

Seidlitz, Arnim; Fitzenberger, Bernd

Conference Paper

Effects of Motherhood and Selection into Motherhood: Revisiting the 2007 Parental Benefits Reform in Germany

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Seidlitz, Arnim; Fitzenberger, Bernd (2021) : Effects of Motherhood and Selection into Motherhood: Revisiting the 2007 Parental Benefits Reform in Germany, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/242435>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Effects of Motherhood and Selection into Motherhood: Revisiting the 2007 Parental Benefits Reform in Germany

Bernd Fitzenberger* Arnim Seidlitz†

March 1, 2021

Preliminary; please do not cite or circulate without authors' permission.

Abstract: This paper studies the causal effect of giving first birth and how that effect changes by the 2007 reform in parental benefits in Germany. We further investigate how the reform affects the selection of women into motherhood and how that selection effect affects labor market outcomes. A large novel data set merging data from the pension insurance with administrative labor market data provides information on all births. We apply a dynamic treatment effect approach which differs from other strategies used so far in most of the literature to estimate the causal effect of motherhood and to evaluate the 2007 reform. The reform has positive medium-run effects on earnings and employment while reducing full-time employment. There are no effects on second-order fertility. While the reform slightly changes the selection of mothers, this has little impact on the reform effect for the causal effect of motherhood.

JEL-Classification: J08, J13, J16, J22.

Keywords: Parental leave reform, causal effect of motherhood, earnings, employment dynamic treatment effect.

*IAB, FAU Erlangen-Nürnberg, IFS, CESifo, IZA and ROA

†Humboldt University Berlin and Berlin School of Economics

1 Introduction

The birth of the first child is still a turning point in the career of many mothers and the so-called child penalty due to motherhood is viewed as a key reason for the persistent gender gap in the labor market (Angelov et al., 2016; Kleven et al., 2019b,a). Employment and earnings drops to (almost) zero for a few months immediately before and after child birth. The process of reentering employment is on average sluggish. After returning to employment first-time mothers often work less hours and often soon have their second child (Sigle-Rushton and Waldfogel, 2007; Fitzenberger et al., 2013).

Most governments provide paid parental benefits and/or unpaid parental leave (henceforth job protection) to support parents after child birth and to enhance the compatibility between work and having children. All OECD countries except the U.S. pay nationwide parental benefits, and in the U.S. some states like California have such programs (OECD, 2019).

Olivetti and Petrongolo (2017) and Rossin-Slater (2018) review different systems of family policies around the world and the economic research literature on this topic. There is evidence that job protection helps mothers to return to their previous employers. However, medium-run effects on labor market outcomes of extending an already existing job protection are often insignificant, and there is concern that longer periods of paid or unpaid parental leave are detrimental for the post-birth career of mothers. For Germany, Schönberg and Ludsteck (2014) investigate various reforms extending job protection from two to 36 months between 1979 and 1992. The study finds sizeable negative short-run effects on post-birth employment of mother but only small negative long-run effects. Angelov et al. (2016), Kleven et al. (2019b), and Kleven et al. (2019a) find sizeable negative long-run effects of child birth on mothers' labor market outcomes in various countries with long unpaid parental leave regulations.

Paid parental benefit systems have also gained interests within the research community. Rossin-Slater et al. (2013) evaluate the introduction of benefits in California offering transfers for up to six weeks. This relatively short duration seems to have positive effects on maternal employment until the third year after giving birth. The international evidence on extending benefits duration include studies on Austria (Lalive et al., 2014) and Norway (Dahl et al., 2016). Both countries had compared to California already a more generous parental benefits system. Increasing the benefit duration further had in both countries

no significant medium run effects on employment outcomes.

Our study contributes to this literature by using novel administrative data and by applying a dynamic treatment effect approach to estimate the causal effect of motherhood which differs from other methodological approaches used so far in most of the literature. A large data merge of data from the pension insurance on fertility with administrative labor market data provides information on all births, independent of the pre-birth employment status. Likewise, second births can be observed. Methodologically, we use a dynamic treatment approach which combines the advantages of a dynamic matching approach and an event study (Fitzenberger et al., 2013; Sianesi, 2008). This estimation involves a control group which is one huge difference to all other papers on the reform and to most papers on the effect of giving birth (for example Angelov et al. (2016); Kleven et al. (2019a,b)). This article further revisits the 2007 reform in parental benefits. The reform decreased for a subgroup the benefit duration but increased amount and duration of benefits for another subgroup conditional on interrupting work during the first year after child birth. We carefully assesses the change in the selection of mothers due to the reform and estimates the effect of the reform on the causal effect of motherhood after accounting for this change in selection. The reform has positive medium-run effects on earnings and employment while reducing full-time employment. There are no effects on second-order fertility. The remainder of the paper contains in section 2 a discussion of the reform and the literature on it, in section 3 and 4 a description of the data and the estimation approach, in section 5 the presentation of our results and in section 6 a comparison to alternative estimation strategies.

2 The 2007 parental leave reform

The 2007-reform made Germany to one of the most generous countries concerning parental benefits worldwide. All mothers of children born from January 2007 on may receive the new "parental allowance" until the 12 month after giving birth. The reform was implemented with a clear cutoff. Mothers of children born until December 2006 faced a different benefit system which paid substantially less benefits to most mothers.

The German system contains generally different institutions. From six weeks before the expected date of delivery until eight weeks after giving birth, the mothers are in so-called

"maternity protection" (Mutterschutz). During that period, it is for them forbidden to maintain a paid work but they receive the "maternity allowance" (Mutterschaftsgeld) which covers 100 percent of the pre-birth earnings¹. Health insurances and employers provide the funding of this "maternity allowance". It is not seen as part of the family policy but to insure the health of the mother and her child. Most importantly, it was not reformed in 2007 such that pre- and post-reform mothers are treated equally (BMFSFJ, 2020b).

Under the old system, the eligibility for further benefits depend on the mother's and potentially her partner's income (henceforth for convenience household income²). Mothers of high income households received after the end of maternity protection – from the third month after giving birth on – no further financial benefit. Medium and low income households were paid 300 Euro monthly so-called "child raising allowance" (Erziehungsgeld). The former group was eligible for this benefit until the sixth month and the latter group until the 24th month after giving birth. Low income households alternatively had the possibility to choose 450 Euro monthly until the end of the first year instead of receiving 300 Euro monthly until the end of the second year (BMFSFJ, 2004). That second option for low-earners was however only chosen by a minority of mothers (Kluve and Tamm, 2013).

From 2007 on, eligibility is universal and the amount of monthly benefits is earning dependent. Mothers receive 65 percent³ of their monthly earnings in the 12 months before entering "maternity protection" but at least 300 and at most 1800 Euro. This benefit is paid until the end of the first year after giving birth. Single-mothers have the possibilities to receive two additional months of "parental allowance"⁴ (Ehlert, 2008).

It is also important to note what was not touched by the reform. The "maternity protection" and the "maternity allowance" directly before and after giving birth remained unchanged. Further, mothers loose eligibility for both, the old "child-raising allowance"

¹To be precise, the pre-birth earnings are calculated as average of the three calendar month before the mother enters pregnancy.

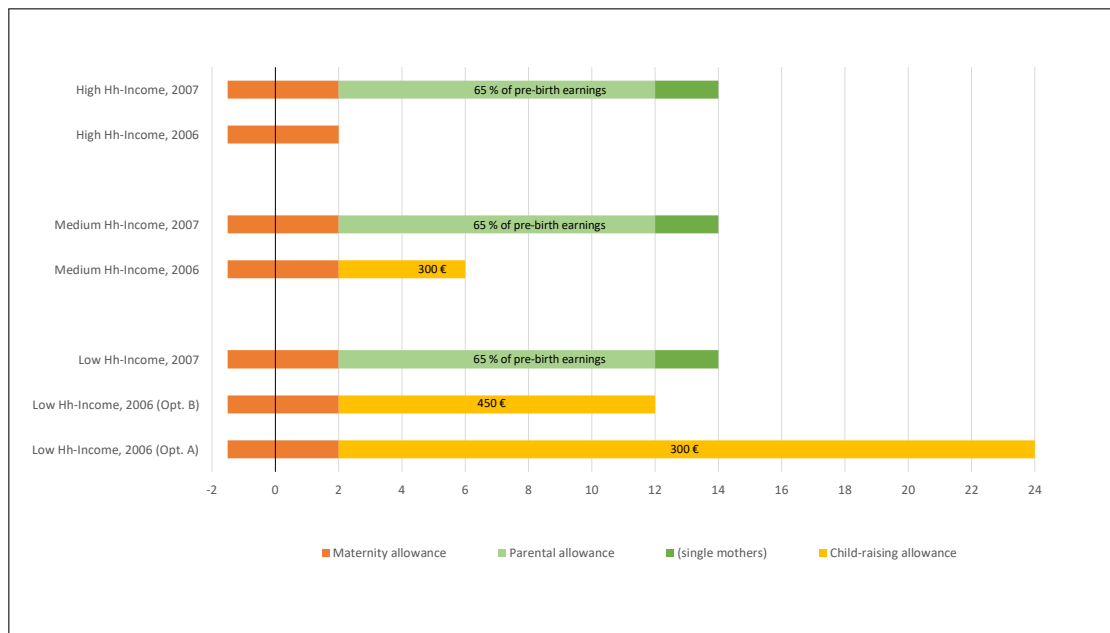
²The relevant quantity was the yearly net-income of the mother if she raised her child as single-mother and the combined income of her and her partner if they lived in one household. For both groups of mothers different income thresholds were used.

³The replacement rate decreases in earnings. The differences are moderate for earnings above 1000 Euro monthly (between 65 and 67 percent), below 1000 Euro it increases to 100 percent for 440 Euro monthly earnings (BMFSFJ, 2020a).

⁴Fathers are also targeted by the reform. From 2007 it is possible that both parents receive benefits. However, effects on them cannot be covered in this projects. The same holds true for adjustments in 2015 which added more flexibility for the receivers of "parental allowance" (BMFSFJ, 2020a)

and the new "parental allowance" when working full-time, working until at most 30 hours weekly is allowed⁵. Mothers also enjoy job protection (Elternzeit) for three years, unchanged since 1992. Within that period mothers may return to their previous jobs (BMFSFJ, 2020a).

Figure 1 shows the pre- and post-reform situation with respect to household income. It is quite easy to distinguish reform-winners from reform-losers. Medium and especially high income household receive unambiguously longer and higher benefits while low-income household may fare worse compared to the old system. To break down all the complex regulations to a simple message, it is the case that Germany changed from a system paying higher benefits to low-earners to one paying higher benefits to high-earners.



