuses on the effect for mothers who were employed before giving birth. Using an RDD approach comparing mothers giving birth shortly before and shortly after the reform, Kluve and Tamm (2013) and Geyer et al. (2015) document a strong fall strongly during the first year after giving birth due to the negative work incentives during benefit receipt (see Figure 1). In contrast, based on a regression approach, Bergemann and Riphahn (2011) find a positive effect on early return to job.

Turning to medium-run effects, the RDD-estimates of Kluve and Schmitz (2018) involve significant positive effects on employment until five years after giving birth which

[10] Single-mothers are eligible for parental benefits for two additional months. Fathers are also targeted by the reform. From 2007 onward, both parents may receive benefits. However, effects on fathers cannot be analyzed here because our data do not allow us to identify fathers.

are driven by medium- and high-earners. The probability to work full-time after five years fell for low-earners but increased for high earners. Using a DiD-approach to control for potential selection issues between first and fourth quarter mothers, Frodermann et al. (2023) find a significant positive reform effect on earnings two years after birth. For low-earners, that is the only significant effect, while for high-earners the positive effect though declining remains significant afterwards. After eight years it is insignificant.[11] Bergemann and Riphahn (2020) undertake an event study. The median mother returns earlier to work but there is no effect on employment after the end of the job protection period (from year four onward). The medium-run effects (after five years) on earnings and employment are lower than the short-run effects but still significantly positive. Generally, the reform winners with high pre-birth earnings are found to drive the positive effects.

Two studies address fertility effects of the reform. Cygan-Rehm (2016) finds a strong temporary decrease in higher-order fertility, an effect driven by the reform-losers, but the negative effect turns insignificant over time. Raute (2019) shows that first and higher order fertility increase for high-earners compared to low-earners. These findings highlight the importance to investigate to what extent changes in the selection of mothers drive the reform effect on labor market outcomes, a point which so far has not been addressed in the literature. For instance, Frodermann et al. (2023) and Kluve and Schmitz (2018) use a DiD and RDD approach, which by construction defines away a possible selection effect of the reform. Bergemann and Riphahn (2020) analyze mothers who give birth out of employment between 2005 to 2008, which means that changes in the selection of mothers could be part of the estimated reform effect. In a robustness analysis, their results do not change much when using a narrower time window.

To the best of our knowledge, we are the first who estimate the child penalty combining an event study approach with a control group and quantify the reform effect on the selection of mothers accounting explicitly for general time trends apart from the reform effect. No attention has been paid in the literature on reform effects for non-employed mothers and on selection issues regarding the possibly heterogeneous child penalty, both of which we also address in this paper.

3 Data

We link two different high-quality administrative data sources, namely fertility data from the pension insurance and administrative employment records.

[11] Geyer et al. (2015) finds the low earners even to benefit above average in the short-run.

3.1 Administrative employment records

The first data source are the Integrated Employment Biographies (IEB) available at the Institute for Employment Research (IAB) involving all employees in Germany except for civil servants and self-employed. The data provide precise information on the timing of employment spells, gross earnings, part-time status, education, and the district of employment.[12] We transform the spell data to a monthly panel with daily earnings.[13] There is considerable overreporting of full-time employment up to 2011 which became apparent by the change in the notification procedure in 2011. We use our correction procedure developed in Fitzenberger and Seidlitz (2020) to correct for this problem.[14]

3.2 Pension insurance data and linkage to IEB

The data source from the pension insurance is the "Versichertenkontenstichprobe" (VSKT) (FDZ-RV, 2021). This includes both spell data on employment and for mothers it mainly involves the exact date of birth of each of their children. The latter information is extremely valuable as it is not recorded precisely in the IEB data. The VKST data are augmented for our analysis by the IEB data using a probabilistic matching approach because no personal identifier can be used.[15] The matching works well because both data sources are based on the same employment spells reported by employers and because the month of birth of the employee is available in both data sources.

Furthermore, we introduce a few additional minimum criteria on the employment history to decrease the risk of mismatches.[16] Arguably, this approach runs the risk that the sample has an above average labor market attachment.[17] At the same time,

[12]For detailed information see the documentation of the SIAB, the scientific use file comprising a two percent sample of the IEB (Antoni et al., 2019).

[13]Therefore, we sum up earnings for up to three different jobs within a month but we exclude employments with a duration of less than 14 days or with earnings below 300 Euros in one month. Earnings are deflated by the annual consumer price index of the Federal Statistical Office (Statistisches Bundesamt, 2019).

[14]See Figure 20 for a graph on this issue and its correction.

[15]The data linkage was conducted within the project "Custom Shaped Administrative Data for the Analysis of Labour Market (CADAL)", as part of the SPP "The German Labor Market in a Globalized World: Challenges through Trade, Technology, and Demographics" which was funded by the German Research Foundation (DFG). The matching algorithm was applied to about 306,000 female observations of the pension insurance data source. We needed to drop those which are born before 1964 or after 1988 and those who had a child before 2004 as it is described in the following section. This left us with 89'000 observations. For 74 percent, a unique match in the IEB was found. Further details are available upon request.

[16]Given the probabilistic matching strategy, it is clear that the accuracy increases with the length of employment history available. We therefore require the observations to have at least worked for two years and to report at least ten spells in total and five employment spells in the IEB. The reporting quality of spells in marginal employment is a further potential issue. Therefore, we require the average earnings to exceed the marginal employment earnings threshold. Further, we exclude all women with an employment history in pre re-unified East-Germany which is a minor concern for the birth cohorts considered.

[17]Our approach to estimate the child penalty requires that mothers and non-mothers, used as control

our sample is much less restrictive than the selection of mothers which arises when motherhood is directly identified from the IEB. Müller and Strauch (2017) show how motherhood can be detected from the information on employment in the IEB data based on the reported reason that an employment spell is interrupted (Schönberg and Ludsteck (2014) use a similar approach). This approach works well for first-time mothers who are employed immediately before entering motherhood. In contrast, our linked data involve information on all births including both mothers who are not employed before birth and higher-order births.

3.3 Descriptives

We consider mothers who have their first child within three years before or after the reform, i.e. between 2004 and 2009, and who are aged between 21 and 40 at the time of child birth (for the mothers and the control observations these are the birth cohorts 1964 to 1988). As a simple comparison group, we keep all women who remain childless until 2010, regardless whether they later have a child.

These restrictions leave us a sample of 51,000 females, a bit more than 10,000 of them have their first child during the period of interest. Table 1 shows some descriptives for our sample. Relative to the average control observations at the time of birth, mothers are a bit better educated: 88 percent hold at least a secondary educational certificate compared to 81 percent for the controls. The differences in age and region of residence are small.

[Table 1 about here.]

Earnings, employment, full-time status, and second-order fertility are the main outcome variables of our analysis. Table 1 shows a dramatic decline in the means of the three labor market outcomes when the pre-birth period is compared to the post-birth period for mothers. Mean earnings fall to less than half of the pre-birth earnings, employment decreases by 30, full-time employment by even more than 40 percentage points (pp).

Furthermore, the group of mothers shows better labor market outcomes before giving birth than so far childless women. Mothers are positively selected in both labor market outcomes and educational achievement, which shows the need to account for pre-birth differences in our estimation approach. Finally, a bit more than 50 percent of the mothers have a second child within five years while only about 22 percent of the so far childless women have a first child.

observations, have worked for some time before or after giving birth. For this reason, we do not think that the sample selection due to the linkage procedure causes a systematic difference between mothers and control observations.

4 Selection effect of the reform

Our analysis of fertility effects builds and extends upon the work by Raute (2019) who showed that the reform increased the propensity to have a child for high-earning women relative to low-earners. One of the goals of the parental leave benefits reform was to increase the low birth rate in Germany, especially for high-earning women with a high labor force attachment. Following Raute (2019), we estimate the DiD regression

$$child_{iy} = \alpha lm\_outcome_{iy} + \beta lm\_outcome_{iy} \times post_y + \gamma X_{iy} + \lambda_y + \epsilon_{iy} \qquad (1)$$

where $child_{iy}$ is a dummy for having a first child in year y, $lm\_outcome_{iy}$ is the labor market outcome of interest (log-earnings or dummy for earnings lieing above threshold of € 8,550), $post_y$ is the postreform dummy, which is equal to one from 2007 onward, X_{iy} are other covariates, and λ_y is the year fixed effect. The interaction effect captures differences in the reform effect on fertility by labor market outcomes relative to the general time trend. We use the second year before giving birth as "pre-treatment" period for the labor market outcomes in both specifications as anticipation effects could confound the values of the year directly before giving birth.

[Table 2 about here.]

The right panel of Table 2 shows the time trend in the birth rate among women, who have not had a child until the year before. There is an upward trend in the birth rate increasing steadily from 4.2% in 2004 to 4.8% in 2009. There is no disproportionate increase in the birth rate in 2007 (+ .1 pp compared to 2006) but a disproportionate increase in 2008 (+ .2 pp compared to 2007). Though the increase in 2008 should not be interpreted as causal, it suggests a positive reform effect.

The left panel of Table 2 involves two regression specifications. The upper specification controls for non-employment and annual earnings in the second calendar year before birth. The birth rate increases due to the reform significantly both for high-earners and non-employed relative to employed women with low earnings. The regression estimates imply that for employed women with earnings above (below) approximately € 17,500, the increase in fertility is larger (smaller) than for non-employed women. The lower specification, following Raute (2019), controls for whether earnings lie above or below € 8,550, because women with earnings above € 8,550 receive higher total benefits after the reform while low-earners may even receive less benefits after the reform depending on the household income (section 2). The findings show a weakly significant increase in fertility for women whose earnings were above the threshold. According to the first specification, the increase in fertility for women at the 75th percentile (€ 30,000) of the earnings distribution is 2.2 pp (an increase of 49 percent compared to the fertility in 2006). That is considerably larger than the increase of 1.9 pp for women earning € 8,550 (39th percentile). Though these effects should not be interpreted as causal, they suggest strong differences in the reform effect on fertility.

Starting from this finding, we next develop our approach to the estimate effects of the parental benefits reform.

5 Econometric Approach

Our econometric approach involes two parts. First, we estimate the causal effect of motherhood on various post-birth outcomes, the child penalty, using a semiparametric event study approach with a control group. Second, we decompose the difference in the post- and pre-reform child penalty into the causal reform effect and the composition effect driven by the changing selection of mothers.

5.1 The child penalty

We estimate the average treatment effect for the treated (ATT) on post-birth outcomes based on discrete time data similar to Fitzenberger et al. (2013). The treatment is 'first child birth at a certain point in time' (a calendar month) against the counterfactual of waiting, which entails the two alternatives having a child at a later date and never having a child.[18] The control group varying by treatment month only includes nontreated women who have not had a child yet (see subsection 5.1.1). This accounts for the staggered treatment, thus preventing forbidden comparisons of first-time mothers with women who had a child before (Roth et al., 2023; Melentyeva and Riedel, 2023), and avoids conditioning on future outcomes. We match treated mothers and nontreated women by observable characteristics twelve months before the actual child birth assuming that the gestation period starts 11 months earlier, accounting for likely anticipation effects.

The key identification assumption for our analyis is a dynamic conditional independence assumption (CIA). It states that conditional on the variables controlled for the assignment to treatment in this time period is random, i.e. independent of the potential outcomes. Specifically, our dynamic CIA stipulates that given the duration of childlessness and given the covariates, having a first birth within the next year does not depend upon nontreatment outcomes. Formally, the dynamic CIA is

$$E(Y_{0,t}(t+\tau)|D_t = 1, D_1 = ... = D_{t-1} = 0, X_t) \quad (2)$$

[18]The treatment effect we estimate is an example for a dynamic treatment effect as considered in the evaluation of active labor market policies by Sianesi (2004) and Biewen et al. (2014).

$$= E(Y_{0,t}(t+\tau)|D_t = 0, D_1 = ... = D_{t-1} = 0, X_t),$$

where D_t is the treatment dummy in month t, X_t involves all - possibly time-varying - covariates and lagged labor market outcomes, t is treatment time (in our case 12 months before birth), and $\tau = 1, 2, 3, ...$ is time relative to the treatment time t. $D_t = 1, D_1 = ... = D_{t-1} = 0$ involves the treatment '(first-time) motherhood in $t + 12$', $D_t = 0, D_1 = ... = D_{t-1} = 0$ involves the counterfactual 'remaining childless until $t + 12$', and $Y_{0,t}(t + \tau)$ is the labor market outcome in period $t + \tau$ in case of no motherhood up to $t+12$, which is the counterfactual for those having a child in $t+12$.

The CIA can be motivated as follows: Our rich administrative data allow to control for a number of socio-economic characteristics and for detailed labor market history (X_t). Conditional on these, the treated women are not likely to differ systematically with regard to their non-treatment outcome $Y_{0,t}(t + \tau)$ from women who stay childless until one year after the birth of the child. This is because the exact timing of birth cannot be planned with certainty and may depend upon random circumstances not reflected in long-run labor market outcomes. At the same time, women differ in their probability to have a child within the next year which is also reflected by the characteristics controlled for.[19]

For estimating the child penalty, we do not exclude the alternative of giving first birth at a later date. This corresponds to the fact that fertility decisions are taken jointly with career decisions and that women repeatedly make choices regarding fertility throughout life, such that each month a woman decides to remain childless until the month the decision is made to have her first child. This results in a dynamic selection of first time mothers at a certain date. Thus, using solely a control group of women who do not give first birth until a much later date or who will never have a child would bias the control group due to further dynamic selection towards women with a low propensity of having a child. This bias is likely to be correlated with labor market outcomes (e.g., women with a strong labor market attachment could be more likely to never have a child).

The treatment group consists of women who have their first child between the age of 21 and 40. The control group varying with calendar time of actual birth consists of women who are still childless when the 'treated group' gives birth. We measure the treatment effect at a monthly frequency. Our analysis uses an estimate of the average counterfactual outcome based on the individual-specific comparison group.

[19]Note that our approach is robust to mothers knowing the probability of having a first birth now versus later and them acting upon its determinants (Abbring and Van den Berg, 2003).

5.1.1 Temporal alignment

Our evaluation approach requires a temporal alignment between treated mothers and controls. We define the event time as the month of giving first birth and the treatment time as the month -12. If women in the control group remain childless until one year after the birth event, they are aligned to the treated mothers.[20] Figure 3 provides a graphical illustration of event and treatment time. All periods are defined relative to the event time, i.e. the birth month for the treatment group. Therefore "12 months before the event" is refered to as month -12 or "treatment period". This reflects that the birth of a child is decided upon at time of conception implying anticipation effects during the gestation period.

[Figure 3 about here.]

When the gap between treatment time and event time, which we denote as classification window, is shortened below one year, the control group also includes women who, at the event time, are pregnant or already have concrete plans to become mothers very soon. As main robustness checks, we investigate the sensitivity of the findings to shortening the classification window to 0 months or to extending it to 2, 3, 4, or 5 years. With 0 months, the control group includes all women giving first birth after the birth event considered. A classification window of x ($x = 2, 3, 4, 5$) years excludes from the control group all women giving first birth during x years after the birth event considered, thus reducing the control group to women with no plans to have a first child within the next x years. A classification window of five years covers the maximum period for which we estimate the child penalty, so that the control group involves only women for whom no post-first-birth outcomes would be included in the analysis.

