Causal effects on employment after first birth
- A dynamic treatment approach -

Bernd Fitzenberger*, Katrin Sommerfeld**, Susanne Steffes***

February 2012

Abstract: The effects of childbirth on future labor market outcomes are a key issue for policy discussion. This paper estimates by age the effect of having the first childbirth now versus waiting on future employment of the mother. We employ the dynamic treatment approach of Sianesi (2004, 2008) and Fredriksson and Johansson (2008). Our implementation uses inverse probability weighting (IPW) as advocated by Busso et al. (2009). We also assess effect heterogeneity by estimating ex post outcome regressions as in Abadie and Imbens (2011). We implement this approach at a monthly frequency using the German SOEP data set from 1991 to 2008. The results show that there are very strong negative employment effects after childbirth. Although the employment loss is reduced over the first five years following childbirth, the negative employment effect does not level off to zero. The employment loss is lower for high-skilled mothers and for them it levels off to zero five years after childbirth. The 2001 reforms, which facilitated part–time work and increased incentives for an earlier return-to-job, reduced significantly the negative employment effects after first childbirth.

JEL-Classification: C14, J13, J22

Keywords: Female labor supply, Maternity leave, Dynamic treatment effect, Inverse Probability Weighting.

---

* University of Freiburg, IFS, IZA, ZEW. E-mail: bernd.fitzenberger@vwl.uni-freiburg.de

** University of Freiburg. E-mail: katrin.sommerfeld@vwl.uni-freiburg.de

*** Centre for European Economic Research ZEW. E-mail: steffes@zew.de

Acknowledgements: We thank Martha Bailey, Annette Bergemann, Steffen Künn, Astrid Kunze, and Jeffrey Smith for fruitful discussions. Moreover, we are grateful to seminar participants at University of Michigan in Ann Arbor, IZA Summer School, ESPE (Essen), University of Freiburg, SOEP user conference (Berlin), SSES (Fribourg), Workshop on Empirical Analysis of Family Economics (Aarhus), Leopoldina Workshop (Munich), Ausschuss für Ökonometrie, and ZEW Mannheim.

1 Introduction

Reconciling work and family life is a key issue in both families and public policy debates (OECD, 2007). Childbirth is associated with a strong reduction in female labor supply right after birth, which in a life-cycle perspective has adverse effects on later labor market outcomes (OECD, 2002; Ondrich et al., 1999; Schönberg and Ludsteck, 2007). Decisions about career objectives and about starting a family have to be taken jointly – which usually occurs between the age of 20 and 40 years. Many western countries, including Germany, exhibit low fertility rates and low female labor market participation rates, both posing great challenges for future economic development. International comparisons show that at least parts of these differences are due to huge disparities in the family policies (Gustafsson et al., 1996). The joint nature of fertility and employment decisions is thus at the heart of various policy reforms during the last decades.

Between the 1970s and the 1990s, Germany strongly expanded maternity leave coverage (Dustmann and Schönberg, 2011) from a period of a couple of months to three years. In light of the long coverage, there has been growing concern that mothers do not return to their job and that their careers are negatively affected by these long employment interruptions. These concerns led to a policy reform in 2001 providing financial incentives for an earlier return-to-job after childbirth and to foster part-time work when the child is young, partly to compensate for the negative incentive effect of means-tested income support during maternity leave. Another reform was implemented in 2007 (Gerlach et al., 2009; Bergemann and Riphahn, 2011b; Kluve and Tamm, 2009). Using a dynamic treatment effects approach, our paper estimates the effect of childbirth for females on employment over the course of five years after childbirth. When constructing the control group for females, who give birth at a certain age, we account for the fact that being childless at this age entails the possibility to have a child at a later age. We analyze the effects of the 2001 reform, but not of the 2007 reform because for births after the 2007 reform there are not yet enough observations on employment.

Most of the existing studies estimating causal effects of births on female labor force participation analyze the impact of policy changes in the maximum duration of parental leave. Ondrich et al. (1996, 2003) as well as Schönberg and Ludsteck (2007) find that in Germany the increases in the maximum duration between 1986 and 1992 led to successive increases in the duration of employment interruptions. A similar result is found by Lalive and Zweimüller (2009) for the extension of Austria's maximum parental leave duration in the year 1990 and by Baker and Milligan (2008) for Canada. Bergemann and Riphahn (2011a) and Kluve and Tamm (2009) analyze the 2007 reform in Germany.

This reform increases the income support during the first year of maternity leave, makes income support during parental leave available to male employees, and cuts the income support during the second and third year of maternity leave. This is to foster an early return-to-job by mothers, possibly via part-time work. Bergemann and Riphahn (2011a) and Kluve and Tamm (2009) find that during the first 12 months (when 67 percent of pre-birth earnings are paid) employment rates decline, while after 12 months – when parental non-means-tested benefits expire – employment probabilities increase. Kluve and Tamm (2009), Schönberg and Ludsteck (2007) as well as Lalive and Zweimüller (2009) estimate causal effects by applying a regression discontinuity design. The advantage of this approach is the random assignment of the treatment, as couples can hardly plan the exact timing of birth and did not know about the reform at the time of conception. However, one may be concerned about the external validity of these results in light of the possibility that individuals adjust their behavior over time when an institutional change is associated with changes in cultural norms (see Fortin, 2009, for the importance of gender role attitudes). Furthermore, Kluve and Tamm (2009) provide some evidence that childbirths were slightly reduced during the last weeks before the reform date and slighty increased during the days afterwards.

For the U.S., Goldin (2006) and Fortin (2009) describe the strong rise of labor force participation of women until the 1990s and the fact that it has leveled-off since. There is a popular discussion as to whether mothers in the U.S. increasingly "opt-out" of the labor market, although there is no clear evidence for such a trend (Goldin, 2006; Fortin, 2009). There exists a small but growing literature for the U.S. estimating specifically the effects of birth on subsequent employment of mothers. Troske and Voicu (2010, 2011) use an complex semi-structural five-equation model to estimate the causal effect of childbirth on the labor supply of mothers in the US. They analyze the effect of timing and spacing of births on employment by estimating a multinomial probit model for different employment states (full time, part time, etc) and, simultaneously, a probit model for the fertility decision. The equations are estimated sequentially with auto-correlated error terms. The advantage of this approach is the simultaneous modeling of labor supply and fertility decisions. Identification is mainly driven by including the number of children of the mother's siblings. However, this identification strategy requires a tight specification of the timing of the decision process. Furthermore, the empirical analysis is restricted to the selective sample of married women who have a first child after marriage. The estimated model contrasts the employment outcome of mothers with the employment outcome of childless women accounting for the correlation between labor market states and fertility. Troske and Voicu (2011) find that delaying the first birth leads to higher levels of labor market involvement before the birth of the first child and it reduces the negative employment effects after birth. Troske and Voicu (2010) emphasize the strong persistence in labor market states resulting in lasting negative effects of childbirth on employment. The study also finds strong effect heterogeneity of birth both regarding observables and unobservables, putting a cautionary note on the effectiveness of government policies to reconcile work and family.

A sequence of papers estimates the effect of delaying age at first birth on subsequent labor market outcomes. Herr (2007) and Miller (2011) instrument age at first birth by naturally occuring fertility shocks. They find that delaying first birth increases employment and wages after birth. The effects are strongest for college educated women and those in professional and managerial occupations (Miller, 2011). Wilde et al. (2010) criticize the use of the instruments used by the former two studies. Wilde et al. take issue with the lack of exogeneity when using time-varying instruments and with measurement error problems. Instead, they use characteristics of parents or events at early age to instrument the age at first birth. The study also finds that delaying first birth improves earnings after birth. All three studies discussed in this paragraph use women who have a child at a later age as the control group for women who have a child at an earlier age and they account for the endogeneity by instrumenting. Thus, these studies estimate a different parameter than the aforementioned studies by Troske and Voicu (2010, 2011).

Wilde et al. (2010) and Miller (2011) explore how the wage and career consequences of motherhood differ by skill and timing. Wage trajectories diverge sharply for high-skilled women after, but not before, they have children, while there is little change for low-skilled women. It appears that the lifetime costs of childbearing, especially early childbearing, are particularly high for skilled women. These differential costs of childbearing may account for the far greater tendency of high-skilled women to delay or avoid having a child altogether. Correspondingly, Miller (2011) finds that delaying birth shows a stronger positive effect for high-skilled women compared to less skilled women.

In this paper, we apply a dynamic treatment effect approach, which does not require complete modelling of the labor market decisions and the fertility decision. Important caveats are that we cannot estimate the causal effect of having a child versus not having a child or of having a child at two different points of time. Our approach has been used to estimate the effect of active labor market policies (Sianesi, 2004, 2008; Fitzenberger et al., 2008) in a situation where nontreated individuals today may be treated in the future. This approach also draws on Fredriksson and Johansson (2008) and is linked to the timing-of-events approach of Abbring and van den Berg (2003). We address the selection into childbirth by contrasting first-time mothers with highly comparable women who display a similar propensity to have a first child soon. This comparison is based on reduced-form estimates of the propensity to have a first child at a certain time and age. We assume that individuals can decide whether to have a child now or wait. After conditioning on labor market history and personal circumstances and values, the exact timing of birth is random, which is the basis of our identification strategy. Conditioning on future birth would result in a highly selective control group and for this reason we instead allow control individuals to have a child later or not. We align treatment and control group dynamically by age in months, which permits us to estimate a monthly treatment effect. This is done by inverse probability weighting (IPW) based on the estimated propensity to have a child now. Inverse probability weighting is preferred for bias and variance reasons in finite samples with very good overlap (Busso et al., 2009), if the weights are normalized to sum up to one (in contrast to Frölich, 2004). Finally, we employ ex-post outcome regressions (Abadie and Imbens, 2011) to analyze the heterogeneity of childbirth in more detail.

As the opportunity costs of childbirth vary with acquired human capital, e.g. higher work experience or education, the timing of first birth has an important impact on future career outcomes (Miller, 2011; Wilde et al., 2010; Troske and Voicu, 2010). Put differently, women of a certain age can decide whether to have a child now or wait – where the expected negative career effects vary by age. This raises two questions which are addressed by our study: 1) Does the *timing* of first birth have an impact on post-birth employment? 2) Is this impact in line with human capital theory, i.e. is the negative treatment effect weaker for women with more human capital? When addressing these questions, we recognize that post-birth employment behavior affects future employment chances because longer breaks and lower working hours reduce the amount of human capital (Lefebvre et al., 2009; Beblo and Wolf, 2002). This in turn is critical in the current debate on poverty among families and single parents. Clearly, the employment behavior in the first years after childbirth depends on the private environment, institutions (for instance availability of childcare, see Kreyenfeld and Hank, 2000; van Ham and Büchel, 2004), but also on pre-birth employment and education. Furthermore, the fertility decision anticipates the expectation about employment after childbirth.

