Punishing Potential Mothers? Evidence for Statistical Employer Discrimination From a Natural Experiment*

Jonas Jessen[†]
Robin Jessen[‡]
Jochen Kluve [§]

November 1, 2018

Abstract

Theory predicts that employers may discriminate statistically and pay female employees of child-bearing age lower wages than their male counterparts, and hire them at lower rates, ceteris paribus. This holds especially true if firms face direct costs of employees' motherhood. We use a natural experiment generated by a German policy reform to test these hypotheses: before January 1st, 2006, large firms (more than 30 employees) were obliged to pay for the generous maternity protection of their female employees; specifically, 14 weeks of 100 per cent wage continuation around the date of delivery, making the firms' costs a direct function of employees' gender and age. From 2006 onwards, each firm contributed to maternity protection by a flat-rate contribution to the statutory health insurance system, where the contribution depends only on the size of the workforce and is thus independent of gender and age. This had been the regulation for small firms already before the reform. We use comprehensive data from linked employer-employee administrative records to estimate the reform impact - i.e. the *switching-off* of the treatment - on female labour market outcomes, thus providing a measure of the extent of statistical employer discrimination pre-reform. The empirical analysis uses difference-in-differences and trend-break models to precisely pin down the size and mechanism of the reform effects. Our results indicate that the reform had a statistically significant treatment effect of 0.65 per cent on female workers' annual wages, which, as a measure of statistical employer discrimination, closely reflects the costs that firms faced under the pre-2006 regulation. We conclude that it is fruitful for policy makers to identify factors that could result in statistical discrimination, and that it is worthwhile for the public to finance, through taxes, costs that occur asymmetrically to mothers only, if the aim is to reduce gender-specific labour market differences.

Keywords Maternity leave · Statistical discrimination · Gender wage gap

JEL Classification J31 · J71

*We thank seminar participants at DIW Berlin and conference participants at EALE 2018 Lyon and COMPIE 2018 Berlin for valuable comments. The usual disclaimer applies.

[†]DIW Berlin, jjessen@diw.de
[‡]RWI, robin.jessen@rwi-essen.de
[§]RWI and Humboldt University Berlin, jochen.kluve@hu-berlin.de

1 Introduction

Theory predicts that employers may discriminate statistically and will pay female employees of child bearing age lower wages than their male counterparts, and hire them at lower rates, ceteris paribus. This discriminatory behaviour will be more pronounced if firms face some of the direct costs of employees' motherhood. In this paper we test these predictions using a natural experiment created by the reform of German maternity protection laws. Before the reform, firms had to pay mothers' wage continuation around childbirth; hence, firms' costs arising from maternity protection wage payments was a direct function of their workforce's probability to become mothers, i.e. it effectively depended on the gender and age of their workforce. The reform introduced legislation according to which all wage continuation to mothers comes centrally from the social security system, and firms' contributions are merely a function of the number of employees, regardless of gender and age.

A driving mechanism behind the gender wage gap is the fact that, on average, mothers take substantially more time off work after childbirth than fathers do. These career interruptions result in lower labour market experience and consequently lower wages (Gangl and Ziefle 2009; Goldin 2014; Adda et al. 2017). However, a substantial pay gap exists even when controlling for observables such as labour market experience or tenure and an - admittedly substantially smaller - gap also exists pre-birth. One reason for this is that maternity is not only punished *ex-post*, but also *ex-ante* via statistical discrimination of women of childbearing age (i.e. potential mothers). Employers in general bear some of the costs of motherhood. They need to find a replacement during leave, the accumulation of firm-specific human capital is disrupted, and existing skills of their employees deteriorate during leave. Also, mothers may not return to their job (full-time) afterwards. This makes profit-maximizing companies reluctant to either hire them in the first place or to promote them. If investing in human capital through on-the-job training also becomes less valuable, this further depresses wages, reduces promotions and decreases tenure. Under standard assumptions, disincentives in hiring potential mothers lead to adjustments along two margins: fewer potential mothers are hired, and if so, they receive lower wages, ceteris paribus. Theory predicts that this effect is stronger if companies have to pay wage continuation directly. Using a natural experiment, we estimate to what extent this effect can be counter-acted if the state pays wage continuation centrally through the social security system.

Before January 1st, 2006, large firms (31 employees or more) were obliged to pay for the generous maternity protection – 14 weeks of 100 percent wage continuation around the date of delivery – of their female employees. From 2006 onwards, each firm contributed to maternity protection through a flat-rate contribution to the countrywide health insurance system, where the flat-rate contribution is paid for every worker in the firm, irrespective of whether the worker is male or female (i.e. firms with an all-male workforce contribute the same as firms with an all-female workforce). This had been the regulation for small firms ($\leq$ 30 workers) already before the reform. In its ruling declaring the previous regime as unconstitutional, the German Federal Constitutional Court stated it was unable to say with certainty "how large the probability is that due to this burden employers who have to pay maternity leave wage replacements decided not to employ women" (Bundesverfassungsgericht 2003a, Section 120). This is precisely our research question. We use comprehensive data – from annual linked employer-employee administrative records – covering over 1 million workers in more than 10,000 firms to estimate the labour market effects of the reform.

The gender wage gap and gender employment differences have received extensive coverage in the literature. The overall reading is that despite signs of a narrowing over time these differences have persisted in all Western economies (e.g Blau and Kahn (2003); Weichselbaumer and Winter-

Ebmer (2005); Olivetti and Petrongolo (2008). For Europe, evidence suggests that the gender wage gap is particularly wide at the top and bottom end of the wage distribution (Arulampalam et al. 2007; Christofides et al. 2013), although this is also the area where a decline in Germany has been more pronounced in the early 2000s (Antonczyk et al. 2010). Card et al. (2015) look at firm-specific pay premiums as a source of (gender) wage inequality and point out that if firms have some control over the wages offered, relative wages of women will be influenced by both a potential sorting of women into higher or lower paying firms and on their relative bargaining power. In Germany, the share of females is higher in smaller firms which on average pay lower wages. Heinze and Wolf (2010) confirm some selection of women into lower paying firms more generally, i.e. for firms of all sizes.

To our knowledge, no previous study links the gender wage gap to *statistical* discrimination of *potential* mothers. So far, the literature has focused on *actual* mothers. However, not only being a mother but already the *possibility* of becoming a mother can have adverse effects in the labour market. This links the topic to the literature on statistical discrimination: the basic idea of statistical discrimination, originally developed by Phelps (1972) and Arrow (1973), is that employers discriminate due to imperfect information on a certain group. In the case of statistical discrimination against potential mothers, firms know that a substantial share of women of child-bearing age will eventually have children (about 80% of women in Germany have a child over the course of their childbearing age), imposing an additional cost on profit-maximizing firms and making them more reluctant to hire potential mothers. Hence, the combination of sex and age is used as a proxy for the additional costs that employers have to shoulder due to prolonged absence of workers, which can make statistical discrimination indeed rational (profit maximizing) and occur in the absence of taste-based discrimination.[1] Whereas in Germany anti-discrimination laws regulate that employers are not allowed to ask (potential) employees about a current or planned pregnancy, it is an evident possibility that a woman of child-bearing age actually will become a mother: the annual average probability to give birth for women in the age bracket 24-35 is 5.5%, peaking at an average of 7% p.a. for women aged 30-32. Adverse labour market prospects for women can then arise without any sexism, and, as Phelps (1972, p. 661) notes, "[d]iscrimination is no less damaging to its victims for being statistical."

In addition to contributing to the literature on statistical discrimination, our study also adds to the rapidly growing literature on labour market effects of parental leave regulations (Specifically for Germany see e.g. Kluve and Tamm (2013); Schönberg and Ludsteck (2014); Kluve and Schmitz (2018); Bergemann and Riphahn (2017); international studies include Baker et al. (2008); Lalive and Zweimüller (2009); Pronzato (2009); Dahl et al. (2016). The emerging set of studies for Germany analyze changes in the parental leave system, while our focus is on the labour market effects of German motherhood protection, which, to the best of our knowledge, has so far not been analyzed. Indeed, studies on paid motherhood protection rather than parental leave are rare as most European countries have had regulations in place for decades with little or no exogenous variation; moreover, maternity protection is often integrated into the parental leave regulation.

[1] Becker (1957)'s classical theory of discrimination is based on (dis)taste or prejudice, plain sexism in this context, and implies that agents do not solely maximize profit, but also take personal taste in their decision making into account. The literature commonly attempts to identify discrimination by identifying the part of the gender wage gap that can be explained by differences in characteristics (such as education, experience, skills etc.) and attributing the unexplained component to be due to discriminatory factors (audit studies are another popular approach (e.g. Neumark et al. (1996) or Goldin and Rouse (1997)) but they typically have low external validity; Altonji and Blank (1999) is a key reference including a review on taste-based discrimination). Becker (1957)'s theory suggests that taste-discriminatory firms forgo profits and should hence not survive in a competitive environment. A similar point is stressed by Arrow (1973). This prediction has been partly tested and confirmed by Weber and Zulehner (2014).