We consider each event (birth) month from January 2004 to December 2009 as separate treatment cohort with calendar month specific control group, similar to Fitzenberger et al. (2013); Cengiz et al. (2019), thus avoiding comparisons of treated mothers with women who had given birth before. For each of these 72 months, we pool treated women, who have her first child in this month, and all potential control women (non-mothers), who remain childless during the classification window considered (one year or for (0, 2, 3, 4, or 5) years as robustness check). We require all observations to be between 21 and 40 years of age at the event time. This way, we generate a $person \times$ *event-calendar-months* $[2004, 1 \ \textit{to} \ 2009, 12] \times$ *month-relative-to-event-time* $[-35 \ \textit{to} \ 59]$ data set used for our analysis. For *month-relative-to-event-time*, 0 denotes the month

[20] We do not impute a random placebo treatment time for control observations (as discussed in Kleven et al. (2019a)). We do not restrict the control group to never-mothers (during the observation window) because at the treatment time this would condition on the future. Only for never mothers, it would be plausible to simulate placebo treatment times.

of birth, -35 means 35 months before birth, and 59 means 59 months after birth. Effectively, the panel data set used for our empirical analysis consisted of women who are duplicated for each (potential) treatment month $t' \in$ [2004,1 to 2009,12] - note that $t' = t + 12$ where t is the treatment time and t' is the event time of child birth. Our approach stacks observations for different event times: A woman may be used in one treatment period as treated and potentially multiple times as control observation earlier, including the case of never having a child. $t' \in$ [2004,1 to 2006,12] represents the pre-reform period and $t' \in$ [2007,1 to 2009,12] the post-reform period.

Take as an example a woman who gives first birth in April 2006. Until April 2005, marked blue in Figure 4, she serves as a control observation. From May 2005 until March 2006, she is not used as then her own child will be born in less than one year, which reflects the classification window. In April 2006, she finally enters the sample as treated and afterwards, for all event months until December 2009, she is not part of the sample - neither as treated nor as control observation - because she is no longer childless. This way we avoid any forbidden comparisons of a treated mother with a women who had had her first child before.

[Figure 4 about here.]

Stacking observations as described results in a large data set. Our original data set involves about 51,000 women (see Table 1). For our analysis, we have 10,727 treatment and about 2.6 million control observations (person-treatment-month observations). The key advantage of the temporal alignment is that the outcomes for both treated and control observations can be analyzed relative to the event time, as in Figure 5. The control group of not-yet-mothers is younger than the group of mothers. We therefore show earnings and employment for the two groups both pre-reform and post-reform with and without age adjustment. Reweighting the control group to have the same age distribution as the group of mothers, the pre-treatment outcome differences between the two groups decrease (see subfigures (b) and (d) of Figure 5).

[Figure 5 about here.]

5.1.2 Inverse Probability Weighting (IPW)

To control the selection of mothers, we use propensity scores for giving birth in a certain calendar month. We estimate separate models for the pre-reform and the post-reform period while pooling across event months within each subperiod. Specifically, we model the probability of having the first child as a function of socio-demographic characteristics and labor market history as observed twelve months before birth.

Under the unconfoundedness of treatment, eq. (2), and perfect overlap in the propensity score, Busso et al. (2014) find that in small samples with estimated propensity score, a modified inverse probability weighting estimator (IPW) performs well. The crucial modification involves the normalization of weights for the nontreated women.

Our analysis has to account for the fact that the group of eligible comparison women changes by month of birth, see Section 5.1.1. Correspondingly, the alignment between treated and nontreated observations changes by age and month of birth. Recall, that we estimate the average treatment effect for the actual mothers (ATT) in different event months t', which is the child penalty on labor market outcomes at different *months-relative-to-event-time*. The estimation by IPW involves steps: First, we estimate the treatment effect separately for each event month t', eq. (3),

$$\hat{\theta}_{t'} = \frac{\sum_{i=1}^{n} T_{i,t'} Y_{i,t'}}{\sum_{i=1}^{n} T_{i,t'}} - \frac{\sum_{i=1}^{n} (1 - T_{i,t'}) \hat{W}_{i,t'} Y_{i,t'}}{\sum_{i=1}^{n} (1 - T_{i,t'}) \hat{W}_{i,t'}} , \tag{3}$$

and second, we weight these treatment effects by the number of treated observations, eq. (4),

$$\hat{\theta} = \frac{\sum_{t'=1}^{T} \sum_{i=1}^{n} T_{i,t'} \hat{\theta}_{t'}}{\sum_{t'=1}^{T} \sum_{i=1}^{n} T_{i,t'}} . \tag{4}$$

n is the total number of women in the data set. $T_{i,t'}$ denote the treatment dummy variable for individuals i in event month t' (treated and non-treated), respectively. The first fraction in equation (3) is the (unweighted) average of the treatment group and the second fraction is the average of the control group with weights $\hat{W}_{i,t'}$:

$$\hat{W}_{i,t'} = E_{i,t'} \times \frac{\hat{p}_{t'}(X_{i,t})}{1 - \hat{p}_{t'}(X_{i,t})}. \tag{5}$$

$E_{i,t'}$ is a dummy variable for eligibility as nontreated observation which takes the value of one if woman i can be used as a control observation for event month t'. Otherwise, $E_{i,t'}$ is set to zero giving this woman zero weight in event month t'. Thus, $E_{i,t'}$ is a dynamic non-treatment dummy. Our baseline analysis sets the classification window to one year. Hence, $E_{i,t'}$ is equal to one if i remains childless for at least one year after the respective birth month. $\hat{p}_{t'}(X_{i,t})$ denotes the estimated probability to have a first child in time period t' as a function of covariates $X_{i,t}$ which can vary over time and display the covariates at the treatment period $t = t' - 12$. The application of the estimated weights $\hat{W}_{i,t'}$ leads to a reweighting of the nontreated women according to the odds–ratio of having a child within the next year (based on $t = t' - 12$).

The IPW reweighting estimator has the advantage of not relying on a tuning parameter.[21] Moreover, it is easy to implement and standard errors are readily obtained by bootstrapping. The probability to give first birth in month t' given the characteristics $X_{i,t}$ is estimated by a probit regression based on the observations in the aforementioned

[21] No trimming of $\hat{p}_{t'}(X_{i,t})$ close to one is necessary in our application.

duplicated data set at *month-relative-to-event-time* $= -12$ months, i.e. one year before giving birth in the duplicated data set. The characteristics we use to determine the chance to be a mother are an indicator for working in former East-Germany ($east_{it}$) and dummies for medium and high education (low education serves as reference category). Further, we include fixed effects for the calender months, years and age at treatment. Arguably, most important are the controls for the employment history. We include three times 24 variables for the earnings, employment and full-time status of the second and third year before giving birth (months -35 to -12 with respect to giving birth). This regression is conducted separately for four age groups (women aged 21 to 25, 26 to 30, 31 to 35 and 36 to 40) and separately for the pre-reform and the post-reform period. We use the following specification for treatment month t:

$$\begin{aligned} P(mother_{i,t'} = 1|X_{i,t}) = \Phi \Bigg(\beta_0 + \beta_1 east_{it} + \sum_{l=22}^{40} \beta_l I(age_{it} = l) + \sum_{j=2}^{3} \gamma_j I(edu_{it} = j) \\ + \sum_{k=-35}^{-12} (\delta_k earn(k)_{it} + \alpha_k empl(k)_{it} + \theta_k ft(k)_{it}) + \lambda_m + \mu_y \Bigg) \end{aligned} \tag{6}$$

with $\Phi(.)$ representing the standard normal distribution function and $X_{i,t}$ comprising the aforementioned partly time-varying covariates considered in eq. (6). The fitted value of $P(mother_{i,t'} = 1|X_{i,t})$ are used as $\hat{p}_{t'(i)}$ in eq. (5).

5.2 From child penalty to reform effect

$\hat{\theta}_{pre}$ and $\hat{\theta}_{post}$ are the estimated ATTs of motherhood in the pre-reform and post-reform period, respectively. Estimating the reform effect by the simple pre-post-difference $\hat{\theta}_{post} - \hat{\theta}_{pre}$ would ignore possible changes in the selection of mothers and could only be used in an RDD approach within a narrow time window around the reform. The reform was intended to increase fertility, thus it can result in a change both in the composition of mothers and in the motherhood effect.

Using a DiD and RDD approach, Frodermann et al. (2023) and Kluve and Schmitz (2018) focus on first time mothers between October 2006 and March 2007. They argue that the mothers giving birth during this six month period were unaware of the reform at the time of conception and that therefore selection into motherhood did not change with the reform. By construction, this local RDD approach estimates the refom effect only for a sample of mothers unchanged by the reform, thereby ignoring the composition effect. For a comprehensive assessment of the reform, the possible selection effects are important and therefore, we outline in the following our approach to estimate both the causal reform effect on mothers and the composition effect caused by a changing selection of mothers.

5.2.1 Composition and reform effect

In the following, we describe our approach to separate the reform effect and the composition effect. The results reported in Section 4 for regressing motherhood on pre-birth labor market history and its interaction with the reform dummy reveal that women with higher pre-birth earnings have an increased probability to become mother in the post-reform period. This difference in pre-birth outcomes can be thought of as the first dimension of the selection effect. It may drive differences in the child penalty between pre- and post-reform periods, a selection effect on post-birth outcomes, the composition effect.

To assess how this composition effect affects the child penalty, we use a second IPW to reweight the sample of pre-reform observations to the sample of post-reform mothers. This reweighting allows to estimate the counterfactual child penalties - the ATT - on post-birth outcomes that would have applied for a sample of mothers as observed in the post-reform period but facing the pre-reform parental benefits rules. This way our approach estimates the pure reform effect on the child penalty for the post-reform sample of mothers and not for the pre-reform sample of mothers as does the RDD approach.

Estimating this counterfactual ATT procedes in two steps:

1. We take the post-reform mothers and separately the sample of pre-reform mothers and the sample of pre-reform control women and the four age groups (21 to 25, 26 to 30, 31 to 35 and 36 to 40). For each of the eight samples, we run a separate Probit regression to estimate the probability that a woman gives birth in the post-reform period as a function of pre-birth characteristics and history as observed 12 months before birth against the alternative of being a pre-reform mother or a pre-reform control person, respectively. This probability $P(post_{i,t'} = 1, mother_{i,t'} = 1|X_i)$ - for a person i in the pre-reform sample - is specified as a function of age, education, region (east, west), and labor market history (earnings, employment, full-time employment) during the second and third year before birth. The labor market outcomes are averaged for the four half-year periods $[-35, -30]$, $[-29, -24]$, $[-23, -18]$, and $[-17, -12]$ (months-relative-to-event).

2. The counterfactual ATT of motherhood is estimated by reweighting pre-reform observations - for both treated and controls separately - with

$$\hat{G}_{i,t'} = E_{i,t'} \times \frac{\hat{P}(post_{i,t'} = 1, mother_{i,t'} = 1|X_{i,t})}{1 - \hat{P}(post_{i,t'} = 1, mother_{i,t'} = 1|X_{i,t})},$$

where $\hat{P}(post_{i,t'} = 1, mother_{i,t'} = 1|X_i)$ are the respective fitted probabilities from the Probit regressions in step 1. The counterfactual pre-reform ATT is then estimated by

$$\hat{\theta}(07/09,\text{pre})_{t'} = \frac{\sum_{i=1}^{n} T_{i,t'}\hat{G}_{i,t'}Y_{i,t'}}{\sum_{i=1}^{n} T_{i,t'}\hat{G}_{i,t'}} - \frac{\sum_{i=1}^{n}(1-T_{i,t'})\hat{G}_{i,t'}Y_{i,t'}}{\sum_{i=1}^{n}(1-T_{i,t'})\hat{G}_{i,t'}}\,, \tag{7}$$

$$\hat{\theta}(07/09,\text{pre}) = \frac{\sum_{t'=1}^{T}\sum_{i=1}^{n} T_{i,t'}\hat{\theta}(07/09,\text{pre})_{t'}}{\sum_{t'=1}^{T}\sum_{i=1}^{n} T_{i,t'}}\,. \tag{8}$$

The approach is implemented analogously to eq. (3). The treatment effect for event month t' is given by the difference between treatment and control average. Both averages are reweighted with weights $\hat{G}$, reweighting the pre-reform sample to mimic the post-reform sample of mothers. The pre-reform observations, event periods from January 2004 to December 2006, receive a weight $\hat{G}_{i,t}$ which is high (low) for those who are (not) comparable to the mothers of the year from 2007 to 2009. Here, $i = 1, ..., n$ denotes the sample of women observed for the pre-reform period. Then $\hat{\theta}(07/09,\text{pre})$ is the estimate on the child penalty of the sample of mothers as it is observed between 2007 and 2009 but under a hypothetical treatment by pre-reform parental benefits.

The difference between the factual ATTs of motherhood in the pre- and post-reform period is the raw difference

$$\text{raw\_diff} = \hat{\theta}_{post} - \hat{\theta}_{pre}\,. \tag{9}$$

However, this is not the causal reform effect because the treated mothers before and after the reform may differ systematically in their socio-economic characteristics and labor market history, which may themselves result in a change in the child penalty. The effect on the change of the selection of mother (composition effect) on the ATT can be estimated by

$$\text{comp\_eff} = \hat{\theta}(07/09,\text{pre}) - \hat{\theta}_{pre}\,. \tag{10}$$

This composition effect quantifies how the counterfactual child penalty for the post-reform sample would have differed in the pre-reform period from the child penalty as observed for the pre-reform sample. This difference arises because of the change in the sample of mothers due to the reform. The pure causal reform effect of the ATT of motherhood – child penalty – for the post-reform sample of mothers is then given by

$$\begin{aligned}\text{reform-eff} &= \text{raw\_diff} - \text{comp\_-eff}\\ &= \hat{\theta}_{post} - \hat{\theta}(07/09,\text{pre})\end{aligned} \tag{11}$$

This reform effect measures for the post-reform sample of mothers how the causal effect of motherhood changes due to the reform.

6 Baseline Findings

Applying the estimation approach described in Section 5, we now present the results on earnings, employment and full-time employment. We also show how the child penalty differs by age group and pre-birth earnings. Finally, we consider second order fertility.

6.1 Overall labor market outcomes

Figure 5 shows labor market outcomes for mothers and non-mothers pre- and post-reform after temporal alignment with and without adjustment for age but without controlling for differences in socio-demographics and labor market history. Apparently, there is a strong positive selection of mothers relative to non-mothers, reflected in the difference for the pre-treatment period (second and third year before giving birth). It corresponds to the positive estimates in Table 2 for the labor market outcomes without interaction.

Figure 6 depicts earnings and employment profiles for the group of treated mothers and the control group of non-mothers. After applying the IPW-approach described in Section 5.1, the control group is aligned to the treated mothers with regard to the covariates considered and labor market history up to month -12. Compared to Figure 5, when showing the profiles after temporal alignment, the two lines for the non-mothers are shifted (dotted light [pre-reform] and dotted dark [post-reform]). In the pre-treatment period until one year before giving birth, the outcomes for the control group match very precisely those of the mothers both pre- and post-reform.[22] These findings make us confident that our econometric specification for the propensity score is sufficiently flexible to achieve comparability between treated mothers and the control group.

[Figure 6 about here.]

