The wage effects of entering motherhood

A within-firm matching approach

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

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

August 2007

Abstract
We analyze the wage effects following employment breaks of women who enter motherhood using a novel matching approach where mothers' wages upon return to the job are compared to those of their female colleagues within the same firm. Using an administrative German data set, we apply a fixed-effects propensity score matching based on information two years before birth. Our results yield new insights into the nature of the wage penalty associated with motherhood: when matching with firm fixed effects we find first births to reduce women's wages by 19 percent, whereas ignoring the firm identifier and matching across all firms would yield a wage cut of 26 percent. A subsequent regression analysis confirms that the wage loss increases with the duration of the employment break.

Keywords
wages, family pay gap, 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.

1 Introduction

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 should be observed between mothers-to-be and women who are not going to have a child even before 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 semi-parametric 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 (the "monozygotic twin colleague"). This paper provides a robust measure of the wage backlog caused by child birth and parental leave based on sensitivity analyses with various matching procedures.

We find first births to reduce women's wages by about 19 percent, when applying a within-firm matching. Ignoring the firm-specific effect and matching across all firms (to a "dizygotic twin colleague"), yields a larger wage cut of 26 percent. Concluding from this result, selection into firms is an important explanatory factor of the family pay gap since women with children are more likely to be found in firms with lower wage growth. However, selection does not explain the whole gap. Even compared to their immediate (monozygotic) firm colleagues are mothers' wages negatively affected upon return to employment. As expected, the wage loss increases with the duration of the employment break as we can show in a subsequent regression analysis.

2 What's new?

The wage penalty or 'family pay gap' has been investigated mostly for the United States (see evidence by Budig and England 2001, Lundberg and Rose 2000, Waldfogel 1998a) and for the United Kingdom (Joshi, Paci and Waldfogel 1999, Waldfogel, 1998b). Studies for Germany show that the wage penalty of motherhood is substantially high (see estimations by

Beblo and Wolf 2002a and 2002b, Ejrnaes and Kunze 2004, Kunze 2002 and Ondrich, Spiess and Yang 2001[1], Schönberg and Ludsteck 2007). At the same time, 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 the formal leave period (Beckman and Kurtz 2001).

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.

Most 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 semi-parametric approach based on matching to determine the wage backlogs of mothers relative to comparable non-mothers within the same firm. This study hence provides two innovations in the field of family pay gaps: (1) we use firm-specific effects to account for differences in the way how firms integrate and promote mothers returning to their jobs and (2) we use a semi-parametric estimation method which imposes no restrictions on the functional form of the relationship between child-related employment breaks and wages.

[1] Schönberg and Ludsteck (2007) are finding also a negative effect, but it is offset by a positive selection effect, resulting in a zero or even positive overall effect.

The challenge of our research question is to determine, which level the wage of a mother would have if she had not given birth and experienced an employment break within a specific observation period. Since this counterfactual outcome cannot be observed, 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 a combined procedure of exact and propensity score matching to determine a useful control group. The exact matching consists of finding a similar colleague, based on the estimated propensity scores, within the same firm. Let us call this match a monozygotic twin colleague (as opposed to a dizygotic twin colleague found by matching across firms). As a result, we compare women who give birth to their first child (treatment group "mothers") and women who do not give birth during the observation period (control group "non-mothers"), but are continuously employed within the same firm as the mothers to accommodate segregation and unobserved firm-specific effects. By propensity score matching based on information two years before birth (including wage level and wage growth), we take into account that mothers-to-be may be less attached to the labor market even before having a child 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.[2]

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 12 months after the end of the break. These dates are determined dynamically 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) working in the same firm and on the same days. The mean difference in wages reflects the average treatment-on-the-treated effect of entering motherhood and experiencing a specific employment break after returning to their former employer. As the wage effect is likely to differ across women due to

[2] To our knowledge, the only study that exploits the econometric methodology of matching to analyze the wages of mothers is provided by Simonsen and Skipper (2005, 2006). In contrast to our approach, they can not assign female employees to their firms and thus only compare women across firms.

heterogeneity in the duration of the employment interruption and other individual characteristics. In a subsequent regression analysis we therefore investigate the differences in the wage losses using the duration of the employment break as explanatory variable.