On the one hand, women with higher human capital or higher labor force attachment have the most to lose from an employment break in terms of human capital and thus career opportunities (Troske and Voicu, 2010; Miller, 2011). We would therefore expect older and more educated first-time mothers to catch up more quickly to the control group than younger or less educated first-time mothers. On the other hand, for older first-time mothers who have had a successful career, the control group also consists of highly career-oriented women which could increase the employment loss caused by the birth of a first child. Thus, the direction of how the employment effect of first childbirth varies with age is an empirical question.

Our empirical analysis uses the German SOEP data for the time period 1991 to 2008 as it provides not only detailed monthly information on employment, but also a large set of control variables. In order to reduce the selectivity of the control group, we limit our analysis to women who have their first child between the age of 24 and 33. Our results show a sharp decline in employment around childbirth. The average treatment effect on the treated slowly converges to zero, implying that employment rates recover as the child grows older. However, we also show that within the observation period of five years the estimated average treatment effects on the treated do not level off. Although the employment effects differ depending on the age at first childbirth, the trends can not be ranked by age. This is because older mothers display higher employment rates, but the same holds for the corresponding control group. As expected from human capital theory first-time mothers with a university degree display weaker treatment effects. Finally, our results suggest that the 2001 reform was effective in fostering employment during the first five years after birth.

The paper proceeds as follows. The institutional background in Germany is described in section 2. Section 3 develops the econometric approach and section 4 describes the data used. The empirical results are discussed in section 5. Section 6 concludes.

2 Institutional Background

Maternity leave coverage in Germany, which provides employment protection after childbirth, was extended to three years in the early 1990s. This is very generous in international comparison (OECD, 2002; Schönberg and Ludsteck, 2007). There is concern that the long maternity leave coverage is one reason for the low employment rates of German mothers (OECD, 2002; Schönberg and Ludsteck, 2007; Ondrich et al., 1999). The German government changed the rules governing income support during maternity leave in 2001 and in 2007 in order to increase the incentive for mothers to return to their jobs sooner.[1] For

[1]The 2007 reform has received more attention in the public debate compared to the 2001 reform. In particular, the income support was changed from a means-tested flat transfer to a transfer which is proportional to the earnings before birth (up to a maximum level of Euro 1800) during the first year after birth. After 2007, all employed mothers are eligible for this new transfer and the level of the transfer is much higher. At the same time, the transfer was eliminated for the time period after 12 months. The 2007 reform also stipulates that the period with income support can be extended to 14 months if both parents take at least 2 months of maternity leave. Two goals of the reform were: First, to increase the incentives for male workers to take part of the maternity leave and who share the child care with the

our empirical analysis, we will only analyze the 2001 reform because for births after the 2007 reform there are not yet enough observations on employment.

Two different laws govern maternity leave coverage (Kreyenfeld, 2001; Schönberg and Ludsteck, 2007). Both laws concern the job protection and the regulation of financial benefits, and they apply to different time periods. The "Maternity Protection Law" (Mutterschutzgesetz) requires a pregnant woman to interrupt working six weeks before and eight weeks after birth in order to protect her and the baby's health.[2] During this period the mother receives her full net labor income which is financed partly by the employer and partly by the health insurance.

The "Parental Leave Law" (*Elternzeitgesetz*) regulates the job protection period and the income assistance during the parental leave period.[3] The job protection involves the right to return to a *comparable* job at the previous employer, which does not have to be exactly at the same workplace as before birth. The job protection period is three years of which up to twelve months may be delayed until the child reaches the age of eight. Effects of the extensions of maternity leave coverage from the 1970s to the early 1990s on children outcomes are analyzed by Dustmann and Schönberg (2011). According to the "Parental Leave Law", parental benefits are paid if the parent on leave does not work more than the allowed amount of hours, lives in the same household and predominantly cares for the child by him- or herself. Before 2001, a parent on parental leave had been allowed to work at most 19 hours a week. In January 2001, this upper bound was extended to 30 hours. At the same time, in 2001, the parent on leave was given the right to choose between receiving 300 Euro per month during the first 24 months or 450 Euro per month during the first 12 months – if eligible. This should provide incentives for a return-to-job right after 12 or 24 months. This maternity benefit was means-tested, i.e. it was paid to families with an annual net income less than 30,000 Euro in two-parent households and 23,000 Euro in single-parent households. Six months after birth these boundaries decreased to 16,500 (13,500) Euro plus 3,140 Euro for each additional child. In addition to maternity benefits, families receive a monthly child allowance (Kindergeld) for dependent children of today 184 Euro for the first and second child, 190 Euro for the third child and 215 Euro for all further children. Table 1 shows the development of the child allowance during our observation period. As an alternative, families can choose a

mother. Second, to help women to combine work and family and to allow for a quicker return to the job after birth.

[2]The post-birth break is mandatory whereas for the pre-birth period women can apply for an exemption. In the case of giving birth to twins the after-birth period is extended to twelve weeks.

[3]Parents can share paternity leave coverage and the cumulated leave periods can be at most three years. Thus, the maximum leave period for fathers is three years minus maternity protection after birth.

tax exemption for dependent children, which is more attractive for high income families. Furthermore, exactly at the same date of the reform discussed so far (January 2001), there was another political reform which facilitated part-time work – not only among parents. The latter reform established a legal claim for part-time work and regulated fixed term contracts ("Teilzeit- und Befristungsgesetz").

Working mothers need formal or informal childcare for their children. In 2007, around 14 percent of all children under age three attended formal or institutional childcare (Mühler, 2010). The coverage is 37 percent in East Germany and only 8 percent in West Germany. In West Germany, the availability of formal childcare for children under the age of three has been growing slowly over time starting from a low level. From age three on, every child in Germany is entitled to formal institutional childcare in a kindergarten. For this age group, the attendance rates amounts to 89 percent. However, the opening hours in existing childcare facilities are often not sufficient (Boll, 2009) and attendance is often on a part-time basis. The lack of suitable formal childcare is often viewed as an obstacle for higher employment levels among mothers of young children.

3 Econometric Approach

Our goal is to estimate the average treatment effect for the treated (ATT) on employment, with the treatment being 'first childbirth at a certain age'. We estimate the treatment parameter of 'childbirth now versus waiting' based on discrete time data. This approach has been suggested by Sianesi (2004, 2008) in the context of estimating the effects of active labor market programs. We assume a dynamic conditional independence assumption and apply inverse probability weighting (IPW) based on the estimated propensity score (Hirano et al., 2003). Following Busso et al. (2009), we implement the IPW estimator for the ATT by normalizing the weights of the members in the comparison group to sum up to one. Because our control group changes by age at first birth and there is attrition both in the control group over time, the normalization of weights changes along the two dimensions age at first birth and elapsed time since birth. We summarize the ATT estimates and analyze effect heterogeneity using weighted outcome regressions as suggested by Abadie and Imbens (2011) for matching estimators. The next subsection introduces and motivates the treatment parameter to be estimated (section 3.1). Then, section 3.2 describes our implementation of the IPW estimator. Finally, we introduce the weighted outcome regression based on IPW (sectin 3.3).

3.1 Dynamic Treatment Approach

The treatment effect that we estimate is the effect of treatment versus waiting as introduced by Sianesi (2004, 2008). Thus, for a certain age, we estimate the ATT of having a first child at this age versus not having a child at this age. The group of nontreated women consists of women with no child at this age and who may or may not have a first child in the future. Treated and nontreated women did not have a child before the considered age.

The treatment effect to be estimated can be viewed as an example for the timing-of-events approach as applied in the context of program evaluation of active labor market policies by Sianesi (2004, 2008) for Sweden or by Fitzenberger et al. (2008) for Germany. These countries have a comprehensive system of active labor market policies implying that with some probability unemployed individuals who have not been treated by a certain point of time may receive treatment later. If one constructs a control group based on individuals who are not treated during the observation window, one runs the risk of conditioning on future outcomes as for example individuals will not participate in treatment in the future if they have found a job before. Put differently, individuals who are eligible for treatment at a certain point of time but receive treatment later would be excluded from the control group. The exit from unemployment to employment and the entry into program participation are two competing risks, and unobserved characteristics are typically correlated with the chances of finding a job (quickly) as well as with program participation. Correspondingly, Sianesi (2004, 2008) and Fredriksson and Johansson (2008) argue that excluding future participants from the control group would lead to biased estimates of the treatment effect because of selection on unobservables.

Our application of the dynamic treatment approach is comparable to the training literature example insofar as the control group remains childless up to the considered point in time (similar to unemployed individuals remaining jobless and/or without training). This highlights the importance of avoiding conditioning on future outcomes. Therefore, in our application, for women giving birth for the first time at a certain age, we do not exclude the alternative of giving first birth at a later age. This corresponds to the dynamic nature of fertility decisions.[4] Using solely a control group of women who do not give first birth until a much later age or who will never have a child would bias the control group towards women with a low propensity of having a child. This is due to selection into motherhood.[5] This bias is likely to be correlated with labor market outcomes (e.g.

[4]As Ciliberto et al. (2010, p. 10) put it "[...] employees make their fertility choices many times in their lives [...]". Nevertheless, this study uses cross-sectional data.

[5]Some articles use non-mothers for comparison and this way include selection into motherhood in

women with a strong unobserved career orientation are more likely to exhibit a higher labor market attachment and lower fertility rates).[6] The endogeneity problem of fertility and employment also arises because there may be labor market shocks and shocks to personal circumstances (relationships, household characteristics) with differential effects on fertility and labor market outcomes. It is plausible that positive labor market shocks are negatively correlated with future fertility or that personal shocks, which reduce the fertility in the future, are also correlated with labor market outcomes.[7] For the purpose of our analysis, these effects are similar in nature to the competing risks approach prevalent in the timing-of-events analysis of active labor market policies. Therefore, the results would be biased if we excluded women who give first birth later (Wilde et al., 2010). This argument assumes that we cannot control unobserved heterogeneity of the treated women if we merely consider nontreated women who do not give birth in the future. Instead we follow Sianesi (2004, 2008) and assume 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 a certain point in time, i.e. at a certain age. 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 24 and 33 and who are not retired or disabled. The control group consists of women who did not have a child at the precise age of the treated mother at the time of birth.[8] Hence, we assign to each treated woman an individual-specific control group, thereby imposing an exact alignment of the age in months. We measure both the treatment effect and the age of the women at a monthly frequency.[9] Our analysis estimates

the estimated effect, as it is an important component (called 'indirect effect' by Simonsen and Skipper, 2006), see also Ejrnaes and Kunze (2011). However, the approach presented here abstracts from selection, considering only the 'direct effect' (Simonsen and Skipper, 2006).