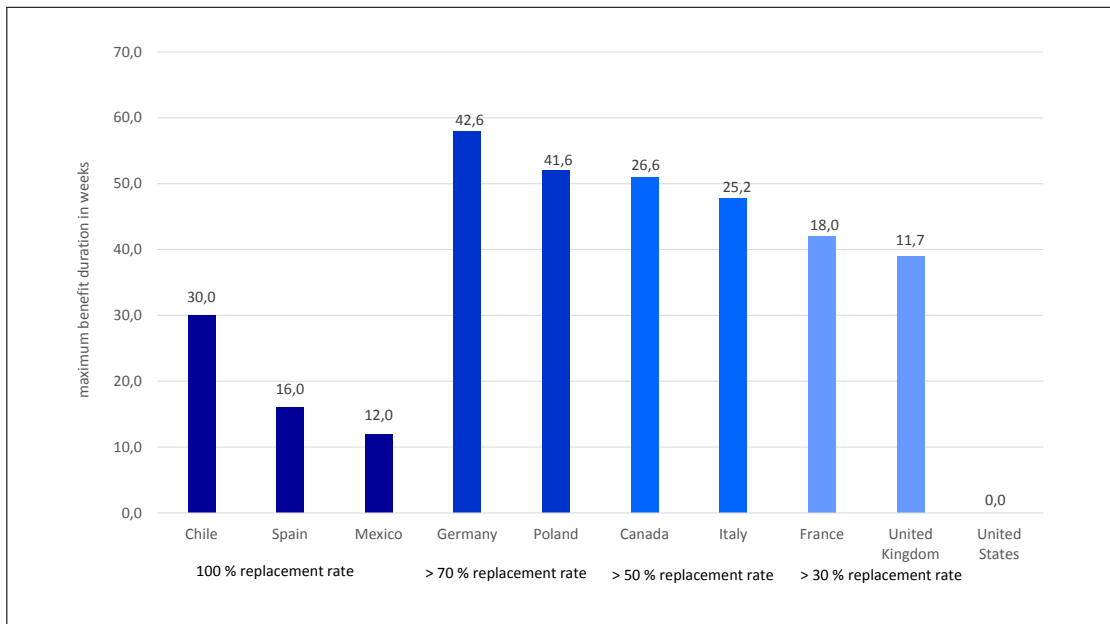
High and medium income households according to the old "child-raising allowance" were households with an yearly net-income above 30000 and 22000 (23000 and 16000 for single-mothers), respectively.
 "Parental allowance" is bound to be at least 300 and at most 1800 Euro monthly.
 Receivers of both, "child-raising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

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

The "parental allowance" is internationally one of the most generous systems. Figure 2 compares the system of some of the larger OECD-countries. The countries are grouped by the average replacement rate in terms of the previous income. The height of the bars gives the maximum duration of benefits. The Spanish speaking countries have institutions comparable to "maternity allowance" in Germany, a replacement of 100 percent but only for a short period. The other European countries and Canada pay benefits for a longer duration but do not replace the earnings entirely. Outstandingly, the US do not have

⁵If mothers work during receivance of "parental allowance", their benefits will decrease to 65 percent of the difference between pre- and post birth earnings (BMFSFJ, 2020a)

parental benefits on national level. Poland and Germany are arguably the most generous countries of the displayed countries. An indication for this claim is the number on top of the bars which gives the product of replacement rate and benefits duration, the theoretical amount of weeks with 100 percent replacement. The investigation of the German reform is thus interesting for most other countries which would have the room for an increment in transfers paid to mothers.



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 2: Comparison of different parental benefit systems

2.1 Literature on the German reform

Shortly after the reform was implemented and data became available, researcher started to investigate its short-term effects. Bergemann and Riphahn (2011) found a positive reform effect on employment for the second year while Kluge and Tamm (2013) found a negative reform effect on employment for the first year after giving birth. Geyer et al. (2015) confirms both – the positive effect in the second year at least for low-earners. These short-term effects are not contradictory. Given the increase of benefits in the first year and the decrease in the second year (figure 1), they are rather expectable.

Meanwhile, there are also some papers studying the medium term outcomes. Bergemann and Riphahn (2020) show that the temporary gains in employment probability do not persist in the medium run. Frodermann et al. (2020) concentrate in their study on earnings

and find a significant positive reform effect two years after giving birth. For low-earners, that is the only significant effect, while for high-earners the positive effects stays significant but diminishes over the following years. After eight years it is insignificant. The working hours do not seem to be significantly affected. This is widely in line with the results of Kluge and Schmitz (2018) who find significant effects on employment until five years after giving birth which are driven by medium and high earners. They show further heterogeneous effects for full-time employment. The probability to work full-time after five years is reduced for low-earners but increased for high earners due to the reform. On average, it is insignificant. It seems thus that the effect in the medium run (after five years) are less striking than the short run effects. There is at least evidence for a positive reform effect on outcomes as earnings and participation which are significant for high-earners.

Two studies work on fertility effects of the reform. Cygan-Rehm (2016) finds temporarily a huge decrease in higher-order fertility. In the medium run this negative effect becomes insignificant. It is driven by the reform-losers. Raute (2019) shows that the propensity to become mother (first and higher order fertility) for high earning women increases compared to low earning women as response to the reform.

This results implies that it is important to investigate the effect of the selection induced by the reform on labor market outcomes. That is something which is so far missing in the literature. Frodermann et al. (2020); Kluge and Schmitz (2018) use a diff-in-diff and regression discontinuity design relying only on observations directly before and after the reform. They need to assume that differences in the selection of mothers do not drive their results and that is for their setting convincing as the reform was announced only six months in advance such that there was no time to adjust fertility behavior. Bergemann and Riphahn (2020), on the other hand, work with mothers of the year 2005 to 2008, two years before and after the reform. Changes in the selection of mothers might be a bigger threat to them although they show that their result do not change much when they restrict themselves to a narrower time window.

However, to our best knowledge, we are the first who try to quantify the effect of the selection induced by the reform on labor market outcomes.

3 Data

We have the great chance to work with the merge of two different high quality administrative data sources from Germany. These are data which originate on one hand from social insurance and on the other hand from pension insurance.

3.1 Social insurance data

The data of the social insurance are the so-called Integrated Employment Biographies (IEB). They are administrated by the Institute for Employment Research (IAB) and contain data on all employees in Germany but not on civil servants and self-employed. For them, there are precise information on employment spells including start and end date, gross-earnings, a part-time status, education level and the county of employment⁶. From these spell data, we form a monthly panel and construct an average daily earnings rate for each month. Therefore, we add potentially the earnings of three different jobs but we do not regard marginally employments with less than 14 days duration or earnings below 300 Euros in one month.

One potential issue of the IEB-data is the top coding of earnings. However, for the population of women only the top four percent of employment spells are distorted (Fitzenberger and Seidlitz, 2020). We hence decided to neglect this issue. The earnings are further deflated by the annual consumer price index of the Federal Statistical Office (Statistisches Bundesamt, 2019).

3.2 Pension insurance data and data merge

The data from the pension insurance are called "Versichertenkontostichprobe (VSKT)". These are also spell data on employment outcomes (FDZ-RV, 2021). Most important for us, they contain for the mothers the concrete month of giving birth for each of her children. This information is extremely valuable as it is not included in the data of the social insurance.

The data merge was conducted within the project "Custom Shaped Administrative Data

⁶For a detailed description of the data which include many more features, we are not using in this project, we would recommend the documentation of the SIAB. The SIAB is a two percent sample of the IEB (Antoni et al., 2019).

for the Analysis of Labour Market (CADAL)"⁷ by researchers at the IAB. As both sources provided their data anonymized, the union of the two data sources is based on a probabilistic matching. This procedure is possible as the data of the pension insurance also contains information on employment spells although not as detailed as the social insurance. The colleagues at the IAB designed for the matching an algorithm which is mainly based on the ratio of employment spells which coincides in terms of start and end date, earnings and status of employment (covered by social insurance, marginal or in apprenticeship) in both data bases and required that both data bases reported the same month and year of birth.

This algorithm was applied to roughly 306000 female observation of the pension insurance. For two third of them a unique match in the social insurance was found. The rest of the data where the algorithm identified either no or more than one candidate match were not used.

For our analysis, we introduced a few additional minimum criteria on the working history and exclude all observations with a working history in the former East German GDR, to decrease the risk of mismatches⁸. Of course this approach contains the danger that our sample has an above average labor market attachment and is less representative for the entire population. However, we want to highlight that our selection is arguable much less restrictive than the selection which arises when maternity is directly identified from social insurance data. Müller and Strauch (2017) described a way how mothers can be detected from the information on working spells in the IEB data. This approach is very helpful and widely used (for example by Frodermann et al. (2020)) but contains a strong restriction. It requires mothers to work directly before entering motherhood while our strategy basically requires mothers to have worked for some time in their life at any time before or after giving birth.

⁷This project was part of the SPP "The German Labor Market in a Globalized World: Challenges through Trade, Technology, and Demographics" (SPP 1764) which was funded by the German Research Foundation (DFG).

⁸Given our probabilistic matching strategy, it is clear that the accuracy increases with the length of the working history. We therefore require our observations to have at least worked for two years within the social insurance and to report at least ten spells in total and five working spells covered by the social insurance. The reporting quality of spells in marginal employment is further a potential issue. Therefore, we require the observations to have earned on average more than the marginal employment threshold in her working spells. Lastly, the algorithm may produces mismatches if an observation has periods which are covered in pension insurances but not in social insurance. Consequently, we exclude observation who have started their working career before 1975 when the social insurance records began (given our observational period that is no severe restriction) and who have a working history in the former GDR which is generally not included in social insurance data.

3.3 Descriptives

To evaluate the effect of the 2007 reform, we regard the mothers who have her first child within three years before or after the reform eg. the first time-mothers of the six years from 2004 to 2009. This means naturally that we exclude all women who have her first child before 2004 from the matched sample. Further, we restrict the sample to a core working and fertility age group of women aged 21 to 40. All observations are required to be in this age group in at least one year between 2004 and 2009. This is equivalent with stating that we only evaluate the population born between 1964 and 1988. For our control group, we keep all women who are childless until 2010, regardless whether they become later mothers.

This restrictions leave us 50000 observation, a bit more than 10000 of them become her first child in the period of interest. Table 1 shows some descriptives for our sample.

Table 1: Average values for outcome and control variables

| | childless (until 2010) | mothers (first child, 2004 -2009) | |
|------------------------|---------------------------|-----------------------------------|--------------------|
| | | before giving birth | after giving birth |
| labor market outcomes: | | | |
| daily earnings | 55.04 | 63 | 27.26 |
| participation | .745 | .813 | .518 |
| fulltime | .553 | .664 | .139 |
| controls: | | | |
| year of birth | 1977.79 | 1976.28 | |
| former East Germany | .042 | .033 | |
| low education | .167 | .111 | |
| medium education | .666 | .712 | |
| high education | .139 | .161 | |
| N | 40167 | 10695 | |

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 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.

We allow for missing value as a "forth" educational category (2.6 percent of the sample). This explains, why the numbers of the three categories low, medium and high education do not add up to one.

High, medium and low education refer to having a tertiary, secondary and no secondary certificate.

We learn that the sample of mothers is bit better educated. 87 percent of them hold at

least a secondary educational certificate compared to 81 percent of the childless women. The differences in age and region of residence are small. The share of East Germans in the sample around four percent is very low which we explain by the selection induced by the matching.

Earnings, participation and full-time status are going to be the main outcome variables of our analysis. Table 1 contains the dramatic decline in means of these variables if we compare for mothers their pre- and post-birth periods. The share of women in full-time is more than 50 percentage points lower after giving birth and mean earnings reduced to less than half of the pre-birth earnings.

We further see that the group of mothers before giving birth shows in all three observed dimensions on average better outcomes than childless women. They seemed to be positively selected in both, labor market outcomes and educational achievement. For any empirical approach using a control group to evaluate the causal effect of giving birth, this results shows the relevance to account for these pre-birth differences. In our approach, we reweight the control group to match the group of mothers by an inverse probability weighting.

4 Econometric Approach

We first estimate the causal effect of motherhood on various post-birth outcomes before and after the 2007 parental leave reform using a dynamic treatment effects approach as in Fitzenberger et al. (2013). In the second step, we estimate the impact of the reform on the selection of mothers with regard to pre-birth outcomes and socio-economic characteristics. We further investigate how the selection effect affects the causal effect of motherhood on post-birth outcomes. Accounting for these selection effects allows us in a third step to estimate the causal reform effect on the effect of motherhood on post-birth outcomes. Now we describe these three steps in detail.