In the US, some states have introduced paid motherhood leave schemes, which are analyzed in Rossin-Slater et al. (2013); Baum and Ruhm (2016). A recent study by Biewen and Seifert (2018) quantifies the association of the probability of parenthood on career transitions for men and women in Germany and finds a negative relationship between the contemporaneous probability to have a child and horizontal career transitions for women (defined as job changes in which the number of subordinates does not change by more than two). The transitions might thus still be associated with substantial wage increases.

Against this background, our paper contributes in three dimensions: first, by looking at labour market effects of the (generous) maternity protection scheme in Germany; and second, by providing evidence on the effects of employers having to pay for wage replacements during motherhood leave. Finally, although the field of economics has produced a rich literature on discrimination against women, evidence is lacking concerning the statistical discrimination against women of childbearing age, a gap we seek to fill with this study.

Given the direct costs that the pre-2006 regulation implied for large firms along with the expected post-reform adjustment mechanism, we conjecture that evidence for statistical employer discrimination should be most visible when comparing male and female wages in large firms. In the empirical analysis, we implement a *difference-in-differences model* to capture the post-reform effect, as well as a *trend-break model* that is able to distinguish between general convergence (or divergence) in outcomes between treatment and control groups over time - for instance, female and male wages - and the divergence from this trend that would be caused by the reform. The existence of such a treatment effect would in turn imply, and measure, the corresponding degree of statistical employer discrimination pre-2006.

Our results indicate that the adjustment process in large firms predicted by theory has indeed taken place in practice: the DID estimates show a statistically significant increase in female wages relative to male wages of 1.1 per cent in the post-reform period. The trend-break model carves out this pattern in more detail and shows that the general time trend in the convergence of the gender wage gap in large firms is estimated to be zero, and the post-2006 trend estimates a statistically significant annual wage effect of 0.65 per cent. This treatment effect of 0.65 indicates that indeed under the pre-2006 regulation employers acted in a statistically discriminating way towards female workers. We also find that the magnitude of the estimated wage effect is close to the ex ante expected wage continuation costs of 0.7 per cent, indicating that the statistically discriminatory behaviour of large firms pre-reform reflects rather closely the costs for the firm arising from the regulation. The findings imply that it is fruitful for policy makers to identify factors that could result in statistical discrimination against potential mothers, and that it is worthwhile for the public to finance, through taxes, costs that occur asymmetrically to mothers only, if the aim is to prevent negative labour market effects *ex-ante* and *ex-post* for (potential) mothers, and to reduce gender-specific earnings differences.

The paper proceeds as follows: Section 2 delineates the institutional background and the way in which the policy reform creates a natural experiment. In Section 3 we discuss the mechanisms through which the reform affects female wages and employment rates; we also calculate ex-ante expected effect sizes. Section 4 describes the data and our empirical approach. Section 5 presents and discusses the results, and Section 6 concludes.

2 Institutional Background

In all OECD countries bar the United States mothers are entitled to paid motherhood leave around the time of their childbirth. Motherhood leave is not to be confused with maternal and more

generally parental leave, which denotes employment-protected leave of absence for employed parents subsequent to motherhood leave. Whereas motherhood leave commonly involves 100 per cent wage replacement, parental leave does not necessarily come with a wage replacement (the "'parental benefit"'), and if so, it is typically determined as a fraction of pre-childbirth labour earnings, and may also be paid for a shorter time period than the employment protection period.[2]

The non-binding ILO convention on maternity leave stipulates a period of at least 14 weeks of employment-protected leave of absence for employed mothers. In Germany, motherhood leave lasts precisely from six weeks before projected childbirth until eight weeks thereafter, where the latter period extends to 12 weeks for multiple births, premature delivery or if a disability of the child is discovered within this period. During those 14 weeks, mothers receive full replacement of their average net earnings of the preceding three months, though the payments are capped at the upper earnings limit in the statutory pension fund (amounting to 5,250 Euros per month gross earnings in West Germany in 2006, the year of the reform we study). In contrast to parental leave benefits, called *Elterngeld* in Germany, those payments can naturally only be claimed by mothers.[3] The maternity protection legislation had been largely unchanged since 1986 until on 1 January 2006 a reform came into effect, which (partly) changed who bears the costs of the maternity protection benefits, and which we will make use of in our empirical analysis.

Before 2006, employers with more than 30 full-time-equivalent employees had to pay a substantial share of the wage replacements of mothers during maternity protection themselves:[4] women who were insured by a statutory health insurance company received a fixed amount of 13 Euros per calender day (roughly 400 Euros monthly) and additionally, if they were employed previously, the difference to their previous net earnings from their employer. In 2001 firms had to subsidize maternity protection benefits by about 1.48 billion Euros (Bundesverfassungsgericht 2003a).[5]

Figure 1 displays the costs firms faced for the 14 weeks of motherhood protection leave for a typical individual as a function of net earnings. Up to the upper earnings limit, the costs are a linearly increasing function of the previous wage (for a particular worker, the precise net earnings corresponding to the upper earnings limit depend on the tax bracket, i.e. the composition of the household, see figure notes). Note that the figure displays firms' direct, effective cost per month for one female worker in maternity protection: the fixed share of 13 Euros per calendar day (i.e. a

[2]A comparison of policies in OECD countries is given in https://www.oecd.org/els/soc/PF2_1_Parental_leave_systems.pdf.

[3]The tax-funded parental leave benefit "'Elterngeld"' has been in place since 2007. In general, it replaces 67% of pre-childbirth labour earnings for a total of 14 months if both mother and father take up up the benefit; the 14 months can be freely distributed between both parents, as long as each takes a minimum of 2 months. If only one parent takes up the benefit, the maximum receipt period is 12 months, see e.g. Kluve and Schmitz (2018)

[4]Specifically, the regulation set the threshold at 20 and granted the statutory health insurers the flexibility to increase that threshold from 20 to up to 30. In order to take into account actual behaviour, we contacted the different regional entities of the largest statutory health insurer, AOK, that were responsible for executing the maternity protection payments: as a result, the majority of them set the limit to 30, such that we use this effective cut-off in our analysis. The relevant measure and threshold of "'full-time equivalent"' FTE is determined by the number of employees weighted by hours worked. A person working less than 10 hours counted 0.25 FTE, a person working 10 to 20 hours 0.5 FTE, 20-30 hours 0.75 FTE and a person working more than 30 hours counted as 1 FTE. In our data (see Section 4) a part-time variable indicates whether a person has worked more or less than 18 hours per week, and the FTE categories have to be approximated: in particular, we use the German Socio-Economic Panel (SOEP) (Wagner et al. 2007) to impute the respective shares of workers falling into the working hour groups. We use information from 2003, in line with our specification of small and large firms (see below), and implement the imputation separately by gender.

[5]Note that individuals with a private health insurance were not covered by this regulation and also not affected by the change in law. However, since only about 10 per cent of Germans are privately insured, and since our data do not contain civil servants (see Section 4), of which many have a private insurance, this is not a major issue. Concerning our estimates, we still identify lower bounds as we cannot determine the type of insurance an individual had.

total of about 1,400 Euros for the 14 weeks) covered by the statutory health insurance is already deduced, and the cost curve therefore begins its upward sloping part at monthly earnings of about 400 Euros. As this 13 Euros daily contribution paid by insurers had remained unchanged since 1968, whereas wages had risen substantially, the share paid by larger firms had increased strongly over time.

For smaller firms, in contrast, statutory health companies paid the entire wage continuation. Those firms had to pay a social security contribution per employee to come up for those costs, a pay-as-you-go system called *Umlage U2 - Mutterschaft* (Contribution U2 - Motherhood). The contribution rate was determined by the health insurance companies. As of early 2006, the rate was between 0.2 and 0.3 percentage points of gross earnings (up to the upper earnings limit) among the ten largest insurers. Firms had to contribute to *Umlage U2* per worker regardless of their gender and age. This set-up was explicitly designed to prevent adverse employment effects for women of childbearing age. Hence, small firms with a higher share of female employees of childbearing age didn't have to bear a higher financial burden than predominantly male firms. In 2003, around 90% of all firms were covered by the *Umlage*, but this included only one-third of female and one-quarter of male employees. In practice, AOK, Germany's largest insurer with regional independent entities covering about a third of the population, was responsible for executing maternity leave payments for employees insured at so-called *Ersatzkassen*, but they were reimbursed by the larger firms. For employees, there were effectively no differences depending on the size of the employer.