[6]The underlying reason for this type of endogeneity is that both, fertility and career are controlled by women (at least in part, see Troske and Voicu, 2010). For different approaches to address endogeneity in this context see Miller (2011), Herr (2007), Gustafsson et al. (1996), and Wilde et al. (2010).

[7]An extreme example may be a work accident of a woman which prevents her both from working and from having a child in the future.

[8]We impose that the control women remain childless for at least three months after the respective age month of birth.

[9]However, for the purpose of presenting our results, we will later pool the month-specific treatment effects into groups (see section 5). For example, there should be little difference between a women who has her first child at 26 years and 11 months or at exactly 27 years.

an average counterfactual outcome for each treated women based on the individual-specific group of individuals not treated so far.

We assume that the rich set of covariates allows us to capture systematic differences in the propensity to have a first child at a certain age. Specifically, we assume that given the duration of childlessness and given the covariates, having a first birth within the next year is random, i.e. the dynamic conditional independence assumption holds (DCIA as discussed in Fitzenberger et al., 2008). This randomness in the exact month of birth is given by nature, as ”conceptions are not perfectly controllable events” (Hotz and Miller, 1988, p. 91). Furthermore, we assume that the timing within the year considered, i.e. the month of birth, is unrelated to the selection into first birth.

The DCIA can be motivated as follows for the age range considered. Our data allow to control for a large number of characteristics regarding education, labor market experience, relationships, and attitudes. One year before birth the treated women are not likely to differ systematically from those women who stay childless at least for 15 more months. 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 rather implausible that woman plan the exact month of firth birth more than a year earlier. At the same time, women differ in their probability to have a child within the next year and that is likely to be reflected by the characteristics controlled for. Thus, controlling for these variables makes women comparable in the probability to have a child. We argue that the DCIA is particularly plausible during the age range 24 to 33 years considered. At age 24, women have realized the major parts of their human capital investments and most of them have started their labor market career. In contrast, in their late 30s it is likely that first time mothers and childless women become less comparable. Because the fecundity starts to drop in the 30s and giving birth becomes more risky at a higher age, most women in their late 30s may either be determined to have a child or decide to stay childless (i.e. the probability to have a child is close to zero).[10] Thus, treated and non-treated females become less comparable at a higher age because many of them may act upon the anticipation as to whether and when they will have a child.

[10]Based on US data, Miller (2011) argues that the decline of fecundity with age is highly nonlinear, and is apparent primarily for pregnancies after age 33. Analogous to Miller (2011), we only consider first child births up to the mother's age of 33. For empirical descriptions of the fertility decline with age, see e.g. Stein, 1985, p. 328 and Imthurna et al., 2008, p. 126.

3.2 Inverse Probability Weighting (IPW)

To control the selection of treated women, we estimate a sequence of quarterly propensity scores.[11] The probability of having the first child within the next year is modeled as a function of human capital and employment history, which is particularly important if decisions on having a child and decisions on the labor market career are taken jointly. As sequential labor market decisions are correlated (Troske and Voicu, 2010), controlling for past labor market career is crucial for successful matching. Moreover, we control for status of the relationship with the partner and for the self reported importance of having a family.[12] We include covariates of the partner, such as income and education, tries to proxy the partner's role in the joint decision process.

Under the unconfoundedness of the treatment and perfect overlap in the propensity score, Busso et al. (2009) conclude that in small samples with unknown propensity score, a modified inverse probability weighting estimator (IPW) performs best in comparison to various matching estimators. This result stands in contrast to the conclusions obtained by the Monte Carlo study in Frölich (2004). The crucial modification of the IPW estimator involves the normalization of weights for the nontreated women. According to the results of Busso et al. (2009), the poor performance of IPW reported by Frölich's (2004) Monte Carlo study is due to the fact that Frölich does not normalize the IPW weights. Doing so strongly improves the performance of the estimator as found by Busso et al. (2009), who suggest to estimate the ATT as follows (Busso et al., 2009, eq. (7)):

$$(1) \quad \hat{\theta}_{BDM} = \frac{\sum_{i=1}^{n} T_i Y_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{j=1}^{n} (1 - T_j) \hat{W}_j Y_j}{\sum_{j=1}^{n} (1 - T_j) \hat{W}_j}$$

$$\text{with weights} \quad \hat{W}_j = \hat{p}(X_j)/(1 - \hat{p}(X_j)) \ .$$

Furthermore, T_i, T_j denote the treatment dummy variables for individuals i, j (treated and non-treated), respectively, and $\hat{p}(X_j)$ denotes the estimated propensity score as a function of covariates X_j. The application of the weights W_j leads to a reweighting of the nontreated women according to the odds–ratio of having a child within the next year. Note that the denominator corresponds to the sum of the weights in the numerator.

For our application, we have to account for the fact that for the estimation of treatment versus waiting the group of eligible comparison women changes by month of age.

[11] The size of the treatment sample is not sufficient to go down to a monthly frequency when estimating the propensity scores. Furthermore, we think that the selection into childbirth does not change strongly from month to month.

[12] In fact, in contrast to matching, the reweighting estimator we will use later requires the propensity score to be a conditional probability (Busso et al., 2009), which is evident in our application. As there is sufficient overlap, we do not require any trimming.

Correspondingly, the alignment between treated and nontreated observations changes as well by month of age. The higher the age, the smaller the 'eligible' control group for first birth. Thus, we modify the estimator in equation (1) to

$$(2) \quad \hat{\theta} = \frac{\sum_{i=1}^{n} T_i \left\{ Y_i - \frac{\sum_{j=1}^{n}(1-T_j)\hat{W}_{i,j}Y_{j,age(i)}}{\sum_{j=1}^{n}(1-T_j)\hat{W}_{i,j}} \right\}}{\sum_{i=1}^{n} T_i}$$

$$\text{with weights} \quad \hat{W}_{i,j} = E_{i,j}\hat{p}(X_j)/(1-\hat{p}(X_j)) \ .$$

$E_{i,j}$ is a dummy variable which take the value of one if woman j can be used as a control observation for treated woman i, i.e. if j remains childless for at least three months after the respective age at first birth for the treated woman i. $E_{i,j}$ is set to zero if woman j has a child earlier than the three months after birth for woman i. $Y_{j,age(i)}$ is the outcome of control woman j aligned to the age at first birth for treated woman i.

This IPW or reweighting estimator has the advantage of not relying on a tuning parameter such as a bandwidth in kernel matching or the number of nearest neighbors in nearest neighbor matching. Moreover, it is easy to implement and standard errors are readily obtained by bootstrapping. Busso et al. (2009) show that this estimator is preferred in terms of bias and variance in finite sample settings with unknown propensity score. However, they stress that this only holds with good overlap and when misspecification of the propensity score is not a concern.

3.3 Outcome regression based on IPW

In order to account for time trends and to assess effect heterogeneity by observable characteristics (e.g. by age, education) including the effect of the 2001 reform, we follow Abadie and Imbens (2011) who suggest to estimate ex post outcome regressions after matching. We estimate weighted linear regressions of the individual outcomes on an intercept, a treatment dummy as well as further covariates and the interactions of their demeaned values with the treatment dummy (Wooldridge, 2002, p. 612). In addition, we use year dummies as main effects to account for a time trend in female employment rates. Because of sample size issues, we do not align treated and controls by calendar year. Thus, all estimated treatment effects reported in the following - including the baseline estimates without further controls - are based on outcome regressions controlling for year dummies.

We adjust the approach by Abadie and Imbens to the IPW case employing a constant weight of $1/\sum_{j=1}^{n}(T_j)$ for a treated woman i and a weight of $(1-T_j)\hat{W}_{i,j}/\sum_{j=1}^{n}(1-T_j)\hat{W}_{i,j}$ for a nontreated woman j with outcome $Y_{j,age(i)}$ aligned in age for $E_{i,j} = 1$ to treated woman i.

Specifically, we run the following regression of the outcome Y after the beginning of treatment

$$(3) \quad Y_{j,age} = \alpha + x_j\beta + \gamma T_j + T_j(x_j - \bar{x})\delta + u_j \,,$$

where x_j denote the observable characteristics considered, i.e. educational dummies, and $\bar{x}$ corresponds to the sample average among the treated. The outcome measurement $Y_{i,age}$ is recorded once for a treated woman i and carries a weight of 1. The outcome measurement $Y_{j,age(i)}$ for a nontreated woman j is recorded as many times as she can be used as a control observation for a treated woman i (each time aligned in age to i). Each single measurement carries the weight $(1-T_j)\hat{W}_{i,j}/\sum_{j=1}^{n}(1-T_j)\hat{W}_{i,j}$ and the weights of eligible control observations sum up to one for each treated woman i. These regression weights mimic the weights used in equation (1). Note that in a weighted regression (3) without the covariates involving x_j the estimate for γ would reproduce $\hat{\theta}_{BDM}$ as in equation (2).

In a weighted regression with the covariates involving x_j, the estimate for γ corresponds to the ATT estimate corrected for the mismatch in observable characteristics. The coefficients β control for the impact of the characteristics on the average outcome variable. If $\delta = 0$ in the regression (3), then there is no linear effect heterogeneity by the level of covariates. The standard errors of the estimated regression coefficients are obtained from the bootstrap procedure for the IPW estimator by rerunning regression (3) for all resamples.

4 Data

The analysis is based on data from the German Socio-Economic Panel (SOEP), a yearly household survey interviewing all persons above age 15. The survey collects data on employment related questions such as working hours, income and many more. It also asks for many other topics such as leisure, health, satisfaction and values. Moreover, it contains information about the monthly employment status and income sources, allowing us to analyze our research question at the monthly frequency. The SOEP has been used in a number of studies on employment of mothers and the effects of maternity leave, see e.g. Bergemann and Riphahn (2011a), Sommerfeld (2009), Vogel (2009), Wrohlich (2004), or Kreyenfeld and Hank (2000).

We use data from 1991 until 2008, because 1991 is the first year households from East Germany were included in the panel. Pensioners, disabled persons, and those women between 24 and 33 years of age, as younger women are often still in education. In 2000, the median age at first birth was 29 years in West Germany (Pötzsch, 2005). Older women pose a problem if the selection of the control group gets too strong, because the share of women in the control group who will have a child later decreases with age. We chose age 33 as upper limit for this study, because fertility is rather stable until the age of 34 and declines strongly from age 35 onwards (Pötzsch, 2005, Stein, 1985, p. 328, Imthurna et al., 2008, p. 126). Importantly, women who already have a child are excluded from the data.