The differences between the treatment group and the control group after IPW estimate the child penalty, i.e. the ATT of motherhood from *time-to-event* 11 months before birth onward. Figure 7 shows the precisely estimated ATTs for **earnings** in the post-reform sample ($\hat{\theta}_{post}$), the pre-reform sample ($\hat{\theta}_{pre}$), and in the reweighted pre-reform sample ($\hat{\theta}(07/09, pre)$). Following a similar qualitative pattern for all three groups, there is a dramatic decline in earnings during the months directly before birth and then a slow recovery after birth. Despite these similarities, there are significant effect differences between the reweighted pre-reform sample and the post-reform sample. These differences represent the estimated reform effect.

[22] Using the weights for the pre-reform sample to mimic the post-reform mothers (eq. (7), not shown in Figure 6, results are available upon request) gives a treatment and control group also matching the labor market history of the post-reform mothers.

[Figure 7 about here.]

Table 3 provides annual aggregates of the monthly effects depicted in Figure 7. The reform has its largest effect in the short run during the first three years after giving birth. As reported in the literature, there is a strong fall in earnings during the first year and a moderate increase in the second and third year. The general job protection for mothers ends after three years and a number of mothers returns to work after month 36. This reduces the still positive earnings effect in the medium run – four and five years after birth. The reform increased daily earnings in the fifth year after giving birth by € 1.86 (an increase of close to € 700 in annual earnings). Furthermore, there is a significant increase in earnings in the first year before birth being visible for the months -9 to -4 and amounting to € 0.60 Euro in daily earnings. To our knowledge, our study is the first to demonstrate this "entitlement effect" for Germany which arises because after the reform mothers have an incentive to remain employed until birth due to parental benefis depending upon previous earnings (see Girsberger et al. (2023) for similar evidence on Switzerland). In total over the six years (starting from month -12), we estimate an insignificant total earnings gain of roughly € 1,100, a rather small cumulative earnings increase. Evidently, the positive reform effect from the second year onward compensates the strong negative earnings effect in the first year after birth. The estimated pattern suggests that the cumulative reform effect increases further beyond year 5, but a rigorous assessment would require further analysis.

[Table 3 about here.]

We do not detect a significant difference between the unweighted and reweighted pre-reform sample with regard to outcomes from month -11 onward. Note that the differences in Figure 7 during months -36 to -12 are very small and insignificant, as they should be according to our identifcation strategy. While the reform is likely to change the composition of mothers towards higher fertility among both high-earning women and zero-earners and there is a small positive entitlement during the year before birth, there is no noticeable average composition effect in eq. (11) concerning the overall reform effect. Below, we will explore whether this finding of no composition effect masks opposing reform effects for different groups of mothers.

Figure 8 and Table 3 show that the effect of motherhood on **employment** is qualitatively very similar to that for earnings. The same also holds for the reform effect, being strongly negative during the first year after birth and positive afterwards, and getting smaller from the fourth year onward. In the post-reform period, mothers are up to 6 pp more likely to be employed in the second and third year and about 2.5 pp in the fifth year after birth compared to the pre-reform period. The employment interruption for some mothers is reduced from 36 months in the pre-reform period to less than 15 months in the post-reform period (Figure 8). There is also a small entitlement effect during the first year before birth of about 1 pp.

[Figure 8 about here.]

After the reform, a large earnings gap between mothers and the control group still remains in the medium run. An important part of that can be explained by the higher part-time rate among mothers. For this reason, we also provide results on **full-time employment** (figure 9 and Table 3). During the first three years the reform effect is similar as for overall employment but its size is much smaller. In the fourth and fifth year, the reform effect is negative and insignificant. The entitlement effect before giving birth is of similar size as the effect on employment indicating that after the reform mothers predominately increase full-time employment before birth.

Altogether, this means that the positive reform effects on earnings and employment from the fourth year onward solely stem from additional part-time employment.

[Figure 9 about here.]

6.2 Heterogeneous effects by age and pre-birth earnings

Results on heterogeneous reform effects by the age of the mother are shown in Figures 10 to 12 and the upper panels of Tables 4 and 5. The child penalty differs considerably by age, with young mothers showing the largest earnings and employment penalty (figures 10,11). The medium-term reform effects on employment and earnings are on the other hand positive for mothers over the age of 26 and drive the positive overall effect. The oldest group, 36 to 40 years, benefits the most. This group shows a significant earnings increase of € 7,500 (roughly 7 times the average effect size) up to five years after birth. The age group 26 to 30 year shows significantly positive effects on overall employment but negative effects on full-time employment. In contrast, the youngest age group loses due to the reform, with negative and insignificant effects on employment and a highly significant decline in cumulative earnings by more than € 4,800.

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Table 4 about here.]

[Table 5 about here.]

Concerning pre-birth earnings, we split the sample into three groups according to their earnings in the third and second year before giving birth (months -35 to -12) based on Section 4. Fertility increases both among women with zero earnings and among women with annual earnings above € 17,500 relative to the group of low-earners with positive earnings below € 17,500. The reform effect is estimated separately for these three groups, see Figures 13 to 15 and the lower panels of Tables 4 and 5.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

The high-earners benefit strongly from the reform. Their earnings gains in the medium run and in the year before birth (entitlement effect) as well as the cumulative gains since conception are above average. In contrast, the low-earners show no significantly positive effects. In fact, the group experiences a fall in full-time employment in the medium run and a significant cumulative earnings loss.

The results for zero-earners are quite interesting because they are a small group that has not been studied separately and that shows very different results from those for low-earners. This small group has the largest confidence intervals. However, the results for this group indicate both a very strong positive entitlement effect and positive medium-run effects on earnings. In both cases, the point estimates are three times larger than for the average mother. The increased incentives result in higher employment of the mothers before birth and a lasting higher labor market attachment after birth.

The findings for low- and high-earners are similar to those reported in the existing literature for previously employed mothers. Bergemann and Riphahn (2011), Frodermann et al. (2023), and Kluve and Schmitz (2018) find that the reform winners (with high pre-birth earnings) benefit above average while the reform losers do not gain in the medium run. However, our findings are at odds with Geyer et al. (2015) arguing that the low-earners driving the positive medium-run reform effects.

The results for the different age groups are partly consistent with those for different levels of pre-birth earnings. Table 6 shows that the youngest age group is the only group with a majority of low-earning mothers while the other three age groups have around 70 percent high-earning mothers. The zero-earners involve only around 5 percent of the observed sample and in all four age groups their share is less than 10 percent. However, certain patterns, such as the very high positive effect for the oldest, are not driven by pre-birth earnings. Therefore, investigating the effects by age and for zero-earners separately contributes important novel findings to the literature.

[Table 6 about here.]

None of the subgroups gives evidence for a large composition effect. The differences in the selection are insignificant for all groups according to age and the pre-birth earnings.

6.3 Findings for second order fertility

[Figure 16 about here.]

Since one of the goals of the reform was to increase fertility, we also investigate how the reform affected the propensity to have a second child. There are no significant effects of the reform, as shown in Figure 16 and Table 3, with only small and insignificant differences in the cumulative incidence of a second birth for the pre- and post-reform cohorts. These results are partly consistent with Cygan-Rehm (2016), who finds temporary but no lasting reform effects on second order fertility. However, a caveat of our and her analysis is that the birth of a second child for most of the pre-reform cohort occurs after the reform date, which may attenuate the estimated reform effect.

7 Alternative estimation strategies

We now contrast our findings with alternative estimation strategies for the child penalty using different control groups and event-study estimates. Concerning the reform effects, RDD-estimates and event-study estimates are presented. We also focus on the precision of the different estimates.

7.1 Child penalty estimates using event study approach

The widely cited studies Kleven et al. (2019b,a) use an event study approach without control group to estimate the child penalty based on a TWFE regression for a staggered treatment, i.e. post-birth outcomes of mothers are contrasted with pre-birth outcomes of future mothers (Roth et al., 2023; Melentyeva and Riedel, 2023).[23] Kleven et al. (2020) extend the framework to estimate reform effects on the child penalty using an interaction of the event-dummies with a post-reform dummy. We implement this

[23] If the motherhood effect varies (e.g. by age), these forbidden comparisons may bias the TWFE treatment effect estimate due to forbidden comparisons. Our baseline approach avoids such forbidden comparisons because the control observations only uses not-yet-mothers.

approach by estimating the following regression

$$
\begin{aligned}
Y_{iym} = & \sum_{\substack{j=-35 \\ j \neq -12}}^{59} \alpha_j I(ym - t_i = j) + \sum_{k=-35}^{59} \beta_k I(ym - t_i = k) \times post_y \\
& + \pi east_{iym} + \sum_{j=1}^{3} \theta_j I(edu_i = j) + \sum_{l=19}^{45} \gamma_l I(age_i = l) + \lambda_y + \delta_m + \epsilon_{iym} \,, \qquad (12)
\end{aligned}
$$

for the sample of first-time mothers only. Here, Y_{iym} is the labor market outcome of interest of mother i in year y and calendar month m (January,...,December) [ym is an integer counter for the observation month] and t_i denotes the month of birth (event time). Hence, $ym - t_i (= j)$ measures time (in months) to birth, α_j ($j = -35, ..., 59; j \neq -12$) measures how the outcome variable varies by time to birth relative to month -12 in the pre-reform period, and β_k how this changes in the post-reform period, which is the reform effect. This measure of the (causal) child penalty takes the difference of post-birth outcomes to the pre-birth outcome at -12, i.e. the latter is taken as the comparison level for the non-birth outcome after controlling for general time and age effects and other time-invariant individual characteristics. This approach works well to account for the strong immediate effects after birth, which are likely to be much larger than the changes in the counterfactual situation. However, as Kleven et al. (2019a) discuss themselves, the event study estimates for the medium-run outcomes may be biased because the counterfactual changes as well, beyond what is captured by the time and age effects. In particular, the counterfactual may differ between soon-to-be mothers and women not having a child any time soon. This implies that the pre-birth outcomes for mothers may not provide an appropriate estimate of the counterfactual.

[Figure 17 about here.]

Figure 17 contrasts our baseline estimates (for the pre-reform period) in Figures 7 and 8 with the event study estimates for α_j. Both for employment and earnings, the event study estimate shows basically zero effects during the second and third year before birth and also matches well the baseline estimates one year pre- and post-birth. However, from the second post-birth year onward, the event study estimates diverge strongly from the baseline estimates. Hence, the event study estimates imply a persistent and much larger child penalty compared to our baseline estimates. This is because the pre-birth outcomes of mothers provide a higher comparison level compared to the baseline control group. Furthermore, the gap increases over time due to the general upward trend in employment and earnings among women pre-birth (section 5.1.1 and Figure 5). In contrast, our baseline control group contains women who will be first-time mothers later on, from which onward the gap to the treatment group is reduced.

7.2 Child Penalty with alternative control groups

Inspired by the DiD literature for staggered treatments (Roth et al., 2023; Borusyak et al., 2021), we compare our baseline results to different variations of the control group. We use five alternative control groups which differ in the classification window. The baseline classification window is one year meaning that the control group contains women who give birth earliest one year after the treated observations. The approach using a classification window of zero years contains all observation who are childless at the month of giving birth for the treatment group,. It results hence a control group which is larger than in the baseline approach. The control groups used for the other four classification windows (of two, three, four, and five years) decrease in the length of the classification. Conceptually the choice determines from which time after the event onward (month after birth) later treated observations are included in the control group and how they influence the child penalty estimates. Using a classification window of five years removes all observations giving birth to a child in the observed five year period such that this provides estimates using a control group of women not having a first child during the observation window.

Figure 18 shows that the short-run differences in child penalties compared to the baseline increase for each of the approaches until the end of the classification window, i.e. when the first control observations enter motherhood. Until the end of the classification window, the child penalty coincides for earnings with the event-study estimates, but it lies in between the baseline and the event-study estimates for employment. After the end of the classification windows, the child penalties converge to the baseline estimates, suggesting that the length of the window matters mostly for the short-run effects.

[Figure 18 about here.]

There are different drivers of the gap between the baseline and an approach using only never-mothers during the observations period (classification window of five years): First a dynamic development in the outcome in absence of motherhood and second the child penalty which reduces the outcome for those in the baseline control group entering motherhood later on.[24] Considering the differences by age groups (not shown here but available upon request), it turns out that the event-study estimates and the approach using a classification window of five years may differ consideraly for earnings. This means the similarity for the average masks strong age differences.

In sum, we conclude that the upward trend in employment and earnings among women pre-birth is the major reason that the event-study estimates of the child penalty are higher than for our baseline estimates. However, when excluding soon-to-be first-time mothers by means of a longer classification windows for our control group, the

[24] A third factor would be the number of control observations entering motherhood.

child penalty increases as well as long the women in the control group do not have a child. After women in the control group start having a child, the child panelty converges to our baseline estimates. Note, however, that a longer classification window conditions on future outcomes and treatment decisions, which is why we prefer a classification window of one year for our baseline estimates.

7.3 Comparison of estimated reform effects

Figure 19 shows estimated reform effects based on our baseline control group estimates, the event study (the estimates for β_k of (12)), and RDD estimates. For RDD, we replicate the approach of Kluve and Schmitz (2018) which is based on births three months before or after the reform implementation and which does not use a control group. The most noteworthy finding is that the baseline results are very precisely estimated and often differ significantly from the point estimates of the two other approaches, when ignoring the much larger statistical variation of the two other approaches. At the same time, the baseline estimates lie within the much larger confidence intervals of the two other approaches. This means that the alternative estimates are unlikely to differ significantly from the baseline estimates.

The RDD estimates of the reform effect prove quite noisy. The estimated reform effect is a bit smaller than the baseline during the first year after birth and afterwards it is quite similar to the baseline estimates. The event study estimates are more precise than the RDD estimates but still more imprecise than the baseline. They suggest a positive earnings effect already in the second and third year before birth and a stronger effect than the baseline during the first year before birth. The event-study effect basically coincides with the baseline effects during the first year after birth for employment and for the first three years for earnings, but it is considerably smaller later on. In fact, the event study estimates do not imply a significantly positive medium-run reform effect whereas the baseline effects prove significantly positive for almost all months from the second post-birth year onward. The approaches with different classification windows yield very similar results as the baseline (available upon request).

Some important methodological conclusions follow. The RDD approach is too noisy to detect significant reform effects. The event-study approach also yields less precise estimates than the baseline estimates. It tends to find stronger positive reform effects before birth than baseline, possibly confusing the reform effect with the general upward trend in earnings and employment, and thus overestimating the entitlement effect. And it underestimates the positive reform effects on employment and earnings after the first post-birth year. Both the RDD and the event-study approach are also unlikely to detect heterogeneous reform effects and miss the impact of the changing selection of mothers through the fertility effect of the reform.

[Figure 19 about here.]

8 Conclusions

This paper estimates the causal effect of first-time motherhood, the child penalty, on various post-birth outcomes in Germany, and it investigates the selection of women into motherhood and the effect of the 2007 parental leave reform. Mothers are positively selected in terms of their pre-birth labor market outcomes. The reform changed the selection of women into motherhood, increasing the fertility of both high-earners and zero-earners relative to low-earners.

The reform has negative effects on labor market outcomes during the first year after birth, the period of benefits eligibility, and afterwards small positive effects on earnings and employment. We estimate that in the fifth year after giving birth the reform increased annual earnings by almost 700 Euros and the employment rate by 2.5 pp (6 and 4 percent of the pre-reform levels). The cumulative child penalty until the fifth year after giving birth is however insignificantly affected. We estimate the earning loss due to motherhood to decrease by only approximately 1000 Euro (1 percent).