It is obvious, that this approach places high demands on the data needed. We hence base our analysis on a data set of process generated information on all employees in Germany covered by social security, which is provided by the Institute for Employment Research (IAB).

The remainder of the paper is structured as follows. Section 3 presents our methodological approach. The data is described in Section 4. Section 5 discusses the results of the matching procedure and the second-step wage gap analysis. The last section concludes and discusses potential extensions of our approach.

3 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 (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 useful 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, if, e.g., these women had been promoted otherwise, their wage scales are tenure based or 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.

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). This implies that:

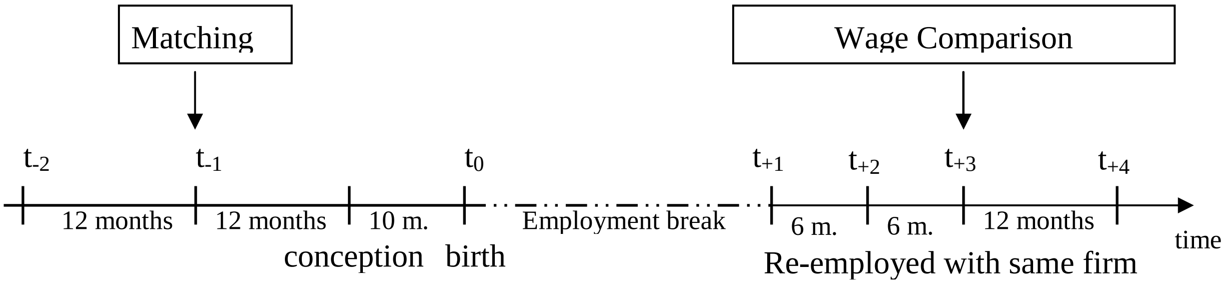
$$E(Y_0 \mid D=1, X) = E(Y_0 \mid D=0, X)$$

For the CIA to hold, the set of variables (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 algorithms. Given our firm-specific information, we will apply 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 (see Ejrnaes and Kunze 2004; analogously to Ashenfelter’s dip in labor market policy evaluation, see e.g. Bergemann, Fitzenberger, Speckesser 2003), we define our first observation point 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[3]: 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.[4]

Figure 1: Time frame of our evaluation approach



3.1 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 a feasible alternative, a combination of exact and propensity score matching, to determine a useful control group.

Exact matching compares people with exactly the same values of observed characteristics X . Since this method works only with a limited number of discrete X -variables or, alternatively,

---

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

4 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. If we assume that firm mobility is positively correlated with the expected wage backlog, our estimates will provide a lower bound for the true wage effect of motherhood. .

value ranges for continuous X-variables, the choice of the relevant X-variables is crucial, 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. Propensity score matching reduces this dimension problem by defining a distance metric on X. 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: $P(X) = \Pr(D=1|X)$.[5]

In our combined matching procedure, we first estimate a parametric probit model to predict the individual propensity score $P(X)$. Exact matching within the firm, based on $P(X)$, then ensures that treated and untreated women underlie the same unobserved fixed effect influencing the wage determination process within the establishment. In our setting, $P(X)$ 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. Controlling for education and occupation is meant to account for unobserved individual heterogeneity affecting the occupational choice. 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. We further account for differences in hourly wage rates 22 months before.

We apply nearest neighbor matching (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. Sensitivity analyses with alternative matching algorithms are discussed in Section 5.3.

[5] 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.

3.2 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 wages right upon return as well as 12 months after the end of 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_{+2}) 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.[6] 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.