4.1 Estimating the causal effect of motherhood

Our goal is to estimate the average treatment effect for the treated (ATT) on post-birth outcomes based on discrete time data. The treatment is 'first child birth at a certain point in time' against the alternative of waiting. The alternative of waiting, i.e. of not having

a first child now, entails both, the possibility of never having a child and the possibility of having a child at a later date. Thus, for a certain point in time, we estimate the ATT of having a first child at this point of time versus not yet having a child.⁹ Both, treated mothers and nontreated women at a certain date are required to not have a child yet and the group of nontreated women consists both of women who never have a child and women who have a child at a later date (see subsection 4.1.1). This definition of the control group avoids conditioning on future outcomes.

This section draws on Fitzenberger et al. (2013) with the difference that Fitzenberger et al. (2013) align treated mothers and nontreated women by age at first birth whereas we align treated and nontreated women by time at first birth, which allows us to clearly separate treated and nontreated women before and after the reform. There is no fundamental difference between the two approaches because either way the second dimension (age or time, respectively) is controlled for when aligning treated and nontreated women through inverse probability reweighting.

The key identification assumption for our analysis is a dynamic conditional independence assumption (Fitzenberger et al., 2013). It states that – conditional on the variables controlled for – until a certain time period the assignment to treatment in this time period is random, i.e. independent of the potential outcomes. Specifically, our dynamic conditional independence assumption stipulates that given the duration of childlessness and given the covariates, having a first birth within the next year is random.

The dynamic conditional independence assumption can be motivated as follows. Our rich administrative data allow to control for a number of socio-economic characteristics and labor market history. One year before birth, the treated women are not likely to differ systematically from those women who stay childless until shortly after the birth of the child. The exact timing of birth cannot be planned with certainty and may depend upon random circumstances not reflected in long-run labor market choices. It is highly implausible that women plan the exact month of first birth more than a year ahead. At the same time, women differ in their probability to have a child within the next year and this is likely to be reflected by the characteristics that are being controlled for. In our analysis, we match treated mothers and nontreated women by observable characteristics

⁹The treatment effect we estimate is an example for the dynamic treatment approach applied in the context of program evaluation of active labor market policies by Sianesi (2004) for Sweden or by Biewen et al. (2014) for Germany.

twelve months before birth, which is about the time when the decision for becoming pregnant is made.

In our application, for estimating the effects of giving birth for the first time, 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 make fertility choices many times in life, having decided not to have a child in all periods before they decide to have their 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 unobserved career orientation may be more likely to exhibit a higher labor market attachment and lower fertility rates).

Our approach assumes that women giving birth to their first child are comparable before the gestation lag, i.e. at the time before pregnancy, to women who do not give birth at this date. This approach assumes that women do not know the exact timing of first birth before the gestation period. But they may know the probability of having a first birth now versus later and they may act upon the determinants of this probability (Abbring and Van den Berg, 2003). Assuming a no-anticipation condition with respect to the precise date of pregnancy before the gestation period allows us to match treated and nontreated women at this date.

The treatment group in our analysis consists of women who have their first child between the age of 21 and 40. The control group 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 for each treated woman based on the individual-specific comparison group of individuals not treated yet.

4.1.1 Temporal alignment

Our evaluation approach requires a temporal alignment between treated mothers and control observations. For the treatment group the treatment time is given by the month (time period) at giving first birth, i.e. calendar time is the treatment time. Women in the control group are aligned to the treated mothers based on not giving first birth in the

calendar month of birth of the mother. We do not impute a random placebo treatment time for control observations (as discussed, e.g., in Kleven et al. (2019a)).¹⁰

As classification window for our benchmark analysis, we consider as suitable control observation a woman who is childless and continues to be so for the following 11 months. When this classification window is reduced, the control group also includes women who are pregnant or who already have concrete plans to become mothers at the date of birth of the treated mother. When this classification window is extended, the control group would be childless for a longer period which conditions on the future thus increasing the dynamic selection of the control group. As robustness checks, we also investigate the sensitivity of our core findings to reducing the classification window to 1 month, i.e. the control group includes all women giving first births after the birth of the treated mother, and to extending the classification window to 24 months, i.e. women giving first births during the second year after the birth of the treated mother are excluded from the control group.

We consider each treatment month from January 2004 to December 2009 separately. For each of these 72 months, we pool treated women, who have her first child in this month, and all potential control women, those who are childless for at least the following 11 months for the benchmark analysis (the classification window is 1 month or 24 months for robustness checks, respectively). Further, we require all observations to be between 21 and 40 years of age at the treatment time (date of birth). In the next step, we pool the data over the 72 treatment months.

This way, we generate a $person \times treatment - calendar - months [2004, 1 \text{ to } 2009, 12] \times month - relative - to - treatment - time [-35 \text{ to } 59]$ data set used for our analysis. For *month-relative-to-treatment-time*, 0 denotes the month of birth, -35 means 35 months before birth, and 59 means 59 months after birth. Effectively, for the panel data set used for our empirical analysis, women are thus duplicated for each (potential) treatment month $t \in [2004, 1 \text{ to } 2009, 12]$. A woman may be used in one treatment period as treated, in the month where she becomes mother and potentially multiple times as control observation, in all months which are at least one year before she has her child. $t \in [2004, 1 \text{ to } 2006, 12]$ represents the pre-reform period and $t \in [2007, 1 \text{ to } 2009, 12]$ the post-reform period.

¹⁰This is because we do not want to restrict the control group to women who do not give birth during the entire observation window. Only for the latter group, it would be plausible to simulate placebo treatment times.

Take as an example a woman who gives first birth in April 2006. In the months until April 2005, those which are marked blue in figure 3, she serves as control observation. She is not used in May 2005 as then her own child has less than one year to come. So, we do not regard her as suitable control observation anymore. The same argumentation holds until March 2006. In April of that year, she give birth to her first child mother and enters the sample as treated. Afterwards, for all treatment months until December 2009, she is not used, as she is already a mother.

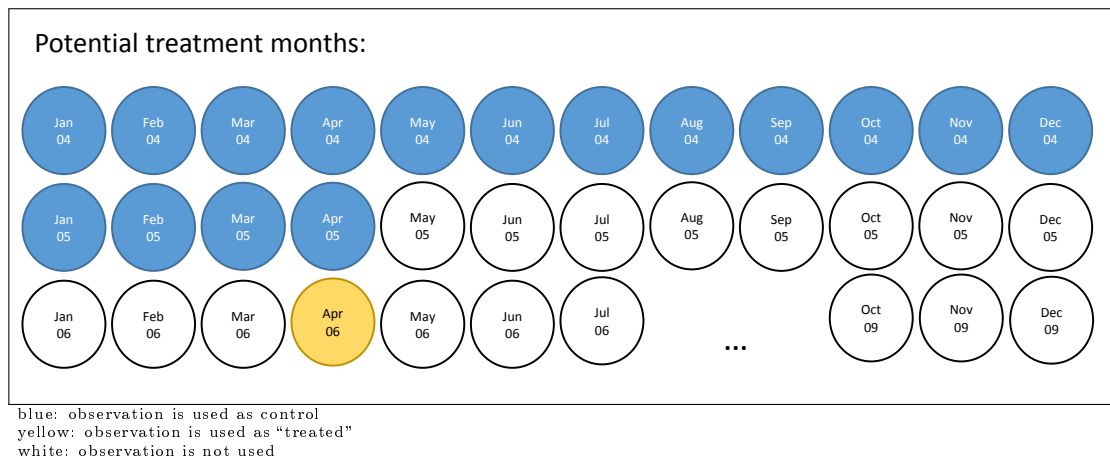


Figure 3: Duplication of observations for temporal alignment

Using this approach results in a very large data set. The approximately 50,000 woman, we have in our sample and which are shown in table 1, add up to 10,695 treatment and more than 2,650,000 control observations because not-yet-treated women are duplicated. Main advantage of the temporal alignment is that the outcomes for both treated and control observation can be now analyzed with respect to the timing of treatment. It is for example possible to plot average daily earnings and labor market participation for treatment and control group and separated by time periods (pre-reform: treatment years 2004 to 2006 and post-reform: treatment years 2007 to 2009).

We determine the average earnings and the employment rate for the treated and non-treated by month with respect to birth (treatment time) after temporal alignment. Non-treated women are considered for all the potential treatment times as long as they not have not had a child. The average outcomes are determined for the pre-reform and the post-reform period which are calculated in two steps. First, for each treatment month the average outcomes are calculated for all months from 35 months before the treatment month up to 59 months after the treatment month. Second, we average these averages by

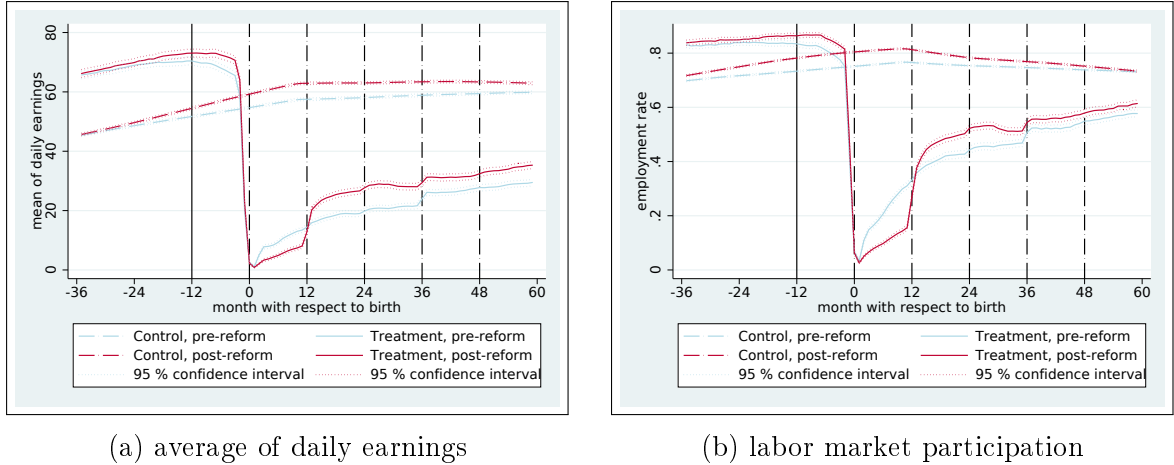


Figure 4: Raw data after temporal alignment

treatment month weighting by the number of treated mother. As to be expected, Figure 4 show that the labor market outcomes for mothers drop sharply to virtually zero at child birth. The recovery afterwards seems to be influenced by the reform as the lines for pre- and post-reform differ significantly. However, we see clearly that we should not interpret the difference between control women and actual mothers. The two group already differ strongly in the pre-treatment period. Mothers earn more and have a higher probability to be employed in the third and second year before giving birth. As there is yet no child neither present nor anticipated, this difference cannot be causal and is a clear evidence for the positive selection of mothers.

4.1.2 Inverse Probability Weighting (IPW)

To control the selection of treated women, we use monthly propensity scores for giving birth at a certain date. We estimate separate models for the pre-reform and the post-reform period while pooling across treatment months within each subperiod. Specifically, we model the probability of having the first child twelve months from now as a function of socio-demographic characteristics and labor market history. Controlling for past labor market career is crucial for successful matching.