Medium-aged mothers (31 to 40 years of age) seem to benefit above average while the youngest age group (21 to 25 years of age) shows worse labor market outcomes after the reform. Concerning pre-birth earnings, we find that high-earners but also zero-earners drive the positive average effects. The reform effects on full-time employment in the medium run and on second-order fertility are insignificant. Hence, an increase in part-time employment drives the positive reform effect on earnings and (overall) employment.

Further, we find a positive reform effect during the year before giving birth. This entitlement effect is significant for all three labor market outcomes, which to our knowledge our study is the first to demonstrate for Germany. This effect arises because the parental leave benefits in the post-reform period depend on earnings of mothers immediately before birth. Our findings suggest that the attachment to the labor market trough the entitlement effect might be a channel for the positive medium-run effect.

The evidence on the reform effect fits quite well the existing literature which finds positive medium-run effects on earnings and employment per se but not for full-time employment (Frodermann et al., 2023; Kluve and Schmitz, 2018). The positive reform effect stems both from mothers with better pre-birth labor market outcomes and from women with zero earnings. Thus, not only women with higher earnings as pointed before (Bergemann and Riphahn, 2020; Frodermann et al., 2023; Kluve and Schmitz, 2018) but also zero-earners benefit from the reform. The losers are employed low-earning mothers, most of whom are still young.

Our econometric approach differs substantially from most of the literature on the child penalty and on assessing the impact of institutional changes. There are three key advantages to our approach: First, we use a flexible semiparametric event-study approach with a control group while spelling out precisely the identification assumptions, thus avoiding conditioning on future outcomes. Second, we use a well defined control group of not-yet mothers at each event time thus avoiding comparisons of mothers with women who had had a child before when estimating the child penalty. Third, our estimation approach goes beyond an RDD approach thus capturing the changing selection of mothers due to the reform.

We contrast our findings to estimates based on alternative approaches. For post-birth outcomes during the first year after birth, the methodological differences do not matter much because findings are driven by the strong dip in employment and earnings for mothers immediately after birth. However, the estimates on the child penalty start to diverge from the second year after giving birth onward compared to an event-study approach as in Kleven et al. (2019b). Our findings using a control group approach in a dynamic treatment setting show a much better medium-run effect of motherhood compared to the event-study approach. Regarding the reform effect, the differences between our approach and other approaches are important as well. While the effects on earnings and employment are insignificant using an event-study or RDD approach, our control group approach implies small significantly positive reform effects on post-birth earnings and employment in the medium run. This means that the choice of the econometric strategy matters in an important way for estimating the effects of the policy reform considered here.

References

Abbring, J. H. and Van den Berg, G. J. (2003). The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517.

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the Gender Gap in Pay. *Journal of Labor Economics*, 34(3):545–579.

Antoni, M., Ganzer, A., and vom Berge, P. (2019). Sample of Integrated Labour Market Biographies Regional File (SIAB-R) 1975 - 2017. FDZ-Datenreport 04/2019 (en), Nuremberg.

Bergemann, A. and Riphahn, R. T. (2011). Female labour supply and parental leave benefits - the causal effect of paying higher transfers for a shorter period of time. *Applied Economics Letters*, 18(1):17–20.

Bergemann, A. and Riphahn, R. T. (2020). Maternal employment effects of paid parental leave. Working Paper Series 2020:6, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Biewen, M., Fitzenberger, B., Osikominu, A., and Paul, M. (2014). The effectiveness of public-sponsored training revisited: The importance of data and methodological choices. *Journal of Labor Economics*, 32(4):837–897.

BMFSFJ (2004). Erziehungsgeld, Elternzeit. Das Bundeserziehungsgeldgesetzt. Technical report, Bundesministerium für Familie, Senioren, Frauen und Jugend.

BMFSFJ (2020a). Elterngeld, ElterngeldPlus und Elternzeit. Das Bundeselterngeld- und Elternzeitgesetz, 23. Auflage. Technical report, Bundesministerium für Familie, Senioren, Frauen und Jugend.

BMFSFJ (2020b). Leitfaden zum Mutterschutz, 16. Auflage. Technical report, Bundesministerium für Familie, Senioren, Frauen und Jugend.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Busso, M., DiNardo, J., and McCrary, J. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, 96(5):885–897.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Cygan-Rehm, K. (2016). Parental leave benefit and differential fertility responses: evidence from a German reform. *Journal of Population Economics*, 29(1):73–103.

Dahl, G. B., Løken, K. V., Mogstad, M., and Salvanes, K. V. (2016). What Is the Case for Paid Maternity Leave? *Review of Economics and Statistics*, 98(4):655–670.

de Chaisemartin, C. and d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, 26(3):C1–C30.

Ehlert, N. (2008). Elterngeld als Teil nachhaltiger Familienpolitik, 3. Auflage. Technical report, Bundesministerium für Familie, Senioren, Frauen und Jugend.

FDZ-RV (2021). Codeplan FDZ-Biografiedatensatz - VSKT 2018. Technical report, Deutsche Rentenversicherung.

Fitzenberger, B. and Seidlitz, A. (2020). The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. *Journal for Labour Market Research*, 54(1).

Fitzenberger, B., Sommerfeld, K., and Steffes, S. (2013). Causal effects on employment after first birth - A dynamic treatment approach. *Labour Economics*, 25(C):49–62.

Frodermann, C., Wrohlich, K., and Zucco, A. (2023). Parental leave policy and long-run earnings of mothers. *Labour Economics*, 80:102296.

Geyer, J., Haan, P., and Wrohlich, K. (2015). The effects of family policy on maternal labor supply: Combining evidence from a structural model and a quasi-experimental approach. *Labour Economics*, 36:84 – 98.

Girsberger, E. M., Hassani-Nezhad, L., Karunanethy, K., and Lalive, R. (2023). Mothers at work: How mandating a short maternity leave affects work and fertility. *Labour Economics*, 84:102364.

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

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child penalties across countries: Evidence and explanations. In *AEA Papers and Proceedings*, volume 109, pages 122–26.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2020). Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation. Working Paper 28082, National Bureau of Economic Research.

Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4).

Kluve, J. and Schmitz, S. (2018). Back to work: Parental Benefits and Mothers' Labor Market Outcomes in the Medium Run. *ILR Review*, 71(1):143–173.

Kluve, J. and Tamm, M. (2013). Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment. *Journal of Population Economics*, 26(3):983–1005.

Lalive, R., Schlosser, A., Steinhauer, A., and Zweimüller, J. (2014). Parental Leave and Mothers' Careers: The Relative Importance of job Protection and Cash Benefits. *Review of Economic Studies*, 81(1):219–265.

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

Müller, D. and Strauch, K. (2017). Identifying mothers in administrative data. Fdz methodenreport, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

OECD (2019). Family Database. Technical report, Organisation for Economic Co-operation and Development.

Olivetti, C. and Petrongolo, B. (2017). The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries. *Journal of Economic Perspectives*, 31(1):205–30.

Raute, A. (2019). Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits. *Journal of Public Economics*, 169:203 – 222.

Rossin-Slater, M. (2018). Maternity and Family Leave Policy. In Susan L. Averett, L. M. A. and Hoffman, S. D., editors, *Oxford Handbook of Women and the Economy*. Oxford University Press.

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

Schönberg, U. and Ludsteck, J. (2014). Expansions in Maternity Leave Coverage and Mothers' Labor Market Outcomes after Childbirth. *Journal of Labor Economics*, 32(3):469–505.

Sianesi, B. (2004). An evaluation of the swedish system of active labor market programs in the 1990s. *Review of Economics and statistics*, 86(1):133–155.

Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics*, 15(3):370–399.

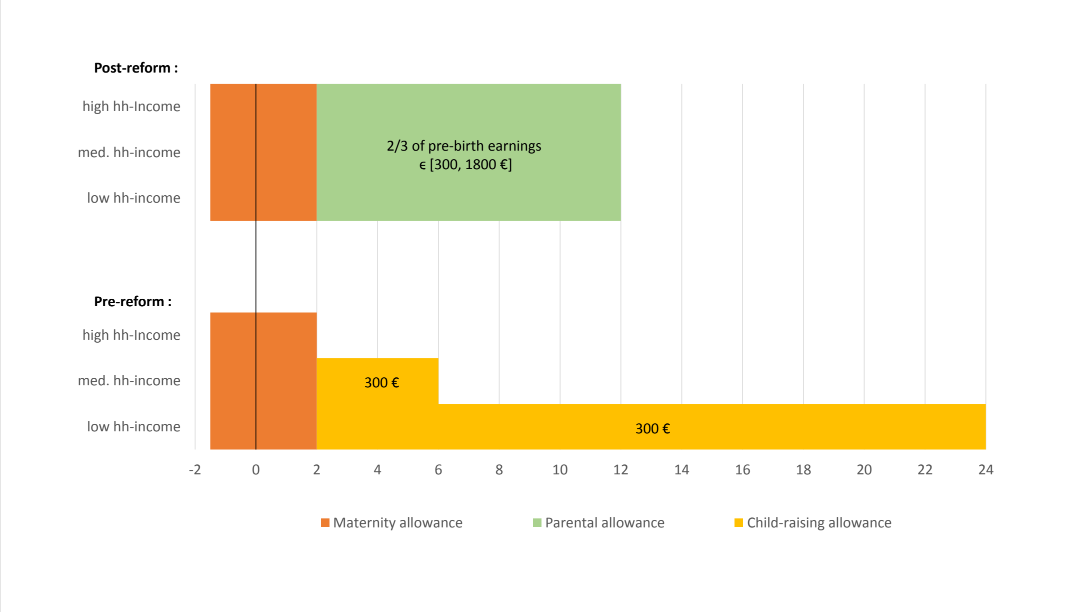
Statistisches Bundesamt (2019). Verbraucherpreisindizes für Deutschland. *Jahresbericht 2018*.

Stearns, J. (2018). The long-run effects of wage replacement and job protection: Evidence from two maternity leave reforms in great britain. *Available at SSRN 3030808*.

Tamm, M. (2013). The impact of a large parental leave benefit reform on the timing of birth around the day of implementation*. *Oxford Bulletin of Economics and Statistics*, 75(4):585–601.

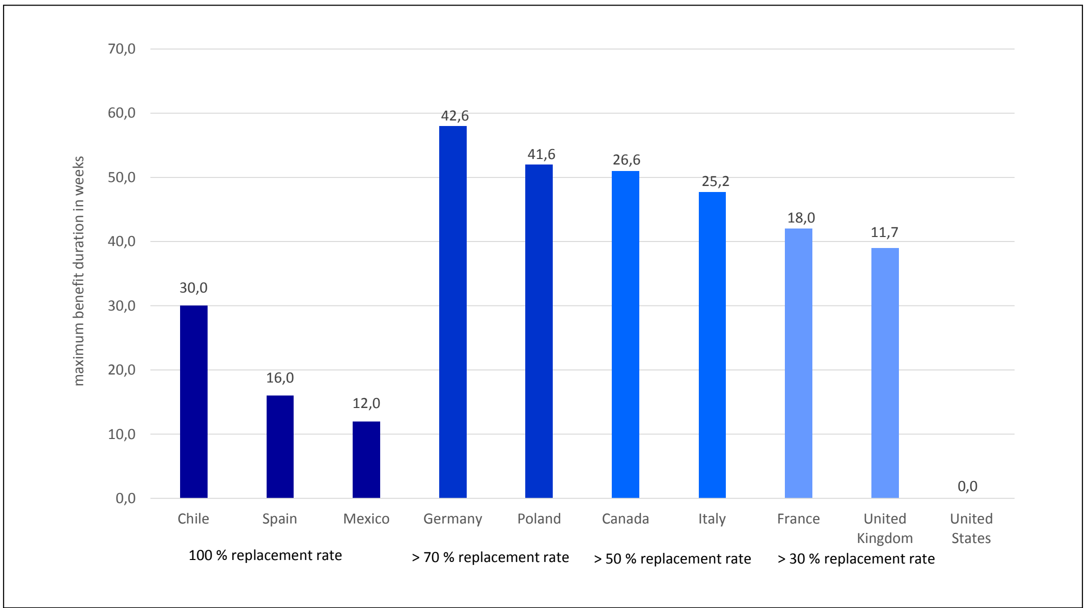
Appendix

Figure 1: Benefits under new and old regimes by household income



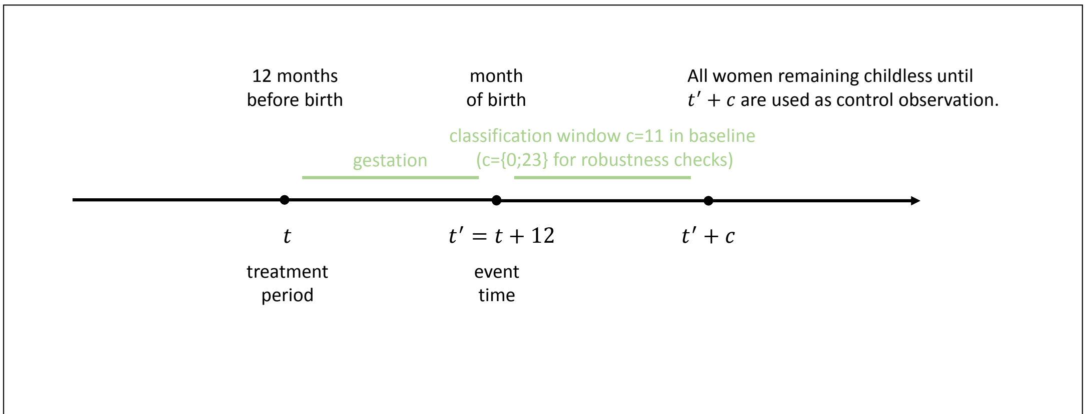
High and medium income households according to the old "child-raising allowance" were households with an yearly net-income above 30000 and 22000 (23000 and 16000 for single-mothers), respectively.
"Parental allowance" is bound to be at least 300 and at most 1800 Euro monthly.
Receivers of both, "child-raising allowance" and "parental allowance" are restricted to work at most 30 hours per week.