One firm with a workforce of around 100 employees, about half of them female, put in a constitutional complaint against the regulation in the early 2000s. The firm refused to pay one mother the benefits of 1705.52 Euro and took the case first to labour courts and then to the constitutional court. The firm argued that the protection of mothers is in the interest of the public, and hence should be tax financed. On 18 November 2003, the German Federal Constitutional Court ruled the then current legislation as unconstitutional. They argued that, since it created a disincentive for larger firms to hire women, it violated the constitutional principles of Article 3 declaring equal opportunities for men and women, and of Article 12 which grants the right to chose the workplace freely (Bundesverfassungsgericht 2003b). The court laid out that smaller firms were relieved from the asymmetrical costs due to the pay-as-you-go system, but an expansion to larger ones was refused by employer representations due to the administrative expenses. It made clear that from a constitutional perspective one could not argue that a large employer is free to choose not to hire women for business considerations due to the additional costs from maternity protection, as this has adverse labour market consequences for women. The original justification for the design of motherhood protection payments, i.e. specifically the rule that only larger firms cover the costs themselves, was that the legislator deemed the burden to be relatively small in relation to the sum of wages for larger firms. In its November 2003 ruling, the constitutional court demanded that a new regulation, in line with the principles of the German constitution, would have to come into effect not later than 1 January 2006. Theoretically, anticipation effects might have played a role from the moment of the ruling onwards, although the court gave the government flexibility in designing a new legislation in line with the constitution. Eventually on 1 January 2006 a new law came into effect regulating that firms of all sizes have to take part in the pay-as-you.go system *Umlage U2*. That is, also large firms have to pay the social security contribution for maternity protection payments - a pure function of the size of their workforce, irrespective of gender and age composition - and the statutory health insurance companies reimburse firms for the wage replacements. Table 1 summarizes the reform of the maternity protection law.

Our prior is that the introduction of the new legislation led to a notable improvement of labour market prospects for women of childbearing age working at large firms. The wage continuation

payment at a 100 per cent rate is economically relevant: if a mother is employed full-time for one year and goes into maternity protection for 14 weeks, she works less than 75 percent of the year, at a full annual salary. Since the flat payment of about 400 Euro per month is paid to every mother from statutory health insurance, and the remaining difference was paid by the firms, the legislation was especially costly for firms employing high earners (recall Figure 1). For instance, for a mother with a monthly pre-birth net income of 2,500 Euros, the firm would have had to contribute an additional 6,780 Euros for the 14 weeks of maternity protection. These costs are substantive in both absolute and relative terms; average hiring costs in Germany, for instance, are about 4,700 Euros (Muehlemann and Pfeifer 2016).

In comparison to an employee working the full year, at the same monthly wage, employers would pay the mother more than 30 percent net more per hour worked. If at the end of the maternity protection period the mother does not return to work immediately, which is almost always the case, firms may have to find a (temporary) replacement. Again, if they hire a new employee for a year, the firm has to calculate with a substantially higher wage as it needs to cover the additional maternity protection benefits besides the wage for the replacement worker. In this study, we quantify to what extent employers take these costs into account. The prediction is that employers who have to directly bear the costs of the wage continuation employ fewer women of child-bearing age, and pay them lower wages than other employees. A potential mechanism for the latter effect, in particular, would be that women's bargaining position in large firms was substantially weaker before 2006, which has been shown to have a negative effect on wages (see, e.g., Card et al. 2015). Moreover, the higher costs may also have led to fewer promotions of affected women.

3 Mechanisms and Expected Effects

In this section we describe the mechanism through which the 2006 reform is expected to impact on wages and employment rates of female workers at large firms. We expect direct effects on the treatment group and spill-over effects on the control groups, in other words, a violation of the stable unit treatment assumption (SUTVA). The sign of spill-over effects on the control group depends on the choice of the control group and the outcome variable. Therefore we can benchmark the estimates from a difference-in-differences specification as upper or lower bounds of the average treatment effects on the treated, similar to Blundell et al. (2004).

When calculating the costs of employing a potential mother, employers take into account the firm's expected total costs of motherhood. These include costs that were not affected by the 2006 reform, including e.g. replacing the mother while she is on leave, or reductions in output during this period. The reform did affect whether companies had to contribute to the wage continuation during maternity protection. The expected annual costs of this wage continuation before the reform were $E[C]$

$$E[C] = \begin{cases} 0 & \text{if} \quad X(W) \leq 13 \times 365 \\ p \times 98 \times \left(\frac{X(W)}{365} - 13\right) & \text{if} \quad X(W) > 13 \times 365 \quad \& \quad W \leq \bar{W} \\ p \times 98 \times \left(\frac{X(\bar{W})}{365} - 13\right) & \text{if} \quad W > \bar{W}, \end{cases} \tag{1}$$

where W denotes the annual gross wage, $X()$ is a function that converts gross wages to net wages, and p is the probability that a potential mother will give birth in a given year, which is on

average 5.5% in our sample.[6] The expression $(\frac{X(W)}{365} - 13)$ defines the daily wage continuation, and 98 is the number of days for which wage continuation is paid.

Consider a woman who earns the sample average in our sample, 2,440 Euro, under the counterfactual: given the 2006 tax regime and single filing her labour income net of taxes and social security contributions is 1,547 Euro.[7] Using Equation (1), the expected wage continuation costs due to motherhood in a given year are about 204 Euro or 0.7% of her annual gross earnings. This figure benchmarks the results for wage adjustment. Under fully inelastic labour supply, the effect on female wages would be similar to this figure. In practice, this is likely to be an upper bound for the wage increase, as the share of employed women at large firms is likely to increase as well. In the following we give a brief overview of expected spill-over effects and the interpretation of the estimated average treatment effect on the treated for different specifications.

Wages of female employees vs. male employees at large firms — The reform decreased the expected costs of employing female workers, which is expected to lead to a wage increase for women. At given expected costs, employers can pay females higher wages, which is expected to lead to an increase in the number of employed females. This in turn puts downward pressure on the wages of male employees. The spill-over on wages of males is thus expected to be negative and the estimated average treatment effect on the treated is an upper bound.

Share of female employees at large vs. small firms — Newly employed women at large firms would come from small firms and from unemployment. Thus, we expect the share of females at large firms to increase relative to the share of females at small firms. As the share of females at small firms is affected negatively by the reform, estimates from a difference-in-differences specification yield the upper bound of the average treatment effect on large firms.

Wages of female employees at large vs. small firms — In the benchmark case of completely separated labour markets for small and large firms, wages in small firms would not be affected by the reform (neglecting general equilibrium effects, such as e.g. through changes of prices of goods). In contrast, if there are no obstacles to moving between large and small firms, female employees at small firms can credibly threaten to move to larger firms, and we would expect the wage increase at small firms to equal that at large firms. The spill-over effect is positive and the difference-in-differences estimator would be thus a lower bound of the ATT.

4 Data and empirical approach

We use two primary data sources for our empirical analysis. First, individual employment spell data taken from social security records. These include detailed information on employees' employment history and gross daily wages or benefits and contain a limited set of social demographic characteristics. As is common for social security data, the wage is censored from above at the upper earnings limit in the statutory pension fund. The threshold is adjusted in most years and differs between East and West Germany, e.g. in 2010 the threshold was 66,000 Euro in West Germany.

[6] We implement a method developed by Müller and Strauch (2017) to identify births in German social security data. The annual births probabilities are in line with those calculated by Raute (Forthcoming), but slightly lower as we restrict the upper age limit to 35 in our calculations.

[7] This figure can be obtained using the tax calculator provided by the Federal Ministry of Finance and applying social security contributions of 0.2 of gross earnings.

The data consist of all people who during the observation period were either employed subject to social security contributions, marginally part-time employed, benefit recipient, job-seeker or participant in an employment or training measure. This means that civil servants, self-employed and participants in higher education are not covered.

The second data source, the establishment data, come from an annual representative survey of establishments in Germany employing at least one employee subject to social security. Through a unique establishment ID the two data sets are then merged (Heining et al. 2016). All individuals that have been employed by an establishment in the panel are included in our data. In the empirical analysis, we use the waves from 2001 to 2010. The linked data set (LIAB LM 9314) is provided by the German Institute for Employment Research (IAB). In total we observe more than 10,000 establishments in each year, which have been linked to individuals, corresponding to more than 1 million individual observations per year.

The individual-level data contain information on gender, age and daily wages of individuals, as well as their full labour market history during the observation period. By law employers are obliged to report the beginning and end of each employment relationship along with an annual report at the end of each year. The reported daily wage comprises all gross earnings including premiums and allowances. The data contain information on full-time, part-time and marginal employment, but do not contain detailed information on hours of work. Hence one cannot conclude unambiguously whether a change in daily wages is due to a change in the hourly wage or in hours worked.

The main analyses are conducted at the establishment level, using the cut-off of 30 full-time equivalent employees defined by the maternity protection legislation before 2006 to distinguish large from small firms. Assignment to the groups of large and small firms, respectively, is based on firm size in the year 2003, prior to the ruling of the constitutional court.[8] In a treatment-control framework, firms with more than 30 employees covered by the pre-2006 regulation constitute the "treatment group" (the upper right-hand cell in Table 1). That is, the analysis is placed in reverse calendar time, considering the *switching-off* of the treatment.