A childless woman can potentially give birth to her first child in any particular month. We therefore construct a different control group for every month, consisting of women without children so far. The definition of non-treatment is crucial and involves the classification window which defines treatment and control group. On the one hand, women of the control group should not have their first child in the near future to be able to sharply distinguish between treatment and control group. On the other hand, the control group should be well comparable to the treatment group such that the probability for having a child is similar. As a compromise, we chose a window of three months during which control group individuals must not have their first child.

We have to address attrition in the data, which the normalization of the IPW-weights must take account of. The size of the treatment and control group is depicted in Figure 1. For the estimation of the treatment effect it is essential to match highly comparable women. This comparability is achieved by the exact alignment of age (in months) as has been described earlier, and by a reweighting procedure which is based on the propensity to have the first child at a certain age. The required pre-birth information is collected at least 12 months before the birth to avoid the anticipation of the exact timing of first birth. However, in order to use relevant information for the time of birth, the interview should not have been too far in the past; therefore we choose a maximum of 15 months before birth based on a window width of three months. Reweighting on the propensity score is required to assure that the dynamic conditional independence assumption (DCIA) holds. To capture all relevant factors which jointly influence fertility and labor supply, we include human capital variables (degree of schooling and training education), the pre-birth labor market history (working hours, wage, job type, full time, part time and unemployment experience) as well as informations on relationships (partnership, marriage). In addition, we include the self-reported importance of family for one's satisfaction as an attitudinal variable affecting the fertility decision.

Employment after first birth is our outcome variable of interest. We do not distinguish between full time or part time employment as part time work may be a stepping stone towards higher working hours (Vogel, 2009; Ziefle, 2004). Full time employment is very low in Germany after childbirth (see e.g. Geyer and Steiner, 2007; Sommerfeld, 2009).

Furthermore, there is no distinction between household work (being out-of-labor force) and job search (being unemployed) as the data do not allow us to distinguish between these states in a reliable way. We abstract from reductions in hours of work around the time of childbirth (Buligescu et al., 2009). We follow our outcome variable employment for 5 years, starting from the month in which the treated woman gives birth to her first child. Extending the analysis to a time period beyond five years severely reduces the sample size due to sample attrition (results are available upon request).

5 Empirical Results

We start with descriptive evidence. Then, we discuss the estimated treatment effects as well as the reform effects, and we perform a sensitivity analysis.

5.1 Descriptive Statistics before and after IPW-matching

Table 2 shows descriptive statistics at the time of matching before ('unweighted') and after IPW-matching ('weighted'). The probit estimates for the propensity score are displayed in Table 3. Table 2 shows that before matching there are some sizeable differences in observables between the treatment and control group (e.g. treated women are more often employed and live more often in a partnership). Hardly, any significant differences remain after matching. Thus, IPW seems to balance the observable characteristics very well.

Figures 2 to 4 show that some differences between treatment and control group exist before matching (i.e. without IPW-weighting) before birth with respect to the employment rate, educational degrees, and partnership, particularly at younger ages. The highest obtained educational degrees also differ somewhat for younger women, as some of them are still in education.[13] These figures also show that the weighting procedure works well to improve comparability between treatment and control group at the point in time before treatment starts. However, the employment rate of the treatment group before birth still remains slightly higher, i.e. young first-time mothers have a higher labor force attachment before birth.

Figure 5 shows employment rates before and after birth with and without reweighting of the control group. Recall that matching takes place about one year before treatment. It can be seen again that the reweighting procedure works well for the alignment of the employment rate before birth, even though the treatment group still exhibits slightly higher employment rates. At birth, the employment rate of the treated observations

[13] Recall that for the treatment effect at age 24 the training degree is measured at age 23 - an age where some educational programs are not completed, yet (particularly university).

drops to zero due to mandatory maternity leave of at least eight weeks after the birth. Afterwards, the employment rate of first-time mothers recovers very slowly, while that of women in the control group decreases slightly. This is due to the fact that these women may also drop out of employment, e.g. due to childbirth.

To shed more light on the comparability, Figure 6 shows the rates at which women in the control group have their first birth and at which treated women have their second birth. It can be seen that after five years, nearly 50% of the control observations have had their first birth. Among the treated, nearly 60% have had a second birth by this time.

5.2 Average Treatment Effect on the Treated (ATT)

Figure 7 depicts the monthly ATTs on employment outcomes during the first five years after first birth. The estimates are pooled across age groups and the confidence intervals are based on 100 bootstrap replications. Recall that the reweighted estimates control for year dummies , see section 3.3. First, the estimated ATT is negative and persistent, implying that within the observation window of five years the first-time mothers never fully catch up to the control group. This holds even though part time employment is also covered by our outcome variable and women in the control group may also drop out of employment, e.g. because of childbirth. The development of the ATT over time clearly shows a strong upward trend, However, the upward trend is interrupted by short periods of decline which are due to mothers who (re)enter into the labor market after childbirth, but then drop out again (Figure 5).

At time zero, which corresponds to the month of birth, the absolute values of the ATT are highest because all mothers have to reduce their labor supply to zero by law. During the first year after childbirth, the ATT increases strongly from -81 percentage points (ppoints) to around -50 ppoints. In the second year the ATT continues to catch up, reaching -30 ppoints. Three years after birth we observe an upward jump in the ATT (to -20 ppoints) which is due to the institutional changes that occur at this point in time (recall the job protection period and childcare availability). Still, the jump after three years is small. Afterwards, the negative employment effect remains fairly constant reaching around -20 ppoints after five years. Our key finding is the large and persistent negative employment loss after first childbirth which does not level off completely within five years.

Figure 7 also shows employment effects before childbirth. Note that we align treated and nontreated women 12 months before birth. Accordingly there is no significant employment effect during the second year before birth, which shows that, after matching, treated and nontreated women are well matched regarding their employment history.

Right before birth, there is a small negative employment effect, which we denote as a lock-in-effect caused by the pregnancy – in analogy to the training literature. The future mothers start to reduce their labor supply. Note that our matching approach does not allow matched women in the control group to have their first child until three months after childbirth by the respective treated mother. Therefore, the control group is locked into non-treatment during this period, which is associated with a higher employment rate compared to women who have a child soon after the treated women.

So far, we have not investigated effect heterogeneity. Presumably, the timing of first birth depends upon work experience including career progression in the past. Education affects the timing of labor market entry, career progression, and the timing of birth. In Germany, young employees with a vocational training degree generally enter the labor market in their early twenties at the latest. Students usually complete their university degree between 25 and 30 years of age (Hochschulrektorenkonferenz (HRK), 2010, p. 32).

Women holding a university degree have spent more time in education and thus have higher opportunity costs for employment breaks. They lose most from an employment interruption due to high wages and larger depreciation of human capital (Miller, 2011; Troske and Voicu, 2010; Wilde et al., 2010). When highly educated women anticipate large wage losses (as found by Wilde et al., 2010; Ejrnaes and Kunze, 2011), it is rational for them to return to the labor market more quickly than others. Troske and Voicu (2010) find that highly educated mothers in the US display a higher than average labor force participation after first childbirth, but they work less hours. For these reasons, we expect a smaller reduction and a stronger recovery in employment rates after birth for university graduate compared to other educational groups, a prediction which is confirmed by Figures 8 and 9. At the child's age of 3.2 years, the employment effect for university graduates is not significantly different from zero anymore. For the two other educational groups, the employment effect remains significantly negative over the course of the five years after child birth. The average difference between medium and high educated women is about 16 ppoints. For the group of women without a completed degree, the effect becomes somewhat erratic and the confidence interval become very large due to the small sample size. Altogether, first-time mothers with a high education level and some work experience exhibit a higher attachment to the labor market compared to other groups. However, the differences are not significantly different from one another (Figure 9).

Table 4 reports cumulative employment losses in person-months. During the first five years after birth, an average of about two years (23 months or 38.5%) are 'lost' in employment due to having a child now instead of waiting. This effect is largest for women with a training degree (27 months or 44.3%) and considerably below the average for university graduates (17 months or 27.7%).

Figure 10 plots the ATT employment effects by two-year age groups, averaging treatment-control differences among first-time mothers in the age group. The estimated ATTs are always negative and show similar patterns as before averaging over the age range 24 to 33 years. Note that this similarity holds although older women work show higher employment rates than younger ones in both the treatment *and* the control group - these differences simply cancel out in the estimation of the ATTs. Two years after childbirth, the employment effects for the different age groups start to diverge with the oldest age group showing the strongest negative employment effects and the youngest age group showing ATTs getting closest to zero five years after birth. This could be due to the fact that the average age at first childbirth is about 28 years, which probably affects the behavior of the control group three to four years after treatment start for the treatment group at age 24 and 25. The differences between the groups of 26–27, 28–29, and 30–31 year-olds are negligible. At the same time, the absolute value of the ATT is very large for the oldest considered group.

Among the results obtained so far, the large negative employment effects for the group of 32–33 year-olds stands out. This contradicts predictions derived from human capital theory saying that the ATT should be lower for older first-time mothers as they have already acquired a high level of human capital through work experience. This group displays a very high labor force attachment before birth and is of great relevance to policy-makers. In fact, the 2001 and the 2007 reforms attempt to increase employment rates after childbirth for previously employed women.

Furthermore, our results are not in accordance with the hypothesis that delayed childbirth leads to higher employment, which some of the literature finds evidence for (Troske and Voicu, 2011; Herr, 2007; Miller, 2011). However, one has to recall at this point that we estimate a different parameter than the literature. Troske and Voicu (2011) estimate the effect of having a child versus not having a child where not having a child does not entail the possibility to have a child in the future. Given that a larger share of women in the control group have a child in the near future for births and that this share is falling with age (Figure 11), we would expect a disproportionately higher employment rate in our control group at higher age compared to Troske and Voicu (2011) rationalizing the stronger negative employment effect at older age in our case. In contrast, Herr (2007) and Miller (2011) provide IV-estimate which only contrast mothers at different age when the child was born, i.e. in their approach the control group effectively used does not involve women who do not have a child. Based on the same argument as above, one would also expect that our results show a stronger negative employment effect at higher age of birth compared to the studies by Herr (2007) and Miller (2011).

Also the characteristics of the treated individuals change with age at birth. Strictly speaking, when employment effects vary with observable and unobservable characteristics, it is not possible to compare the ATT estimates obtained for different age groups because the two treatment groups are not balanced in their characteristics. It would be possible to reweight the two to make them comparable regarding observables. However, it is likely that the two groups also differ in unobservables which would require further assumptions and a different estimation approach to balance. Thus, the difference between our results and the literature may also be caused by differences in the composition of mothers by age at childbirth.