Figure 2: Comparison of different parental benefit systems



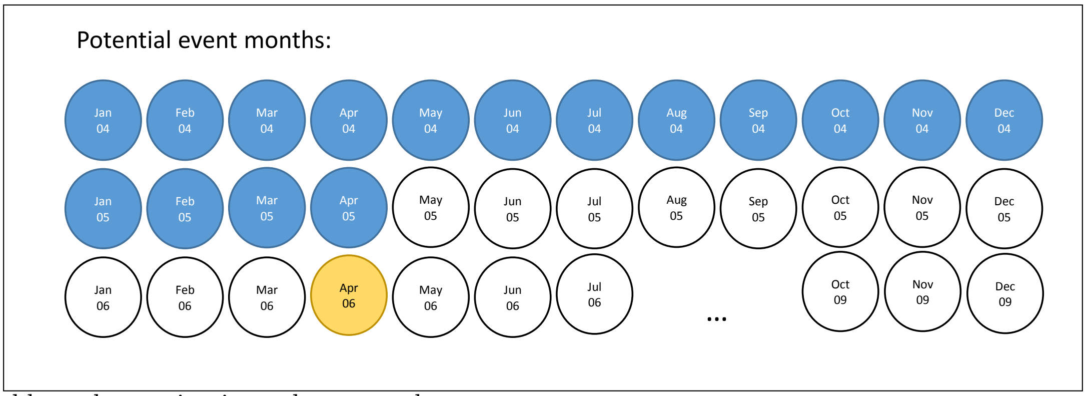
Source: OECD Family Database (OECD, 2019).
Countries are grouped by the average level of replacement rate. The number at the top of the bar is the product of the replacement rate and the maximum duration in weeks. It can be interpreted as the theoretical number of weeks with 100 percent replacement.
For Germany: 14 weeks x 100 % (maternity allowance)+ 44 weeks x 65 % (parental allowance)= 42,6

Figure 3: Timing of treatment and event



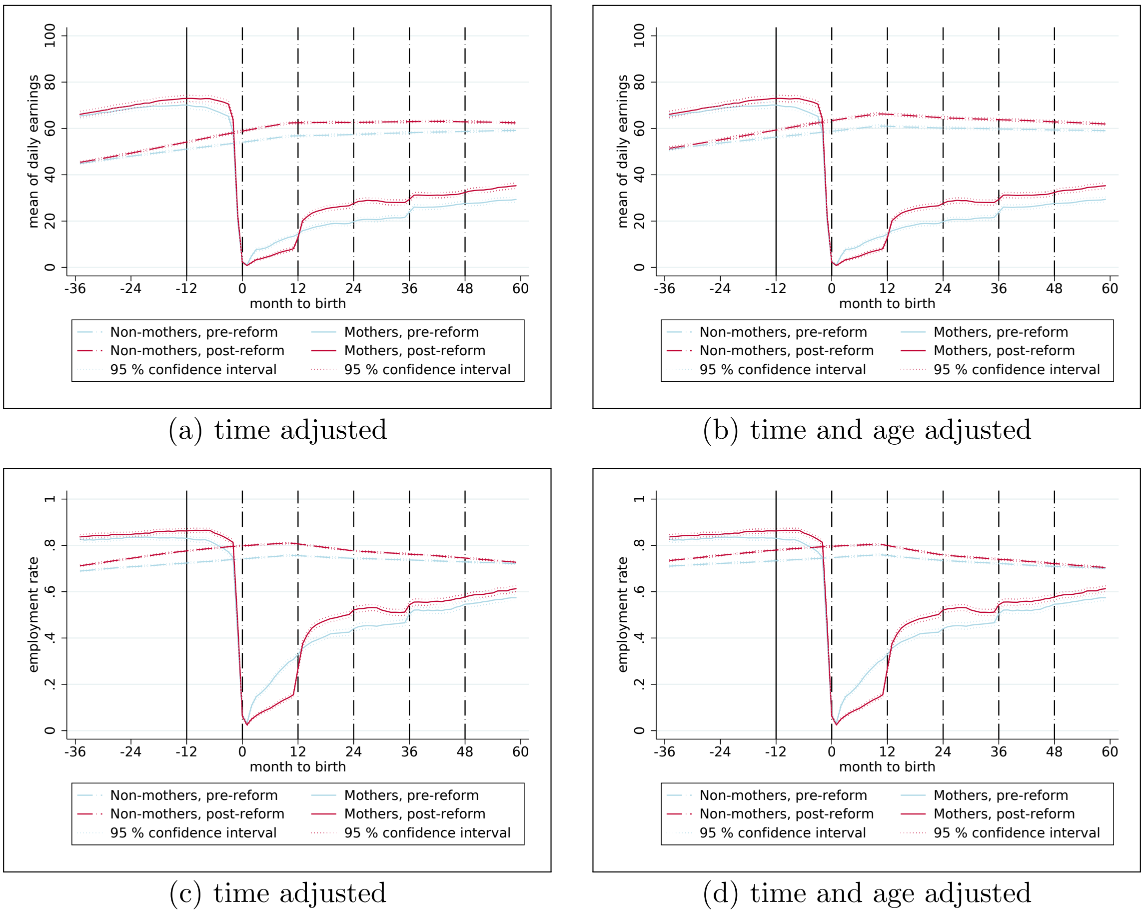
Notes: To illustrate this for the example in Figure 4, consider a woman givin birth in (event) month April 2006. Her treatment period is then April 2005. For the baseline estimates a classification window of one year is used, i.e. all women remaining childless before April 2007 are used as control observation.

Figure 4: Duplication of observations for temporal alignment



blue: observation is used as control
yellow: observation is used as treated
white: observation is not used

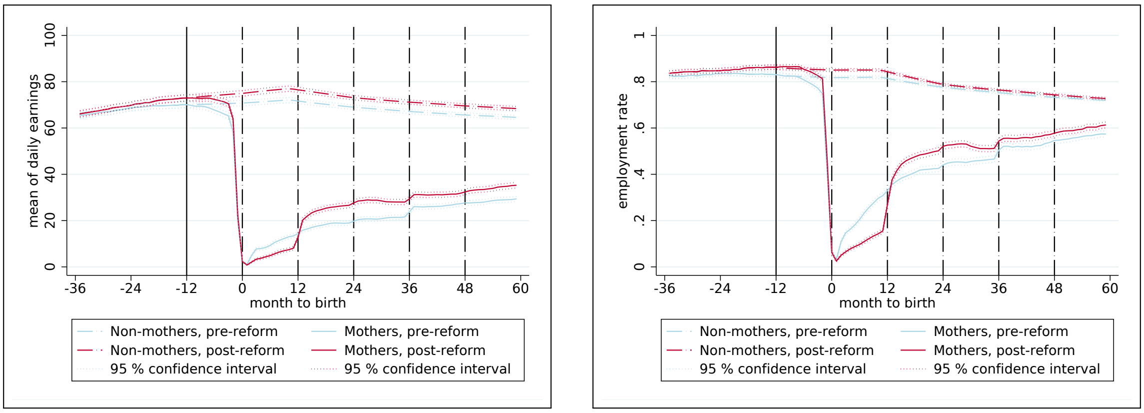
Figure 5: Means of earnings and employment rate after temporal alignment



(a) time adjusted (b) time and age adjusted

(c) time adjusted (d) time and age adjusted

Figure 6: Mothers (treatment group) and non-mothers (control group) after IPW



(a) earnings (b) employment

Figure 7: Child penalty on daily earnings

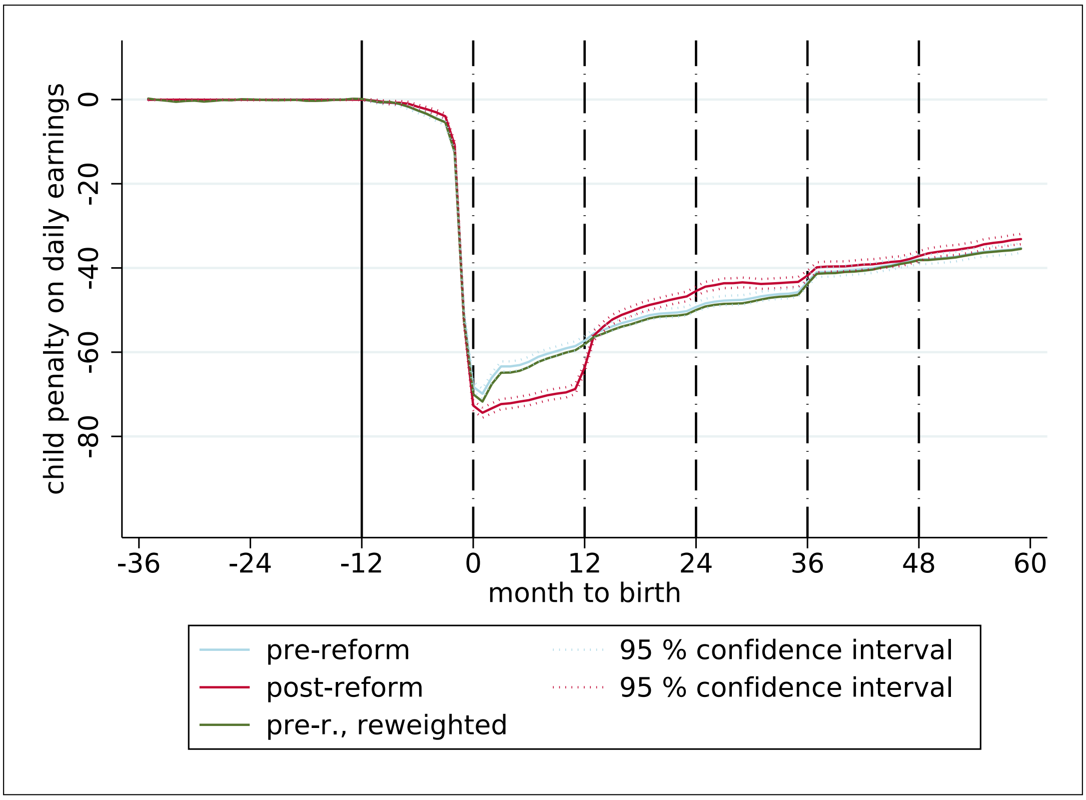


Figure 8: Child penalty on employment rate

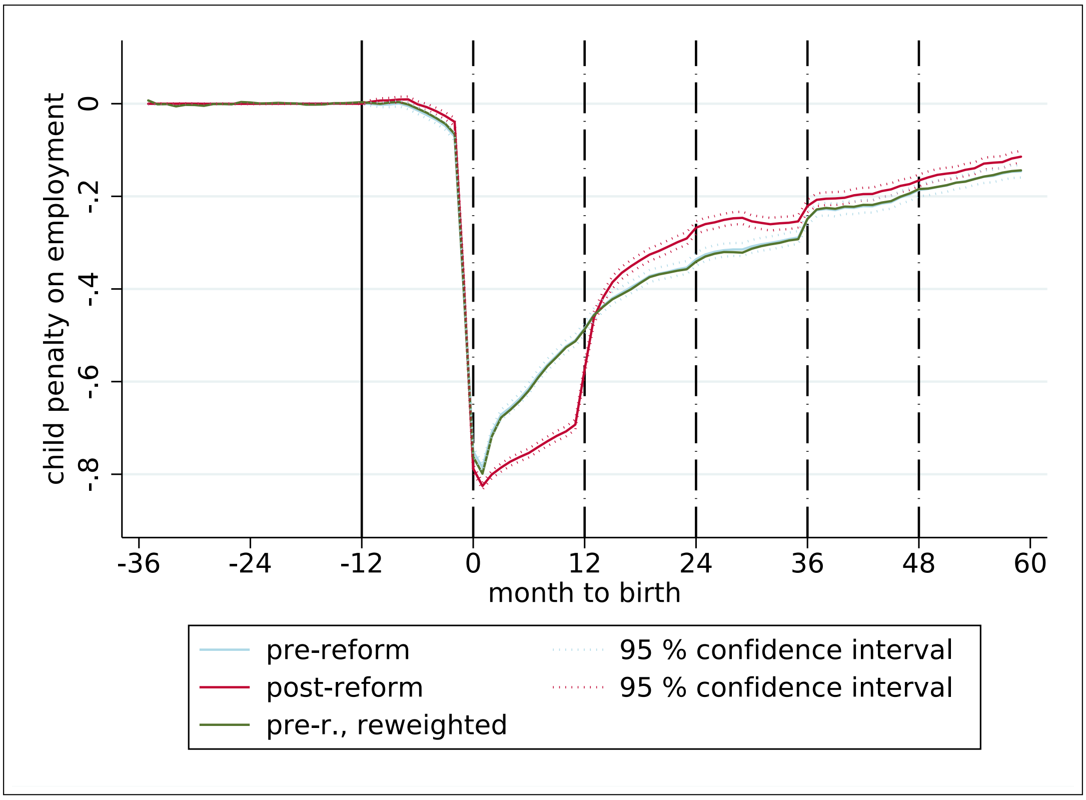


Figure 9: Child penalty on full-time rate

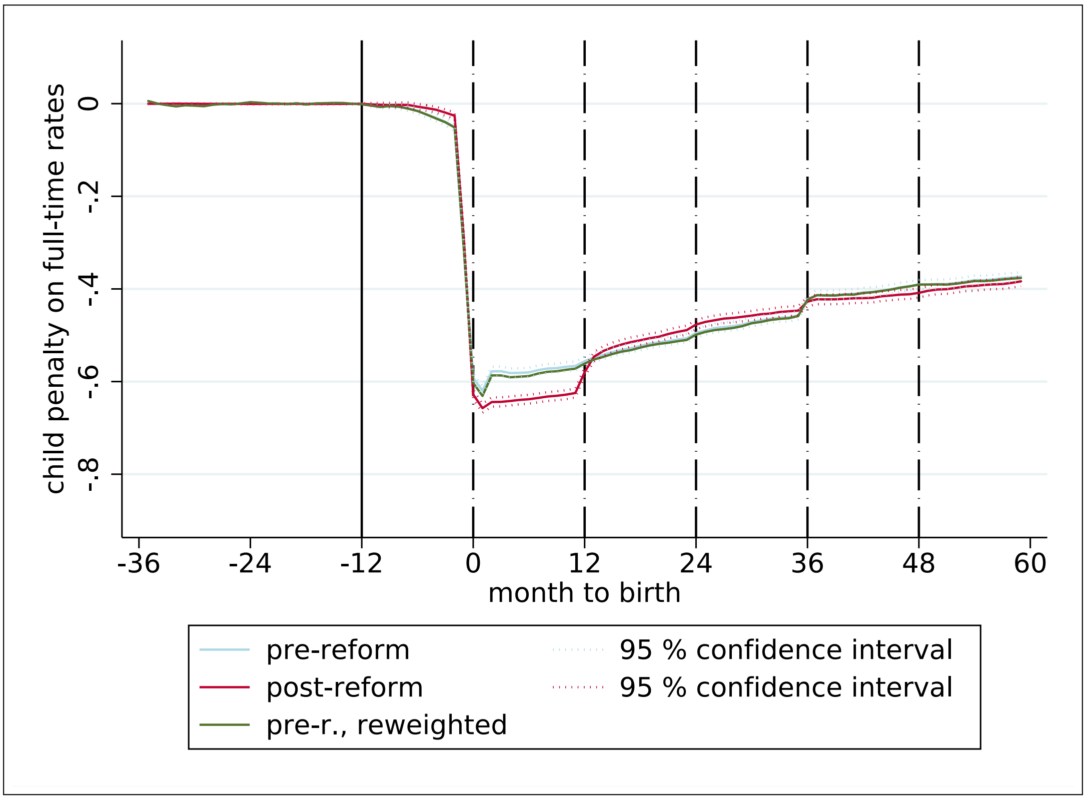
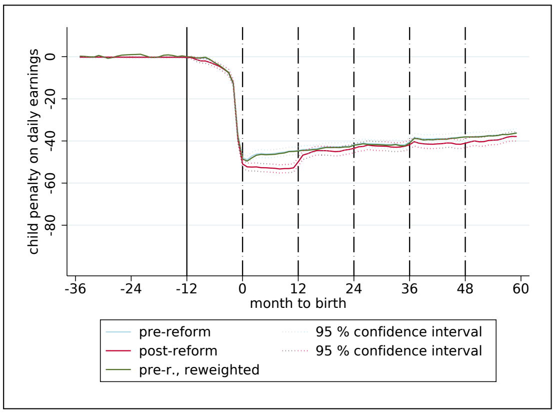
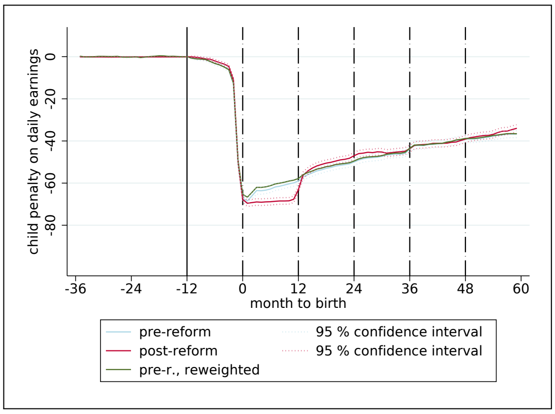