We exclude firms with 9 employees or fewer, since these very small firms arguably are too different from large ones and many of them have very little fluctuation in their workforce (hirings per year). Moreover, we specifically use regularly employed workers to calculate firm-level indicators, in order to allow for a reasonable comparisons of daily wages. We exclude person-level observations with daily wages below 1 Euro, which we assume are either dormant employment relationships or spells attributable to measurement error. We keep both regularly and marginally employed individuals, because the definition of marginal employment changed in 2003, transforming some regular employment relationships into marginal ones. Excluding marginally employed individuals would lead to a marked drop in the observed average wage in 2003. Finally, we include only individual spells that cover June 30 of a given year, as this is the point in time when the surveys at the establishments are conducted. After imposing those preparatory steps, we retain more than 28,000 firm-year observations covering more than 5 million individual-year observations.

In principle, one alternative would have been to assign workers to treatment and control groups depending on the size of their firm in 2003. This, however, would lead to several problems; first, one could then only include those individuals that were employed in 2003 at either type of firm, which by construction would lead to a more strongly ageing sample than the working population, and therefore also to stronger wage increases over time. Additionally, workers switch between

[8]The Federal Constitutional Court declared the then current legislation as unconstitutional in autumn of 2003, as described above. Hence, anticipatory effects could have played a role from this moment onwards; for instance, if an employer believes that a newly hired woman is unlikely to have a child within the next two years, her motherhood protection period would fall under the new regime.

firms: thus, when a worker is first assigned to the treatment group in 2003 she remains in that group even if she switches to a larger firm later, and vice versa. Again by construction, this leads to a convergence of wages between treatment and control group over time.[9] Whereas the research design for testing the hypothesis on statistical discrimination is thus based on the firm level, we still make use of the many advantages of merged worker-establishment data: e.g., the combined individual-firm information is used to calculate key establishment variables of interest, in particular the average wage, the share of female workers, and new hires.

Table 2 displays summary statistics of large and small firms for the observation period 2001 to 2010. We focus on this period in the analysis in order to take into account a sufficiently long enough period prior to the 2006 reform (and the ruling of the constitutional court); at the same time, given the mechanisms laid out in the previous section, any treatment effect will have occurred, and adjustment process completed, within four years post-reform. The table also presents summary statistics by main sectors, in particular classifying workers and firms into blue collar, white collar, and public sector using a three-digit sector classification.

The share of female workers in all firms is in the range of one third to 40 per cent, with large firms showing a smaller average share of female workers (33 per cent) than small firms (41 per cent). This pattern by firm size is maintained across the main three sectors shown in Table 2, albeit at different levels: female workers make up about one fourth of the workforce in blue collar occupations, about half of the workforce in white collar firms, and about two thirds in the public sector. These descriptive findings are also the case specifically for females of childbearing age (younger than 35), with a somewhat narrower large-small firm difference across sectors (e.g. 40 vs. 35 per cent for all firms). The 35-years-of-age threshold is chosen based on the childbirths identified in our sample: the smoothed age distribution is plotted in Figure 2 and indicates that by the age of 35 only a small share of women is likely to have a birth in the upcoming years.

The average age of employees is around 43 years, with very little variation by firm size or sector. Looking at the mean monthly gross full-time wage in Table 2, several pronounced patterns are apparent: first, male average monthly wages are consistently higher than female average wages, with an absolute difference of around 200 to 300 Euros, around 10 per cent in relative terms. Second, for both men and women, average wages are substantially higher in large firms than in small firms: in the full sample, the respective differences large vs. small are about 915 Euros (approx. 50 per cent) for women and about 860 Euros (approx. 40 per cent) for men. These differences, in turn, vary significantly by sector. In the manufacturing sector, the large-small firm differential is more than 1,000 Euros in absolute terms (for both men and women), in white collar occupations it is around 550 Euros, and in the public sector it is around 400 Euros.

The share of female workers working full-time, unsurprisingly, is consistently lower than the share of male workers working full time. *Within* gender, i.e. looking at the share of women in the firm who work full-time, this share is substantially *higher* in large firms than in small firms: the difference amounts to 10 percentage points in the full sample (74 per cent vs. 64 per cent); it is of similar size in the manufacturing (79 vs. 67 per cent) and public sectors (63 vs. 54 per cent), but not visible, or slightly reversed, in white collar occupations (where 67 per cent of women work full-time in large firms, and 69 per cent in small firms). *Across* gender, however, i.e. looking at the share of women among all full-time employees, this share is typically *lower* in large firms than in small firms. Specifically, the difference is 7 percentage points in the full sample (28 per cent in

[9] As an illustration of the fundamental difference between an analysis at the firm vs. the individual level: suppose an individual works for a firm with 18 employees in 2003 and switches to a firm with 35 employees in 2004 (unchanged from 2003). At the individual level she remains in the control group, whereas at the firm level she contributes to the average wage for a control group establishment in 2003 and a treatment group establishment in 2004.

small and 35 per cent in large firms), and varies in size and level by sector: 7 percentage points in manufacturing (17 per cent - the smallest share overall - in large vs. 24 per cent in small firms), 6 percentage points in white collar occupations (37 per cent vs. 43 per cent), and 1 percentage point in the public sector (56 per cent vs. 57 per cent).

Table 2 also shows that, evidently in line with firm size, the average number of new hires per year is (much) higher in large firms than in small firms. The overall average share of female workers among new hires is just under 50 per cent, regardless of firm size. Again, differences are visible by main sector: in manufacturing, the share of new hires who are female is around one fourth to one third, in white collar jobs it is about half, and in the public sector around 60 per cent.

To identify the causal effect of discontinuing the regime of letting large firms pay for motherhood protection leave, i.e. *switching off* the treatment, we implement two specifications. As was evident in Table 2, firms of different sizes differ fundamentally. A difference-in-differences specification allows for differences among the groups, but crucially relies on the common trend assumption, which states that in absence of the regime change, the groups would have followed the same trend, i.e. the difference between them would have remained constant. By definition this assumption cannot be tested. The following specification gives the difference-in-differences estimator for a given outcome y for individual i in firm j at time t.

$$y_{ijt} = \gamma_1 \text{treat}_{ijt} + \gamma_2 \text{treat}_{ijt}\text{post}_t + \beta X_{ijt} + \epsilon_{ijt} \tag{2}$$

β is a vector of coefficients and X_{ijt} is a vector of control variables, which includes firm-specific variables, year dummies and a constant. treat_{ijt} is a dummy that indicates whether individual i is in a treatment group firm. Depending on the specification, this might mean that she is working at a company in the treatment group in year t or that she is a woman. post_t is a binary indicator that takes on the value 1 from 2006 onwards. The coefficient γ_2 is the difference-in-differences estimator. We allow for the error term ϵ_{ijt} to contain firm-specific fixed effects. Omitting individual-specific subscripts from equation (2) yields the estimator for firm-level outcomes.[10]

The difference-in-differences estimator relies on the common trend assumption. However, in the case of the gender wage gap, for instance, this assumption might be problematic: instead of a common trend there could be some convergence over time, even in absence of the reform. Therefore we propose an alternative estimator, which relies on more general assumptions. Specifically, we assume that the rate of convergence (or divergence) in the outcome between treatment and control group would be constant in the absence of the reform. The reform effect manifests itself in a break of this trend. The *trend-break model* is specified as follows:

$$y_{ijt} = \delta_1 \text{treat}_{ijt} + \delta_2 \text{treat}_{ijt}\text{trend}_t + \delta_3 \text{treat}_{ijt}\text{post}_t\text{posttrend}_t + \beta X_{ijt} + \upsilon_{ijt}. \tag{3}$$

Here $\text{trend}_t = \text{year}_t - 2000$ and $\text{posttrend}_t = \text{year}_t - 2005$ if $t > 2005$ and 0 otherwise for the latter expression. The coefficient δ_2 thus gives the annual convergence or divergence between the control and treatment groups. The coefficient δ_3 gives the diversion from this longer-term trend for the post-reform years. In case of a common trend in the years before the reform, δ_2 would equal zero. δ_3 indicates the average annual effect of the reform on the treated. The trend-break specification has two major advantages. First, it does not rely on the common trend assumption, and second it may be more in line with the expectation that wages and shares of employees adjust gradually to new regimes.

We implement equations (2) and (3) for the two key outcomes, wages and shares of female employees, and two dimensions of control and treatment groups, females vs. males and large vs.

[10]For firm-level estimations, controlling for firm fixed effects of course leads to an omission of the variable treat_j.

small firms. As described above, firm size is measured in 2003 and assignment to treatment and control is based on this measure, to avoid an assignment that is endogenous to the reform.