4 Data

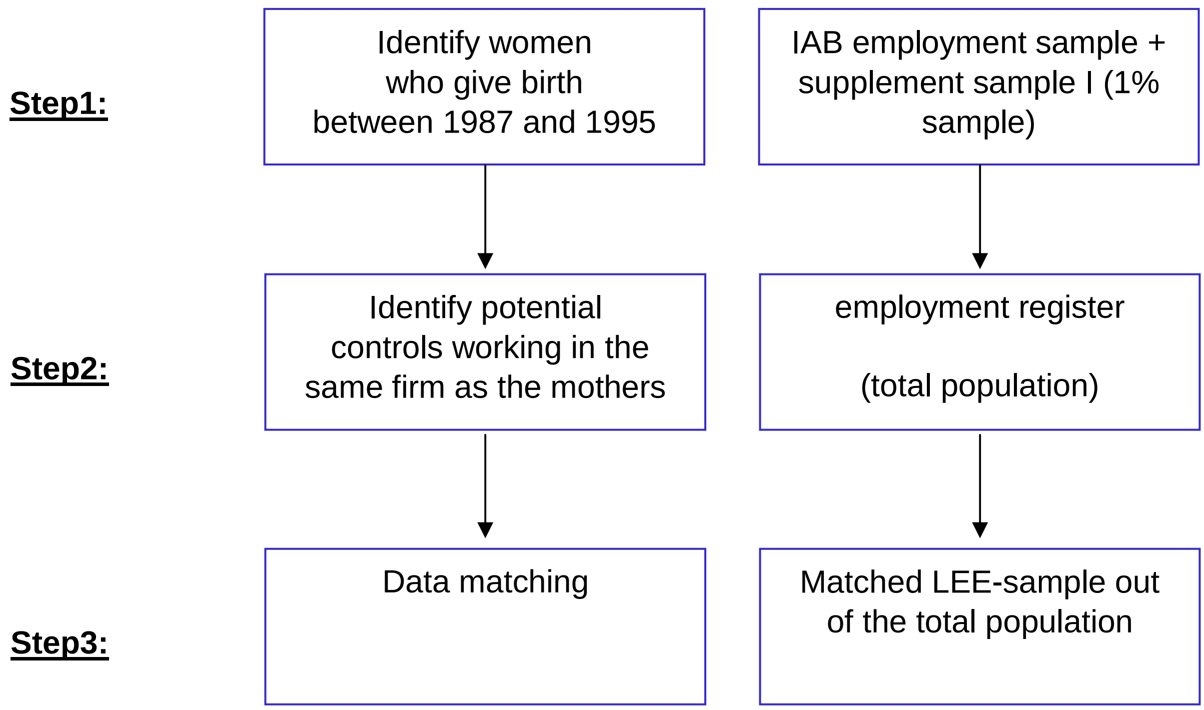
The merit of our empirical analysis is nourished by the combination of several data sets that allow longitudinal comparisons between mothers and non-mothers within the same firm. We draw on process generated data provided by the Institute for Employment Research (IAB). These German register data are 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

[6] We chose t_{+2} to compare wage rates after the employment break because – due to administrative reasons – the identification of part-time employment is more reliable when the calendar year has changed.

employees are not covered by the data: (i) (temporary) civil servants or self-employed persons and (ii) women who are employed in East Germany or abroad.

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.[7]

Figure 2: Sampling procedure



We use two different samples of these register data (see Figure 2). In Step 1 we combine the IAB employment sample with additional administrative data assembled at the state pension authorities (IAB employment supplement sample I).[8] 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,

[7] 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.

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.[9]

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 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 until 1999. After deleting observations with missing values, we remain with 1,390 observations of mothers.

As described in Section 3.2 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 are registered in the social security system (Step 2). The following procedure describes our strategy to identify all female colleagues of our treatment group:

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} in the identified firms and did not experience an employment break within this period.

After this selection and matching of the two groups in Step 3, the data set consists of 307,541 observations of potential control women.[10] Due to missing observations of selected variables,

---

8 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).

9 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. Schönberg and Ludsteck (2007) are using the supplementary file, too to impute times of parental leave into the total population.

10 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

1,357 mothers and 298,822 female colleagues enter the propensity score estimation. For the purpose of meaningful wage comparison, we further restrict our sample to women in full-time employment one year after the mothers' return to the job (in t_{+2}) because we do not have information about the number of working hours in part time jobs.[11] We are left with 561 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. Note also that the control group may contain mothers, under the condition that they delivered their children before 1997 and were continuously employed during our observation period.

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 1). While 9.5 % of all treated observations have no female colleague – and hence have to be ignored – and 7.8 % mothers have just one control observation, there is one case where we are able to identify 7159 potential control observations for one specific mother. According to Table 1 only 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.

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.

[11] One advantage of this selection is, that we are sure to have "additional" payments like the Christmas bonus in the wages of the mothers and the comparison group. So we do not have selectivity on the wages for the two groups.

