Discussion Paper No. 06-053

The Wage Effects of Entering Motherhood

A Within-firm Matching Approach

Miriam Beblo, Stefan Bender and Elke Wolf

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 06-053

The Wage Effects of Entering Motherhood

A Within-firm Matching Approach

Miriam Beblo, Stefan Bender and Elke Wolf

Executive summary

This paper addresses the question why women with children are observed to have lower wages than women without children. This 'family pay gap' is commonly attributed to differences in employment experience – lower human capital formation, respectively human capital depreciation, during child-related employment breaks – and differences in job flexibility or effort between mothers and non-mothers. An alternative explanation is segregation, that is, selection of women who will eventually have children into more family-compatible occupations and establishments at the price of lower wage earnings. In this case, a pay gap would be observed between mothers-to-be and women who will never have children even before the event of a child birth and subsequent career intermittence.

This paper tries to disentangle the segregation effect from the wage effect caused by a child-related employment break by drawing on longitudinal data of female employees which include wages before and after child break. We make use of firm specific effects, as we are able to identify colleagues within the same firm. Hence, by applying a semiparametric estimation method based on matching, we compare the wage rate of each female employee who experienced a child-related employment break with that of a continuously employed but otherwise similar colleague of the same firm. By use of various matching procedures this paper provides a robust measure of the wage backlog caused by child birth and parental leave.

We find first births to reduce women's wages by 16 to 19 percent, regardless of the matching procedure applied. However, neglecting the firm-specific effect and matching across all firms, yields a wage cut of 30 percent. Concluding from this result, selection into firms is a major explanatory factor of the family pay gap – women with children are more likely to be found in firms with lower wage growth – but it does not tell the whole story. Even compared to their immediate firm colleagues are mothers' wages negatively affected upon return to the job. As expected, the wage loss increases with the duration of the employment break as we can show in a subsequent regression analysis.

The wage effects of entering motherhood

A within-firm matching approach

Miriam Beblo[*], Stefan Bender[~] and Elke Wolf[#]

[*] *Berlin School of Economics (FHW) and Centre for European Economic Research (ZEW)*
[~] *Institute for Employment Research (IAB)*
[#] *Munich University of Applied Sciences and Centre for European Economic Research (ZEW)*

July 2006

Abstract
We analyze the wage effects of employment breaks of women entering motherhood using a novel within-firm matching approach where mothers' wages upon return to the job are compared with those of their female colleagues within the same firm. Using an administrative German data set we investigate three different matching procedures based on information two years before birth: (1) exact matching on individual characteristics, (2) propensity score matching and (3) a combined procedure of exact and propensity score matching. Our results yield new insights into the nature of the wage penalty associated with motherhood, since we find first births to reduce women's wages by 16 to 19 percent, regardless of the matching procedure applied. Neglecting the firm identifier and matching across all firms, however, yields a wage cut of 30 percent. Furthermore, we can show that the wage loss increases with the duration of the employment break.

Keywords
wages, parental leave, matching

JEL-Codes
J13, J31, C14

Acknowledgments
This paper has been written as part of the project “Familienpolitik – Erwerbstätigkeit – Fertilität: Ökonomische und soziale Aspekte der Vereinbarkeit von Familie und Beruf“, commissioned by the Landesstiftung Baden-Württemberg. We thank Bernd Fitzenberger, Hermann Gartner, Reinhard Hujer, Michael Lechner, Jeff Smith and Thomas Zwick for helpful remarks. All remaining errors are our own responsibility.

Introduction

It is an empirical fact that, on average, women with children have lower wage rates than women without children. This wage penalty or ‘family pay gap’ has been investigated vastly in the United States (see recent evidence by Budig and England 2001, Lundberg and Rose 2000, Waldfogel 1998a) and in the United Kingdom (Joshi, Paci and Waldfogel 1999, Waldfogel, 1998b). Lower wages of mothers may be caused by career intermittence due to child birth and child rearing, but also by a reduced job attachment, hence, a decrease in effort of working mothers. Another prominent source for pay differences may be the occupational segregation of mothers-to-be into lower paying jobs or establishments with family-friendly job or firm characteristics. As the underlying effects are manifold and complex, the size of the causal wage loss due to motherhood is difficult to measure.

In this paper we examine the backlog of mothers’ wage rates caused by parental leave breaks using a novel within-firm-semiparametric approach based on matching. That is, we present two innovations: (1) the use of firm specific effects and (2) the use of a nonparametric estimation method. We compare the wage rate of each female employee who experienced a child-related employment break with that of a continuously employed but otherwise similar colleague of the same firm. Due to within-firm matching, unobserved firm-specific heterogeneity can be fully taken into account. Selection in observable and unobservable individual characteristics is accommodated by different matching and estimation algorithms. The effect of the employment break duration on the treatment effect is investigated in a subsequent regression analysis of the wage differences between the matched pairs.

Germany is known as a country with one of the most extensive parental leave legislations, comprising a mother protection period of 14 weeks and a parental leave period of up to 3 years during which the leave taker’s job is protected against dismissal. Although both parents are eligible for the leave and parents are allowed to switch the leave taker several times, 98 percent of those on leave are women. In 2000 only 53 percent of mothers in West Germany and 70 percent in East Germany were re-employed right after the formal leave period (Beckman and Kurtz 2001). The wage effects of such career interruptions have been found to be substantial in Germany (see estimations by Beblo and Wolf 2002a and 2002b, Ejrnaes and Kunze 2004, Kunze 2002 and Ondrich, Spiess and Yang 2001). All of these studies use extended wage estimations to determine the average wage differential between women with employment breaks and continuously employed women. This procedure involves two main problems. First, wage regressions represent a parametric approach which relies on the assumption that the functional form is linear in parameters. Second, the estimated wage effect of employment breaks is based on observed wage differentials of women working in *different* firms. Considering that not only the wage level, but also the distribution of wages differs across firms (see e.g. Davis and Haltiwanger, 1991; Bronars and Famulari, 1997; Abowd, Kramarz and Margolis, 1999), standard wage regressions ignoring these firm-specific effects on wages may lead to biased results. We try to overcome these shortcomings by applying a matching approach to estimate the wage backlogs of mothers relative to comparable non-mothers and assess the differences between colleagues within the same firm. To our knowledge, the only study that exploits such an evaluation approach to analyze the wages of mothers is provided by Simonsen and Skipper (2005, 2006), who do not have a firm identification and thus only compare women across firms.

The challenge of our research question is to determine what the wage rate of a mother would be if she had not given birth and experienced an employment break within a specific observation period. Since this counterfactual outcome is not observable, we have to identify a control group of females without employment breaks which is comparable to our selection of females giving birth with respect to the distribution of all variables that affect the wage determination process. A perfect counterpart for a mother would be a childless female colleague who works in the same company, in a comparable job, is of comparable age, has experienced the same career path, achieved the same educational level and exhibits the same unobservable characteristics – such as ability or motivation - potentially affecting the wage rate. As such an ideal counterpart is difficult to find, we propose three alternative matching procedures to determine a useful control group. In all three cases we compare women giving birth to their first child (mothers) and women not giving birth during the observation period ("non-mothers"), but being continuously employed within the same firm to accommodate firm segregation and unobserved firm-specific effects. Furthermore, we apply matching procedures that produce exact matches with respect to the working time status and occupation of the women. These matching procedures take into account that, following Polachek (1981), mothers-to-be may be less attached to the labor market on average and therefore choose jobs or occupations with rather flat experience profiles but smaller expected wage cuts due to discontinuous employment patterns. This way, our matching is meant to control for observable and unobservable features of mothers-to-be and their employers.

Once the control group is determined we compare the wage rates of mothers and non-mothers before and after the mothers' employment break. We have information on wages right upon return as well as 6 months, 12 months and 24 months after the end of the break. These dates are dynamically determined by the duration of the interruption chosen by the mother. We compare her wage rate with that of the respective (set of) control colleague(s) who is (are) observed working in the same firm and on the same effective days. The mean difference in wages reflects the average treatment-on-the-treated effect of entering motherhood and experiencing a specific employment break. We furthermore compare the differences in wages of mothers and control colleagues before and after a break which reflects the conditional difference-in-difference estimator. The wage effect, however, may differ across women due to heterogeneity in the duration of the employment interruption. In a subsequent regression analysis we therefore investigate the differences in the wage losses using the duration of the employment break as explanatory variable.