5 Results

5.1 Wages

This section presents the empirical results. Given the specifics of the maternity protection regulation and reform (Section 2), and the way it translates into cost implications for firms (Section 3), we would expect to see the clearest adjustment mechanism - and, hence, evidence for statistical employer discrimination - when looking at female vs. male wages in large firms. Figure 3 begins with a descriptive investigation, plotting log monthly wages for men and women (Panel (a)) and the respective male-female differential (Panel (b)) in large firms over the observation period. Panel (a) of the figure shows a parallel - i.e. flat - development of the respective male and female wages during the pre-reform period, and an increase in female wages vs. a continued flat curve for male wages during the post-reform years 2007 through 2010. Panel (b) illustrates this pattern for the gender wage differential, and again shows its narrowing during the post-reform years. Figure 3 therefore gives some indicative evidence for the expected adjustment mechanism, and statistical discrimination by employers.

Figure 4 repeats this exercise specifically for new hires in large firms. Given the resulting smaller sample sizes, the patterns shown are somewhat noisier than in Figure 3. Panel (a) indicates a slightly decreasing trend in wages for both male and female new hires - a trend that would possibly be in line with a German labour market that was at the time (early to mid 2000s) characterized by the highest unemployment rate since the 1950s, and by wage moderation. Perhaps more importantly for this study, however, Figure 4 again shows that during the pre-reform period also for new hires male and female wages in large firms display a parallel development, and that the corresponding gender wage differential is narrowing slightly during the post-reform years 2006-2010. The confidence intervals in panel (b) indicate that this is somewhat less precisely estimated than for all workers, but the main pattern is still visible.

We formalize this analysis in the next step, by correspondingly implementing the difference-in-differences (DID) and trend-break (TB) models of Equations (2) and (3), respectively. The main estimation results for the full sample are presented in Table 3. This table and the subsequent tables follow the same structure: we present estimates of the respective treatment effect coefficients for all employees for the difference-in-differences model in columns (1) and (2), and for the trend-break model in columns (3) and (4), where in each case the first specification includes year dummies, and the second specification in addition firm fixed effects and control variables. This basic four-column structure is then repeated in the right-hand side panel within each table, looking specifically at new hires.

The estimation results in Table 3 indicate that the adjustment process in large firms anticipated in theory has indeed taken place in practice: first, the DID estimates in columns (1) and (2) show an increase in female wages relative to male wages of 2.6 per cent and 1.1 per cent (full specification), respectively, statistically significant at the 1 per cent level. The trend-break model carves out the pattern more finely: the general time trend in the convergence (or divergence) of the gender wage gap in large firms is estimated to be zero, and the post-2006 trend estimates a statistically significant annual wage effect of 0.65 per cent for the full specification (column (4); the estimate in the parsimonious specification in (3) is of similar magnitude but less precisely estimated,

hence not statistically significant). The coefficient estimate of 0.65 is the treatment effect we are interested in: it indicates that indeed under the pre-2006 regulation employers showed statistically discriminating behaviour against women, depressing their wages, which was then reversed with the reform. Moreover, note that the magnitude of the estimated wage effect is quite close to the ex ante expected wage continuation costs of 0.7 per cent (recall the calculation in Section 3), indicating that the statistically discriminatory behaviour of large firms pre-reform may have reflected rather closely the costs for the firm arising from the regulation.

The panel on the right hand side of Table 3 presents estimation results for the wages of new hires. The point estimates for the full specification in both the DID (column (6)) and the TB cases (column (8)) are similar in size to those for all employees; however, due to the substantially smaller sample size for new hires they are less precisely estimated. The following three tables investigate further the main effect found in Table 3. First, Tables 4 and 5 distinguish large firms that pay high wages from large firms that pay lower wages. Specifically, we classify firms as above and below median wage if their average wage was above or below the sample median in the year 2003 (such that this classification, again, is independent of the reform). For both firm types, the respective DID specifications - columns (1) and (2) in Tables 4 and 5 - show a positive and statistically significant impact on female wages after the reform. Looking at the trend-break model, the results indicate that the overall pattern found in Table 3 is determined to a larger extent by firms that pay above median wages: columns (3) and (4) of Table 4 estimate that the annual convergence (divergence) between treatment and control group is zero, but that due to the reform there was a positive annual average wage effect of 0.9 per cent for women, statistically significant, narrowing the gender wage gap (for firms paying below median wages - Table 5 the coefficient estimates are positive, but small in size and imprecisely estimated). In turn, this estimate implies that large firms were statistically discriminating against female workers. And again, this empirical finding echoes ex ante expectations: Section 2 has shown that the costs of maternity protection for the firm are higher for high-wage earners, and this feature of the regulation appears to have led firms to statistically discriminate against this group accordingly.

Additional results on wage effects are presented for the three main sectors in Table 6. The table again distinguishes between all employees and new hires, but focuses on the main specification using the full set of control variables. Several patterns are worth noting: first, estimation results from the DID model across sectors again point towards a statistically significant treatment effect. Second, estimation results from the refined trend-break model indicate different types of adjustment mechanisms (and, implicitly, of statistically discriminating behaviour pre-reform). In both the manufacturing and public sector, the time trend of annual convergence (divergence) between male and female wages is again estimated to be zero (columns (2) and (10), respectively). At the same time, again the estimated treatment effect is of similar magnitude as in the full sample - 0.65 per cent annually in manufacturing (statistically significant), and 0.73 per cent in the public sector (marginally significant), pointing to wage convergence due to the reform, and corresponding statistical discrimination before. In white collar occupations, on the other hand, the adjustment mechanism (and previous statistical discrimination) appears to have affected new hires specifically: the trend-break model in column (8) indicates a negative time trend in the gender wage gap of new hires in this sector (marginally significant), which is then overcompensated by a comparatively large treatment effect of 2.46 per cent (statistically significant) on the post-2006 trend diversion.

5.2 Female Wages by Firm Size and Share of Female Workers

In addition to the main research design looking at female wages in large firms - the most immediately affected group and outcome given the maternity protection regulation - and using male wages as the control dimension, we also estimate reform effects using a comparison of small vs. large firms, and investigate the share of female employees as an outcome variable. Table 7 presents estimation results comparing female wages in small and large firms. The point estimate for the full DID specification (column (2)) shows a relatively large treatment effect, of 4.2 per cent (statistically significant at the 1 per cent level). This is a somewhat unexpected result, given that the theoretical considerations (recall Section 3 place the DID estimator in this case, in the presence of positive spill-overs, as a lower bound of the ATT. One possible explanation of this finding is that the spill-overs between small and large firms are not positive as expected (which would have implied that wage increases in large firms lead to wage increases in small firms, since female employees can credibly threaten to move), but that the post-refom wage increases in large firms make these jobs now more attractive for female workers, perhaps even at the expense of small firms. The corresponding point estimate for new hires is also large in size - 3.3 per cent, column (6) - but less precisely estimated due to the smaller sample size. Such a negative spill-over effect, even though unexpected, would also be borne out by the results from the trend-break model in columns (3) and (4), which indicate that the post-reform trend actually diverges negatively (for small firms) from the generally converging trend of female wages in small and large firms.

Similar to the assessment of wage effects in the previous subsection, Figures 5 and 6 begin with a descriptive analysis of the share of female employees by firm size (large: treatment group; small: control group) over the observation period. As Figure 5 indicates, the two curves run relatively parallel during the pre-reform period: the annual averages for the large firms are much more precisely estimated due to the large sample size, while the annual averages for the small firms are somewhat noisier. The post-2006 part of Panel (a), and slightly more so Panel (b), do suggest some convergence of the share of female employees between treatment and control groups since the reform came into effect. Looking specifically at new hires in Figure 6 sample sizes are substantially smaller, such that the curves are even noisier with wide confidence intervals, and no additional insights emerge (besides perhaps a generally increasing convergence towards the same rate of female new hires in small and large firms).

In line with our previous analysis, Tables 8 - for the full sample - and 9 - by main sectors - formally investigate the outcome *share of female employees* for the treatment-control design of large vs. small firms. The main result from the DID specifications in Table 8 is that there is a statistically significant increase in the share of female employees in large versus small firms in the post-reform period; the effect is particularly pronounced for new hires - a point estimate of 2.58 percentage points in column (6). Given the absence of any significant patterns in the estimates from the trend-break model for all employees (columns (3) and (4)) and the fact that the coefficients of the trend-break model for new hires virtually cancel each other out (positive significant general trend pre-reform, negative diversion from that trend post-reform, columns (7) and (8)), in our view this DID effect post-reform cannot be conclusively tied to statistically discriminating behaviour in the pre-reform period. The results for the manufacturing and white collar occupations in Table 9 virtually replicate these results from Table 8, whereas for the public sector none of the estimated coefficients in the DID or TB models is statistically different from zero.