Table 1: Distribution of control persons

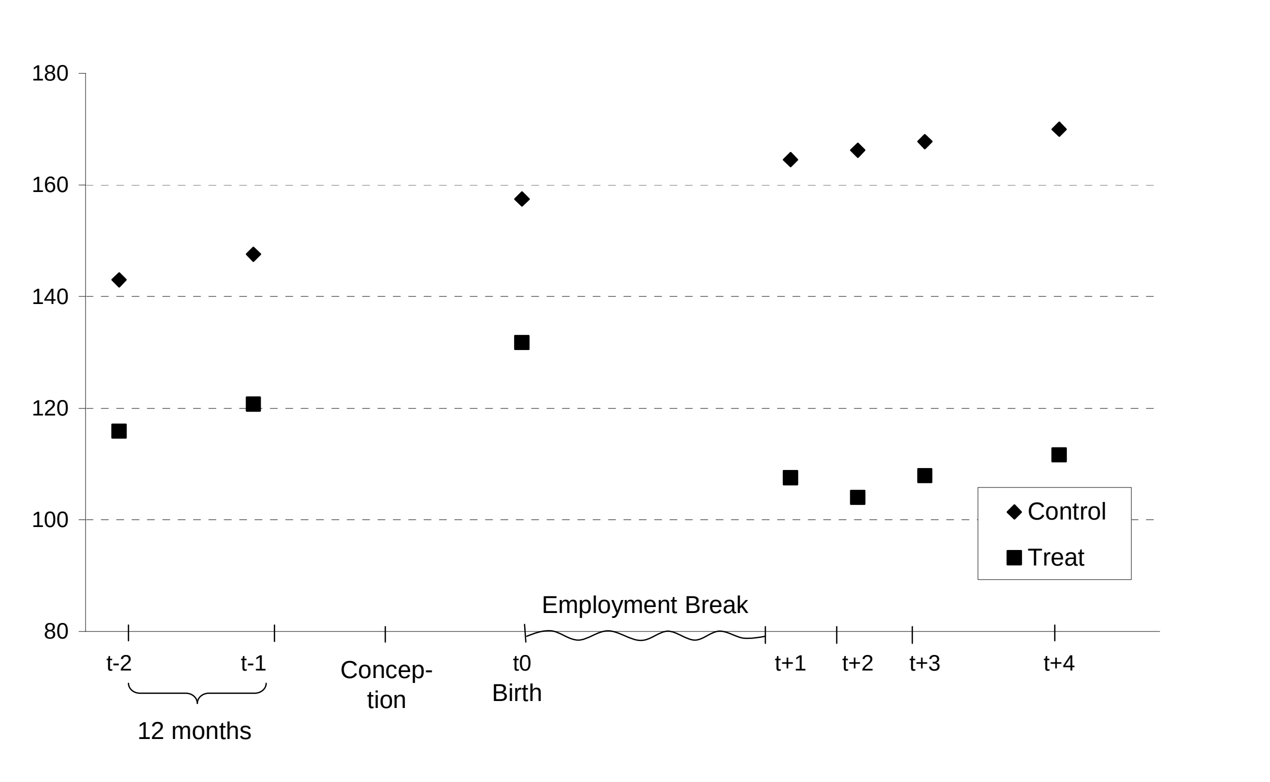
Control observations	Percent of mothers with ... control observations
0	9.54
1	7.77
2	4.95
3	3.53
4	2.83
5	1.41
6	1.41
7	1.24
8	1.06
9	1.77
10	1.41
>10	63.07

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.

Figure 3 illustrates the average wage rates of the selected mothers and female colleagues before and after the mothers' employment break.[12] 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 hardly reach the level from 22 months before birth (t_{-1}) even two years after the end of the employment break (t_{+3}) .