The remainder of the paper is structured as follows. Section 2 presents our methodological approach. The data is described in Section 3. Section 4 discusses the results of the alternative matching procedures and the second-step wage gap analysis. The last section concludes and discusses potential extensions of our approach.

1 Our econometric approach

The goal of this paper is to determine the average treatment effect on the treated (ATT) on the wage rate, that is, the average expected effect of entering motherhood and experiencing an employment break for all employed mothers-to-be. We follow Rubin (1974) and identify the causal effect of the "treatment" by comparing the wage rate of a mother after her parental leave period with the hypothetical situation of the same woman if she had not entered the stage of motherhood.

Let Y_1 denote the wage rate of mothers after returning to their former employer and let Y_0 denote the wage rate of women who did not interrupt their career due to child bearing. Let D be an indicator variable which equals one if a woman experienced a parental leave employment break and equals zero if not. Then, the ATT is given by:

$$E(Y_1 | D=1) - E(Y_0 | D=1).$$

Since the hypothetical situation $E(Y_0 | D=1)$ cannot be observed for mothers, we have to find alternative ways to estimate the average wage of mothers with parental leave experience if they were continuously employed. According to Heckman, LaLonde, Smith (1999), two alternative approaches may be applied to estimate the average non-treatment outcome, that is in our case, the wage rate of a continuously employed non-mother: (i) a before-after comparison of mothers or (ii) a comparison with a control group of non-mothers. The first approach assumes a constant average non-treatment outcome over time for the treated. In other words, this approach requires that mothers would have experienced a constant wage rate, had they remained childless. This assumption does not hold, e.g. if the women would have been promoted otherwise, if their wage scales are tenure based or if macroeconomic shocks have taken place. Another fundamental problem which applies to both approaches is the potential selection bias which occurs if mothers differ from both, mothers-to-be and non-mothers, due to observable and unobservable characteristics. Due to these selection effects, the wage levels of mothers and non-mothers may be different before the treatment already – for whatever reason. In this case, neither a simple cross-section regression of wages depending on past parental leave experience nor a before-after comparison would yield unbiased results. Hence, the definition of an appropriate control group must be chosen very carefully. We will now briefly discuss some approaches to control for selection on observable and unobservable characteristics.

1.1 Controlling for selection on observable characteristics

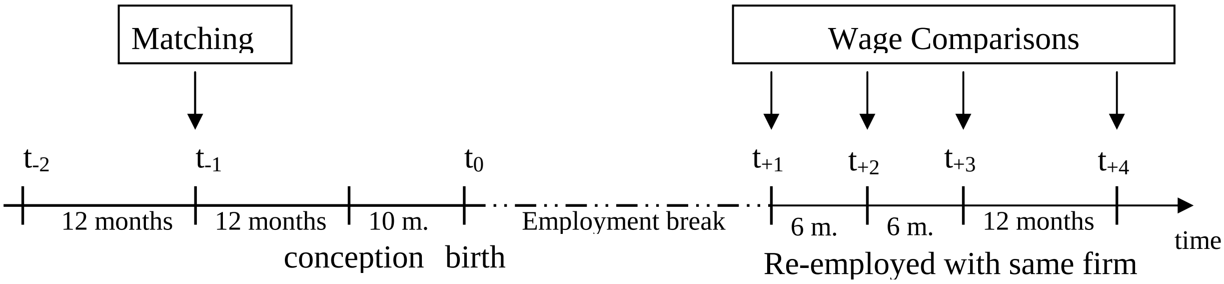
To account for differences in observable characteristics, we refer to the Conditional Independence Assumption (CIA). Under CIA, it does not matter whether we estimate the average outcome of continuous employment based on information about mothers or non-mothers provided that they have similar observable characteristics (Imbens 2004):

$$E(Y_0 \mid D = 1, X) = E(Y_0 \mid D = 0, X)$$

It is therefore important to choose this set of observable characteristics carefully. Basically, (X) should include the wage rate before treatment and all wage determining characteristics. Based on the choice of *(X)*, one can select the appropriate control group by means of propensity score matching or exact matching algorithms. As described in Section 2.3, we will apply both strategies as well as a combination of the two. But when is the appropriate point in time to compare the selected characteristics of mothers and non-mothers? Of course, the definition of the control group should be based on information before the observed career intermittence of mothers. Considering that becoming pregnant is not a fully exogenous event and mothers-to-be may be more likely to substitute wage income for flexible working conditions (which are difficult to observe in general), we should compare mothers-to-be and non-mothers with respect to their wage rate and all wage determining characteristics when the employment break is not yet a certain event. Since the shape of the wage profile just before the first birth might already be affected by the future event (analogously to Ashenfelter's dip in labor market policy evaluation, see e.g. Bergemann, Fitzenberger, Speckesser 2003), we define our first observation point 12 months before conception, respectively 22 months before birth.

Figure 1 illustrates the time frame of our evaluation approach. At t_0, the mother gives birth to her child. To account for differences of women with and without maternity leave breaks, we match mothers and non-mothers at time t_{-1}, assuming that the future pregnancy has not been anticipated yet, at least not in a way related to wages or wage-determining characteristics. The employment break due to motherhood lasts from t_0 to t_{+1} and differs between individuals. At t_{+1}, the mother returns to her former employer[1]: t_{+1}, just as t_{+2}, t_{+3} and t_{+4} are alternative observation points for wage comparisons with the mother's female colleagues (i.e. her matching partners) who are - still or again - working at the same firm.[2]

Figure 1: Time frame of our evaluation approach



1.2 Controlling for selection on unobservable characteristics

Given that the fertility choice is very complex, the correction of the selection bias based on observable characteristics might not be sufficient to yield a consistent estimate of the ATT, because unobservable characteristics may be of significant importance. On one hand, women with a preference for raising children may anticipate their non-employment spells and hence are less willing to invest in their human capital, taking into account that the resource "time" of this investment may be very limited. On the other hand, it is also conceivable that women whose career has come to a halt for some reason are more likely to choose motherhood as a kind of fallback way of life than ambitious and career-oriented women who have found full satisfaction in their work life. In consideration of these facts, the underlying mechanisms affecting fertility and subsequent parental leave spells are not fully observable to us.

---

[1] The return to the job is defined as an employment spell of at least 3 months length.

[2] We are aware that this set-up gives rise to yet another source of selection bias since the analysis is based on a comparison of firm-stayers only, as regards both mothers and control observations. We are planning to perform sensitivity analyses including also firm-movers to assess the potential bias associated with our selective sample.

Provided that the selection effect due to unobservables is time-invariant and linear, Heckman, LaLonde and Smith (1999) propose the difference-in-difference-estimator (DiD) to yield unbiased results. This estimator extends the simple before-after comparison by contrasting the before-after difference in wage rates of mothers to the wage change of non-mothers within the same observation period. According to Heckman, Ichimura and Todd (1998), the DiD estimator combined with non-parametric matching, the so-called conditional difference-in-difference-estimator, proves very useful in controlling for both, selection on observed and unobserved characteristics.

1.3 Definition of the control group

The challenge with the measurement of the ATT is to determine the wage rate of a mother if she had not given birth to a child and interrupted her employment career for this reason. Given that this hypothetical outcome is not observable, we have to identify a control group of non-mothers which is comparable to the mothers with respect to the distribution of all variables which affect the wage determination process. As mentioned above, a perfect counterpart for a mother would therefore be a female colleague without children who works in the same company in a comparable job, has about the same age, has experienced a comparable past career path, achieved the same educational level and exhibits the same unobservable characteristics potentially affecting the wage rate. It is obvious that the ideal counterpart is difficult to find, even if we had full information on all female colleagues. Hence, we propose three feasible alternatives to determine a useful control group.

The most straightforward method of matching, that is exact matching, compares persons with exactly the same values of observed characteristics X . Note that this method works only with a limited number of discrete X -variables or, alternatively, value ranges for continuous X-variables. The choice of the relevant X -variables is delicate, because it is subject to a trade-off denoted as the “curse of dimensionality”. The higher the number of variables selected and the larger the range of values these variables may take, the lower is the probability to find an exact match. But, the lower the number of selected variables and the less values these variables may take, the more vague is the match, that is, the more unequal might be the matched pairs.