6 Conclusion

Despite substantial improvements in women's labour market prospects in the past decades, female workers are still employed at lower rates and receive, on average, lower wages than men. Motherhood has been identified as a driving mechanism of these differences in labour market outcomes. Whereas a large literature has examined the *ex-post* career cost of motherhood, theory predicts that maternity may also be punished *ex-ante* through statistical discrimination by employers. Indeed, employing (potential) mothers is costly for firms: they have to seek and find a replacement during maternity leave periods, and there may be uncertainty at what point in time a mother returns to work, and what number of working hours she might be willing to supply. In addition, a more pronounced cost factor arises if firms face additional direct financial costs of employing potential mothers through the obligation to pay for wage continuation during maternity protection. This could give rise to negative wage and employment effects from statistical employer discrimination, which in turn could have a negative effect on all women of childbearing age, independent of whether they will eventually have a child or not.

To test the hypothesis of statistical employer discrimination, we use the natural experiment of a motherhood protection reform in Germany. Until 2006, large firms in Germany had to pay directly for the largest part of the generous (100 per cent) wage continuation mothers receive during their motherhood protection period, making their employee cost function a direct function of gender and age. In addition, the regulation implied that high-wage earning females were more costly to the firm than low-age earning females. We present calculations of firms' expected costs and find that these are indeed sizable. Small firms, on the other hand, contributed via the countrywide statutory health insurance systems, where contributions depend merely on the size of the firm's workforce. Since payments are therefore independent of an employee's gender, small firms did not face higher costs if they had a larger share of female workers, or paid them higher wages. In 2006, a new regime came into effect in which firms of all sizes contribute to the statutory insurance system, effectively abolishing the system in which larger firms faced substantially larger direct costs from motherhood than smaller firms. To evaluate the extent of statistical discrimination caused by the pre-2006 regulation, we use comprehensive administrative data that link person-level social security records to firms.

Given the direct costs that the pre-2006 regulation implied for large firms along with the expected post-reform adjustment mechanism, we conjecture that evidence for statistical employer discrimination should be most visible when comparing male and female wages in large firms. In the empirical analysis, we implement a difference-in-differences model to capture the post-reform effect, as well as a trend-break model that is able to distinguish between general convergence (or divergence) in outcomes between treatment and control groups over time - for instance, female and male wages - and the divergence from this trend that would be caused by the reform. The existence of such a treatment effect would in turn imply, and measure, the corresponding degree of statistical employer discrimination pre-2006.

The empirical results from our main specification indicate that the adjustment process in large firms anticipated in theory has indeed taken place in practice: first, DID estimates show a statistically significant increase in female wages relative to male wages of 1.1 per cent in the post-reform period. Second, the trend-break model carves out this pattern in more detail: the general time trend in the convergence (or divergence) of the gender wage gap in large firms is estimated to be zero, and the post-2006 trend estimates a statistically significant annual wage effect of 0.65 per cent. This treatment effect of 0.65 indicates that indeed under the pre-2006 regulation employers showed statistically discriminating behaviour against women, depressing their wages,

which was then reversed with the reform. Moreover, it is important to note that the magnitude of the estimated wage effect is quite close to the ex ante expected wage continuation costs of 0.7 per cent, indicating that the statistically discriminatory behaviour of large firms pre-reform may have reflected rather closely the costs for the firm arising from the regulation. Additional analyses show that the treatment effect - and thus, the statistical employer discrimination - is driven predominantly by firms paying above median wages (for whom the pre-2006 regulation was effectively more costly), and is visible in the manufacturing and white collar occupations, but not in the public sector.

This evidence for statistical discrimination does not fall short on policy conclusions. Since we find that labour market prospects improved significantly due to the 2006 reform, this suggests that it is fruitful for policy makers to identify factors that could result in statistical discrimination against potential mothers. A prominent example is the German parental leave legislation, where - besides social norms - the current incentive structure leads women, who on average earn less than their partners, to take longer leave periods. Moreover, our findings support that it is worthwhile for the public to finance, through taxes, costs that occur asymmetrically to mothers only, if the aim is to prevent negative labour market effects *ex-ante* and *ex-post* for (potential) mothers, and to reduce gender-specific earnings differences.

References

Adda, J., C. Dustmann, and K. Stevens (2017): "The career costs of children," *Journal of Political Economy*, 125(2), 293–337. Cited on page 1.

Altonji, J. G., and R. M. Blank (1999): "Race and gender in the labor market," *Handbook of labor economics*, 3, 3143–3259. Cited on page 2.

Antonczyk, D., B. Fitzenberger, and K. Sommerfeld (2010): "Rising wage inequality, the decline of collective bargaining, and the gender wage gap," *Labour Economics*, 17(5), 835–847. Cited on page 2.

Arrow, K. (1973): "The theory of discrimination," *Discrimination in labor markets*, 3(10), 3–33. Cited on page 2.

Arulampalam, W., A. L. Booth, and M. L. Bryan (2007): "Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution," *ILR Review*, 60(2), 163–186. Cited on page 2.

Baker, M., J. Gruber, and K. Milligan (2008): "Universal child care, maternal labor supply, and family well-being," *Journal of political Economy*, 116(4), 709–745. Cited on page 2.

Baum, C. L., and C. J. Ruhm (2016): "The effects of paid family leave in California on labor market outcomes," *Journal of Policy Analysis and Management*, 35(2), 333–356. Cited on page 3.

Becker, G. S. (1957): *The economics of discrimination*. University of Chicago press. Cited on page 2.

Bergemann, A., and R. T. Riphahn (2017): "Maternal employment effects of paid parental leave," *SOEPpapers on Multidisciplinary Panel Data Research*, (No. 900). Cited on page 2.

Biewen, M., and S. Seifert (2018): "Potential Parenthood and Career Progression of Men and Women–A Simultaneous Hazards Approach," *The BE Journal of Economic Analysis & Policy*, 18(2). Cited on page 3.

Blau, F. D., and L. M. Kahn (2003): "Understanding international differences in the gender pay gap," *Journal of Labor economics*, 21(1), 106–144. Cited on page 1.

Blundell, R., M. C. Dias, C. Meghir, and J. Reenen (2004): "EVALUATING THE EMPLOYMENT IMPACT OF A MANDATORY JOB SEARCH PROGRAM," *Journal of the European Economic Association*, 2(4), 569–606. Cited on page 6.

Bundesverfassungsgericht (2003a): "1 BvR 302/96," . Cited on pages 1 and 4.

——— (2003b): "Press release 98 / 2003-12-09," . Cited on page 5.

Card, D., A. R. Cardoso, and P. Kline (2015): "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women," *The Quarterly Journal of Economics*, 131(2), 633–686. Cited on pages 2 and 6.

Christofides, L. N., A. Polycarpou, and K. Vrachimis (2013): "Gender wage gaps,'sticky floors' and 'glass ceilings' in Europe," *Labour Economics*, 21, 86–102. Cited on page 2.

Dahl, G. B., K. V. Løken, M. Mogstad, and K. V. Salvanes (2016): "What is the case for paid maternity leave?," *Review of Economics and Statistics*, 98(4), 655–670. Cited on page 2.

Gangl, M., and A. Ziefle (2009): "Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States," *Demography*, 46(2), 341–369. Cited on page 1.

Goldin, C. (2014): "A grand gender convergence: Its last chapter," *The American Economic Review*, 104(4), 1091–1119. Cited on page 1.

Goldin, C., and C. Rouse (1997): "Orchestrating impartiality: The impact of" blind" auditions on female musicians," Discussion paper, National bureau of economic research. Cited on page 2.

Heining, J., W. Klosterhuber, P. Lehnert, S. Seth, et al. (2016): "Linked-Employer-Employee-Daten des IAB: LIAB-Längsschnittmodell 1993-2014 (LIAB LM 9314)," Discussion paper, Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany]. Cited on page 8.

Heinze, A., and E. Wolf (2010): "The intra-firm gender wage gap: a new view on wage differentials based on linked employer–employee data," *Journal of Population Economics*, 23(3), 851–879. Cited on page 2.

Kluve, J., and S. Schmitz (2018): "Back to Work: Parental Benefits and Mothers' Labor Market Outcomes in the Medium Run," *ILR Review*, 71, 143–173. Cited on pages 2 and 4.

Kluve, J., and M. Tamm (2013): "Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment," *Journal of Population Economics*, 26(3), 983–1005. Cited on page 2.

Lalive, R., and J. Zweimüller (2009): "How does parental leave affect fertility and return to work? Evidence from two natural experiments," *The Quarterly Journal of Economics*, 124(3), 1363–1402. Cited on page 2.

Muehlemann, S., and H. Pfeifer (2016): "The Structure of Hiring Costs in Germany: Evidence from Firm-Level Data," *Industrial Relations: A Journal of Economy and Society*, 55(2), 193–218. Cited on page 6.

Müller, D., and K. Strauch (2017): "Identifying mothers in administrative data," Discussion paper, Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany]. Cited on pages 7 and 20.

Neumark, D., R. J. Bank, and K. D. Van Nort (1996): "Sex discrimination in restaurant hiring: An audit study," *The Quarterly Journal of Economics*, 111(3), 915–941. Cited on page 2.