Under the unconfoundedness of the treatment and perfect overlap in the propensity score, Busso et al. (2014) find that in small samples with unknown propensity score, a modified inverse probability weighting estimator (IPW) performs well. The crucial modification of the IPW estimator 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 4.1.1. Correspondingly, the alignment between treated and nontreated observations changes as well by age and month of birth, respectively. Recall, that we estimate the average treatment effect on the treated (ATT), i.e. the average treatment effect for the actual mothers. This treatment effect is the effect of giving birth in different months t on labor market outcomes at different *months-relative-to-treatment-time* which can be estimated by IPW as

$$\hat{\theta} = \frac{1}{\sum_t \sum_{i=1}^n T_{i,t}} \cdot \sum_t \left\{ \sum_{i=1}^n T_{i,t} \left(Y_i - \frac{\sum_{j=1}^n \hat{W}_{j,t} Y_{j,t}}{\sum_{j=1}^n \hat{W}_{j,t}} \right) \right\} \quad (1)$$

with weights $\hat{W}_{i,t} = E_{j,t} \hat{p}_t(X_j) / (1 - \hat{p}_t(X_j))$. n is the total number of women in the data set. $T_{i,t}$ denote the treatment dummy variable for individuals i in treatment month t (treated and non-treated), respectively, and $\hat{p}_{t(i)}(X_j)$ denotes the estimated probability to have a first child in time period $t(i)$ as a function of covariates X_j . With slight abuse of notation, X_j can vary over time and X_j denotes the covariates of the nontreated woman in period t . The application of the weights $W_{i,j}$ leads to a reweighting of the nontreated women according to the odds-ratio of having a child within the next year. $E_{j,t}$ is a dummy variable for eligibility as nontreated observation which takes the value of one if woman j can be used as a control observation for treatment month t . Otherwise, $E_{j,t}$ is set to zero. Thus, $E_{j,t}$ is a dynamic non-treatment dummy. Our benchmark analysis sets the classification window to 12 months, i.e. for at least 11 months after the respective time period at first birth for the treated woman i .

The IPW reweighting estimator has the advantage of not relying on a tuning parameter. 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 is estimated by a probit regression based on the observations in the aforementioned duplicated data set at *month – relative – to – treatment – 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 the three educational categories (missing value serves as reference category). Further, we include fixed effects for the calendar month, years and age at treatment. Arguably, most important are the controls for the employment history. We include three times 24 variables for the earnings, participation and full-time status of the second and third year

before giving birth (month -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 :

$$P(\text{mother}_{it} = 1|X_i) = \Phi \left(\beta_0 + \beta_1 \text{east}_{it} + \sum_{l=22}^{40} \beta_l I(\text{age}_{it} = l) + \sum_{j=1}^3 \gamma_j I(\text{edu}_{it} = j) + \sum_{k=-35}^{-12} (\delta_k \text{earn}(k)_{it} + \alpha_k \text{parti}(k)_{it} + \theta_k \text{ft}(k)_{it}) + \lambda_m + \mu_y \right) \quad (2)$$

with $\Phi(\cdot)$ representing the standard normal distribution function and X_i comprising the aforementioned partly time-varying covariates considered in eq. (2). For $\hat{p}_{t(i)}$ in eq. (1) we use the fitted value of $P(\text{mother}_{it} = 1|X_i)$.

4.2 Reform Effect

The reform effect can manifest itself in the change in the selection of mothers and in the change in the effect of motherhood on post-birth outcomes. A simple comparison of the post-birth motherhood effects on outcomes will confound those two effects. In the following, we describe our approach to separate the two effects. Furthermore, we show the reform effect on the selection of mothers with respect to their socio-economic characteristics and pre-birth outcomes, i.e. the variables we use to control the selection of women into motherhood, when estimating the causal effect of motherhood (see previous section), and the implied selection effect on the causal effect of motherhood on post-birth outcomes. Accounting for the selection effect of the reform, we then determine the remaining reform effect on the causal effect of motherhood.

The reform effect on the selection of mothers is determined by estimating the average changes in socio-economic characteristics and pre-birth outcomes between the pre-reform period and the post-reform period. We investigate both the absolute change for mothers and the change in the difference between mothers and nontreated women who are temporally aligned, i.e. the difference-in-differences estimator. We will present graphical evidence on this selection effect.

To assess the second dimension of the reform effect on the selection of mothers with regard to the causal effect of motherhood on post-birth outcomes, we use IPW to reweight the

sample of mothers in the pre-reform period to the sample of mothers in the post reform. This reweighting allows to estimate the counterfactual causal effect of motherhood - the ATT - on post-birth outcomes that would have applied for a sample of mothers as observed in the post-reform period.

Estimating this counterfactual ATT - in the following denoted at *pre-reform [mothers], reweighted [to post-reform mothers]* is implemented in two steps:

1. We take the total sample of mothers for the pre- and post-reform period and run a Probit regression to estimate the probability that a mother gives birth in the post-reform period as a function of pre-birth characteristics and history as observed 12 months before birth. This probability $P(post_i = 1|X_i)$ is specified as a function of age, education, region (east, west), and labor market history (earnings, part-time 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-treatment).
2. Then, we estimate the counterfactual ATT of motherhood for *pre-reform [mothers], reweighted [to post-reform mothers]* by reweighting pre-reform mothers with

$$g_i = \frac{\hat{P}(post_i = 1|X_i)}{1 - \hat{P}(post_i = 1|X_i)},$$

where $\hat{P}(post_i = 1|X_i)$ are the fitted probabilities for the Probit regression in step 1. The counterfactual ATT is then given by

$$\hat{\theta}(\text{post-sample,pre-effect}) = \frac{1}{\sum_t \sum_{i=1}^{npre} g_i T_{i,t}} \cdot \sum_t \left\{ \sum_{i=1}^{npre} g_i T_{i,t} \left(Y_i - \frac{\sum_{j=1}^{npre} \hat{W}_{j,t} Y_{j,t}}{\sum_{j=1}^{npre} \hat{W}_{j,t}} \right) \right\} \quad (3)$$

with weights $\hat{W}_{i,j}$ as in eq. (1). Here, $i, j = 1, \dots, npre$ denotes the sample of women observed for the pre-reform period (birth takes place before the reform and control observations in the pre-reform period).

The pre-reform observations of mothers, the ones for the treatment periods from January 2004 to December 2006, receive a weight g_i which is high for those mothers who are comparable to the mothers of the year from 2007 to 2009.

Let the factual ATTs of motherhood in the pre-reform period and the post-reform period - as given by eq. (1) - be denoted by $\hat{\theta}(\text{pre-sample,pre-effect})$ and $\hat{\theta}(\text{post-sample,post-effect})$, respectively. The raw difference between pre-reform and post-reform period is given by the differences of these ATTs, i.e. by

Then, the raw difference between pre- and post-reform is given by

$$\text{raw_diff} = \hat{\theta}(\text{post-sample,post-effect}) - \hat{\theta}(\text{pre-sample,pre-effect}) . \quad (4)$$

However, this is not the causal reform effect because the treated mothers before and after the reform also differ systematically in a socio-economic characteristics and labor market history. These difference in the selection of mother may itself result in a change in the causal effect of motherhood. The effect on the change of the selection of mother (henceforth, the selection effect) on the ATT can be estimated by

$$\text{sel_eff} = \hat{\theta}(\text{post-sample,pre-effect}) - \hat{\theta}(\text{pre-sample,pre-effect}) . \quad (5)$$

This selection effect quantifies how the counterfactual effect of motherhood for the post-reform sample of mother would have differed in the pre-reform period from the effect of motherhood as observed for the pre-reform sample. This difference arises because of the change in the sample of mothers between the pre-reform sample and the post-reform sample.

The causal reform effect of the ATT of motherhood for the post-reform sample of mothers is then given by

$$\begin{aligned} \text{reform-eff} &= \text{raw_diff} - \text{sel_eff} \\ &= \hat{\theta}(\text{post-sample,post-effect}) - \hat{\theta}(\text{post-sample,pre-effect}) . \end{aligned} \quad (6)$$

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

The results of the inverse probability weighting are shown in figure 5. Compared to the graph on the raw data, the values for the treatment group (solid blue and red line) have not changed at all as the mothers have not been reweighted. However, the two lines for the control group are shifted. In the pre-treatment period, the left hand part of the graph until one year before giving birth (month -12), the controls results match

very precisely the treatment results. This gives us confidence that we are indeed able to achieve comparability in absence of the treatment for the two groups.

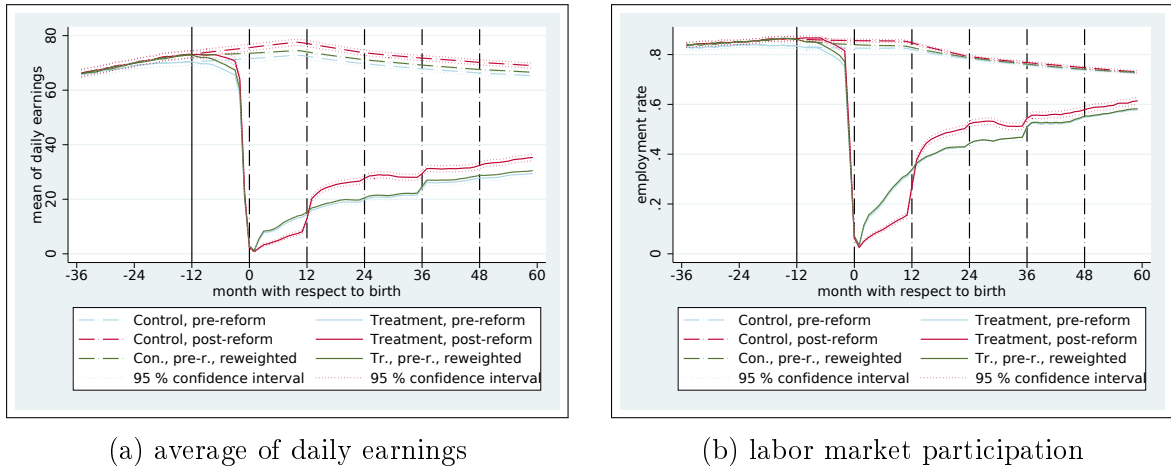


Figure 5: Treatment and control group after IPW

The same holds true for the reweighted pre-reform sample using the weights of (3) (the green lines in figure 5). This group matches in the pre-treatment period, the results of the post-reform group. After giving birth the values for that group are close to the one of the unweighted pre-reform group. The selection induced by the reform seems hence rather small but we will have a closer look on that in the following section.

5 Results

In this project, we regard earnings, employment and full-time employment as labor market outcomes. Additionally, the effects on second order fertility are evaluated. In the end of this section, we also show how the effects differ for different age groups of women.

5.1 Findings on labor market outcomes

The effects of giving birth on earnings are shown in figure 6. The effect is given for the post-reform period, for the pre-reform period and for the reweighted pre-reform sample. The three lines are the difference between the respective lines for treatment and control group of graph 5. These effects of giving birth follow for all three groups similar pattern, until one year before giving birth there is unsurprisingly no effect at all, shortly before giving birth, there is a dramatic decline in earnings and than a slow recovery. There are

yet also significant differences between pre- and post-reform groups. This is the reform effect.