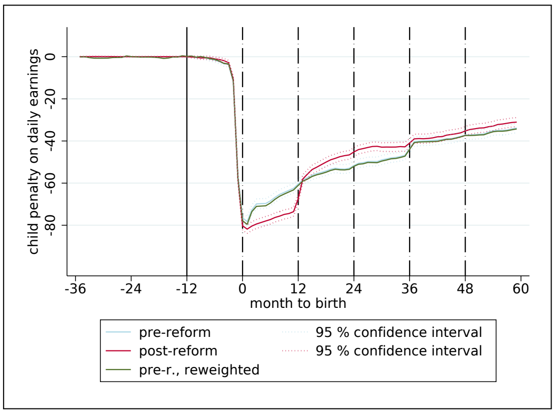
Figure 10: Child penalty on daily earnings according to age groups



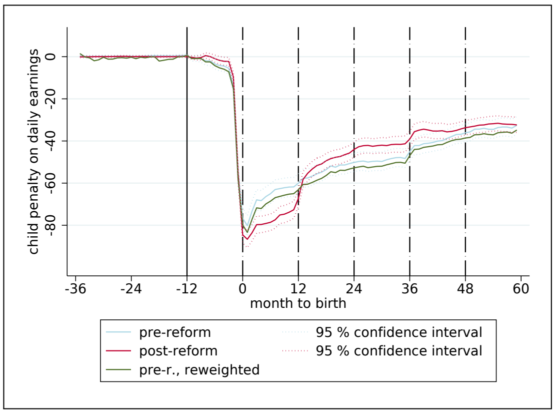
(a) estimates for mothers aged 21 to 25 years



(b) estimates for mothers aged 26 to 30 years

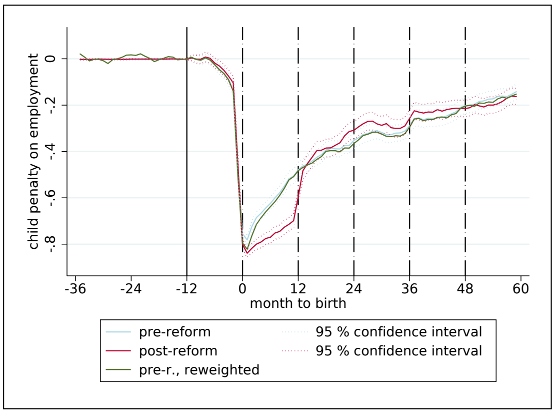


(c) estimates for mothers aged 31 to 35 years

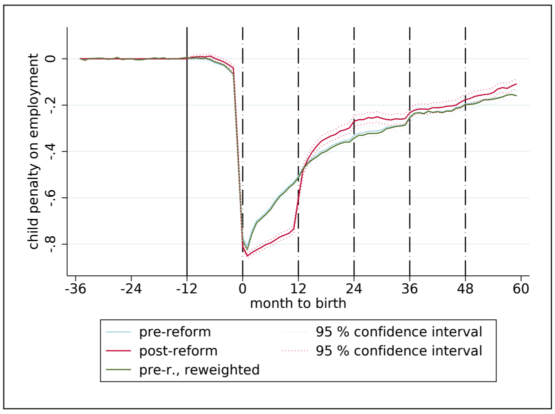


(d) estimates for mothers aged 36 to 40 years

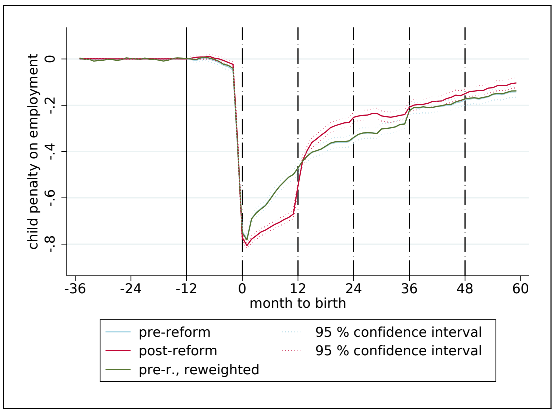
Figure 11: Child penalty on employment rate by age group



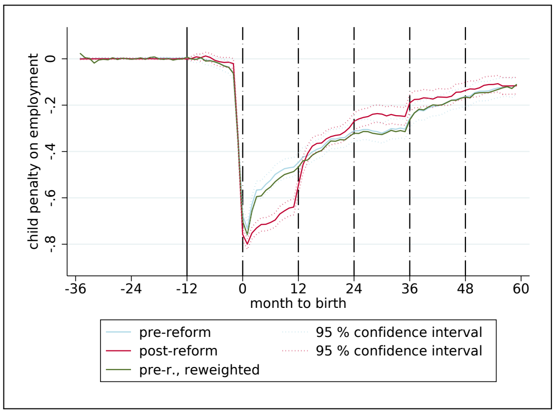
(a) estimates for mothers aged 21 to 25 years



(b) estimates for mothers aged 26 to 30 years

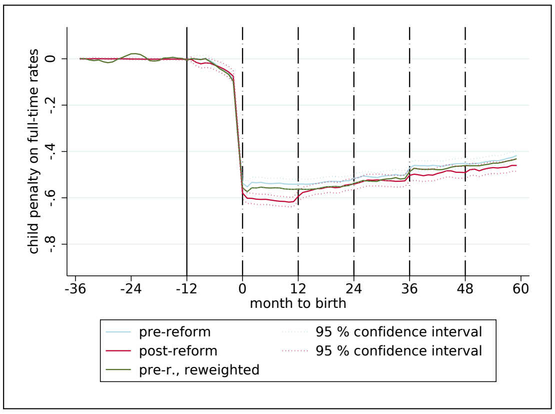


(c) estimates for mothers aged 31 to 35 years

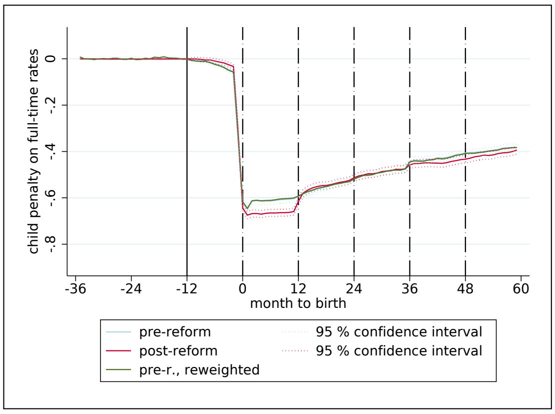


(d) estimates for mothers aged 36 to 40 years

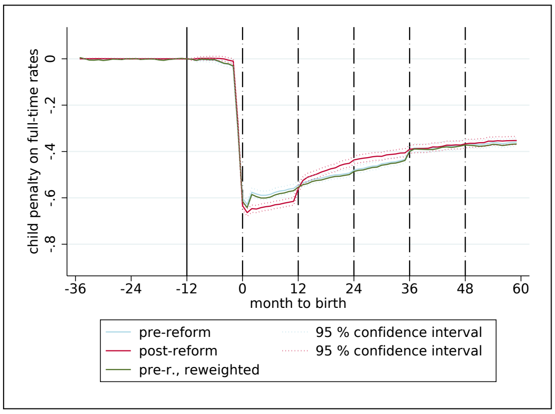
Figure 12: Child penalty on full-time rate according to age groups



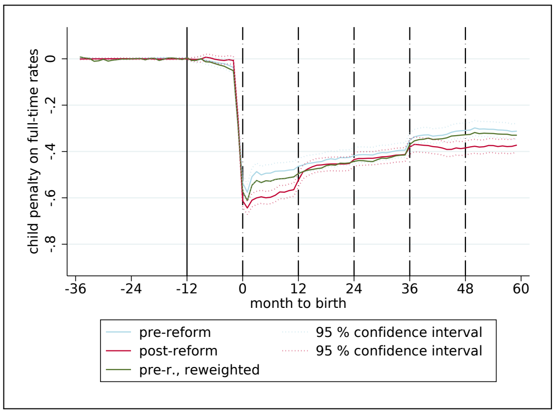
(a) estimates for mothers aged 21 to 25 years



(b) estimates for mothers aged 26 to 30 years

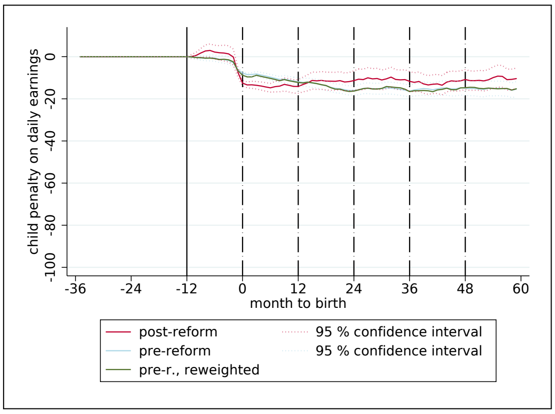


(c) estimates for mothers aged 31 to 35 years

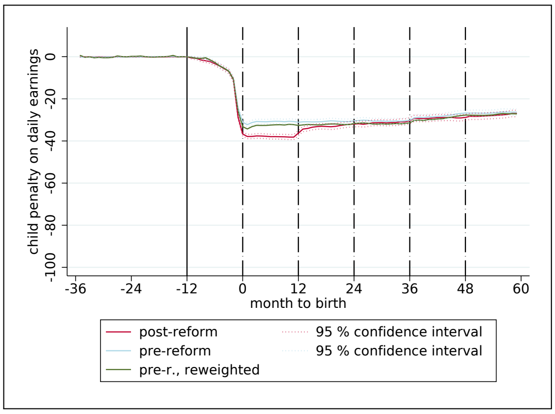


(d) estimates for mothers aged 36 to 40 years

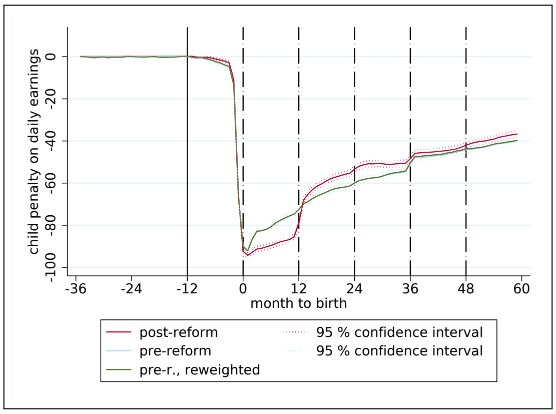
Figure 13: Child penalty on earnings by pre-birth earning groups



(a) $earnings = 0$ in months -35 to -12

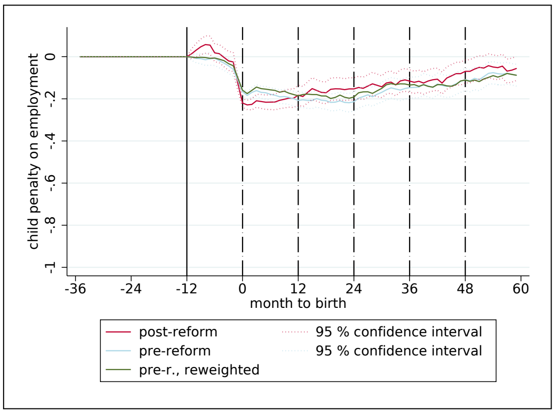


(b) $0 < earnings < 33000$ in months -35 to -12

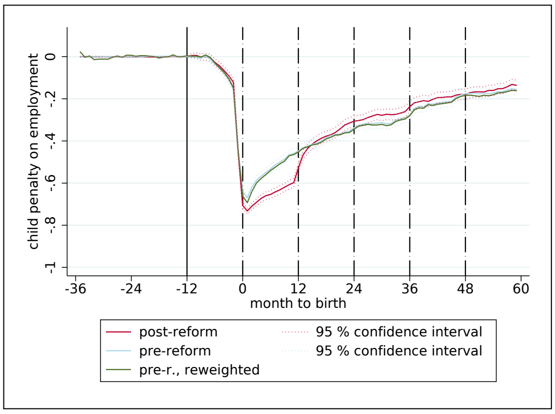


(c) $earnings > 33000$ in months -35 to -12

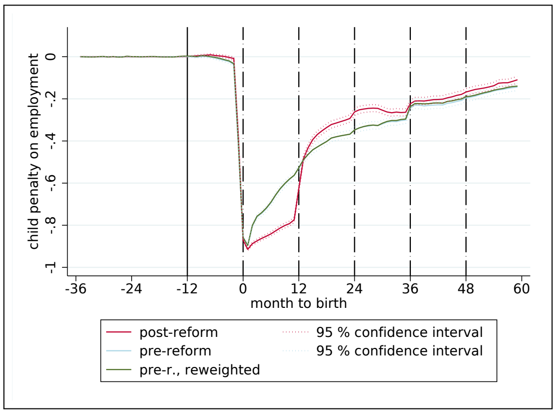
Figure 14: Child penalty on employment by pre-birth earning groups



(a) $earnings = 0$ in months -35 to -12

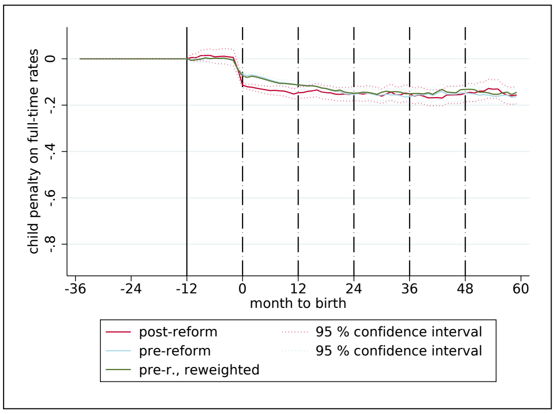


(b) $0 < earnings < 33000$ in months -35 to -12

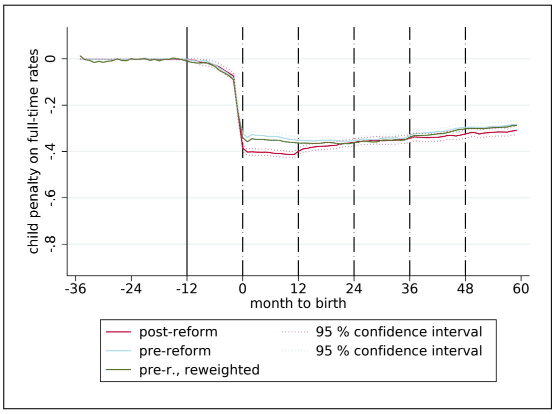


(c) $earnings > 33000$ in months -35 to -12

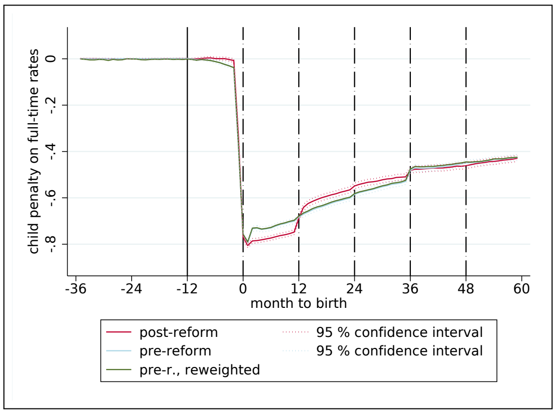
Figure 15: Child penalty on full-time employment by pre-birth earning groups