For our first matching procedure we decided to select “exact” matches with respect to the establishment, occupation (80 categories), age (with a maximum deviation of 2 years), education (3 categories), working time status (full/part time), total work experience (with a maximum deviation of 20 percent or 30 percent if no colleague could be identified otherwise) and daily gross earnings (with a maximum deviation of 10 percent respectively 20 percent).

The information which enters the matching procedure refers to t_{-1} in Figure 1 which is 22 months before entering motherhood. We denote this procedure as MATCH 1. Exact matching with respect to the establishment ensures that treated and untreated women underlie the same unobserved fixed effect influencing the wage determination process within the establishment. Matching on education and occupation is meant to account for unobserved individual heterogeneity affecting the occupational choice in the sense of Polachek (1981). By conditioning on age and experience, we try to account for different stages in the life cycle associated with the likelihood of maternity and labor market attachment. Education levels and working time serve to make daily wages comparable. Since there were still a considerable number of controls for most treated observations, we further accounted for differences in hourly wage rates one year before conception. In the case that more than one female colleague match the criteria, a control observation is generated by calculating averages for all variables across all selected colleagues. The quality of this matching procedure and the percentage of treated observations with an appropriate control and the average number of control observations – conditional on finding at least one control - are discussed in Section 4.1.

Due to the curse of dimensionality, exact matching is not capable of providing an appropriate control in all dimensions and for all mothers. We therefore use an inexact matching procedure in the second application (MATCH 2). Propensity score matching reduces the dimension problem by defining a distance metric on X and subsequent matching is based on the distance metric rather than the X. Rosenbaum and Rubin (1983) illustrate, that the distance metric may be defined as: $P(X) = \Pr(D = 1 | X)$.[3] Hence, we estimate a parametric probit model to predict the individual propensity score $P(X)$. In our setting, this figure describes the likelihood of becoming a mother and returning to a full-time job within the observation period for each individual in the sample. The vector *(X)* hence includes all variables presumably affecting motherhood and subsequent employment. Given the limited information about the household context in our data, we basically include information on age, education, the current occupation and the past employment history. The next issue we have to deal with is the choice of the appropriate matching algorithm. The most common form is the nearest neighbor matching (NNM) selecting one or more untreated observations whose $P(X)$ is closest to that of the mother. A very appealing alternative which makes use of all potential comparison observations, hence holding variance low, is kernel matching (KM).

[3] The intuition behind the propensity score matching is that individuals with the same probability of "participation", that is becoming a mother, can be paired for purpose of comparison.

Given that the number of comparison observations is small for some mothers – namely those working in small establishments (see Table 1) – we try both matching algorithms. First, we apply NNM with replacement in order to keep the bias small. As the choice of the number of nearest neighbors is subject to a trade-off between bias and variance, we choose one neighbor, being aware that the variance may be high. Note that all pairs have to belong to the *same* establishment in order to control for unobserved firm-specific effects influencing the wage determination process. To restrict the differences between the nearest neighbors – which tend to be larger in smaller firms – we define a caliper of 0.5. Matching more than one nearest neighbor increases the bias, while the variance of the match becomes smaller. Therefore, we secondly apply a KM with an Epanechnikov kernel[4]. Considering that comparison observations – that is all female colleagues of mothers – are numerous in the full sample but asymmetrically distributed across firms– mothers in small firms have fewer potential counterparts whereas mothers in bigger firms are more likely to have more adequate matches – kernel matching is especially helpful because it exploits additional data when available but it does not rely on bad matches where close neighbors do not exist. The results serve to check the robustness of the NNM results, but are not discussed in detail.

As a third matching algorithm, we combine the two traditional matching approaches described into one procedure (MATCH 3). In a first step, we select exact matches with respect to the establishment, working time, occupation and total work experience (with a maximum deviation of 20 percent). Obviously the number of untreated counterfactual observations is much bigger than in MATCH 1. In a second step, we use the propensity score determined by the MATCH 2 procedure in order to define a comparable control group beyond the pre-matching from step 1 by nearest neighbor matching. The advantage of this third procedure is that not only unobserved firm characteristics affecting the wage determination process within the establishment are accommodated, but also unobserved individual heterogeneity affecting the occupational choice and the overall career path. Note that all of the three matching algorithms allow the control group to contain mothers, under the condition that they delivered their children before 1997 and were continuously employed during our observation period.

1.4 Wage comparison

Once the control groups are determined, we calculate the difference in the wage rates of mothers and non-mothers at different points in time. As illustrated in Figure 1, we consider

4 A normal kernel is less appropriate in our setting, because it would rely on all potential control observations – irrespective whether they work in the same firm or not.

wages right upon return as well as 6 months, 12 months and 24 months after the break. The timing of these dates is determined by the duration of the career interruption of the mother. In this analysis, we compare her wage rate 12 months after reentering the labor market (in t_{+3}) with that of the respective (set of) control colleague(s) – defined by the matching process in t_{-1} – who is (are) still working in the same firm.[5] The differences in individual wages determine the average treatment-on-the-treated effect (ATT) of being a mother.

The treatment effect, however, may depend on the duration of the employment interruption. In a regression analysis, we therefore investigate the effect of duration of the employment break on the average wage differential between mothers and women without employment breaks.

2 Data

The merit of our empirical analysis is significantly nourished by the uniqueness of our data set that allows longitudinal comparisons between mothers and non-mothers within the same firm. We draw on process generated data provided by the Institute for Employment Research (IAB). These German register data are generated by an integrated notifying procedure for the public health insurance, statutory pension scheme and unemployment insurance which was introduced in 1973. By law, employers have to provide information to the social security agencies for employees acquiring claims to the social security system. These notifications are required at the beginning and ending of any employment relationship. In addition, employers are obliged to provide an annual report for each employee who is employed on December 31st of each year and covered by social insurance. The reports include information on sex, year of birth, nationality, occupation, qualification and gross wage rate of the employee. Furthermore, each spell includes information on the industry and a unique firm identifier of the establishment where an individual is employed. According to the obligation to register with the state pension authorities, this data encompasses all persons who have paid contributions to the pension system or who have been covered by the pension system through contributions by the unemployment insurance or by being a parent. As a consequence, certain groups of employees are not covered by the data:

- (Temporary) civil servants or self-employed persons
- Women who are employed in East Germany or abroad.

---

[5] We chose t_{+3} to compare wage rates after the employment break because – out of administrative reasons - the identification of part-time employment is more reliable when the calendar year has changed.

The latter selection is necessary because the supplementary information on the nature of employment breaks is available for workers employed in West Germany only. Nevertheless, the sample represents still about 80% of all employees on the labor market.[6]

We use two different samples of these register data. We combine the IAB employment sample with additional administrative data assembled at the state pension authorities (IAB employment supplement sample I).[7] Both data sets can be linked by the social security number. The matched file contains a 1% random sample of the total German population having been gainfully employed at least for one day between 1975 and 1995 (for details see Bender, Haas, and Klose 2000). Based on the supplement sample, we have exact information about the individuals' entire working lives that allows us to distinguish between different types of "non-working" periods, namely, unemployment, formal parental leave, illness, disability, care for other people, full-time education, military or civil service and other out-of-the-labor-force spells. Furthermore, these data allow us to identify the fertility history of all women. Since the birth of children increases the pension entitlement of the mother, IAB employment supplement sample I provides exact information about the number of children as well as the month of birth.[8]

Based on the exact information about fertility and employment history, we select our treatment group, that is, women who have given birth to their first child between 1987 and 1995. Since we are interested in the wage effects of parental leave periods, we further restrict the sample to women who have been working ten months and 22 months before the birth of their first child and, after the employment break, returned to the same firm for at least three months within our observation period, that is until 1999. After deleting observations with missing values, we remain with 1,390 observations of mothers.