To understand our results, it proves also important to consider second births in the treatment group. One would expect that older or more productive groups space their births more closely (Ejrnaes and Kunze, 2011; Troske and Voicu, 2010). This holds in our data for the age groups up to the 30–31-year old first-time mothers, but not for the groups of the 32–33 years old (Figure 11). That is up to an age of 31 years at firth birth, there are more second births during the five years after birth which suggests that the estimated employment loss increases by age. However, for the 32–33 years old this does not apply and it is just that the control group has a higher employment rate (not displayed). Thus, the reason for the larger absolute value of the ATT of the oldest studied age group is the high participation rate of the control group. This by itself is in accordance with human capital theory whereas the behavior of the mothers at higher age seems to be less in accordance with human capital theory.

5.3 Effects of the 2001 Reform

We now use our ATT estimates to evaluate the effects of the policy changes in 2001 which are directed at increasing the employment of mothers after birth. As described in section 2, starting in 2001 parents were given two options: to work up to 30 instead of 19 hours per week while being in parental leave to claim increased parental leave benefits in the first year they do not apply for during the second year after birth. Furthermore, there was a change to entitle full-time workers to change to part-time work at their jobs, a change which was not restricted to parents of newborn (the title of the new law was *"Teilzeit- und Befristungsgesetz"*). Both legal changes were introduced at the same time and we will estimate their joint effect on the ATT of first birth on employment from 2001 onwards.

The reform effect is estimated by means of regression (3) where we interact dummy variables for year since birth with a dummy variable for the time period starting in 2001. Figure 12 depicts the ATTs before 2001 as baseline estimates and the estimated coefficients of the interaction variables, where the latter estimate the reform effect by year since birth. The reform effects are all significantly positive (also jointly). The point estimates grow from 8.3 ppoints when the child is one year old to 11.2 ppoints at age four.

Note that the reform effects are estimated based both on births before 2001 (as long as the first five years after birth overlap with the postreform period) and births in 2001 or afterwards. To disentangle the effects, Figure 13 displays two separate effects which should be interpreted as a decomposition of the overall reform effect shown in Figure 12. For the second group of children (born between 1996 and 2000), the policy effect is shown as a function of the age of the child when the policy became effective. The results suggest positive employment effects induced by the reform in the order of 6 percentage points. Although year specific effects are mostly not significant, there is joint significance over the five years due to the uniformity of the effects. Also the effect is significantly positive for children born after 2001 when they are one year old. Altogether, the 2001 reform had a lasting significantly positive effect on employment rates of mothers with a first child up to age 4.

5.4 Sensitivity Analysis

In order to check the sensitivity and robustness of our results and to analyze what drives our results, we now discuss some alternative estimates for comparison.

First, Figure 14 contrasts simple OLS estimates with our baseline ATT estimates dicussed above in section 5.2. The OLS results depict the monthly treatment effects from a pooled OLS regression together with the confidence interval based on standard errors which are clustered at the individual level.[14] The OLS estimates lie significantly above the ATT estimates obtained from the dynamic treatment approach. OLS would therefore understate the employment loss. The OLS estimates do not exclude soon-to-be mothers in the first months after birth and they do not account for the drop in employment rates during the last six months before birth. This rationalizes why the OLS estimates exceed the IPW-estimates. Furthermore, the OLS estimates contrast the employment of mothers only with childless women, thus excluding a comparison with women who have a child in the future after the birth of that child. We note that simple OLS estimation does not clearly define a dynamic treatment effect and it yields results which differ from the IPW results and which are difficult to interpret.

[14]For this regression, all women between the age of 24 and 38 who do not have a child above the age of five years have been included. The coefficient estimates on monthly dummy variables for month 1 to 60 after childbirth are depicted. The regression controls for calendar years and for the same set of covariates used above for the estimation of the propensity score.

Second, we implement the dynamic approach from section 3.1 above, but without inverse probability weighting (section 3.2), i.e. we only align mothers and not-yet-mothers by age at childbirth and then contrast mean employment rates by age of the child, see Figure 14. Despite small changes, the qualitative results are unchanged. Thus, selection on observables only plays a negligible role.

Third, we vary the control group first by using only not-yet-mothers (conceptually similar to Herr, 2007; Miller, 2011; Wilde et al., 2010) and second by using only childless women (conceptually similar to Troske and Voicu 2010, 2011). Note that both control groups condition on future outcomes (Fredriksson and Johansson, 2008). Specifically, we first use a control group with only women who give first birth between 3 and 24 months after the childbirth by the respective treated woman. Second, we use a control group with only women who do not give birth to a child within 5 years after the childbirth by the respective treated woman. The results are presented in Figure 15. In the first case, a lock-in effect of 12 months shortly after birth is observed. During the second year there is a slightly positive effect of up to 11 ppoints. Afterwards, no significant effect can be observed. This means that the treatment and the control group exhibit similar employment rates in the long-run. In the second case, the pattern looks similar to the overall ATT but the negative effect remains above 40 ppoints until the end of the observation period. Here, women who never have children probably drive the result. We conclude that despite the high incidence of a second childbirth among the treated women, the employment pattern of women who do not have a child within the next five years drive the results.

Fourth, we use a different selection of the treatment group to check the sensitivity of results. Figure 16 excludes all women from the treatment group who have a second child during the observation period of five years. We find that the employment loss gradually declines and becomes zero after five years. Although this definition of the treatment group conditions on future outcomes, the results suggest that the remaining significant employment loss after five years reported in section 5.2 is driven by the fact that many mothers have a second child during the first five years after first childbirth.

Last but not least, we investigate as to whether are sensitive to panel attrition. Figure 17 limits the sample to those women who are still observed in the data 5 years after first childbirth by the respective treated woman. Constrasting Figures 7 and 17, it is quite reassuring that the pattern of the results does not change in a substantive way.

6 Conclusions

In a critical phase of a woman's life, between age 20 and up to about age 40, crucial career and fertility decisions have to be taken jointly. The impact of childbirth on career outcomes varies greatly with human capital and, thus, with age. The literature suggests that delaying the first childbirth to a later age could pay off positively in terms of career progression. This paper studies the question what is the effect on subsequent employment of having a first child at a certain age against the alternative to delay having a child at that age.

To address this question, we start from the individual decision whether to have the first child now or wait. This is modeled in a dynamic treatment effects framework in analogy to Sianesi (2004, 2008) and Fitzenberger et al. (2008). We then reweight these dynamically constructed control groups by inverse probability weighting (IPW), based on the finding that this procedure has a better performance with respect to bias and variance if the weights are normalized to sum up to one, as found by Busso et al. (2009). Finally, the ex-post outcome regressions proposed by Abadie and Imbens (2011) allow us to explicitly consider effect heterogeneity for different educational groups.

Our results involve large and persistent negative causal effects of first childbirth on subsequent employment. Although the employment loss is reduced during the first five years following childbirth, the treatment effects do not reach at the end of that period, i.e. the employment rates of the treatment group do not catch up to the control group. In other words - and in accordance with the bulk of the literature: having a child is associated with a sizeable employment loss. The strong negative effects for the oldest considered age group of 32–33 year-olds stand out as they contradict expectations from human capital theory according to which these women face the highest opportunity costs of leaving the labor market. Furthermore, our findings indicate that university graduates face significantly smaller employment losses after childbirth compared to other skill groups, a finding which is in accordance with human capital theory. Moreover, we estimate the effects of political reforms which facilitated part-time work and increased incentives for an earlier return-to-job from 2001 onwards. These reforms indeed reduced significantly the negative employment effects after first childbirth.

Finally, we implement a number of sensitivity checks. These suggest that the qualitative pattern of our results are driven by the employment pattern of those women in the control group, who do not have a child within the first five year after first childbirth by the treated women, and the still negative employment effect after five years is driven by those treated women who have a second child during the first five years after having the first one.

References

Abadie, A. and Imbens, G. W. (2011). Bias Corrected Matching Estimators for Average Treatment Effects. *Journal of Business & Economic Statistics*, 29(1):1–11. vormales Abadie2007 als Mimeo.

Abbring, J. and van den Berg, G. (2003). The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica*, 71:1491–1517.

Baker, M. and Milligan, K. (2008). How does job-protected maternity leave affect mothers' employment? *Journal of Labor Economics*, 26:655–691.

Beblo, M. and Wolf, E. (2002). How Much Does a Year Off Cost? Estimating the Wage Effects of Employment Breaks and Part-Time Periods. *Brussels Economic Review – Cahiers Econonomiques de Bruxelles*, 45:191–217.

Bergemann, A. and Riphahn, R. T. (2011a). Female Labor Supply and Parental Leave Benefits - The Causal Effect of Paying Higher Transfers for a Shorter Period of Time. *Applied Economics Letters*, 18:17–20.

Bergemann, A. and Riphahn, R. T. (2011b). The Introduction of a Short-Term Earnings-Related Parental Leave Benefit System and Differential Effects on Employment Intentions. *Schmollers Jahrbuch*, 131:315–325.

Boll, C. (2009). Lohneinbußen durch Erwerbsunterbrechungen - fertilitätstheoretische Fundierung, Quantifizierung auf Basis von SOEP-Daten und familienpolitische Implikationen. *SOEPpapers on Multidisciplinary Panel Data Research*, 160.

Buligescu, B., de Crombrugghe, D., Menteşoğlu, G., and Montizaan, R. (2009). Panel estimates of the wage penalty for maternal leave. *Oxford Economic Papers*, 61:i35–i55.

Busso, M., DiNardo, J., and McCrary, J. (2009). New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators. *IZA Discussion Paper*, 3998.

Ciliberto, F., Miller, A. R., Nielsen, H. S., and Simonsen, M. (2010). Playing the Fertility Game at Work. *Mimeograph*.

Dustmann, C. and Schönberg, U. (2011). The Effect of Expansions in Maternity Leave Coverage on Children's Long-Term Outcomes. *American Economic Journal: Applied Economics*, (forthcoming).

Ejrnaes, M. and Kunze, A. (2011). Work and wage dynamics around childbirth. *IZA Discussion Paper*, 6066.

Fitzenberger, B., Osikominu, A., and Völter, R. (2008). Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany. *Annales d'Economie et de Statistique*, 91-92.

Fortin, N. (2009). Gender role attitudes and women's labor market participation: Opting-out, aids, and the persistent appeal of housewifery. *Unpublished Manuscript, University of British Columbia, Vancouver.*

Fredriksson, P. and Johansson, P. (2008). Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data. *Journal of Business and Economic Statistics*, 26:435–445.

Frölich, M. (2004). Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators. *The Review of Economics and Statistics*, 86(1):77–90.

Gerlach, I., Schneider, H., and Juncke, D. (2009). Elternzeit und -geld als familienpolitische Instrumente: Entwicklung, Zielsetzung und empirische Befunde aus deutschen Utnernehmen. *Sozialer Fortschritt*, 58(12):273–282.