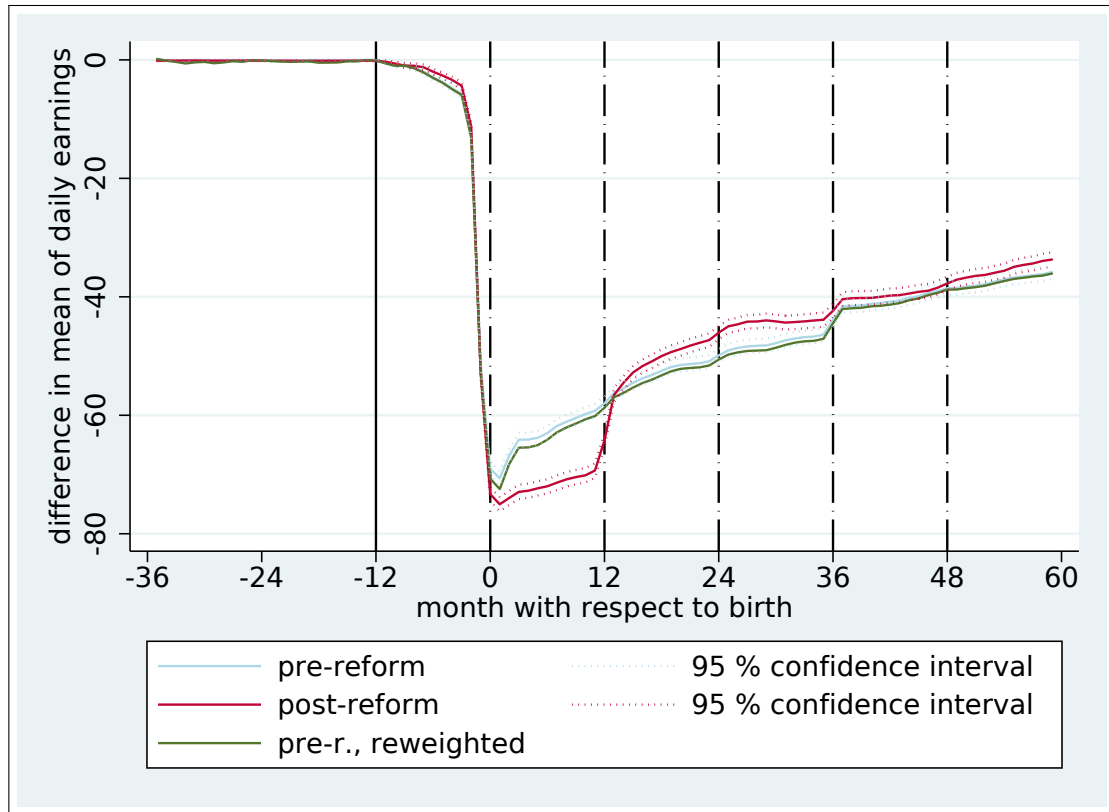


Figure 6: Effects of giving birth on daily earnings

The reform has its largest effects in the short run – the first two years after giving birth. Here we replicate the results of the literature. A decrease in earnings in the first year after giving birth and an increase in the second year. There is further a visible increase in earnings after the end of the third year after giving birth. This appears in pre- and post-reform periods and marks the end of job protection.

In the medium run, the effect is significant as well. We estimate that the reform increased daily earnings on average in the fifth year after giving birth by 1.91 Euro (an almost 700 Euro increase in yearly net-earnings)(see for the estimates the top panel of table A.1 in appendix). Further, we see a significant increase in earnings in the first year before giving birth. This effect has not been noted so far in the literature to our knowledge. It is convincing as the received benefits after reform depend on earnings while their were flat before. Mothers now have incentives to maintain their work until the begin of job protection. This "entitlement effect" is significant on four percent level. According to graph 6, it seems to stem from the ninth to fourth month before giving birth and its size

is estimated to be 0.77 Euro in daily earnings (280 Euro yearly).

We do however not detect a significant difference between the unweighted and reweighted pre-reform sample. The latter is reweighted to match the characteristics of the post-reform sample to simulate a situation where the reform did not influence the selection of mothers. As shown in figure 4 and 5, there are differences between pre- and post-reform mothers already before becoming pregnant. There is hence some selection to be at work. However, these differences in the selection do not seem to be important for the effect of giving birth. The selection induced by the reform is insignificant in the after birth earning outcomes.

For employment, the effects are generally very similar. We see the same patterns for the short run effects, the importance of job protection and a positive effect five years after giving birth (figure 7 and the top panel of table A.2). Mothers of the post-reform group are 2.6 percentage points more likely to be employed compared to the pre-reform group. Interestingly, this effect is quite stable in the fourth and fifth year after giving birth which would be an evidence that the effects do not fade out. Again, we find also the entitlement effect in the first year before giving birth. Mothers employment probability increases significantly by 1.3 percentage points while the effect of the selection induced by the reform is again barely visible.

The effects on full-time employment are shown in figure 8 and table A.3. The entitlement effect before giving birth is of similar size as the effect on employment rate indicating that full-time jobs are predominately maintained until maternity protection. The short-term effects show in the same direction as for employment but are smaller in size. Concerning medium run effects, we see surprisingly a significant negative effect. Women are estimated to have a 2.9 percentage points lower probability of being full-time employed.

The reform increased in the medium run earnings and employment. It decreases the probability to work full-time. This would lead to the conclusion that the positive effects on earnings and employment stem from part-time jobs.

5.2 Findings on second order fertility

It is conceivable that the parental benefits reform also affects the propensity to have a second child. Cygan-Rehm (2016) investigates this outcome and finds temporary but no lasting effects. We also introduce the incidence of having another child as outcome

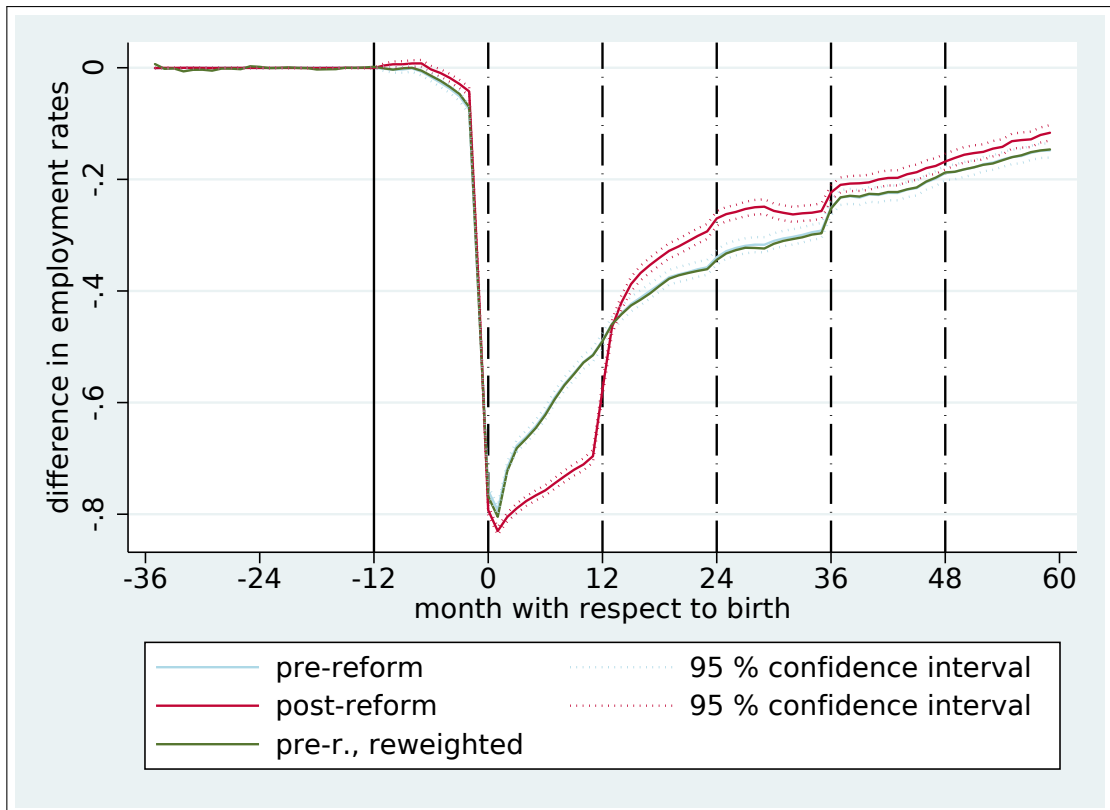


Figure 7: Effects of giving birth on participation rate

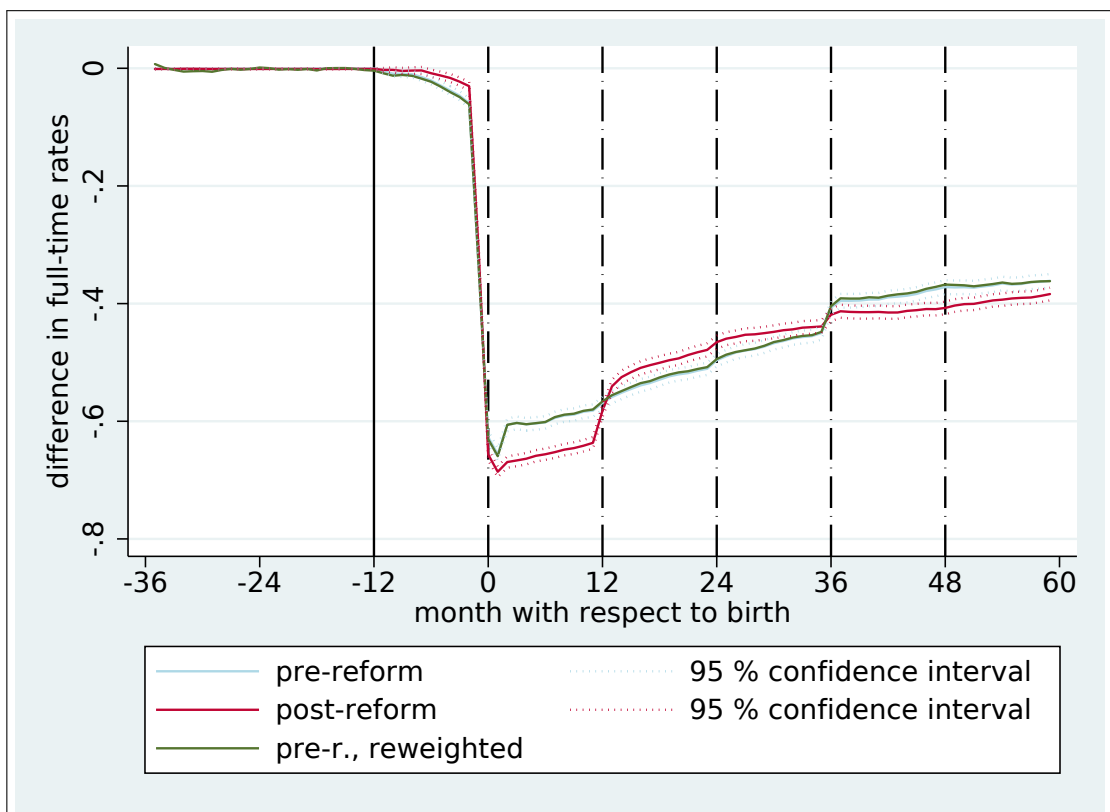


Figure 8: Effects of giving birth on full-time rate

variable. The results are shown in figure 9 and table A.4. They are almost entirely insignificant. Only in the fifth year after giving birth, there is a positive estimate which is significant at the ten percent level. However, we do not want to put too much emphasis on this marginal significant result. A positive effect on second-order fertility would be anyway rather surprising given the positive effects on labor market outcomes.