(a) $earnings = 0$ in months -35 to -12



(b) $0 < earnings < 33000$ in months -35 to -12



(c) $earnings > 33000$ in months -35 to -12

Figure 16: Effects of motherhood on second order fertility

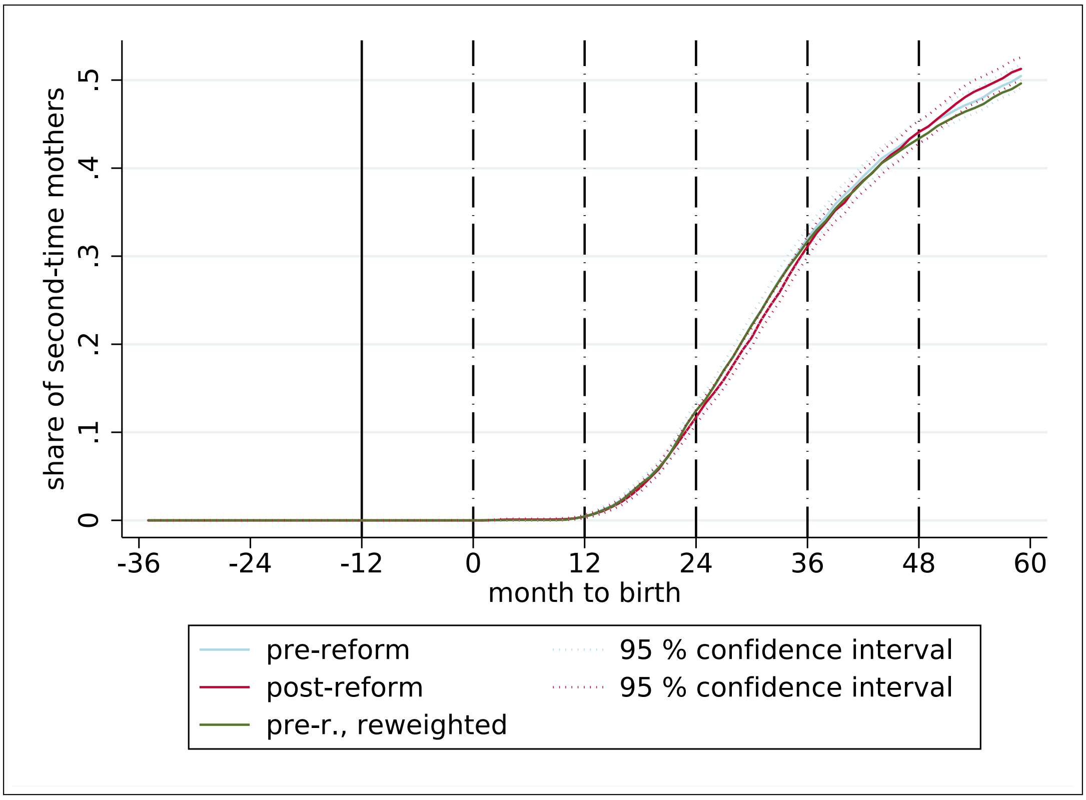
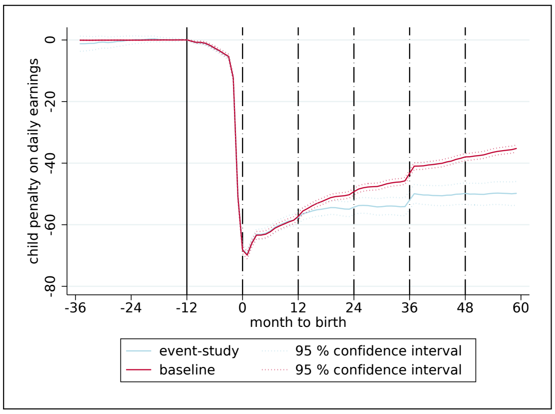
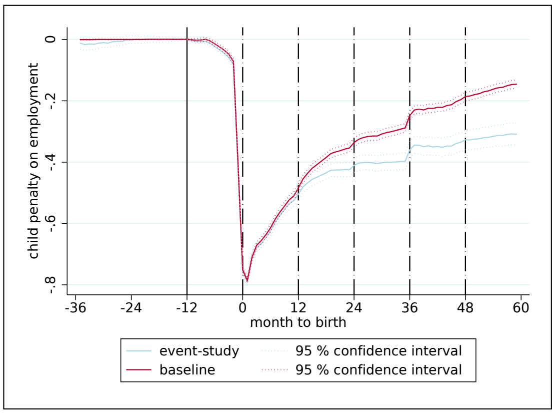


Figure 17: Child penalty, baseline compared to event study estimates

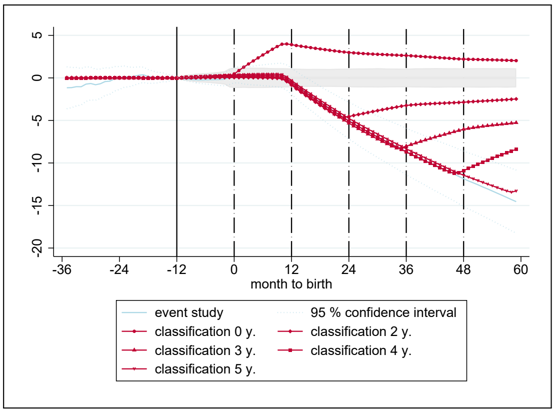


(a) earnings

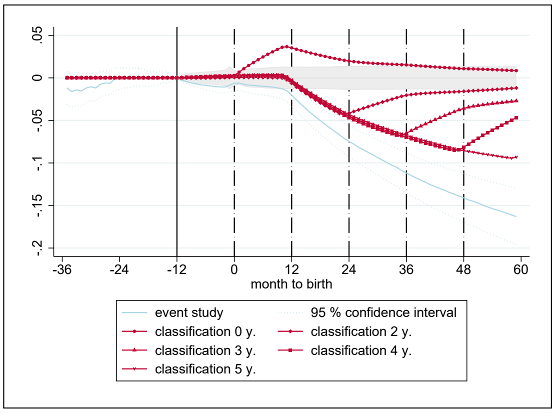


(b) employment

Figure 18: Differences in child penalty estimates compared to baseline



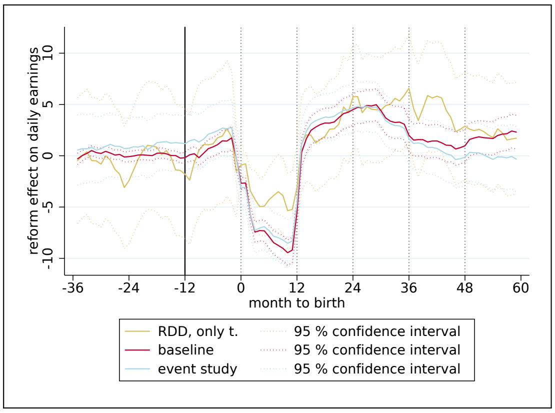
(a) earnings



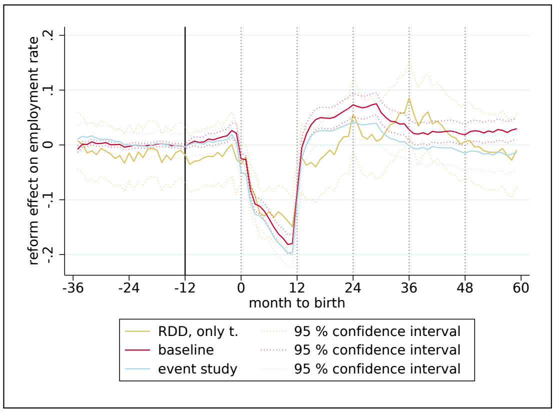
(b) employment

Notes: The graphs show the differences between baseline estimates and the estimates for event-study and variations of the baseline (classification windows 0, 2, 3, 4 and 5 years).
The gray area shows the 95-confidence interval for the baseline estimates.

Figure 19: Estimates for reform effects, baseline compared to RDD (only treated) and event-study



(a) earnings



(b) employment

Table 1: Average values for outcome and control variables

	childless (2004-2009)	mothers (first child, 2004 -2009) before giving birth	after giving birth
labor market outcomes:			
daily earnings	54.34	62.84	27.17
employment	.735	.811	.516
(corrected) full-time	.477	.583	.122
(raw full-time)	.488	.601	.136
further fertility	.216	.513	
covariates:			
year of birth	1977.66	1976.26	
former East Germany	.058	.05	
medium education	.665	.714	
high education	.139	.161	
N	40719	10727	

The left hand part of the table contains the averages for those women who are childless in our period of interests. The right hand part of the table contains the averages for those women who have their first child in our period of interest, three years before and after the reform (2004 - 2009).

The averages on the labor market outcomes are based on a monthly panel including the period 2001 - 2014 as we regard the time from three years before until five years after giving birth.

Further fertility is the share of women who give birth to a second child within five years after first birth for the group of mothers and the share who give first birth within five years after 2010 or ultimately leaving the pool of control observations by turning 41 years of age for the group of childless women.

High and medium education refer to having a tertiary and secondary certificate, respectively.

Table 2: Share of women entering motherhood and selection into motherhood

	Difference-in-differences regression	
	(1)	(2)
log earnings	0.0145***	0.0104***
	(0.0005)	(0.0005)
log earnings × postreform	0.0016**	0.0021***
	(0.0007)	(0.0007)
non-employment	0.1136***	0.0759***
	(0.0047)	(0.0049)
non-employment × postreform	0.0157**	0.0205***
	(0.0069)	(0.0069)
earnings > 8.550 €	0.033***	0.0261***
	(0.0011)	(0.0011)
(earnings > 8.550 €) × postreform	0.0024	0.0028*
	(0.0016)	(0.0016)
year FE	yes	yes
controls		yes
N	239467	239467

Descriptive share of women entering motherhood		
year	share	N
2004	.042	41294
2005	.043	41131
2006	.044	40601
2007	.045	39941
2008	.047	38952
2009	.048	37548
total	.045	239467

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The left panel (first two columns) shows estimates for two difference-in-differences regressions of the probability of giving birth on the labor market history (1). The earnings variable measures log cumulated earnings during the second calender year before giving birth (zero earnings are coded as zero and as one for the non-employment identifier). For the bottom panel, the labor market outcome evaluated is an identifier whether the cumulated earnings exceed € 8,550 in the second year before giving birth. The covariates used in column (1) are the respective labor market outcome, the interaction term and year fixed-effects. Column (2) includes additionally an identifier for former East-Germany, controls for medium and high education and age fixed-effects. The standard error are clustered at the individual level.
The right panel shows the share of women who give first birth in the year of observation.

Table 3: Causal reform effects

	before giving birth	after giving birth				
	1st year	1st year	2nd year	3rd year	4th year	5th year
earnings:						
daily	0.60*	-7.16***	2.21***	4.14***	1.30*	1.86**
	(.36)	(.43)	(.67)	(.74)	(.76)	(.78)
yearly	221*	-2614***	809***	1513***	477*	678**
	(130)	(158)	(245)	(271)	(277)	(285)
cumulative	221*	-2394***	-1585***	-72	405	1083
since month -12	(130)	(222)	(415)	(631)	(840)	(1060)
employment:						
overall	.011***	-.121***	.033***	.058***	.023**	.025***
	(.004)	(.006)	(.009)	(.009)	(.009)	(.009)
full-time	.011***	-.049***	.013***	.018***	-.010*	-.010
	(.004)	(.003)	(.005)	(.006)	(.006)	(.007)
second order fertility		.000	-.002	-.010	-.000	.014
		(.000)	(.003)	(.007)	(.009)	(.009)

Average causal reform effect for the respective year.
Standard errors in parentheses based on 1'000 bootstrap iterations.
* $p < 0.10$,** $p < 0.05$,*** $p < 0.01$ indicate significance of the average effect for the respective years.
Earnings are giving in daily and yearly earnings. The cumulative earnings give the yearly earnings since month -12 (one year before giving birth). Overall (Full-time) employment is an indicator whether the observation is (full-time) employed. Full-time is corrected using the approach presented in Fitzenberger and Seidlitz (2020). Second order fertility is one if the mother has a second child.

Table 4: Heterogeneous reform effects on earnings

		before giving birth	after giving birth				
		1st year	1st year	2nd year	3rd year	4th year	5th year
Heterogeneity with respect to age of mothers:							
women	daily	-0.68	-5.92***	-1.77*	-0.77	-2.06*	-2.03
age 21-25		(.68)	(.56)	(1.02)	(1.16)	(1.22)	(1.33)
	cumulative	-247	-2409***	-3055***	-3336***	-4090***	-4831***
		(249)	(343)	(618)	(930)	(1236)	(1578)
women	daily	0.80*	-6.91***	0.90	1.48	-0.18	0.96
age 26-30		(.47)	(.50)	(.89)	(.99)	(1.03)	(1.11)
	cumulative	291*	-2235***	-1906***	-1364*	-1431	-1082
		(173)	(276)	(522)	(806)	(1093)	(1402)
women	daily	0.19	-7.59***	3.61***	6.41***	1.95	3.11**
age 31-35		(.60)	(.82)	(1.28)	(1.34)	(1.44)	(1.47)
	cumulative	69	-2702***	-1383*	959	1672	2806
		(217)	(401)	(776)	(1162)	(1555)	(1959)
women	daily	2.47	-7.81***	5.47**	9.41***	6.71**	4.31
age 36-40		(1.51)	(1.71)	(2.45)	(2.69)	(2.74)	(2.81)
	cumulative	902	-1951**	47	3483	5935*	7511*
		(551)	(881)	(1580)	(2373)	(3152)	(3996)
Heterogeneity with respect to pre-birth earnings:							
zero earnings	daily	2.26*	-3.51***	2.28	4.41	3.38	4.44
		(1.37)	(1.24)	(2.44)	(2.86)	(2.87)	(3.18)
	cumulative	824	-457	376	1988	3221	4844
		(502)	(791)	(1516)	(2338)	(3101)	(3941)
low earnings	daily	-0.41	-5.26***	-0.97	0.36	0.30	-0.33
		(.61)	(.49)	(.80)	(.86)	(.93)	(1.03)
	cumulative	-149	-2070***	-2426***	-2292***	-2184**	-2305*
		(223)	(325)	(550)	(789)	(1034)	(1303)
high earnings	daily	0.93**	-8.65***	4.83***	5.52***	1.39	2.78***
		(.40)	(.64)	(.92)	(1.00)	(1.01)	(1.04)
	cumulative	342**	-2820***	-1057*	958	1465	2480*
		(147)	(296)	(560)	(846)	(1117)	(1404)

Average causal reform effect for the respective year.
Standard errors in parentheses based on 1'000 bootstrap iterations.
* $p < 0.10$,** $p < 0.05$,*** $p < 0.01$ indicate significance of the average effect for the respective years.
The tables shows as indicated the effects on daily and cumulative (yearly) earnings since month -12 (one year before giving birth).
The groups of zero, low and high earnings refer to the second and third year before giving birth (months -35 to -12 with respect to birth). Low (high) earnings indicate positive earnings below (above) 35,000 Euro in this 24 months period.