As described in Section 2.3, the innovation of our analysis is to measure the backlog of mothers' wages by comparing mothers' and non-mothers' wages within the same firm. Hence, the control group has to be drawn from a sample of all colleagues of these 1,390 mothers selected in the first step. To do so, we make use of the so-called Employment Statistics Register, which includes information about the total population of all people who

---

6 Due to the nature of the data we do not have any information on the household background, such as the household income, the partner's employment status etc.

7 For first descriptive analyses with these data see Prinz (1997), for an analysis of the wage penalties of heterogeneous employment biographies see Beblo and Wolf (2002 a and b) and for the effects of entry into motherhood on women's employment dynamics see Bender, Kohlmann and Lang (2003).

are registered in the social security system. The following procedure describes our strategy to identify all female colleagues of our treatment group, that is, women who became mothers between 1987 and 1995 and were employed before and after their parental leave spell:

1. We identify the treatment group in the Employment Statistics Register.

2. We identify the unique firm number of every observation in the treatment group.

3. We select all women, who were employed in t_{-1} and t_{+3} (or t_{+4}) in the identified firms.

After this selection, the data set consists of 307,541 observations of potential control women.[9] Due to missing observations of selected variables, 1,357 mothers and 298,822 female colleagues enter the propensity score estimation, which is required for two of our estimation procedures. For the purpose of wage comparison, we further restrict our sample to women in full-time employment one year after the mothers' return to the job (in t_{+3}). We end up with 566 mothers and 233,358 female colleagues, for whom we have information on wages after the employment break as well as wages and individual characteristics 22 months before birth. Being aware that our population is very selective in terms of the attachment to the labor market our results may be interpreted as a lower bound to the overall short-run wage effects of entering motherhood.

Figure 2 illustrates the average wage rates[10] of the selected mothers and female colleagues before and after the mothers' employment break. It is obvious that already 22 months before birth mothers-to-be earn lower wages on average than their colleagues. Presumably, this wage differential is caused by differences in observed characteristics for the most part. Interestingly, pre-birth wage growth does not seem to differ between future mothers and their control group. After the employment break, the gap between mothers and women without comparable employment breaks becomes even greater. While the female colleagues experience an almost linear wage growth, mothers' wage profiles exhibit a sharp decline and

---

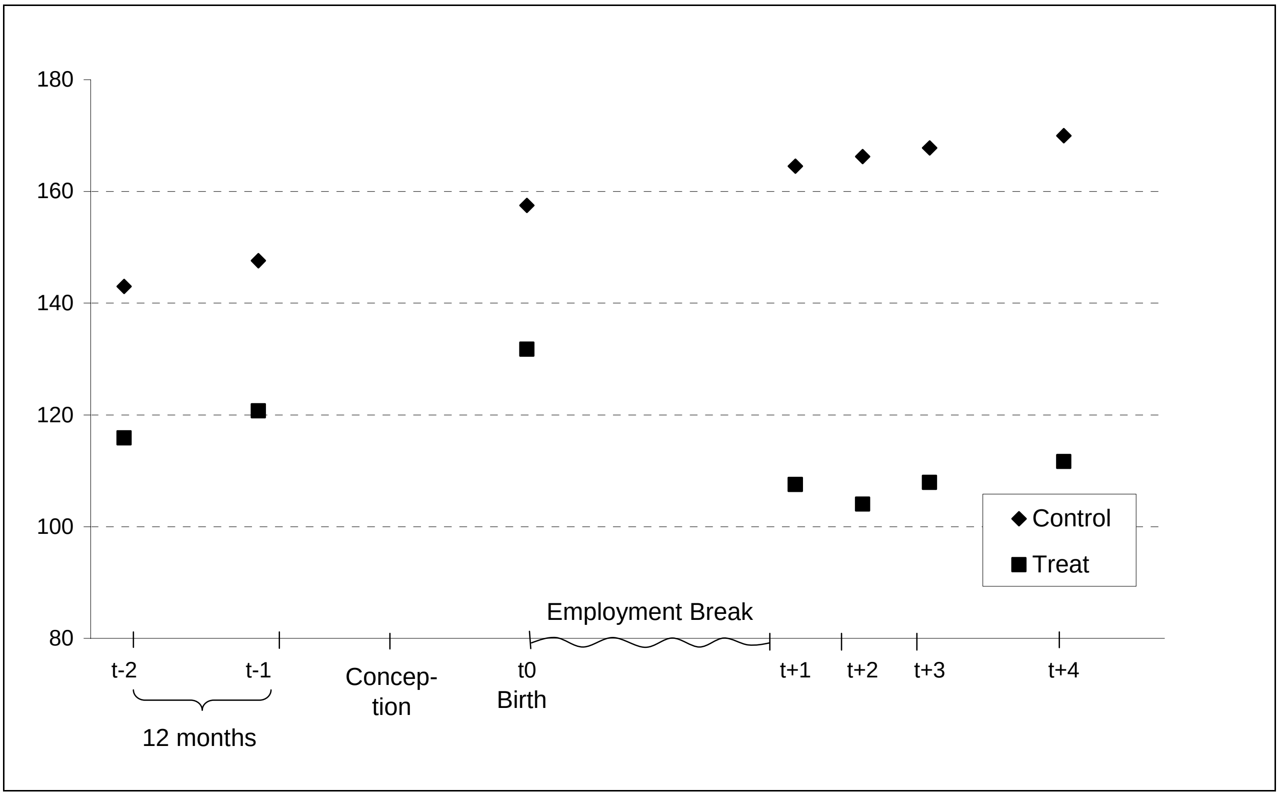
8 Under very restrictive assumptions, it is possible to interpret specific gaps in the IAB employment sample as interruptions due to parental leave or national service (see e.g. Kunze 2002: 11). An exact identification of a child birth, however, is only possible with the supplementary file.

9 Since the selection of our treatment group is based on a 1% sample and the group of potential controls is drawn from the total population, this sampling procedure yields an oversampling of control observations. Given that the 1% sample of the total population (that is the supplement sample I) is completely random, we do not require weights to consistently estimate the probability of entering motherhood, such as in the case of choice-based sampling of the treatment group.

10 Wage information in the employment register is censored at the upper bound. Estimation strategies may be used to impute wages above this ceiling (see for example Gartner 2005). We are using the original wage information, because we do not have many women above the threshold level. This way, we may underestimate the wage losses of mothers who start above this level, if their wages are falling below after the

hardly reach the level from 22 months before birth (t_{-2}) even two years after the end of the employment break (t_{+4}) .

Figure 2: Average wages of mothers (Treat) and their female colleagues (Control) before and after birth



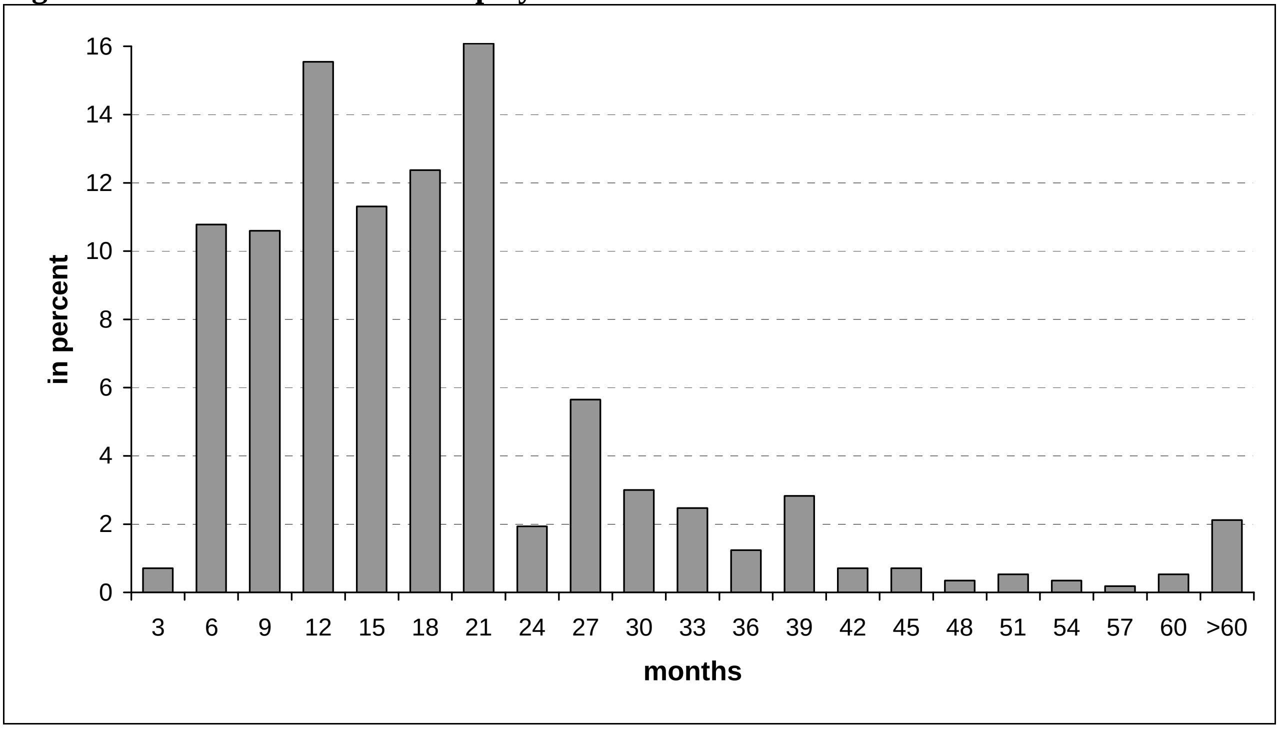
Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233,358 female colleagues, drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. Wages in DM (German marks). 1 DM equals 0.51 euro.