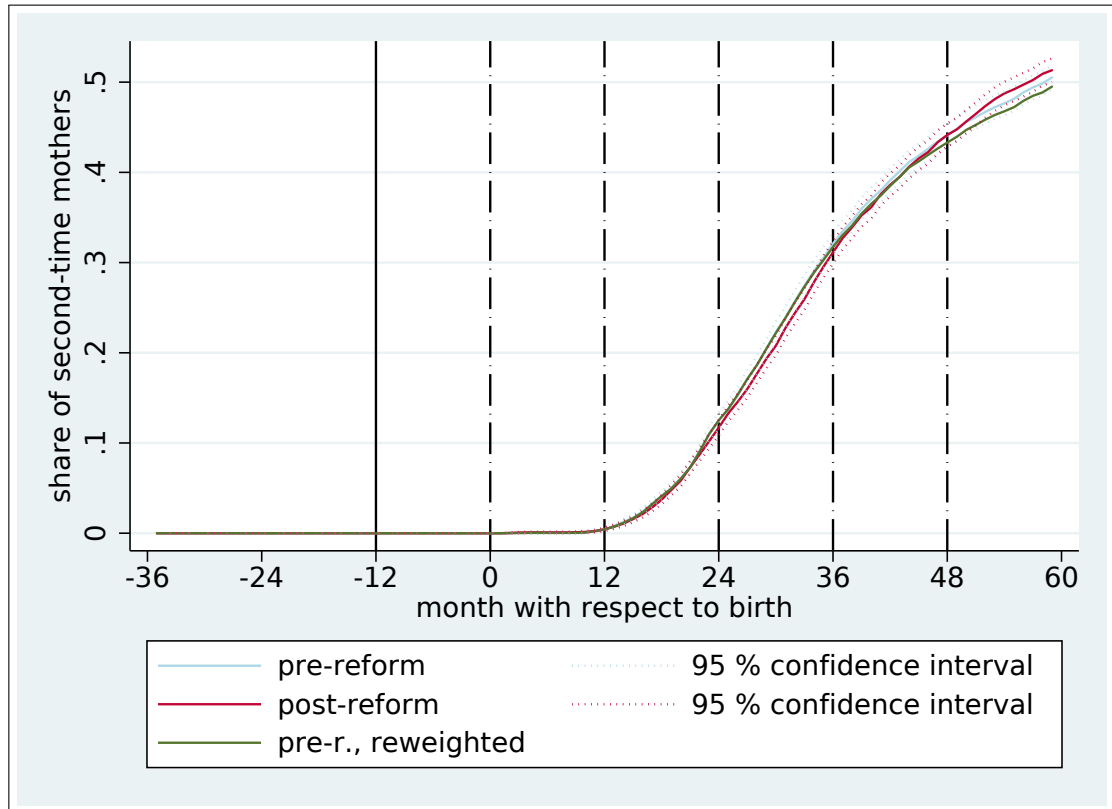


Figure 9: Effects of giving birth on second order fertility

5.3 Heterogenous effects by age of mothers

We also investigate how the effects differ by the age women. The results for labor market outcomes are shown in figures A.1 to A.3 and the lower panels of tables A.1 to A.3.

Generally, the medium run reform effect is arguably most beneficial for mothers aged 31 to 35. This means that they drive the positive effects on earnings and employment. For the full-time rate the effect is insignificant. The reform is on the other hand negative for the youngest age group in all three medium run outcomes. Mothers aged 21 to 25 have a significantly reduced probability to work full-time in the medium run. The negative estimates are insignificant for earnings and employment.

These finding can be easily related to the existing literature which separates results by pre-birth earnings (reform winners and losers). Bergemann and Riphahn (2011), Frodermann et al. (2020) and Kluge and Schmitz (2018), all find that the reform winners (with high pre-birth earnings) benefits above average while the reform losers do not gain in the medium run. If one assumed that pre-birth earnings are lowest for youngest mothers than our findings would agree with that literature.

Further, we show in the lower panel of table A.4 that the effect on second order fertility is insignificant for mothers of all age groups.

6 Comparison to alternative estimation strategies

A popular simple approach to estimate the effect of motherhood on post-birth outcomes and to assess the effect of a reform (involving changes in institutional regulations) amounts to estimating an event study regression using a nonparametric specification in time-relative-to-treatment (Kleven et al., 2019a,b). The event study regression uses the sample of mothers and estimates how post-birth outcomes differ from pre-birth outcomes while controlling for calendar time effects and socio-economic characteristics. This approach does not use a control group and rather relies on the changes between pre-birth outcomes and post-birth outcomes while controlling for other determinants of the outcome variable. We contrast our findings with the estimates from the following event study regression

$$Y_{iy_m} = \sum_{\substack{j=-35 \\ j \neq -12}}^{59} \alpha_j I(y_m - t = j) + \sum_{k=-35}^{59} \beta_k I(y_m - t = k) \text{postreform}_i \\ + \pi \text{east}_{iy_m} + \sum_{j=1}^3 \theta_j I(\text{edu}_{ism} = j) + \sum_{l=19}^{45} \gamma_l I(\text{age} = l) + \lambda_y + \delta_m + \epsilon_{iy_m}, \quad (7)$$

where Y_{iy_m} is the outcome for individual i observed in year y and month m . t is the month of birth and y_m is the year and month observed, thus $y_m - t$ is time relative to the birth month. This regression uses a series of dummies for month relative to birth $I(y_m - t = j)$ from 35 months before birth to 59 months after birth. The omitted category is 12 months before birth and the coefficients α_j measures the difference of the outcome in month j relative to month -12 , which is also our time of alignment for our control group approach

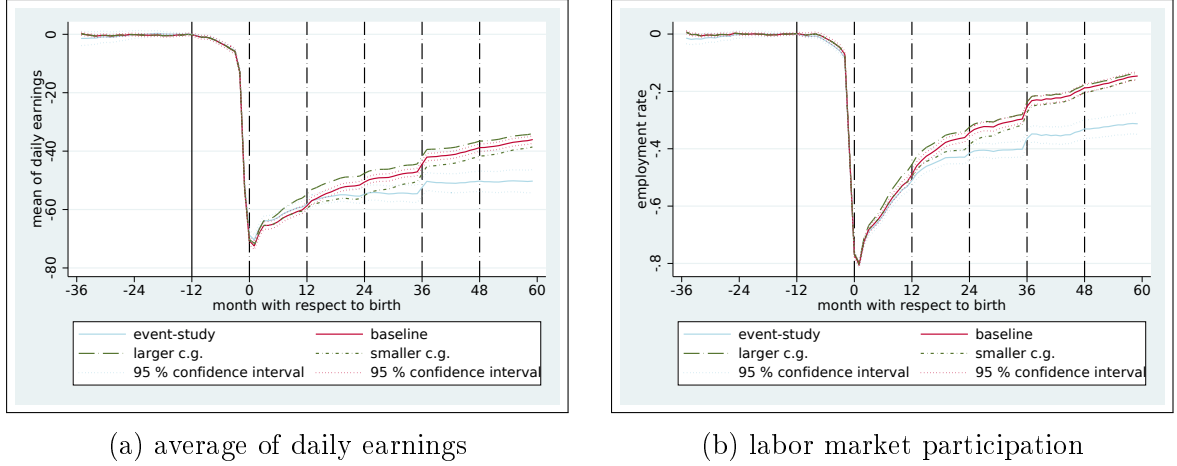
in section ???. α_j are the effect estimates for the the pre-reform period since the reform also involves interactions of the $I(ym - t = j)$ dummies with the post-reform period. The coefficients of the interactions β_k represent the changes of the motherhood effects between the pre-reform period and the post-reform period.

The identification of the immediate motherhood effect right after the birth relies on the implicitly assumed continuity of the outcome over time when accounting for calendar time (Kleven et al., 2019a). This assumption implying small changes over time in the nontreatment case of no motherhood is plausible over short periods of time, i.e. shortly after birth. However, it also relies on the assumption that pre-treatment outcomes for mothers mimic nontreatment outcomes for the counterfactual case of no motherhood. This means that the effects of time and socio-economic characteristics on the counterfactual are captured by the estimates on these covariates in eq. (7). The identifying assumptions underlying the event study approach are less plausible for estimating the effects after a while since treatment, as Kleven et al. (2019a) point out themselves. In addition, the assumption of continuity in time is less plausible when treatment effects are estimated relative to an earlier point of time - as -12 months - with a considerable gap to the treatment time to account for anticipation effects of treatment. Lastly, the event study regression in eq. (7) estimates the calendar time effect for the years after the pre-reform period based on post-birth outcomes for the pre-reform mothers and on outcomes for the post-reform mothers. Since the selection of mothers differs between the pre- and post-reform period, this may introduce a bias in the calendar time effects. Probably even more importantly, the calendar time effects after the year 2009 are solely based on the post-birth outcomes of mothers making it difficult to distinguish these from the causal effects of motherhood.

A further concern is that the estimated reform effect in eq. (7) does not account for differences in pre-birth labor market history apart from the outcome effect at month -12. In particular, this means that changes in the selection of mothers with regard to labor market history due to the reform are not accounted for when estimating the reform effect based on the event study regression in eq. (7).

To explore the importance of these concerns, we contrast the estimated treatment effects and reform effects based on our control group approach to the estimates for eq. (7).

For the pre-reform period, Figure 10 depicts the estimated coefficients α_j based on the

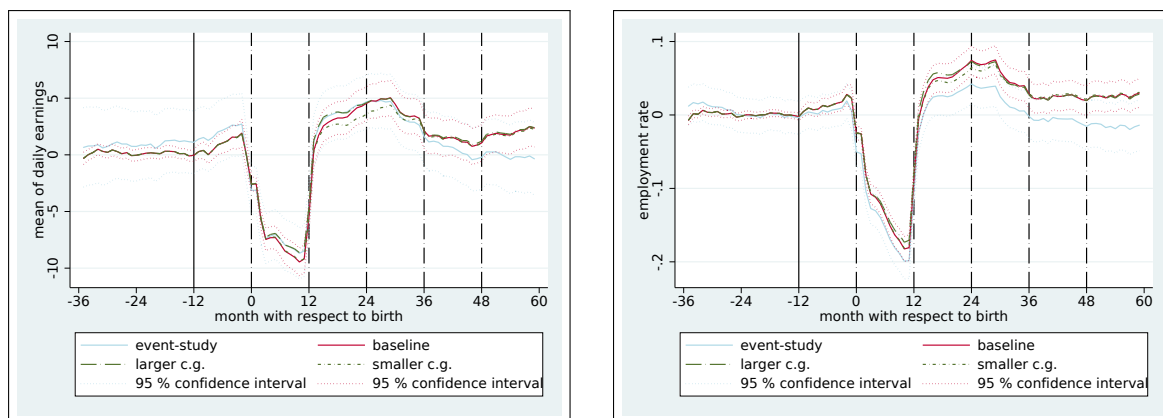


(a) average of daily earnings

(b) labor market participation

Figure 10: Effect of giving birth with event-study, baseline and two alternative control groups

event study regression (denoted by event-study) and the estimated treatment effects based on our control group approach (denoted by baseline [classification window of 12 months], smaller control group [classification window of 24 months], larger control group [classification window of 1 month]). The graph to the left shows the findings for earnings and the graph to the right for labor market participation. The effect estimates pre-treatment and during the first year after treatment are quite similar between the baseline and the event study approach. Thus, the event study approach provides a very good alignment before -12 months and captures well the decline in outcomes during the year preceding birth the effects during the first year after birth. However, the event study approach strongly overestimates the negative earnings and employment effects later on. In particular, the control group approach estimates a continuous recovery in earnings and employment over the course of five years after birth while the event-study approach misses most of the recovery after the first year. One reason for this finding are the better employment and earnings outcomes of the post-reform mothers provide a higher benchmark compared to the outcomes for the control group of our dynamic treatment approach, see Figure 4. Now, we turn to the reform effects. Figure 11 contrasts the estimated reform effects for the event study regression (the β_k 's in eq. (7)) and the reform effects based on the control group approach (baseline, smaller/large control group). Again we find that the effects are very similar between the different approaches before birth and for the first year after birth. However, the event study approach underestimates the positive employment effect from the second post-birth year onward. The difference is growing over time and it amounts



(a) average of daily earnings

(b) labor market participation

Figure 11: Reform effects with event-study, baseline and two alternative control groups

to about five percentage points in the fifth post-birth year. While the reform effects for earnings are still quite similar in the second and third year, the event study approach also underestimates the positive earnings effect in the fourth and fifth post-birth year. What are the reasons for these differences? In addition to the aforementioned reasons for the differences in the estimates of the causal motherhood effect, it is important to note that the event study regression misses changes in the selection of mothers regarding pre-birth outcomes between the pre-reform and the post-reform period, while our control group approach accounts for these.