Olivetti, C., and B. Petrongolo (2008): "Unequal pay or unequal employment? A cross-country analysis of gender gaps," *Journal of Labor Economics*, 26(4), 621–654. Cited on page 2.

Phelps, E. S. (1972): "The statistical theory of racism and sexism," *The american economic review*, 62(4), 659–661. Cited on page 2.

Pronzato, C. D. (2009): "Return to work after childbirth: does parental leave matter in Europe?," *Review of Economics of the Household*, 7(4), 341–360. Cited on page 2.

Raute, A. (Forthcoming): "Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits," *Journal of Public Economics*. Cited on page 7.

Rossin-Slater, M., C. J. Ruhm, and J. Waldfogel (2013): "The effects of California's paid family leave program on mothers' leave-taking and subsequent labor market outcomes," *Journal of Policy Analysis and Management*, 32(2), 224–245. Cited on page 3.

Schönberg, U., and J. Ludsteck (2014): "Expansions in maternity leave coverage and mothers' labor market outcomes after childbirth," *Journal of Labor Economics*, 32(3), 469–505. Cited on page 2.

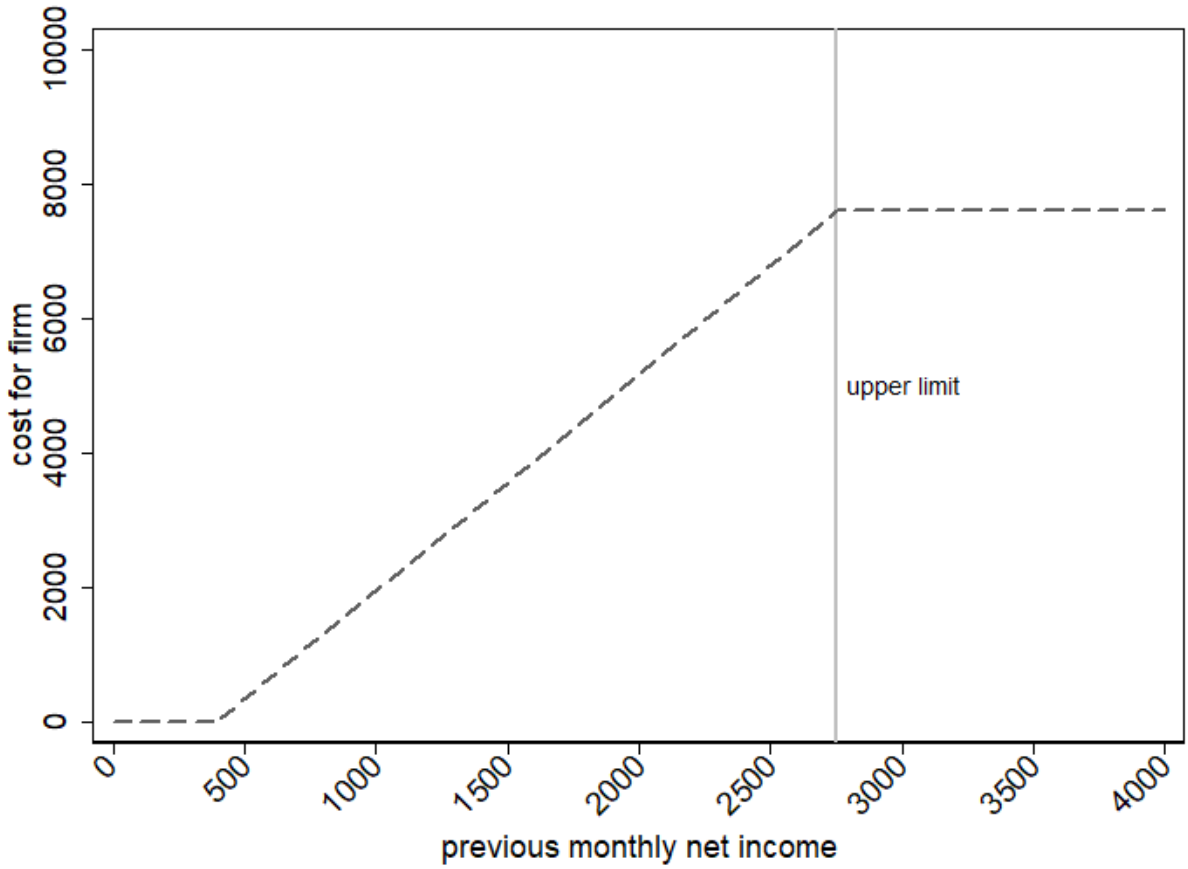
Wagner, G., J. Frick, and J. Schupp (2007): "The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements," . Cited on page 4.

Weber, A., and C. Zulehner (2014): "Competition and gender prejudice: Are discriminatory employers doomed to fail?," *Journal of the European Economic Association*, 12(2), 492–521. Cited on page 2.

Weichselbaumer, D., and R. Winter-Ebmer (2005): "A meta-analysis of the international gender wage gap," *Journal of Economic Surveys*, 19(3), 479–511. Cited on page 1.

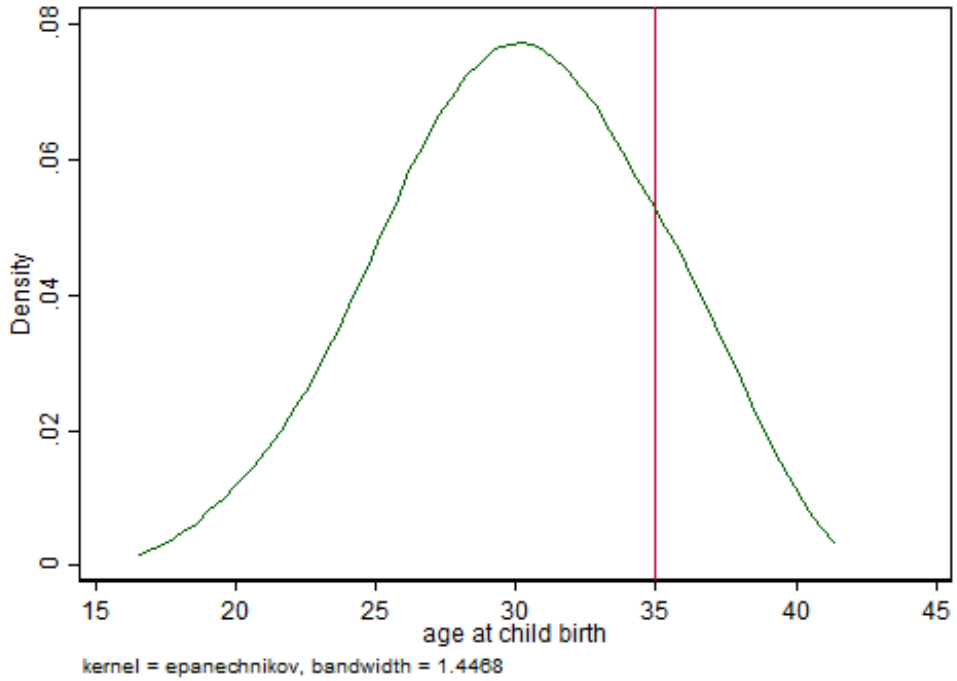
A Figures

Figure 1: Firms' costs per maternity protection period



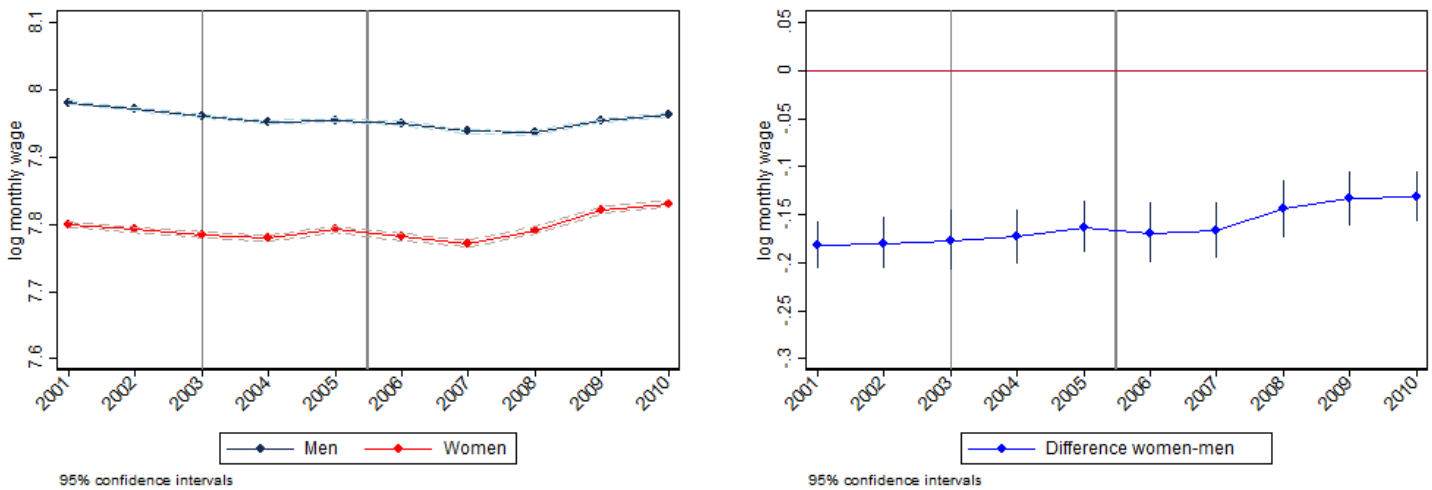
Note: Costs of motherhood protection imposed on employer for different levels of employee's monthly net income in Euro. The "'Upper limit"' indicates the net earnings corresponding to the upper earnings limit in the statutory pension fund (monthly gross earnings of 5,250 Euro); the case illustrated here uses the tax rate for a single worker, or a married couple without children with equal gross earnings.