Geyer, J. and Steiner, V. (2007). Short-Run and Long-Term Effects of Childbirth on Mother's Employment and Working Hours Across Institutional Regimes: An Empirical Analysis Based on the European Community Household Panel. *IZA Discussion Paper*, 2693.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American Economic Review, Papers and Proceedings*, 96:1–21.

Gustafsson, S. S., Wetzels, C. M., Vlasblom, J. D., and Dex, S. (1996). Women's labor force transitions in connection with childbirth: A panel data comparison between Germany, Sweden and Great Britain. *Journal of Population Economics*, 9:223–246.

Herr, J. L. (2007). Does it Pay to Delay? Understanding the Effect of First Birth Timing on Women's Wage Growth. *University of California - Berkeley Working Paper.*

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica*, 71(4):1161–1189.

Hochschulrektorenkonferenz (HRK) (2010). Statistiken zur Hochschulpolitik 2/2010.

Hotz, J. V. and Miller, R. A. (1988). An empirical analysis of life cycle fertility and female labor supply. *Econometrica*, 56(1):91–118.

Imthurna, B., Maurer-Majorb, E., and Stillera, R. (2008). Sterilität/ Infertilität – Ursachen und Abklärung. *Swiss Medical Forum*, 8(7):124–130.

Kluve, J. and Tamm, M. (2009). Now daddy's changing diapers and mommy's making her career: Evaluating a generous parental leave regulation using a natural experiment. *Ruhr Economic Papers*, 145.

Kreyenfeld, M. (2001). *Employment and Fertility – East Germany in the 1990s.* PhD thesis, Max-Planck Institute for Demographic Research, Rostock.

Kreyenfeld, M. and Hank, K. (2000). Does the availability of child care influence the employment of mothers? Findings from Western Germany. *Population Research and Policy Review*, 19:317–337.

Lalive, R. and Zweimüller, J. (2009). How Does Parental Leave Affect Fertility and Return-to-Work? Evidence from Two Natural Experiments. *Quarterly Journal of Economics*, 124(3):1363–1402.

Lefebvre, P., Merrigan, P., and Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a Canadian natural experiment on low-fee universal child care. *Labour Economics*, 16:490–502.

Mühler, G. (2010). Consequences of Mixed Provision of Child Care - An Overview on the German Market. *ZEW Discussion Paper*, 08-077 (2nd version).

Miller, A. R. (2011). The Effects of Motherhood Timing on Career Path. *Journal of Population Economics*, 24(3):1071–1100.

OECD, editor (2002). *OECD Employment Outlook, Chapter 2, Women at work: who are they and how are they faring?* Organisation for Economic Co-operation and Development, Paris.

OECD, editor (2007). *Babies and Bosses. Reconciling Work and Family Life. A Synthesis of Findings for OECD Countries*. Organisation for Economic Co-operation and Development, Paris.

Ondrich, J., Spieß, C. K., and Yang, Q. (1996). Barefoot and in a German kitchen: Federal parental leave and benefit policy and the return to work after childbirth in Germany. *Journal of Population Economics*, 9:247–266.

Ondrich, J., Spieß, C. K., Yang, Q., and Wagner, G. (1999). Full time or part time? German parental leave policy and the return to work after childbirth in Germany. *Research in Labor Economics*, 18:41–74.

Ondrich, J., Spieß, C. K., Yang, Q., and Wagner, G. G. (2003). The liberalization of maternity leave policy and the return to work after childbirth in Germany. *Review of Economics of the Household*, 1(1-2):77–110.

Pötzsch, O. (2005). Unterschiedliche Facetten der Geburtenentwicklung in Deutschland. *Wirtschaft und Statistik*, 6:569–581.

Schönberg, U. and Ludsteck, J. (2007). Maternity Leave Legislation, Female Labor Supply, and the Family Wage Gap. *IZA Discussion Paper*, 2699.

Sianesi, B. (2004). An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *The Review of Economics and Statistics*, 86(1):133–155.

Sianesi, B. (2008). Differential Effects of Swedish Active Labor Market Programs for the Unemployed. *Labour Economics*, 15(3):370–399.

Simonsen, M. and Skipper, L. (2006). The costs of motherhood: An analysis using matching estimators. *Journal of Applied Econometrics*, 21:919–934.

Sommerfeld, K. (2009). Older Babies - More Active Mothers? How Maternal Labor Supply Changes as the Child Grows. *Schmollers Jahrbuch: Journal of Applied Social Science Studies*, 129(2):227–240.

Stein, Z. A. (1985). A woman's age: Childbearing and child rearing. *American Journal of Epidemiology*, 121(3):327–342.

Troske, K. R. and Voicu, A. (2010). Joint estimation of sequential labor force participation and fertility decisions using markov chain monte carlo techniques. *Labour Economics*, 17(1):150 – 169.

Troske, K. R. and Voicu, A. (2011). A Panel Data Analysis of Racial/Ethnic Differences in Married Women's Labor Supply. *IZA Discussion Paper*, 5729.

van Ham, M. and Büchel, F. (2004). Females' Willingness to Work and the Discouragement Effect of a Poor Local Childcare Provision. *Applied Economics Quarterly*, 4:363–377.

Vogel, C. (2009). Teilzeitbeschäftigung - Ausmaß und Bestimmungsgründe der Erwerbsübergänge von Frauen. *Zeitschrift für ArbeitsmarktForschung - Journal for Labor Market Research (ZAF)*, 42(2):170–181.

Wilde, E. T., Batchelder, L., and Ellwood, D. T. (2010). The Mommy Track Divides: The Impact of Childbearing on Wages of Women of Differing Skill Levels. *NBER Working Paper*, 16582.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Wrohlich, K. (2004). Child care costs and mother's labor supply: An empirical analysis for Germany. *DIW*, 412.

Ziefle, A. (2004). Die individuellen Kosten des Erziehungsurlaubs: Eine empirische Analyse der kurz- und längerfristigen Folgen für den Karriereverlauf von Frauen. Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Appendix

Table 1: Child allowance by birth order in Euro

Year	1st child	2nd child	3rd child	Higher parity
1992 - 1995	35.79	66.47	112.48	122.71
1996	102.26	102.26	153.39	178.95
1997 - 1998	112.48	112.48	153.39	178.95
1999	127.82	127.82	153.39	178.95
2000 - 2001	138.05	138.05	153.39	178.95
2002 - 2008	154.00	154.00	154.00	179.00

Table 2: **Characteristics at time of matching, i.e. 1 year before birth**

	Treatment group		Control group			
			Unweighted		IPW-weight	
	Value	[CI]	Value	[CI]	Value	[C
Outcome-variable: Employed	82.4%	[79.0 ; 85.8]	73.2%	[72.0 ; 74.4]	80.7%	[78.8 ;
Wage	9.98	[9.59 ; 10.38]	9.32	[9.06 ; 9.57]	9.51	[9.14 ;
Working hours	35.3	[34.0 ; 36.6]	31.7	[31.2 ; 32.2]	34.8	[33.8 ;
Part time	7.8%	[5.2 ; 10.5]	9.9%	[9.1 ; 10.6]	7.0%	[5.2 ; 8
Job status:						
Not employed	7.2%	[5.4 ; 9.1]	6.3 %	[5.6 ; 7.0]	6.7%	[5.4 ; 8
In Training	5.9%	[3.7 ; 8.0]	15.6%	[14.5 ; 16.6]	8.5%	[7.0 ; 9
Blue collar	11.9%	[9.0 ; 14.7]	9.7%	[8.8 ; 10.6]	12.5%	[10.0 ;
Self-employed	2.6%	[1.2 ; 3.9]	2.8%	[2.1 ; 3.4]	2.4%	[1.3 ; 3
White collar	64.9%	[60.4 ; 69.4]	60.2%	[58.5 ; 61.8]	63.1%	[60.4 ;
Civil servant	7.6%	[5.1 ; 10.1]	5.5%	[4.7 ; 6.2]	6.8%	[5.2 ; 8
Work experience (in years) in...						
... Full time	5.16	[4.87 ; 5.45]	4.18	[4.05 ; 4.32]	4.92	[4.71 ;
... Part time	0.68	[0.54 ; 0.83]	0.83	[0.76 ; 0.90]	0.64	[0.56 ;
... Unemployment	0.29	[0.23 ; 0.35]	0.33	[0.29 ; 0.37]	0.28	[0.24 ;
Education:						
University degree	16.4%	[13.2 ; 19.6]	15.8%	[14.5 ; 17.1]	14.5%	[12.2 ;
Training degree	66.1%	[62.4 ; 69.9]	58.5%	[56.6 ; 60.4]	65.3%	[61.9 ;
No degree	10.6%	[7.9 ; 13.4]	9.0%	[7.8 ; 10.2]	12.2%	[10.0 ;
Currently in education	6.8%	[4.6 ; 9.1]	16.7%	[15.6 ; 17.9]	7.9%	[6.5 ; 9
Currently in partnership	85.6%	[82.3 ; 88.8]	73.7%	[72.5 ; 75.0]	85.9%	[83.8 ;
Partner info missing (no partner)	25.9%	[21.9 ; 29.9]	57.8%	[56.0 ; 59.7]	25.2%	[22.7 ;
Partner's earnings in Euro	1749	[1600 ; 1898]	954	[903 ; 1005]	1734	[1616 ;
Importance of family...						
... very high	48.2%	[44.1 ; 52.2]	42.5%	[40.5 ; 44.5]	47.9%	[44.6 ;
... high	8.7%	[6.2 ; 11.2]	13.5%	[12.1 ; 14.8]	9.0%	[7.1 ; 1
... low or very low	0.4%	[-0.2 ; 0.9]	1.7%	[0.6 ; 2.8]	0.3%	[-0.1 ;
... missing	42.8%	[38.5 ; 47.1]	42.3%	[40.5 ; 44.1]	42.8%	[39.4 ;

Note: Percentage values refer to Dummy variables. CI denotes the confidence interval.