In sum, we conclude that the event study approach provides quite different estimates starting from the second year after birth for both the causal effects of motherhood and the reform effect.

7 Conclusions

This paper estimates the causal effect of first-time motherhood on various post-birth outcomes in Germany and then 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. In the third and second year before giving birth, they earn around 40 percent more than the women in the control group and their employment rate is more than 10 percentage points higher. While the employment rate and the earnings before birth were higher after the reform than before, the reform had little impact on the pre-birth differences between mother and the control group.

While there are significant negative effects of the reform on labor market outcome during the first year, during which the new benefits are paid, the medium-run effects on earnings and employment are significantly positive. We estimate that the reform increased earnings by almost 700 Euros and the employment rate by 2.6 percentage points in the fifth year after giving birth. However, the reform reduced full-time employment in the medium run among mothers. The most positive effects are found for medium aged mothers (31 to 35 years of age) while the youngest age group (21 to 25 years of age) shows worse labor market outcomes relative to the control group after the reform. The reform effect on second-order fertility is small and insignificant.

Further, we find a positive effect during the year before giving birth. This "entitlement effect" is significant for all three labor market outcomes. This finding is plausible because the parental leave benefits in the post-reform period depend on the earnings of mothers immediately before birth. To our knowledge, this is the first study to establish this effect. Our evidence on the reform effects on post-birth outcomes fits quite well to the existing literature which finds positive medium-run effects on earnings and employment per se but not for full-time employment (Frodermann et al., 2020; Kluge and Schmitz, 2018). The positive reform effects stem from mothers with better pre-birth labor market outcomes. These "reform winners" are employed women with higher earnings (Bergemann and Riphahn, 2020; Frodermann et al., 2020; Kluge and Schmitz, 2018). Our results separated by age groups show that the post-birth labor market outcomes for mothers giving birth in their 30s increase after the reform while we do not find such positive reform effects for mothers giving birth in their early 20s. Both pre-birth employment and earnings are considerably higher among the former group compared to the latter group. Consequently, the share of "reform winners" is much larger in the for the older age group.

The econometric strategy of our project differs substantially from most of the literature on estimating the causal effect of motherhood on post-birth outcomes and on assessing the impact of institutional changes in this context. For short-run outcomes post birth, the methodological differences do not matter much because findings are strongly driven by the strong dip in employment and earnings for mothers immediately after birth. However, the results on the effect of giving birth start to diverge from the second year after giving birth onward compared to a event study approach as in Kleven et al. (2019b). Our findings using a control group approach in a dynamic treatment setting shows better medium-run

effects of motherhood compared the event study approach. Even though there is some ambiguity regarding the choice of the appropriate control group, the differences between alternative definitions of the control group are minor and do not matter in the medium run. Regarding the reform effect, the differences between our approach and an event study approach are also important. While the effects on earnings and employment are insignificant using an event study approach, our control group approach implies positively significant reform effects on post-birth earnings and employment in the medium run. This means that the choice of the econometric strategy matters when assessing the effects of the 2007 parental leave reform in Germany.

Bibliography

- 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 Bundeserziehungsgeldgesetz. 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.
- 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.
- 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? *The Review of Economics and Statistics*, 98(4):655–670.
- 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. (2020). Parental Leave Reform and Long-Run Earnings of Mothers. Discussion Papers of DIW Berlin 1847, DIW Berlin, German Institute for Economic Research.
- 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.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child penalties across countries: Evidence and explanations. (25524).

- Kleven, H., Landais, C., and Sogaard, J. E. (2019b). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4).
- Kluge, 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.
- Kluge, 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.
- Müller, D. and Strauch, K. (2017). Identifying mothers in administrative data. Fdz meth-odenreport, 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.
- Rossin-Slater, M., Ruhm, C. J., and Waldfogel, J. (2013). The effects of california's paid family leave program on mothers' leave-taking and subsequent labor market outcomes. *Journal of Policy Analysis and Management*, 32(2):224–245.
- 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.
- Sigle-Rushton, W. and Waldfogel, J. (2007). Motherhood and women’s earnings in anglo-american, continental european, and nordic countries. *Feminist Economics*, 13(2):55–91.
- Statistisches Bundesamt (2019). Verbraucherpreisindizes für Deutschland. *Jahresbericht 2018*.

Appendix

Tables

Table A.1: Average reform effects on mean earnings

| | 1st year before giving birth | | 1st year 2nd year 3rd year 4th year 5th year after giving birth | | | | | | | | | |
|-----------------|---------------------------------|-----------------|--|------------------|-------------------|------------------|-----------------|-------------------|-------------------|------------------|------------------|------------------|
| | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE |
| entire sample | .77** (.033) | .52* (.059) | -7.15*** (0) | -8.37*** (0) | 2.26*** (.001) | 1.52** (.042) | 4.22*** (0) | 3.47*** (0) | 1.4* (.062) | .97 (.183) | 1.91** (.017) | 1.66** (.035) |
| women age 21-25 | -.61 (.368) | -.46 (.421) | -6.07*** (0) | -6.07*** (0) | -1.82* (.074) | -1.87 (.143) | -.65 (.578) | -.95 (.467) | -1.71 (.184) | -2.28* (.097) | -1.65 (.242) | -2.05 (.151) |
| women age 26-30 | 1.09** (.014) | .82** (.041) | -6.83*** (0) | -5.37*** (0) | 1.04 (.228) | 1.72* (.094) | 1.65* (.089) | 1.91* (.076) | -.01 (.991) | -.03 (.979) | 1.04 (.344) | 1.16 (.31) |
| women age 31-35 | .34 (.586) | .28 (.562) | -7.54*** (0) | -8.39*** (0) | 3.58*** (.006) | 3.14** (.028) | 6.38*** (0) | 5.97*** (0) | 1.93 (.161) | 1.54 (.273) | 3.05** (.035) | 2.69* (.063) |
| women age 36-40 | 2.51* (.099) | 1.55 (.161) | -7.91*** (0) | -11.02*** (0) | 5.7** (.02) | 3 (.262) | 9.69*** (0) | 7.29*** (.008) | 6.84*** (.009) | 5.03** (.049) | 4.44 (.117) | 1.97 (.471) |

Average effects as difference between the post-reform and pre-reform (reweighted) results for the reform effect (RE) and between post-reform and pre-reform results for the "full reform effect (full RE)". The former excludes and the latter includes potential reform effects on the selection of mothers.

P-values in parentheses refer to a t-test on significance of the average effect for the respective years.

***, ** and * indicate significance on 99, 95 and 90 percent level.

Table A.2: Average reform effects on participation rate

| | 1st year before giving birth | | 1st year 2nd year 3rd year 4th year 5th year after giving birth | | | | | | | | | |
|-----------------|---------------------------------|-------------------|--|-----------------|-------------------|-------------------|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE |
| entire sample | .013*** (.002) | .014*** (0) | -.121*** (0) | -.126*** (0) | .034*** (0) | .031*** (0) | .059*** (0) | .054*** (0) | .025*** (.006) | .026*** (.002) | .026*** (.006) | .026*** (.004) |
| women age 21-25 | .012 (.298) | .013 (.205) | -.113*** (0) | -.129*** (0) | .018 (.455) | .011 (.616) | .042* (.09) | .036 (.117) | .03 (.2) | .027 (.214) | -.013 (.549) | -.017 (.414) |
| women age 26-30 | .014** (.015) | .014*** (.006) | -.13*** (0) | -.139*** (0) | .032** (.018) | .024* (.072) | .055*** (0) | .046*** (.001) | .02 (.166) | .016 (.267) | .033** (.026) | .029** (.043) |
| women age 31-36 | .011* (.081) | .013** (.016) | -.114*** (0) | -.113*** (0) | .046*** (.002) | .049*** (.001) | .065*** (0) | .066*** (0) | .021 (.15) | .024 (.107) | .032** (.034) | .035** (.017) |
| women age 36-40 | .013 (.346) | .01 (.359) | -.13*** (0) | -.152*** (0) | .017 (.525) | .007 (.781) | .067** (.015) | .06** (.025) | .041 (.12) | .041* (.1) | .021 (.422) | .017 (.494) |

Average effects as difference between the post-reform and pre-reform (reweighted) results for the reform effect (RE) and between post-reform and pre-reform results for the "full reform effect (full RE)". The former excludes and the latter includes potential reform effects on the selection of mothers.

P-values in parentheses refer to a t-test on significance of the average effect for the respective years.

***, ** and * indicate significance on 99, 95 and 90 percent level.

Table A.3: Average reform effects on full-time rate

| | 1st year | | 1st year | | 2nd year | | 3rd year | | 4th year | | 5th year | |
|-----------------|---------------------|------------------|--------------------|--------------------|-------------------|-------------------|------------------|-------------------|--------------------|--------------------|--------------------|--------------------|
| | before giving birth | | after giving birth | | | | | | | | | |
| | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE |
| entire sample | .016*** (0) | .013*** (0) | -.053*** (0) | -.053*** (0) | .022*** (.001) | .025*** (0) | .02*** (.004) | .021*** (.003) | -.027*** (0) | -.023*** (.001) | -.029*** (0) | -.027*** (0) |
| women age 21-25 | .018 (.17) | .006 (.549) | -.051*** (0) | -.051*** (.002) | .016 (.285) | .013 (.45) | .01 (.558) | .006 (.746) | -.032* (.081) | -.037** (.039) | -.036** (.045) | -.043** (.018) |
| women age 26-30 | .019*** (.002) | .014** (.011) | -.056*** (0) | -.05*** (0) | .015 (.131) | .019* (.086) | .003 (.763) | .004 (.747) | -.029*** (.007) | -.031*** (.004) | -.033*** (.002) | -.033*** (.002) |
| women age 31-35 | .012** (.049) | .012** (.016) | -.046*** (0) | -.05*** (0) | .039*** (0) | .038*** (.001) | .046*** (0) | .046*** (0) | -.01 (.444) | -.01 (.464) | -.003 (.78) | -.006 (.624) |
| women age 36-40 | .015 (.252) | .008 (.407) | -.072*** (0) | -.106*** (0) | -.006 (.785) | -.031 (.18) | -.002 (.933) | -.023 (.312) | -.071*** (.003) | -.082*** (0) | -.087*** (0) | -.1*** (0) |

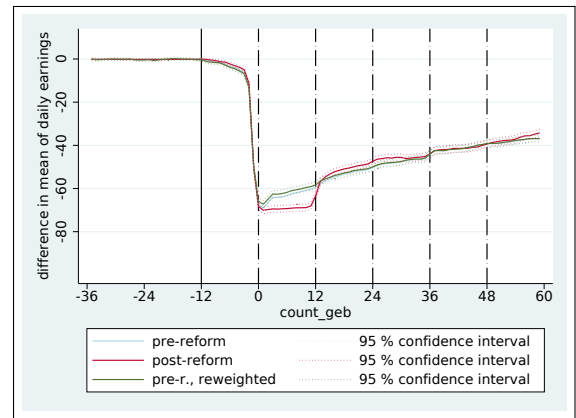
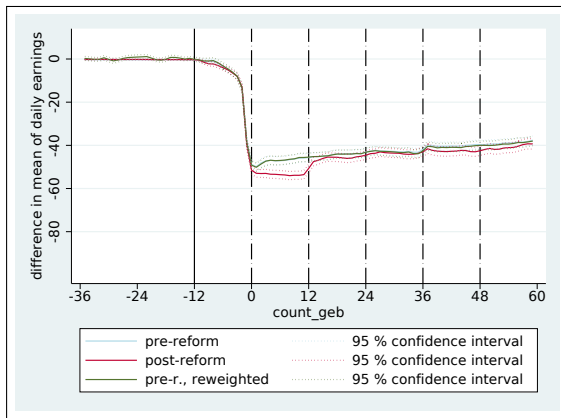
Average effects as difference between the post-reform and pre-reform (reweighted) results for the reform effect (RE) and between post-reform and pre-reform results for the "full reform effect (full RE)". The former excludes and the latter includes potential reform effects on the selection of mothers. P-values in parentheses refer to a t-test on significance of the average effect for the respective years. ***, ** and * indicate significance on 99, 95 and 90 percent level.