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

Figure 3: Average wages of mothers (Treat) and their female colleagues (Control) before and after birth

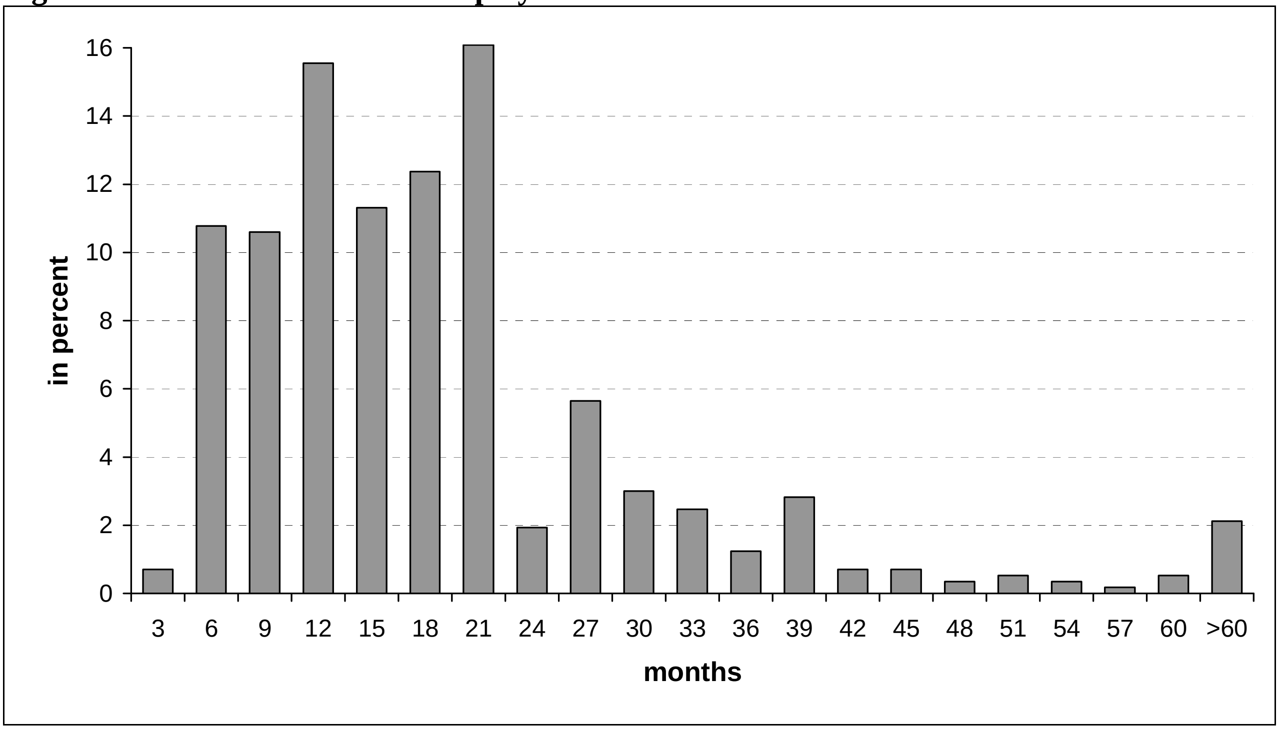


Source: Sample of 566 mothers (child birth between 1987 and 1995) and 233,358 female colleagues, drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. Wages are in DM (German marks). 1 DM equals 0.51 euro.

Figure 4 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 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.

These relatively short leaves, for German standards, underline a sample selection of women who seem to be more job-oriented than the average. Furthermore, all mothers in the sample work full-time (within a year upon return), which bears another source of selection towards women with a mix of generally higher-paying characteristics. When drawing conclusions from the estimated wage effects of motherhood we will consider these sources of systematic sample selection.

Figure 4: 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.

5 Matching results

5.1 Propensity score estimation

In Table 2, 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 in several splines, most of which are statistically significant. Interestingly, college and university graduates have a higher probability of belong to the treatment group than other skilled employees. This result seems plausible once we take into account that we consider only mothers who return to a 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.]