Table 3: Propensity to have first child next year (probit regression results)

Variable	**Coefficient**	**(Std. Err.)**
Wage	-0.001	(0.003)
Actual working hours	0.002	(0.002)
Part time	-0.048	(0.070)
Blue collar	0.038	(0.049)
Self-employed	0.118	(0.110)
Civil servant	0.102	(0.066)
Employer mainly publicly owned	0.011	(0.046)
Experience full time	0.007	(0.073)
Experience part time	-0.067	(0.116)
Experience unemployment	0.213	(0.199)
Experience full time squared	0.000	(0.002)
Experience part time squared	-0.002	(0.003)
Experience unemployment squared	0.002	(0.008)
No training degree	0.159	(0.435)
University degree	-0.652	(0.514)
In education (or educational info missing)	-0.108	(0.062)
Living in East	-0.063	(0.178)
Partnership	-0.197	(0.051)
Partner's age	-0.039	(0.027)
Partner's age squared	0.000	(0.000)
Age difference	0.059	(0.012)
Partner high education	0.396	(0.289)
Partner medium education	0.407	(0.288)
Partner low education	0.395	(0.289)
Partner's gross earnings	0.000	(0.000)
Partner's earnings missing	-0.028	(0.069)
No partner info	-1.398	(0.560)
Age*Importance of family very high	-0.003	(0.003)
Age*Importance of family high	-0.005	(0.003)
Age*Importance of family low or very low	-0.014	(0.011)
No. of years since when importance of family was last asked * ...		
...Importance of family very high	-0.081	(0.080)
...Importance of family high	-0.047	(0.137)
...Importance of family low or very low	0.146	(0.754)
...Age*Importance of family very high	0.003	(0.003)
...Age*Importance of family high	0.003	(0.005)
...Age*Importance of family low or very low	-0.005	(0.028)
Age 23	0.039	(0.068)
Age 24	0.102	(0.067)
Age 25	0.047	(0.073)
Age 26	0.202	(0.076)
Age 27	0.223	(0.086)
Age 28	0.266	(0.098)
Age 29	0.208	(0.117)
Age 30	0.320	(0.137)
Age 31	0.130	(0.168)
Age 32	0.200	(0.198)
Age*Married	0.016	(0.001)
Age*Experience full time	0.000	(0.003)
Age*Experience part time	0.003	(0.004)
Age*Experience unemployment	-0.008	(0.008)
Age*No training degree	-0.004	(0.017)
Age*University degree	0.024	(0.018)
Age*In education	-0.005	(0.018)
After law change	-0.318	(0.097)
Unemployment rate	0.034	(0.016)
Intercept	-1.314	(0.571)

Note: Year dummies, federal states, and firm size have been controlled for.

Figure 1: Size of treatment (left) and control group (right), by age

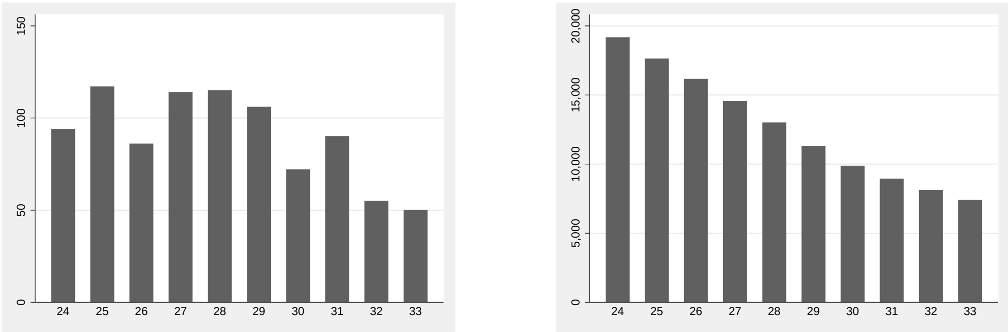


Figure 2: Completed training degrees one year before birth by age at birth

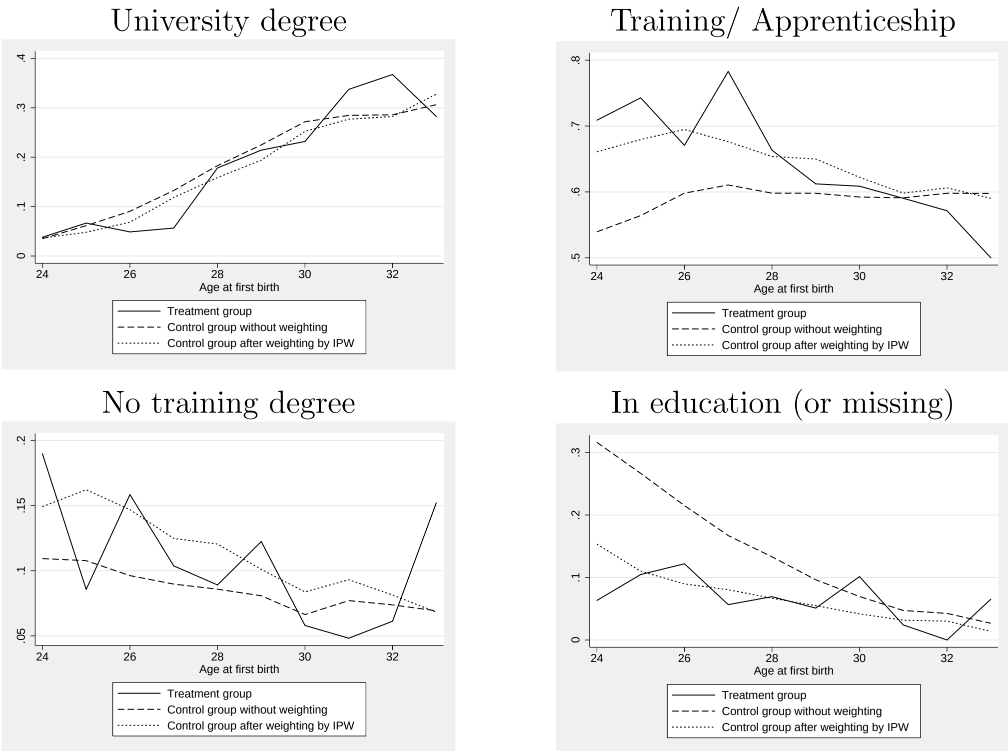


Figure 3: Partnership one year before birth by age at birth

Figure 4: Employment one year before birth by age at birth

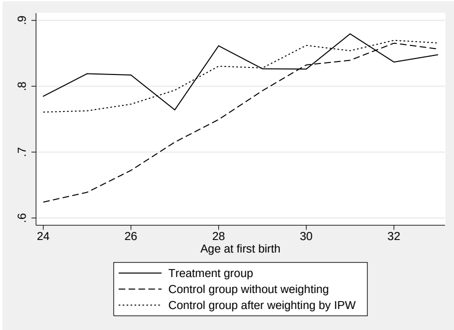


Figure 5: Employment rates before and after reweighting

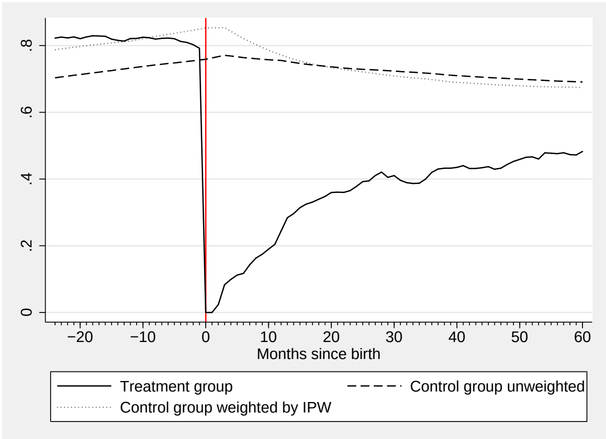


Figure 6: Childbirth rates for treatment and control group (IPW-weighted)

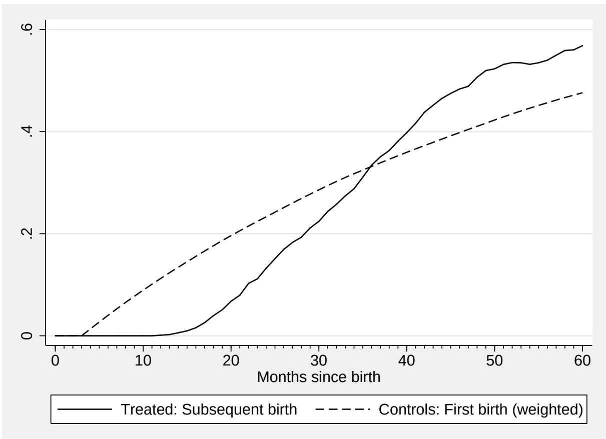


Figure 7: Average Treatment Effect on the Treated (ATT)

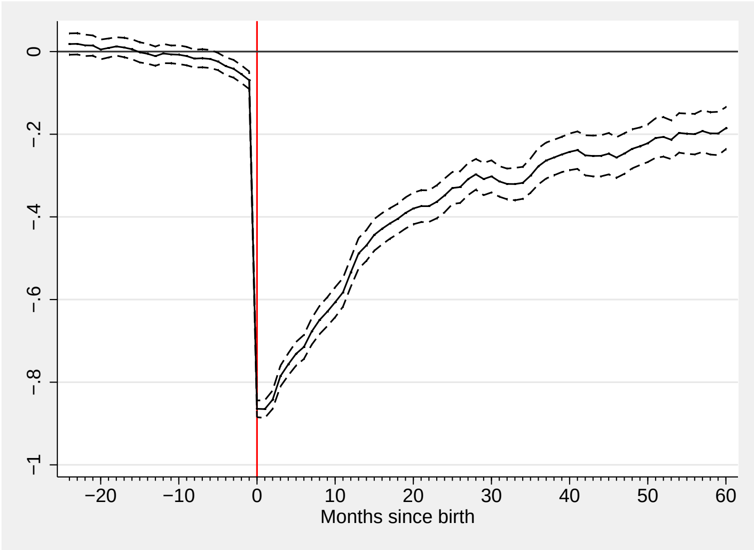
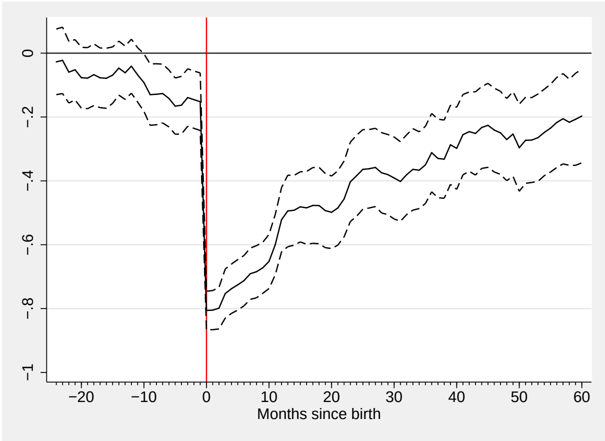