Table 5: Heterogeneous reform effects on employment

		before giving birth	after giving birth				
		1st year	1st year	2nd year	3rd year	4th year	5th year
Heterogeneity with respect to age of mothers:							
women	overall	.013	-.113***	.015	.044*	.028	-.010
age 21-25		(.011)	(.015)	(.024)	(.024)	(.023)	(.023)
	full-time	.001	-.048***	-.005	-.004	-.021	-.022
		(.012)	(.007)	(.012)	(.014)	(.015)	(.015)
women	overall	.012**	-.130***	.032**	.053***	.019	.032**
age 26-30		(.006)	(.009)	(.014)	(.015)	(.014)	(.014)
	full-time	.014**	-.053***	.003	.003	-.016	-.017*
		(.006)	(.005)	(.008)	(.010)	(.010)	(.010)
women	overall	.009	-.113***	.045***	.064***	.020	.031**
age 31-35		(.006)	(.009)	(.014)	(.015)	(.015)	(.015)
	full-time	.010*	-.041***	.032***	.043***	.006	.014
		(.006)	(.006)	(.009)	(.010)	(.011)	(.011)
women	overall	.012	-.128***	.017	.070**	.038	.020
age 36-40		(.014)	(.018)	(.026)	(.028)	(.026)	(.027)
	full-time	.018	-.059***	.002	.007	-.036*	-.053**
		(.013)	(.013)	(.018)	(.019)	(.021)	(.022)
Heterogeneity with respect to pre-birth earnings:							
zero earnings	overall	.029	-.049**	.025	.018	.028	.043
		(.02)	(.02)	(.031)	(.036)	(.038)	(.041)
	full-time	.009	-.041***	-.017	-.006	-.016	.001
		(.014)	(.009)	(.016)	(.02)	(.023)	(.026)
low earnings	overall	.005	-.104***	.013	.039**	.026	.021
		(.01)	(.012)	(.017)	(.017)	(.016)	(.017)
	full-time	.009	-.054***	-.011	-.001	-.014	-.02*
		(.008)	(.005)	(.008)	(.009)	(.01)	(.011)
high earnings	overall	.008***	-.146***	.056***	.06***	.017	.025**
		(.003)	(.007)	(.011)	(.012)	(.011)	(.011)
	full-time	.014***	-.047***	.031***	.026***	-.011	-.008
		(.004)	(.005)	(.007)	(.008)	(.009)	(.009)

Average causal reform effect for the respective year.
Standard errors in parentheses based on 1'000 bootstrap iterations.
* $p < 0.10$,** $p < 0.05$,*** $p < 0.01$ indicate significance of the average effect for the respective years.
The table shows as indicated the effect on overall and full-time employment. Full-time is corrected using the approach presented in Fitzenberger and Seidlitz (2020).
The groups of zero, low and high earnings refer to the second and third year before giving birth (months -35 to -12 with respect to birth). Low (high) earnings indicate positive earnings below (above) 35000 Euro in this 24 months period.

Table 6: Number of observed mothers for subgroups according to age at birth and pre-birth earnings

	aged 21 to 25	aged 26 to 30	aged 31 to 35	aged 36 to 40	
zero earnings	59	175	270	121	625
low earnings	831	1119	842	224	3016
high earnings	631	2773	2799	893	7086
	1521	4067	3911	1228	10727

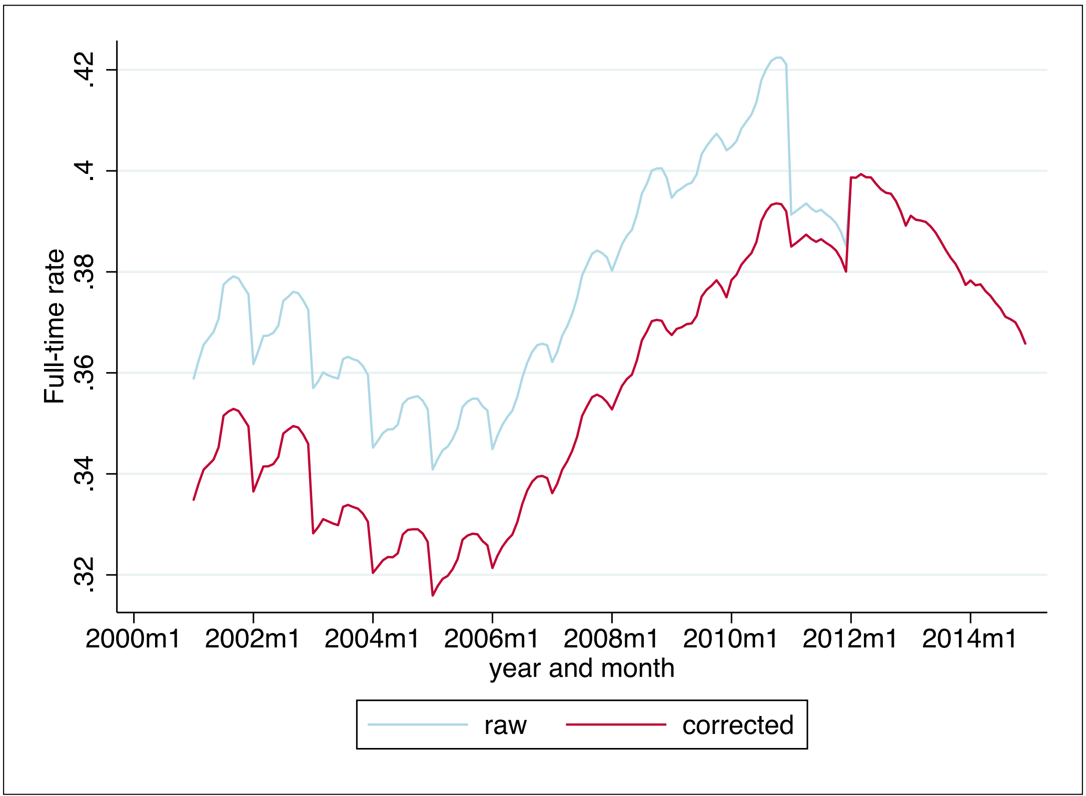
The groups of zero, low and high earnings refer to the second and third year before giving birth (months -35 to -12 with respect to birth). Low (high) earnings indicate positive earnings below (above) 33,000 Euro in this 24 months period.

Additional appendix

Full-time correction

[Figure 20 about here.]

Figure 20: Share of full-time employed over time



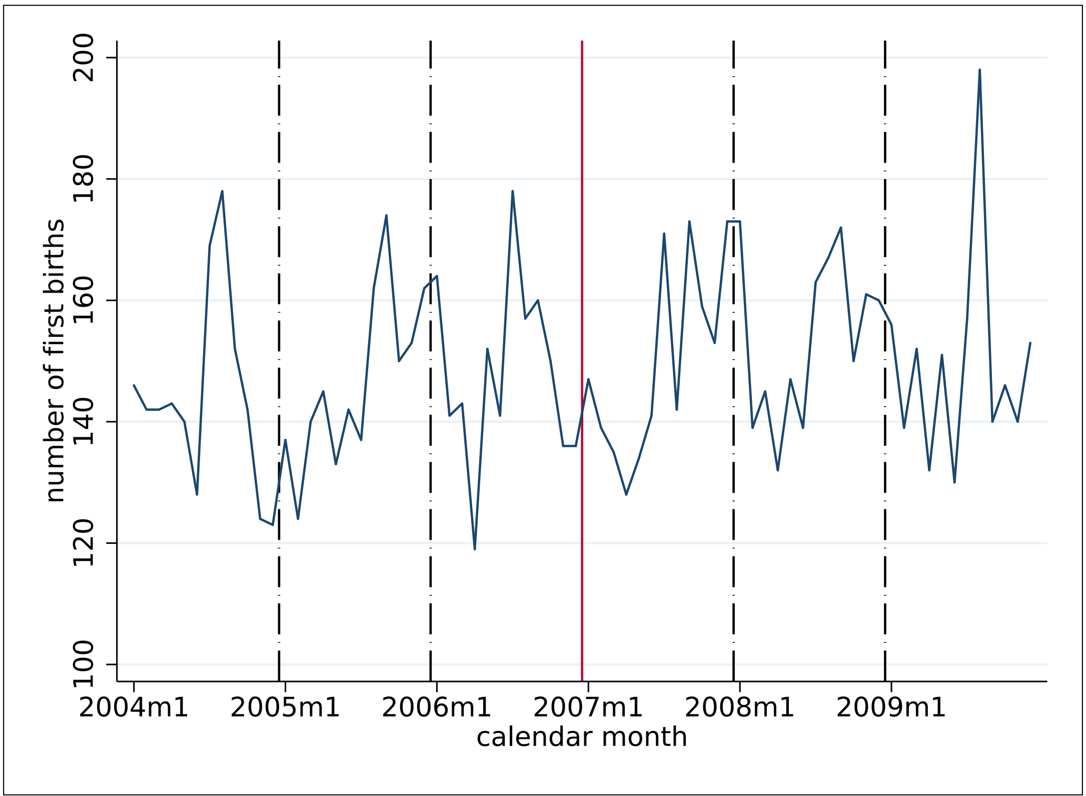
Notes: Raw and corrected full-time in the monthly panel.
For the correction, we used the approach proposed in Fitzenberger and Seidlitz (2020).

Selection into post-reform group

A selection into the post-reform group by manipulating the date of birth might be one concern towards our identification strategy. Tamm (2013) estimates that indeed around 8 percents of the potential births in the last week of December 2006 were shifted to 2007. However, we are confident that such behavior has no crucial impact on our results for three reasons. First, postponing the event of birth is a difficult process, not many parents will have the possibilities to do so. Forcing babies to arrive earlier is known to be easier but there is no evidence for this effect (Tamm, 2013). Secondly, the distribution of births in our sample does not look suspicious, see Figure 21. There are indications for certain seasonal patterns, but no kink around the reform cutoff. Thirdly, we are using a very broad sample with all birth between 2004 and 2009. This concern would be a much severe threat to an identification based on RDD. Our approach which allows and controls for differences in the selection into motherhood should be robust to manipulation around the cutoff.

[Figure 21 about here.]

Figure 21: Number of observed first births in the sample by calender months



Semiparametric event-study approach and RDD

To combine the semiparametric event-study approach with control group and the idea of RDD, we use the same narrow time range of three months before and after the reform (October 2006 to March 2007) as in Frodermann et al. (2023) and Kluve and Schmitz (2018). These studies argue that these mothers were unaware of the reform at the time of conception such that the reform did not change the selection of mothers in this sample. The RDD estimate is based on a small sample of 840 mothers pre- and post-reform, while our baseline analysis uses a sample of 10,700 mothers.

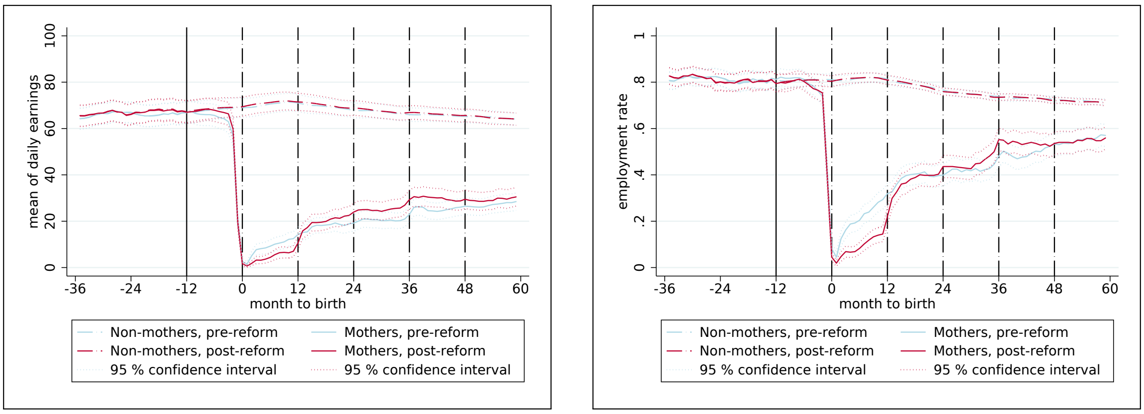
RDD estimates provide valid local reform estimates for an unchanged selection of mothers around the reform date thus missing the selection effect of the reform. The estimation is usually only based on treated observations but it is possible to implement RDD estimates using our control group approach. We do so providing a test of our baseline child penalty. Regarding the reform effects, using a control group of non-mothers or only treated observations should not make a difference in the RDD-setting. In Section 7.3 on the reform effects, we use only treated observations to achieve comparability to Kluve and Schmitz (2018). In this section, we discuss the RDD results obtained when using a control group.

Figure 22 depicts earnings and employment for the mothers pre- and post-reform revealing that the selection of mothers did not change with regard to the labor market history before conception. Because selection is unchanged, no reweighting of pre-reform mothers is needed to mimic the post-reform sample of mothers for the estimation of reform effects.

[Figure 22 about here.]

[Figure 23 about here.]

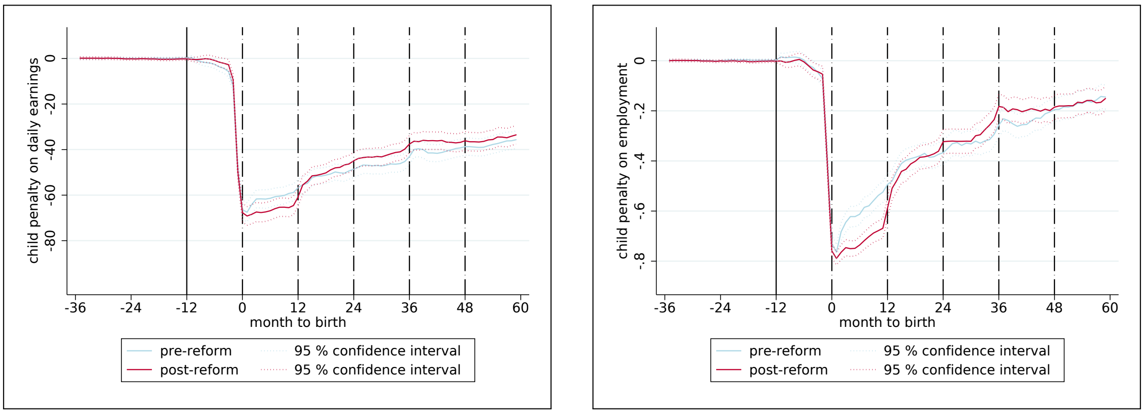
Figure 22: Treatment and non-mothers after IPW, RDD-sample



(a) earnings (b) employment

The difference between treatment and control group, the latter reweighted based on eq. (5), represents the estimated ATT of motherhood, given by the differences between the two lines in Figure 23. In line with the literature and our baseline findings, there is a strong negative reform effect during the first year after giving birth. For earnings, the effect turns positive shortly after, becoming significant in the third and fourth year after birth. Afterwards the earnings effect loses significance but shows roughly the same size, implying that there are small but lasting gains. For employment, the reform effects from the second year onward are quite small and vary in sign. Furthermore, the RDD

Figure 23: Child penalty on earnings and employment, RDD-sample



(a) earnings (b) employment

approach yields quite noisy effect estimates which differ noticeably from the baseline estimates during the first year after birth for earnings and during the first three years for employment, but which are very similar to the baseline estimates in the medium run.