Table 2: Probit estimation results of having a birth at time t_0

	Coeff. estimate	Std. error
Age splines		
<25	.0317	.0237
25-28	**.0675**	.0185
28-31	**-.0447**	.0176
31-34	**-.1358**	.0220
34-37	**-.0859**	.0293
37-40	**-.2246**	.0505
>40	**-.6922**	.3056
Education level (ref. apprenticeship)		
No apprenticeship	-.0017	.0403
College/Univ. graduate	**.4304**	.0693
Wage rate splines		
<80	-.0028	.0079
80-120	.0002	.0004
Tenure splines (in months)		
<20	.0038	.0044
20-40	-.0011	.0035
40-60	-.0092	.0037
60-80	.0146	.0037
80-100	-.0013	.0039
100-120	.0088	.0039
>120	.0089	.0016
Female employees splines (in N of empl./100)		
<0.1	**-18.3304**	1.5021
0.1-0.2	-1.4956	1.0733
0.2-0.5	**-1.8573**	.3028
0.5-1	**-.5236**	.1517
1-2	**-.2933**	.0695
2-5	**-.1209**	.0234
5-10	**-.0479**	.0161
10-20	-.0193	.0102
20-40	**-.0121**	.0056
>40	-.0043	.0046
Work experience in past 4 years	**-.0209**	.0032
Employment breaks (in month) >93 days	**-.0305**	.0043
No. of employment breaks (>31 days)	**-.1822**	.0377
Average wage growth in past 4 years	-.0128	.0371
Constant	1.0591	.7411

Pseudo R squared	.3973
$\chi^2(24)$	
No. of observations	247,081

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. Other control variables: occupational groups.

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 (measured in splines) 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.

Table 3 compares the mean characteristics of the selected mothers and their potential and effective control colleagues before treatment, that is, at time t_{-1}, 22 months before birth. Evidently, the matching algorithm contributes to a balancing of the samples with respect to the relevant variables. While the before-birth wage rate of mothers is substantially lower than the average wage rate of their colleagues who do not interrupt their career due to child bearing (potential controls), the selected control observations within the same firm even seem to earn slightly lower wage rates than the selected mothers 22 months before the date of birth. [The fact that the potential control group exhibits even more and longer employment breaks within the past five years than the mothers-to-be, leads one to suppose that part of the female colleagues have 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 6.]

Table 3: Descriptive statistics for mothers and control groups

	Raw data		Matched data (caliper = 0.5)	
	Mothers	Potential Controls	Mothers	Selected Controls
Characteristics (at time t-1)	Mean	Mean	Mean	Mean
Daily wage in DM	120.72	147.61	127.17	125.02
Average wage growth in past 4 years			.30	.27
Age	28.73	32.90	28.71	28.83
No apprenticeship	0.20	0.30	.24	.22
Apprenticeship	0.76	0.66	.72	.74
College/Univ. graduate	0.04	0.04	.03	.03
Work experience last 4 years	44.41	44.71	44.59	45.04
Past employment breaks (in months)	2.02	2.92	2.00	1.92
No. of past employment breaks	0.24	0.30	.21	.20
Tenure	60.92	70.19	62.82	62.50
Unskilled worker, full time	0.20	0.27	.24	.23
Skilled worker, full time	0.06	0.04	.04	.05
White collar, full time	0.73	0.66	.70	.70
No. of obs.	566	233,358	434	434

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.

5.2 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. In the raw data set the control group receives a higher average wage rate than the mothers-to-be whereas in the matched sample wage levels hardly differ.

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 11 DM between t_{-1} and t_{+3}, the potential control colleagues' wage rates increase by 19 DM on average. The unmatched wage difference between mothers and controls in t_{+3} amounts to more than 60 DM (30.50 €) which is about 30% of the controls' wage level. The matched gap is 28 DM, which translates into an average wage cut of about 19 percent with respect to the control colleague's wage.

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 our propensity score estimation but ignoring firm specific fixed effects. That is, we match across all potential control women[13] and not only within the same firm (see Table 5).[14] The ATT significantly increases in this specification. That is, compared to all female employees across firms, mothers loose almost 39 DM per day, which amounts to a wage drop of about 26 %.

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

	Raw data		Matching within firms		Matching across firms	
	Before	after	before	after	before	after
Control group	149.82	169.17	125.02	143.57	149.82	146.87
Mothers	121.14	108.11	127.17	115.58	121.14	108.11
ATT in German marks		-61.06		-27.99		-38.76
ATT in percent						
# mothers		561		434		561
# control persons		215,819		434		561

Source: Sample of 561 mothers (child birth between 1987 and 1995) and 215,819 female colleagues (full-time employed in t_{-2} and t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. 1 German mark equals 0.51 euro.