Figure 3 describes the duration of employment breaks of first-time mothers in our sample. The majority of these women drop out of the labor market for up to 21 months and the average takes 18 months off – conditional on returning to a full-time position with the same employer thereafter. Only a negligible fraction of mothers returns to work within the first 3 months following birth, which is mostly due to the maternity protection period of 8 weeks. Remember that the maternity leave legislation changed substantially during our observation period. Starting from 10 months in 1987 the maximum leave duration increased up to 36 months as of 1992. On average, about 30% of mothers stay away from the firm for more than the relevant maximum legal parental leave period with a guaranteed return to a status-adequate job. It is interesting to note that the share of women prolonging the parental leave

break. Likewise, we will underestimate the wage increases of non-mothers if their wages rise above the ceiling.

beyond the job-protected period differs tremendously by year. While the majority of women who first became mother between 1989 and 1991 did not return to their former job within the parental leave period, this fraction declined to less than 14 percent in 1992.

Figure 3: Duration of mothers' employment breaks



Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233,358 female colleagues, drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register.

3 Results

3.1 Matching

To perform the second and third matching procedure, a distance metric for the propensity of entering motherhood is required. In Table 1, the estimation results of a probit estimation of the likelihood of becoming a mother at time t_0 conditional on individual characteristics at time t_{-1} are presented. Due to the lack of data, no information on the household background such as household composition, partner's employment status or earnings etc. can be considered. To determine differences with respect to the family situation, we exploit all available individual information which might correlate with the likelihood of having children. Age enters the equation with a linear and a quadratic term, both of which are statistically significant. Women without an apprenticeship training certificate have a lower probability of having a child than skilled employees. This likelihood is greatest for college and university graduates. This result seems plausible once we take into account that we only consider mothers who return to their full-time job within our observation period of nine years. The wage rate at t_{-1} is negatively related to future motherhood, indicating that opportunity costs do matter. Married women are more likely to become mothers whereas those working part time are less likely.

Table 1: Probit estimation results of becoming a mother at time t_0

	Coeff. estimate	Std. error
Age	0.5515	0.031
$Age^2/100$	-0.9428	0.050
Education level (ref. apprenticeship)		
No apprenticeship	-0.1081	0.030
College/Univ. graduate	0.5184	0.057
Occupational status (ref. unskilled blue collar)		
Skilled blue collar	0.3373	0.063
Skilled white collar	0.1556	0.057
Occupational group (ref. unskill. manual labor)		
Agriculture	0.4126	0.159
Qualified manual labor	0.0469	0.060
Technicians	0.1318	0.075
Engineering	0.2309	0.139
Services	0.1860	0.058
Qualified services	0.3741	0.069
Semiprofessionals	0.1870	0.065
Professionals	0.2746	0.141
Business - Accounting	0.2270	0.064
Business - Administration	0.2574	0.063
Management	0.7956	0.116
Part time	-0.3222	0.059
Daily wage rate	-0.0081	0.000
Work experience in past 4 years	-0.0215	0.002
Employment breaks (in month) >93 days	-0.0301	0.003
No. of employment breaks (>31 days)	-0.1395	0.026
Tenure	0.0021	0.000
Average wage growth in past 4 years	0.0043	0.029
Constant	-8.6536	0.487
Pseudo R squared	0.1222	
$\chi^2(24)$	2122	
No. of observations	300,179	

Source: Sample of 1,357 mothers (child birth between 1987 and 1995) and 298,822 female colleagues, drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register.

The negative coefficient for the part time dummy may be explained by our selection of the sample as we only study first births. Hence, most part timers are women who have children already and serve as control persons. We find significant differences between occupational groups as well as between blue and white collar workers. We finally include a set of variables describing the past employment history. These are meant to account for selection into motherhood, as mothers-to-be and women who do not plan to have children may follow differing employment paths from the start of the career. However, the results are ambiguous. Not surprisingly, intermittent work histories tend to reduce the likelihood of entering motherhood and returning to the same firm. Active labor market participation during the past four years (as a proxy for career orientation) has a negative effect on entering our treatment group, while tenure within the same firm increases the propensity to become a mother and return to the former employer. The average yearly wage growth during the last four years does not significantly affect the probability to belong to the treatment group.

Due to data restrictions, we evaluate the wage effects of temporary labor market drop-outs due to motherhood one year after the mother's return to her job.[11] For the time being, we furthermore restrict our analysis to full-time employees, because we do not have information about the number of working hours in part time jobs. As mentioned before, we end up with 566 mothers and 233.358 female colleagues, for whom we have information on wages after the employment break as well as wages and individual characteristics 22 months before birth.

Since mothers in small firms are likely to have only few female colleagues whereas mothers in large firms tend to have more female colleagues, the number of potential control observations per treated observation is very unequally distributed (see Table 2). While 9.5 % of all treated observations have no female colleague – and hence drop out – and 7.8 % mothers have just 1 control observation, there is one case where we identify 7159 potential control observations for one specific mother. According to Table 2 about 63 % of the treatment group are employed in firms where we can identify at least 10 potential control observations. Because of this ratio between mothers and potential control persons in the same firm we do not expect to find a comparable female colleague for each mother.

[11] Since the employer‘s record on whether an employee changes from full-time to part-time is not mandatory during the year, this type of information is more reliable after a new employment spell has started, that is in general, at the beginning of the next calendar year.

Table 2: Number of control persons

Control observations	Percent of mothers with ... control observations
0	9.54
1	7.77
2	4.95
3	3.53
4	2.83
5	1.41
6	1.41
7	1.24
8	1.06
9	1.77
10	1.41
>10	63.07

Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233.358 female colleagues (full-time employed in t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register.

Table 3 compares the mean characteristics of the selected mothers and their alternative control groups before treatment, that is, at time t_{-1}, 12 months before conception. Note that the numbers of observations in the group of mothers differ between the different matching approaches. Due to the strict matching criterion in MATCH 1, based on occupation, age, education, work experience and before break gross earnings, the number of appropriate control observations is much smaller compared to MATCH 2 and MATCH 3. While the exact matching provides appropriate control observations for 196 mothers only, the number of matched pairs based on MATCH 2 amounts to 449.[12]

While the average wage rate of mothers-to-be is continuously lower than the wage rate of women who do not interrupt their career due to child bearing (see column 1 in Table 3 and Figure 2), the selected control observations within the same firm seem to earn lower wage rates than the selected mothers 22 months before the date of birth. Comparing the different matching algorithms, it turns out that MATCH 1 and 3 seem to better balance the pre-birth wage differentials than MATCH 2. With respect to the other characteristics, such as age, education level or occupation, the superiority of any of the matching algorithms is not

[12] In Section 4.3, we will expound how the number of matched pairs depends upon the specific matching procedures.

obvious. MATCH 1, for instance, turns out to be very effective in balancing age differences and MATCH 3 performs best with respect to tenure. Both the education level as well as job status are matched quite well in all three models. The fact that the potential and selected control groups exhibit even more and longer employment breaks within the past five years than the mothers-to-be (except for the control group in MATCH 3) leads one to suppose that part of the selected colleagues entered motherhood already before the start of our observation period and therefore experienced a less continuous employment path on average. We will consider this peculiarity in the assessment of our estimation results in Section 5.