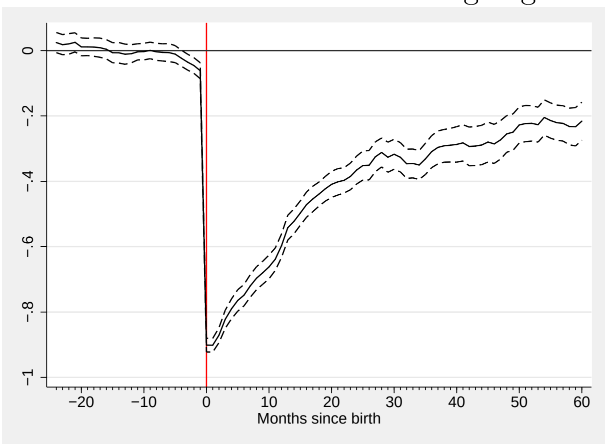
Figure 8: ATTs for different educational levels

Figure 9: ATTs for different educational levels

Effect for mothers without training degree



Effect for mothers with training degree



Effect for mothers with university degree

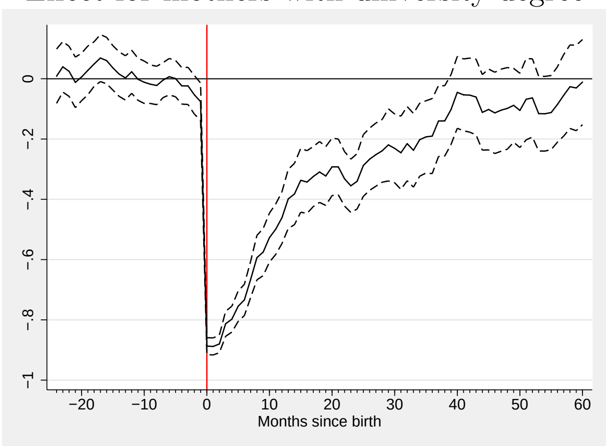
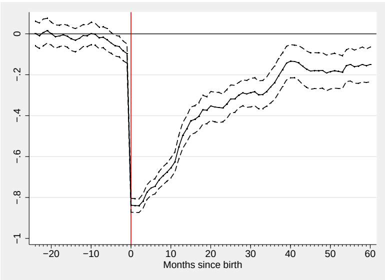


Table 4: **Cumulative employment losses (in person-months)**

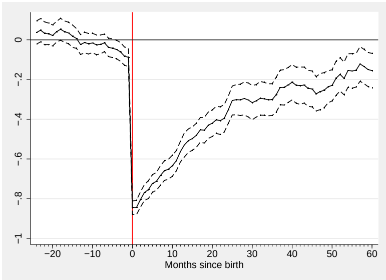
	Pooled		By Training Degree		
	Absolute	In Percent	No training	Training	University
After 1 year	-8.71	-72.5 %	-8.64	-9.20	-8.61
After 2 years	-13.77	-57.4 %	-14.40	-14.73	-12.76
After 3 years	-17.57	-48.8 %	-18.88	-18.78	-15.69
After 4 years	-20.60	-42.9 %	-22.14	-22.25	-16.91
After 5 years	-23.10	-38.5 %	-25.10	-24.98	-17.87

Figure 10: ATTs for different age groups

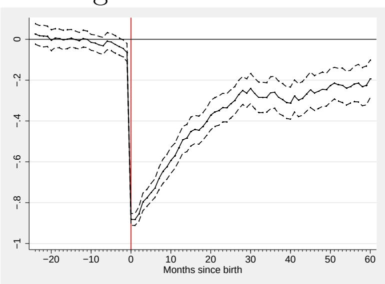
ATT age 24 and 25



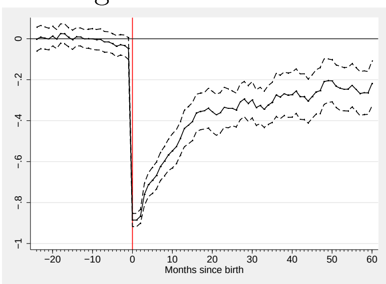
ATT age 26 and 27



ATT age 28 and 29



ATT age 30 and 31



ATT age 32 and 33

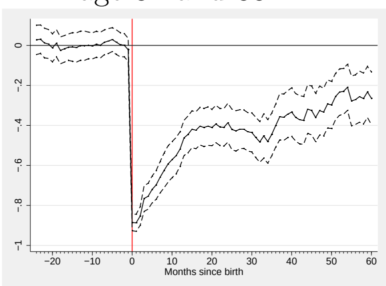


Figure 11: Childbirth rates by age

Treatment group | Control group, IPW-weighted

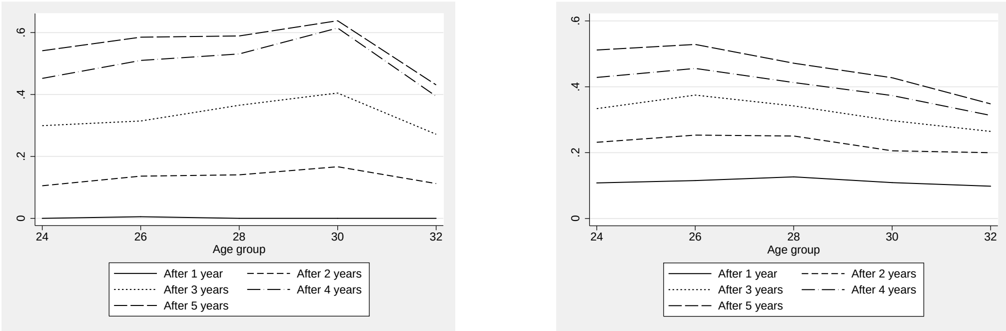


Figure 12: Evaluating a policy change

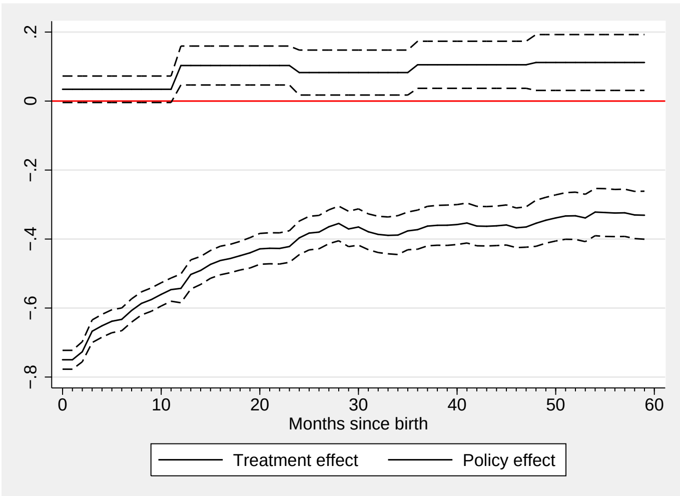


Figure 13: Evaluating a policy change: Effects for subgroups

Children born in 2001 or afterwards | Children born before 2001 and still affected by reforms

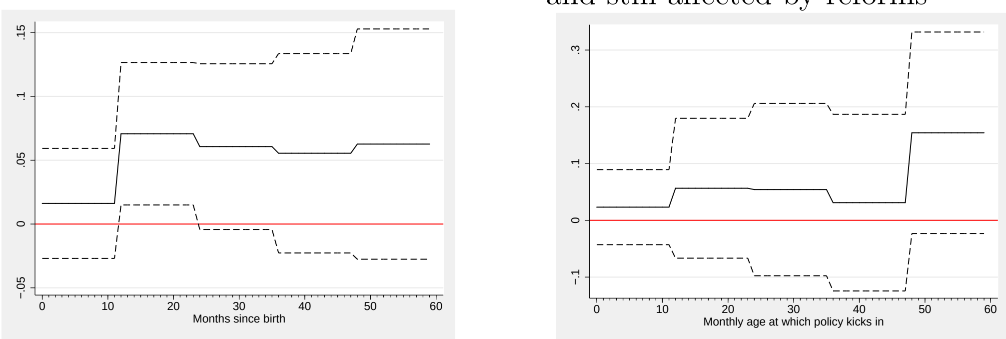
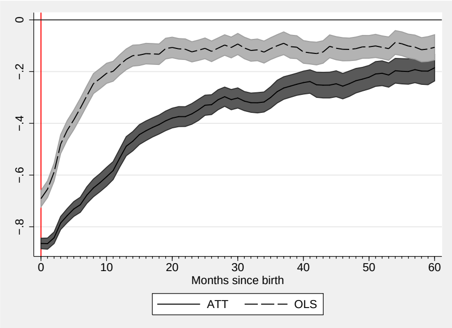


Figure 14: Sensitivity Analysis: Alternative Estimates

OLS



No reweighting

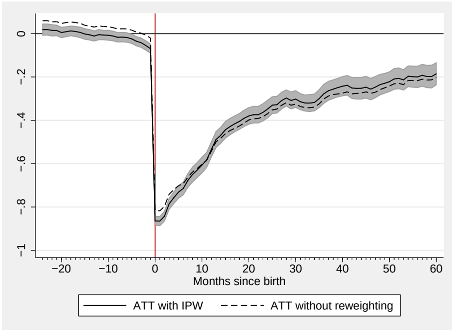
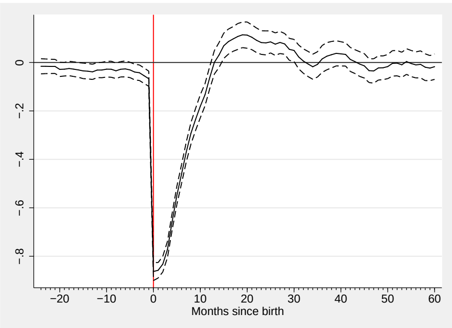


Figure 15: Sensitivity Analysis: Varying control group selection

Control group: Have first child soon (i.e. within two years)



Control group: Have first child late or never (i.e. after 5 years or more)

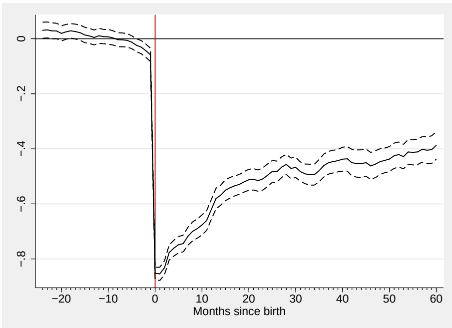


Figure 16: Sensitivity Analysis

No second births in treatment group

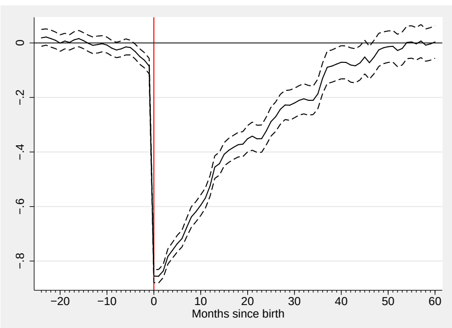


Figure 17: Sensitivity Analysis

Only balanced panel