Until now, we considered only average effects across all women who become 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

---

13 We match out of the 298,822 non-mothers and not out of the whole universe of all non-mothers in Germany. So we are sure to have a comparable result, because the number of firms and the variation between the firms are constant for the two matching procedures.

14 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.

the wage differences after matching are conditioned on the time out of work. In principle, this procedure allows us to calculate group-specific treatment effects for mothers with different characteristics, e.g. employment breaks of different length. Table 5 presents the coefficient estimates of a linear regression of the individual wage differences on the mothers' characteristics. The main conclusion is that the longer the interruption the lower the relative wage on return. [Furthermore, age is positively correlated with the mother's wage rate, hence, negatively correlated with the wage backlog due to motherhood.] College und university graduates experience a significantly lower wage loss due to a child-related leave. The occupation dummies (not presented in the table) hint at significantly lower wage losses in jobs which require manual skills and lower qualification levels.

Table 6: OLS estimation results of the within-firm wage effects

(dep. var.: (control's wage – mother's wage/ control's wage))

	Coeff. estimate	Std. error
Duration of employment break		
0.5-1 years	**.2050**	.0798
1-2 years	**.2474**	.0746
2-3 years	**.2796**	.0930
> 3 years	**.3162**	.1038
Age		
22-<24	.0524	.0669
24-<26	-.0047	.0733
26-<28	-.1873	.0970
Education level (ref. apprenticeship)		
No apprenticeship	-.0245	.0619
College/Univ. graduate	**-.3703**	.1337
Constant	.1553	.0936
Pseudo R squared	.0845	
No. of observations	434	

Source: Sample of 434 mother-control pairs (child birth between 1987 and 1995) drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. Additional control variables include: occupational group dummies.

5.3 Sensitivity analyses

In order to see, how the wage effects change if we allow more heterogeneity among mothers and their colleagues, we now present the results on an exact matching as an alternative specifications. Furthermore, we check the sensitivity of the propensity score matching results with respect to different matching algorithms.

Given that the number of comparison observations is small for some mothers – namely those working in small establishments (see Table 1) – we first try kernel matching (KM) as an alternative matching algorithm. In our baseline matching we applied the most commonly used nearest neighbor matching 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 chose one neighbor, being aware that the variance may be high. To restrict the differences between the nearest neighbors – which tend to be larger in smaller firms – we defined 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 kernel matching (KM) with an Epanechnikov kernel to make sure that still only women within the same firm are selected as appropriate matches.[15] Considering that control 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 second column of Table 6 describes the wage effect based on our baseline matching algorithm with an Epanechnikov kernel and a bandwidth of 0.5. The results indicate that our baseline model (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 with no caliper applied), all mothers working in a firm with at least one female colleague are taken into account and the ATT increases by almost 4 German marks compared to the baseline matching. This rather large difference seems plausible if we think of a mother-to-be in a small firm with only one female colleague to be compared to. 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 reducing the resulting wage differential.

---

[15] 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.

For our sensitivity analysis with respect to the type of matching, we compare the results of an exact matching. Exact matches are selected 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. 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. Due to the curse of dimensionality, exact matching is not capable of providing an appropriate control in all dimensions and for all mothers. Therefore the number of observations in the matching sample reduces to 196.

Table 6: Results of alternative matching specifications

	Fixed-effects matching (kernel) (1)	Fixed-effects matching (NN no caliper) (2)	Fixed-effects matching (caliper=0.1) (3)	Exact matching (strict) (4)	Exact matching (loose) (5)
ATT in German marks	-26.76	-31.55	-27.81	-23.13	-23.48
ATT in percent	18.8	21.9	18.8	16.3	17.0
# mothers	434	507	336	196	266

Source: Total sample of 561 mothers (child birth between 1987 and 1995) and 215,819 female colleagues (full-time employed in t_{+3}), drawn from the IAB employment sample, IAB employment supplement sample I, Employment statistics register. 1 German mark equals 0.51 euro.

The fourth column of Table 6 gives the results of the strict exact matching algorithm and the fifth column those of a less strict matching specification. In the latter, we do not balance the samples of mothers and selected control observations with respect to employment experience. Compared to the stricter exact matching, the number of observations increases by 35% but the matching results hardly change.