Table 3: Descriptive statistics for mothers and control groups

	Raw data		MATCH 1		MATCH2 (caliper = 0.5)		MATCH 3 (caliper = 0.5)	
	Mother	Potential Controls	Mother	Selected Controls	Mother	Selected Controls	Mother	Selected Controls
Characteristics (at time t-1)	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Daily wage in DM	120.72	147.61	128.38	126.64	126.76	117.41	129.33	125.31
Age	28.73	32.90	28.14	28.17	28.75	29.50	28.64	29.10
No apprenticeship	0.20	0.30	0.24	0.24	0.24	0.23	0.27	0.24
Apprenticeship	0.76	0.66	0.76	0.76	0.73	0.74	0.71	0.74
College/Univ. graduate	0.04	0.04	0.01	0.01	0.03	0.03	0.03	0.01
Work experience last 4 years	44.41	44.71	45.38	44.79	44.64	44.56	45.73	46.01
Past employment breaks (in months)	2.02	2.92	1.50	1.82	1.98	2.49	1.38	1.09
No. of past employment breaks	0.24	0.30	0.18	0.26	0.23	0.29	0.18	0.12
Tenure	60.92	70.19	67.66	66.02	62.70	58.18	69.44	70.31
Unskilled worker, full time	0.20	0.27	0.29	0.29	0.23	0.25	0.27	0.28
Skilled worker, full time	0.06	0.04	0.03	0.03	0.05	0.06	0.03	0.04
White collar, full time	0.73	0.66	0.68	0.68	0.71	0.67	0.69	0.68
Mean standardized bias			3.87		6.80		3.03	
No. of obs.	566	233,358	196		449		233	

Source: Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233.358 female colleagues (full-time employed in t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. 1 DM (German mark) equals 0.51 euro.

To get an idea of the quality of the three matching procedures, we apply different tests. One way to evaluate the performance of the match is to calculate the mean standardized bias (MSB) among the covariates of mothers-to-be and non-mothers 22 months before birth (that is, at time t_{-1}). This measure is given by the absolute difference in means divided by the square root of the average of the two associated variances and multiplied by 100. Taking the average of all variables yields an indicator that is decreasing with the match quality. The standardized bias can be interpreted as bias in percent of the average standard deviation. Since there is no fixed threshold saying whether the applied matching procedure is doing well or not, we use this indicator to compare the three alternatives amongst themselves (see e.g. Lechner, 2002). To generate a comparable measure for all three matching algorithms, we refer to all variables used to model the probability of treatment at time t_{-1} (see Table 2). The MSB amounts to 3.03 for MATCH 3, which represents the relatively best performance. MATCH 2 performs much worse with MSB of 6.80, while MATCH 1 reduces the MSB to 3.87.

3.2 Wage effects

As can be seen in Table 4, the average wages of the mother samples and the respective control samples differ quite remarkably between the raw data and the selected individuals after matching according to the three matching procedures. The wage difference before treatment even changes size: in the raw data set the control group receives a higher average wage rate than the mothers-to-be whereas in MATCH 1 wages hardly differ and in MATCH 2 and MATCH 3 the higher wage earners are found among mothers.

Table 4: Daily wages before and after treatment (in DM)

	Raw data		MATCH 1		MATCH 2 (NN, caliper 0.5)		MATCH 3 (NN, caliper 0.5)	
	Before	after	before	after	before	After	before	after
Control group	147.61	167.78	126.64	141.91	117.41	136.72	125.31	143.90
Mothers	120.72	107.92	128.38	118.78	126.76	115.19	129.33	120.23
Wage difference (ATT)		-59.85		-23.13		-21.52		-23.68
Conditional DiD		-32.96		-24.87		-30.87		-27.69
# mothers		566		196		449		233
# control persons		233,358		196		449		233

Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233,358 female colleagues (full-time employed in t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. 1 DM (German mark) equals 0.51 euro.

A look at the average wage rates of mothers and their corresponding control colleagues, one year after re-entry into the job, indicates that the post-treatment outcome of mothers is substantially lower compared to their controls. While mothers' daily wage rates fall by 13 DM between t_{-1} and t_{+3}, the potential control colleagues' wage rates increase by 20 DM, which equals about 28 percent (see also Figure 1). The unmatched wage difference between mothers and controls in t_{+3} amounts to almost 60 DM (30.50 €) which is about 35%. The absolute values of the ATT differ slightly by the matching procedure applied. The matched gap is 23 DM (MATCH 1), respectively 21.50 DM (MATCH 2) or almost 24 DM (MATCH 3), which translates into an average wage cut of about 16-17 percent with respect to the control colleague's wage. Even if the ATTs do not vary that much across the specific matching algorithms, the results provide some indication that MATCH 1 and MATCH 3 result in higher wage effects of child-related employment breaks.

The measures presented so far provide a consistent estimate of the ATT if selection into motherhood is based on observable characteristics only. Given that the propensity score matching in MATCH 2 does not fully eliminate the differences between mothers with employment breaks and continuously employed women, we calculate the conditional difference-in-difference estimate (cDiD). This before-after comparison of the wage differences between mothers and their controls takes account of a sorting process due also to unobservable characteristics or characteristics poorly measured by our variables at hand. Since the pre-birth wage rates of the selected mothers are higher than those of their counterfactual female colleagues – this is especially true for MATCH 2 – the cDiD wage effects are constantly larger than the ATTs. The cDiD ranges between 25 DM (MATCH 1) and 31 DM (MATCH 2) and thus suggests a negative impact of unobservables on the relative after-break wages of mothers. However, this measure only provides a more accurate estimate of the treatment effect of entering motherhood than the basic ATT if the unobservables are time-invariant and linear. In the case when young women, prior to having their first child, are particularly ambitious in reaching a certain position within the firm or when women, after just having been promoted, decide to enter motherhood, the assumption of time-invariant unobservables might not be the most appropriate. In the lack of full information on this complex field of fertility choice (and the partner matching process) we prefer to interpret both measures, the basic ATT and the cDiD as providing us with an interval of the possible treatment effect of entering motherhood and experiencing an employment break. Hence, we conclude that the treatment effect lies somewhere between 16.5 and 19 percent. After all, this appears to be a comparatively robust measure, no matter whether we draw on pure propensity score matching, which might have the drawback of not fully capturing unobserved heterogeneity between treatment and control group, or on exact matching, which suffers from a very small sample size, or on the combined procedure of exact and propensity score matching.

In contrast to most other studies measuring the wage effect of entering motherhood, our data allow us to accommodate firm-specific fixed effects. This aspect may be important if firms differ with respect to their average individual wage growth. If, for example, women with a high likelihood of becoming mother select into booming firms, which offer a variety of career ladders and whose jobs are regarded as stepping-stones, ignoring firm-specific heterogeneity would tend to overestimate the true backlog in wages. In contrast, the expected wage loss of entering motherhood is underestimated if mothers-to-be select into firms whose employees have rather stable wage rates. To test these hypotheses, we calculated the ATT based on MATCH 2 ignoring firm specific fixed effects, that is, we match across all potential control women and not only within the same firm (see Table 5).[13] The ATT significantly increases in this specification. That is, compared to all female employees across firms, mothers loose almost 44.50 DM per day, which amounts to a wage drop of about 29 %. The cDiD also increases from 31 to 39 DM. This leads one to suppose that women who are to get children are more likely to work in firms with lower wage growth rates.

Table 5: Wage effects within and across firms (in DM)

	MATCH 2 within firms (NN, caliper 0.5)		MATCH 2 across firms (NN, caliper 0.5)	
	before	after	before	after
Control group	117.41	136.72	126.25	152.40
Mothers	126.76	115.19	120.72	107.92
Wage difference (ATT)		-21.52		-44.47
Conditional DiD		-30.87		-38.95
# mothers		449		566
# control persons		449		566

Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233.358 female colleagues (full-time employed in t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. 1 DM (German mark) equals 0.51 euro.