Table A.4: Average reform effects on second order fertility

| | 1st year | | 2nd year | | 3rd year | | 4th year | | 5th year | |
|-----------------|--------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------------|
| | after giving birth | | | | | | | | | |
| | RE | full RE | RE | full RE | RE | full RE | RE | full RE | RE | full RE |
| entire sample | .0004 (.31) | .0004 (.328) | -.0022 (.472) | -.0027 (.361) | -.0101 (.169) | -.0107 (.127) | -.0001 (.995) | -.0057 (.512) | .0154* (.095) | .0064 (.487) |
| women age 21-25 | .0022 (.147) | .0023 (.127) | -.0039 (.688) | -.0016 (.853) | -.0199 (.318) | -.0143 (.44) | -.0222 (.374) | -.0187 (.432) | .0031 (.902) | .0028 (.91) |
| women age 26-30 | -.0001 (.891) | 0 (.958) | -.0047 (.32) | -.0053 (.261) | -.0173 (.157) | -.0166 (.163) | -.0103 (.494) | -.0103 (.49) | .0085 (.582) | .0085 (.578) |
| women age 31-35 | .0008 (.161) | .0007 (.178) | -.0003 (.952) | .0003 (.954) | -.0009 (.942) | .0012 (.916) | .0159 (.29) | .0165 (.271) | .0236 (.14) | .0232 (.145) |
| women age 36-40 | -.001 (.32) | -.0015 (.264) | .0008 (.918) | -.0015 (.848) | -.0088 (.651) | -.0088 (.633) | .002 (.938) | .001 (.966) | .0216 (.418) | .02 (.428) |

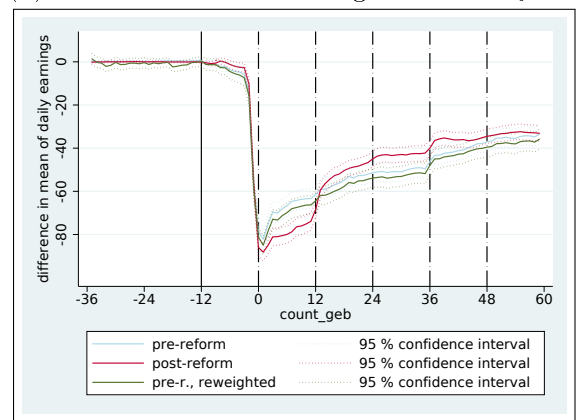
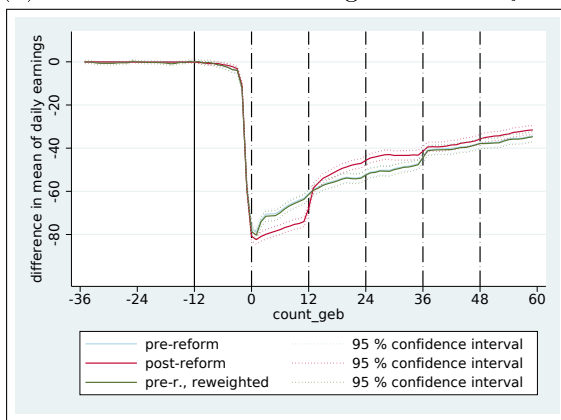
Average effects as difference between the post-reform and pre-reform (reweighted) results for the reform effect (RE) and between post-reform and pre-reform results for the "full reform effect (full RE)". The former excludes and the latter includes potential reform effects on the selection of mothers. P-values in parentheses refer to a t-test on significance of the average effect for the respective years. ***, ** and * indicate significance on 99, 95 and 90 percent level.

Graphs on heterogenous effects by age groups



(a) estimates for mothers aged 21 to 25 years

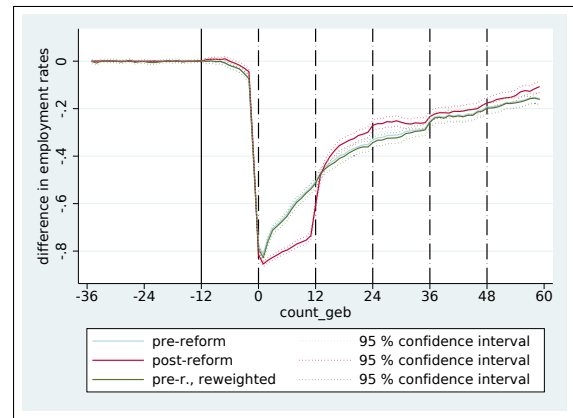
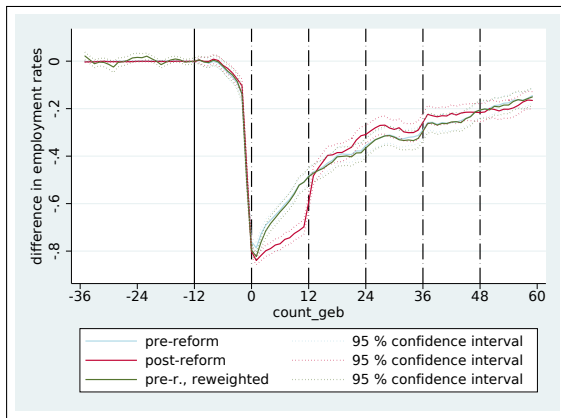
(b) estimates for mothers aged 26 to 30 years



(c) estimates for mothers aged 31 to 35 years

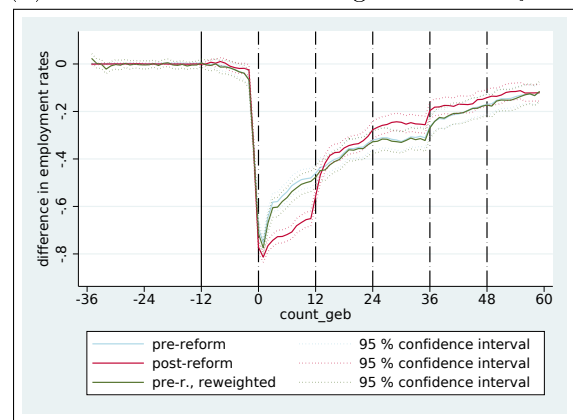
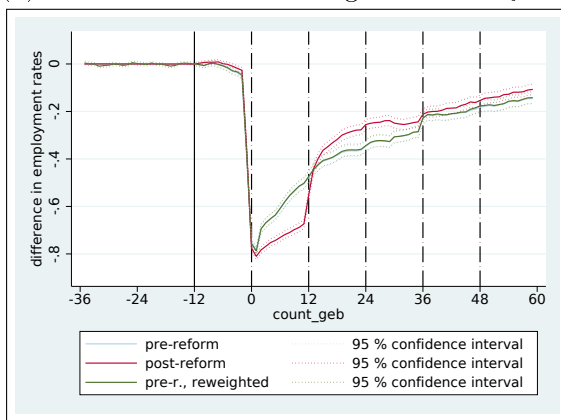
(d) estimates for mothers aged 36 to 40 years

Figure A.1: Effects of giving birth on daily earnings according to age groups



(a) estimates for mothers aged 21 to 25 years

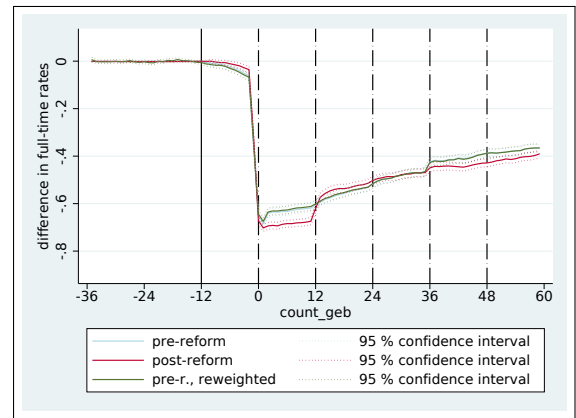
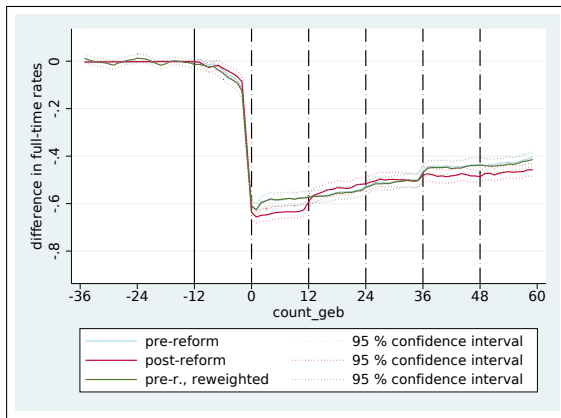
(b) estimates for mothers aged 26 to 30 years



(c) estimates for mothers aged 31 to 35 years

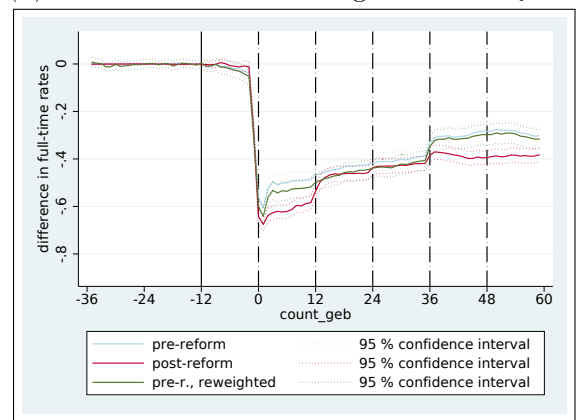
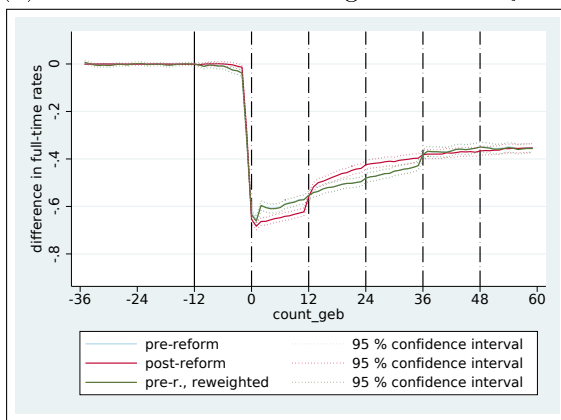
(d) estimates for mothers aged 36 to 40 years

Figure A.2: Effects of giving birth on participation 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 A.3: Effects of giving birth on full-time rate according to age groups