Based on this variety of sensitivity analyses, we conclude that the treatment effect of entering motherhood and having an employment break lies somewhere between 16.5 and 22 percent - 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 small sample size. After all, our baseline matching model seems to provide a comparatively robust measure.

It may be of concern that our strict matching criterions in the exact matching 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. Hence, being stricter on the matching criteria (e.g. applying a larger caliper) tends to reduce both, the number of observations as well as the estimated wage effect. Alternative specifications, however, do not affect the results. We therefore conclude that our model specification does not really suffer from small sample size. Relaxing the matching criterion to a reasonable degree does in fact increase the number of observations, but does not change the general result that women entering motherhood have a 23 to 32 lower daily wage (in German marks) one year after they returned to their job.

6 Conclusions

In this study we examine the backlog of mothers' wage rates due to the birth of their first child using a novel semi-parametric approach based on matching with fixed firm effects. 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.

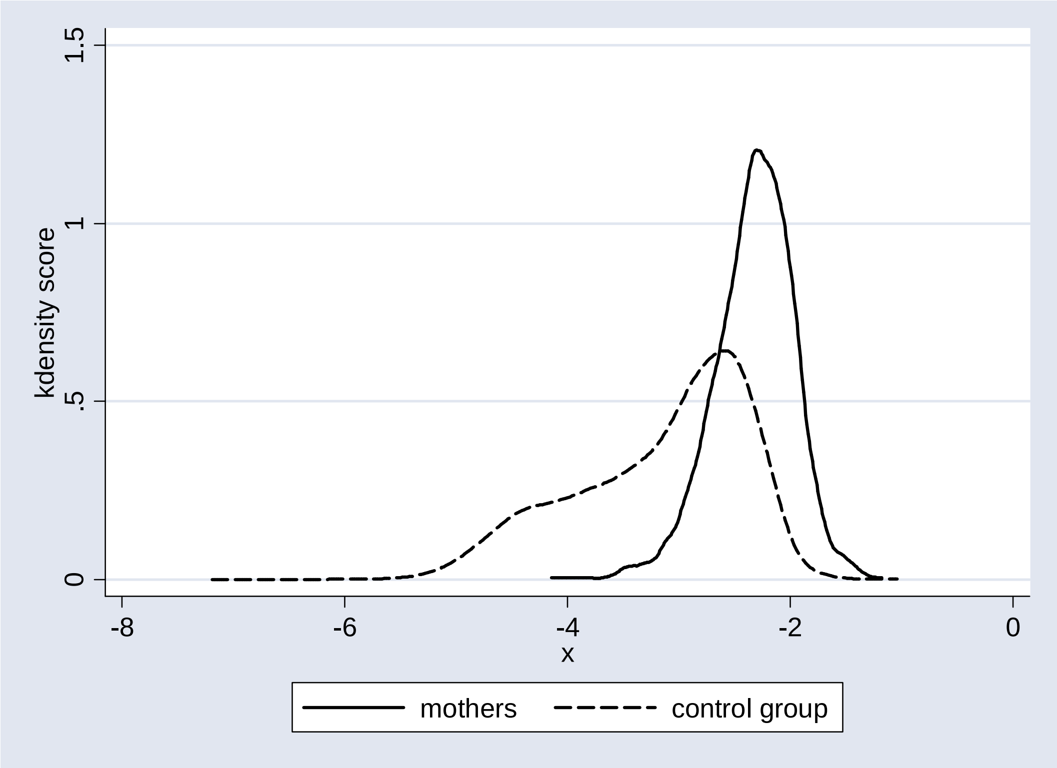
Our findings point to a substantial wage cut of mothers upon return to their job. Even when confining the comparison to women returning to full-time employment, mothers' wages are about 19 percent lower relative to those of their female colleagues with comparable characteristics 22 months before entering 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 with regard to its labor market attachment as we only consider women who worked prior to the birth of their first child and returned 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 that same woman's wage before entering motherhood.

Interestingly, the pre-treatment wages of mothers-to-be are equal or even slightly 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 yields 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.

Schönberg, U. and Ludsteck, J (2007): *Maternity Leave Legislation, Female Labor Supply, and the Family Wage Gap.* Department of Economics, University of Rochester.

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.