Until now, we considered only average effects across all women who became mother between 1987 and 1995. One major reason for differences in the child-related wage cut is the amount of time spent out of the labor market (see e.g. Figure 2). A simple way to see how the duration of an employment break is related to a mother's wage cut is to run a linear regression, where the wage differences after matching are conditioned on the time out of work:

$$w_{t+4}^{c} - w_{t+4}^{t} = \alpha + \beta \cdot break^{t} + \varepsilon.$$

The term $w_{t+4}^{c} - w_{t+4}^{t}$ denotes the differences in wage rates between mothers and their matched colleagues, where t refers to the treated mothers and c to the selected control observations. $break^{t}$ represents the duration of the employment break (in years) of the mother. α and β are parameters to be estimated and ε is the error term. In principle, this procedure allows us to calculate group-specific treatment effects for mothers with

[13] The distribution of the propensity score values of mothers and their potential controls are presented in the appendix. This graph does, however, only refer to the matching across all potential control women. The support problem in the case of within-firm propensity matching is addressed by applying a caliper of 0.5.

employment breaks of different length. The estimation result indicates that every year out of work reduces the daily wage rate by almost 6 DM (β equals 5.89 with a t-value of 2.24).[14] Alternative specifications, allowing a non-linear relationship between the wage loss and the duration of the employment break (e.g. by using dummy variables for different break durations or adding a squared term), seem to suggest that the marginal wage effect does not change with duration. That is, the first year out of work is more or less as "expensive" as the fifth year.

3.3 Sensitivity analyses

It may be of concern that our strict matching criterions in MATCH 1 and MATCH 3 lead to a considerable reduction of the number of observations, associated with smaller estimated wage losses due to the better matching. Applying a caliper in the NNM also yields better matches on the one hand, but reduces the number of observation on the other. In order to see, how the effects change if we allow more heterogeneity among mothers and their colleagues, we now present the results on alternative matching specifications. Furthermore, we will check the sensitivity of the propensity score matching with respect to the matching algorithm, that is, we also present kernel matches in MATCH 2 and MATCH 3. To make sure that only women within the same firm are selected as appropriate matches, we choose an Epanechnikov kernel.

Table 6: Results of alternative matching specifications

	MATCH 1 (less strict) (1)	MATCH 2 (kernel) (2)	MATCH 2 (NN no caliper) (3)	MATCH 3 (less strict) (4)	MATCH 3 (kernel) (5)	MATCH 3 (NN no caliper) (6)
ATT	- 23.48	- 22.69	- 28,12	- 22.54	-24.10	- 36.96
cDiD	- 25.05	- 31.03	- 30.85	- 29.02	- 29.15	- 30.22
# mothers	266	449	512	361	236	512

Source: Total sample of 566 mothers (child birth between 1987 and 1995) and 233,358 female colleagues (full-time employed in t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. 1 DM (German mark) equals 0.51 euro.

The first column of Table 6 gives the results of the exact matching algorithm (MATCH 1) with less strict matching criterions. In contrast to our initial model, we do not balance the samples of mothers and selected control observations with respect to employment experience. Compared to the stricter exact matching presented in Table 4, the number of observations increases by 35% but the results hardly change.

[14] Alternatively, we used the cDiD wage gap between mothers and their colleagues. The impact of the employment break on the wage rate is less obvious in this specification (β equals 2.57 with a t-value of

The second column describes the wage effect based on MATCH 2 with Epanechnikov kernel and a bandwidth of 0.5. The results indicate that our baseline model of MATCH 2 (NNM with caliper 0.5) and the kernel matching yield very similar effects. This implies that the trade-off between bias and variance does not seem to be that severe in our case. If, however, the nearest neighbor is not restricted to be within a certain range (see column 3), all mothers working in a firm with at least one female colleague are taken into account and the ATT respective cDiD increases by almost 7 respective 6 DM. These rather large differences compared to the baseline model of MATCH 2 become plausible if we think about a mother-to-be in a small firm with only one female colleague. If their propensities to become a mother within the next 22 months differ noticeable, we would also expect that differences in their observed and unobserved characteristics yield significant wage differentials. Applying a caliper means skipping these observations and hence reduces the resulting wage differential.

Finally we compare different specifications of MATCH 3. First, we relax the criterions of the exact pre-matching; more precisely, we skip the experience criterion and select all colleagues who are in the same occupation and work the same number of hours as those women who enter motherhood (see column 4). The number of observations increases by about 55 %. Whereas the ATT becomes smaller by more than 1 DM, the cDiD increases by about the same amount. This contradictory result leads one to suggest that the less strict matching procedure is not able to fully account for the heterogeneity among mothers and controls. Hence, interpreting the cDiD seems to be more appropriate. Even though, the difference between the strict pre-matching and these results are not tremendous. Second, we apply an Epanechnikov kernel with bandwidth 0.5 (see column 5). Compared to our baseline model of MATCH 3 (see Table 4), the wage gap due to a child-related employment break increases by 0.4 DM (ATT) respectively 1.40 DM (cDiD). Hence, the differences can be regarded as small. As in the case of MATCH 2, abandoning the caliper increases the estimated wage effects dramatically (see column 6 of Table 6).

Based on the variety of sensitivity analyses, we conclude that applying a rather large caliper of 0.5 reduces both, the number of observations as well as the estimated wage effect. Alternative specifications with respect to the matching algorithm or the pre-matching criterions, however, do not affect the results. We therefore conclude that our model specifications presented in Table 4 – aiming at reducing the heterogeneity between women entering motherhood and their female colleagues – do not really suffer from small sample sizes. Relaxing the matching criterion to a reasonable degree does in fact increase the number

1.72).

of observations, but does not change the general result that women entering motherhood have a 25 to 30 DM lower daily wage one year after they returned to their job.

4 Conclusions

In this study we examine the backlog of mothers' wage rates due to the birth of their first child using a novel within-firm-semiparametric approach based on matching. With data of the IAB employment sample and additional administrative data assembled at the state pension authorities (IAB employment supplement sample I), we can identify all women working in the same firm. Hence, we match each female employee who experienced a child-related employment break with a female colleague of the same firm without a comparable employment break. Due to within-firm matching, unobserved firm-specific heterogeneity can be fully taken into account. Selection in observable and unobservable individual characteristics is accommodated by different matching and estimation algorithms.

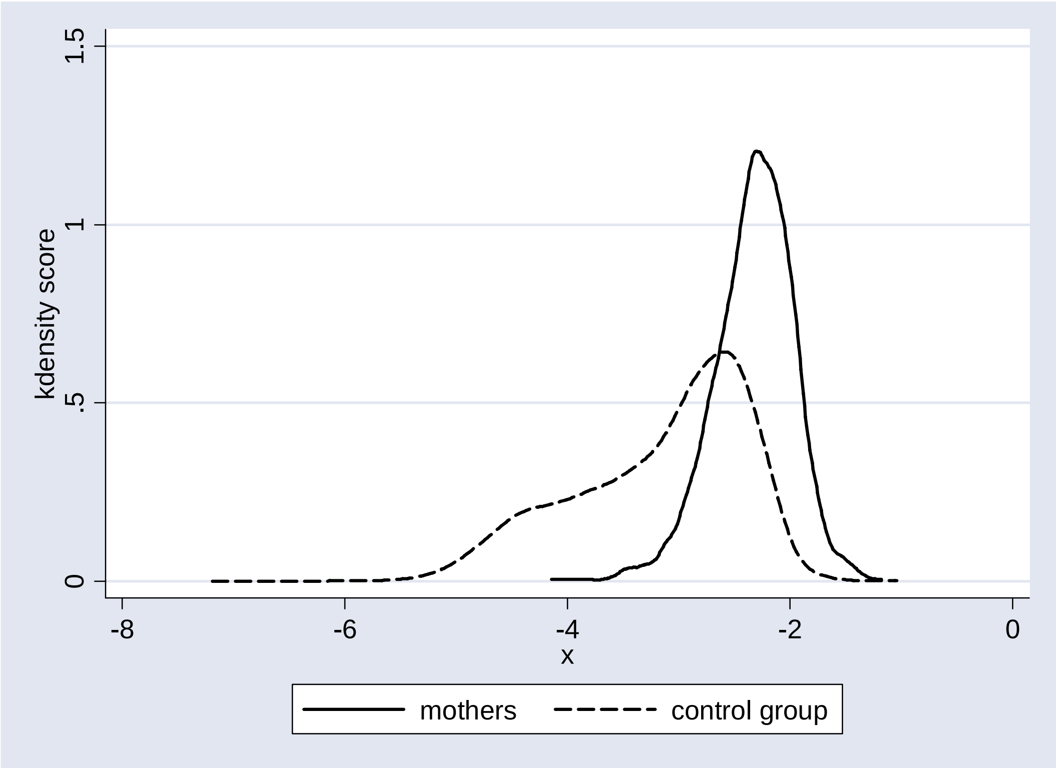
Our findings point to a substantial wage cut of mothers upon return to their job. Even when confining the comparison to women returning to full-time employment, mothers' wages are by 16.5 to 19 percent lower relative to those of their female colleagues with comparable characteristics 22 months before entering motherhood. We interpret our results as a lower bound to the overall short-run wage effects of entering motherhood for two reasons. First, our treated population is very selective in terms of labor market attachment because we only consider women who worked before the birth of their first child. Second, they have to return to a full-time position within our observation period. Note that most of the wage loss of women entering motherhood results from the fact that their female colleagues experience significant wage growth in the meanwhile. But even in absolute terms, the average real wage after a maternity break is slightly lower than before entering motherhood.

Interestingly, the pre-treatment wages of mothers-to-be are equal or higher than their control groups' once we apply our matching procedures. This finding hints at a negative selection into motherhood based on observable characteristics and a positive selection based on unobservable characteristics with respect to the wage level. As a result, the before-after comparison of the wage rates of mothers and non-mothers yield even larger wage cuts due to motherhood and a career interruption. Our firm-specific information provides further insight into the sorting of women into firms. Since the ATT is significantly higher as soon as we ignore firm-specific fixed effects, we conclude that women who plan to get children are more likely to work in firms with lower wage growth rates, be it because of anticipative sorting into these firms or sorting into motherhood.

Appendix

Figure A1 illustrates the predicted linear index from the propensity score estimation for the sample of mothers (black line) and the sample of all possible control persons (dashed line). The likelihoods of entering motherhood and taking parental leave of the group of mothers and the potential control group do not overlap over the whole range of values. However, since the propensity scores of future mothers are distributed quite narrowly, they are covered by the scores taken by the control colleagues for the major part of the distribution.

Figure A1: Kernel densities of the propensity scores of all mothers and possible controls



Note: Propensity scores based on the estimation results presented in Table 1.

Source: Sample of 1,390 mothers (child birth between 1987 and 1995) and 307,541 female colleagues, drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register.

References

Abowd, J. M., Kramarz, D. und D. N. Margolis (1999): High Wage Workers and High Wage Firms, *Econometrica*, Vol. 67(2), 251-333.

Beblo, M. and E. Wolf (2002a): How Much Does a Year off Cost? Estimating the Wage Effects of Employment Breaks and Part-Time Spells, *Cahiers Écononomiques de Bruxelles* 45(2), 191-217.

Beblo, M. and E. Wolf (2002b): *Wage Penalties for Career Interruptions: An Empirical Analysis for West Germany*, ZEW-Discussion Paper No. 02-45, Mannheim.

Beckmann, P. und B. Kurtz (2001), Erwerbstätigkeit von Frauen: Die Betreuung der Kinder ist der Schlüssel, IAB-Kurzbericht 10, Nürnberg.

Bender, S., A. Kohlmann and S. Lang. (2003): *Women, Work, and Motherhood: Changing Employment Penalties for Motherhood in West Germany after 1945 – A Comparative Analysis of Cohorts Born in 1934-1971*, MPIDR Working Paper WP 2003-006.

Bender, S., A. Haas and C. Klose (2000): *The IAB Employment Subsample 1975-1995. Opportunities for Analysis Provided by the Anonymised Subsample*, IZA Discussion Paper No. 177, Bonn.

Bergemann, A., B. Fitzenberger and S. Speckesser (2003): *Evaluating the Employment Effects of Public sector Sponsored Training in East Germany: Conditional Difference-in-Difference and Ashenfelter's Dip*, Working Paper, Version August 2001, University of Mannheim.

Budig, M. and P. England (2001): The Wage Penalty for Motherhood, *American Sociological Review* 66, 204-225.

Bronars, S. G. und M. Famulari (1997): Wage, Tenure, and Wage Growth Variation Within and Across Establishments, *Journal of Labor Economics* , Vol. 15(2), 285 – 317.

Davis, S. J. und J. Haltiwanger (1991*):* Wage Dispersion Between and Within US Manufacturing Plants, 1963 – 1986, *Brookings papers on economic activity*, 115 – 180.

Ejrnaes, M. and A. Kunze (2004): *Wage Dips and Drops Around First Birth*, Working Paper 2004-01, Centre for Applied Microeconometrics, University of Copenhagen.

Fitzenberger, B. and H. Prey (1999): Berufliche Weiterbildung und die Stabilität der Beschäftigung: Eine Evaluation auf der Basis des Sozio-ökonomischen Panels, in: Beer, D., B. Frick, R. Neubäumer and W. Sesselmeier (eds.): Die wirtschaftlichen Folgen von Aus- und Weiterbildung, München/Mering, 139-164.

Gartner, H. (2005): *The imputation of wages above the contribution limit with the German IAB employment sample*. FDZ Methodenreport Nr. 02/2005, Nürnberg.

Heckman, J., Ichimura, H. and P. Todd (1998): Matching as an Econometric Evaluation Estimator, *Review of Economic Studies* 65 (2), 261-294.

Heckman, J., LaLonde, R. J. and J. A. Smith (1999): The Economics and Econometrics of Active Labor Market Programs, in: Ashenfelter, D. and D. Card (eds.): *Handbook of Labor Economics*, Amsterdam, 1865-2097.

Imbens, G.W. (2004): Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review, mimeo.

Joshi, H., P. Paci and J. Waldfogel (1999): The Wages of Motherhood: Better or Worse?, *Cambridge Journal of Economics* 23, 543-564.

Kunze, A. (2002): *The Timing of Working Career and Depreciation of Human Capital*, IZA Discussion Paper No. 509.

Lechner, M. (2002): Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies, *The Review of Economics and Statistics* 84(2), 205-220.

Lundberg, S. and E. Rose (2000): Parenthood and the Earnings of Married Men and Women, *Labour Economics* 7 (6), 689-710.

Ondrich, J., Spiess, C.K. and Q. Yang (2001): *The Effect of Maternity Leave on Women's Pay in Germany 1984-1994*, Working Paper, German Institute for Economic Research (DIW), Berlin.

Polachek, S. W. (1981): Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure, *Review of Economics and Statistics* 63, 60-69.

Prinz, K. (1997): Versicherungsverläufe von Frauen und Männern der Geburtsjahrgänge 1931 bis 1960. *Deutsche Rentenversicherung* 3-4, 220-241.

Rubin, D. B. (1974): Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology*, 66 (5), 688-701.

Rosenbaum, P. R.and D. B. Rubin (1983): The central role of the propensity score in observational studies for causal effects, *Biometrica* 70, 41-55.

Simonsen, M. and L. Skipper (2006): *An Empirical Assessment of Effects of Parenthood on Wages*. Accepted for publication in D. Millimet, J.A. Smith & E. J. Vytlacil (eds), Advances in Econometrics, vol. 21.

Simonsen, M. and L. Skipper (2005): *Estimating the Costs of Motherhood: An Analysis Using Matching Estimators*. Stanford Institute for Economic Policy Research (SIEPR), forthcoming in *Journal of Applied Econometrics*.

Waldfogel, J. (1998a): Understanding the „Family Gap“ in Pay for Women with Children, *Journal of Economic Perspectives* 12(1), 137-156.

Waldfogel, J. (1998b): The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference?, *Journal of Labor Economics* 16(3), 505-545.