Figure 2: Density plot of mothers' age at childbirth



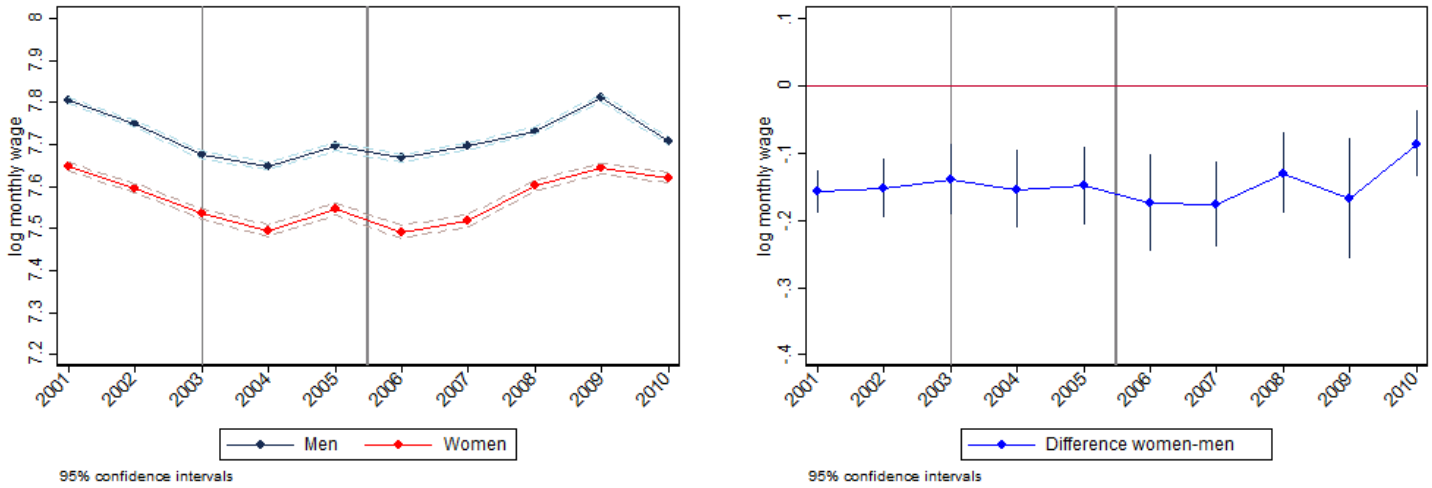
Notes: Authors' calculations based on the administrative data described in section 4; births are identified in the data using a corresponding method developed by (Müller and Strauch 2017). Only childbirths for mothers who were employed subject to social security contributions (pre-birth) are considered. Pooled over the years 2001 to 2010.

Figure 3: Full-time wages of men and women in large firms (> 30 employees)



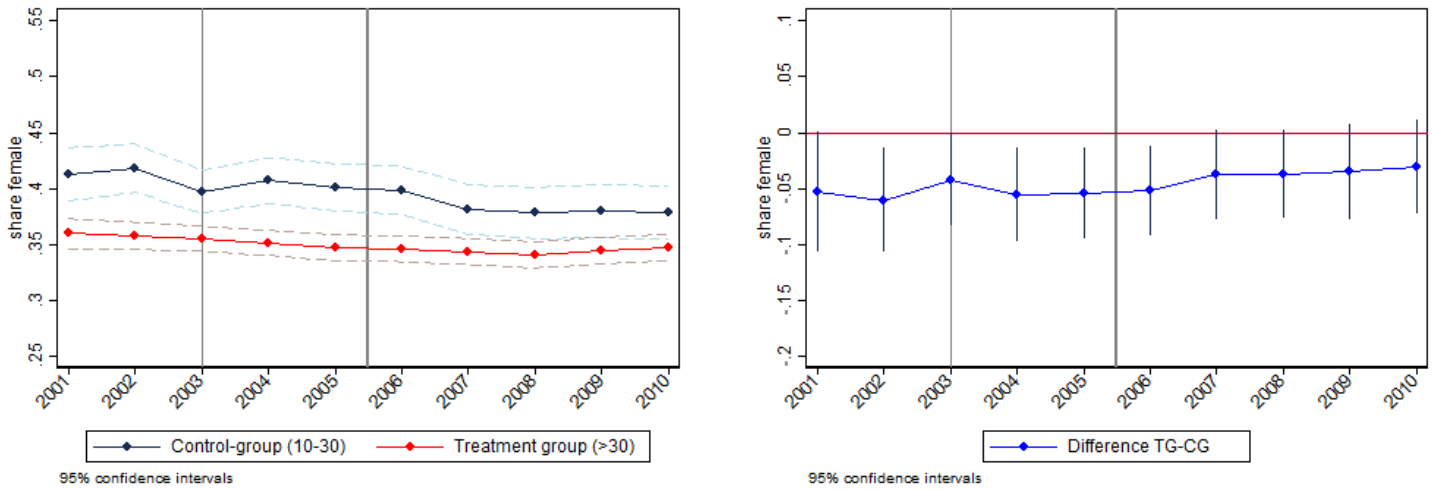
(a) (b)

Figure 4: Full-time wages of newly hired men and women in large firms (> 30 employees)



(a) (b)

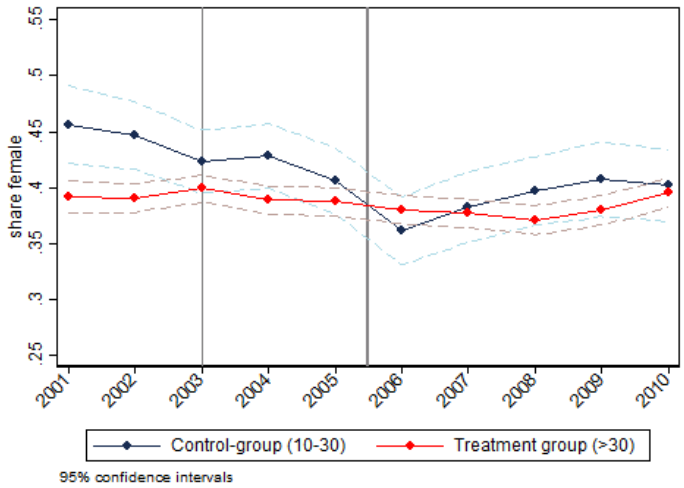
Figure 5: Share of female employees in firms of different sizes



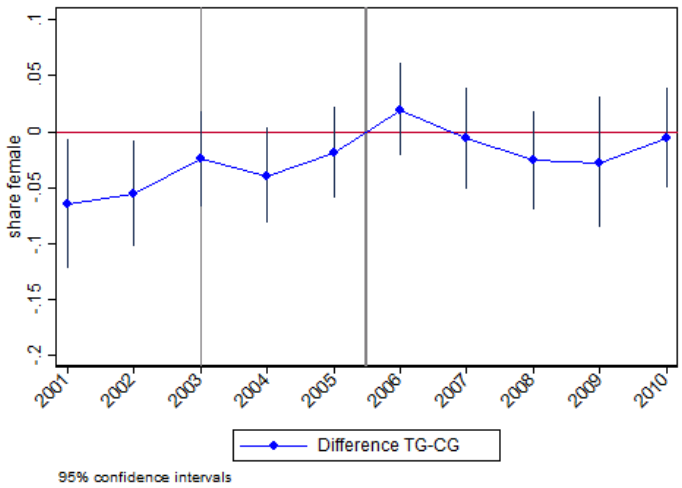
(a) (b)

Notes: Firms are weighted by their number of employees. Number of employees denotes full-time equivalent employees.

Figure 6: Share of female employees among new hires in firms of different sizes



(a)



(b)

B Tables

Table 1: Obligations to pay for maternity protection wage replacement, by firm size and regulation

	Firm size	
	up to 30 employees	more than 30 employees
Until 31 Dec 2005	contribution per employee (Umlage U2), firms' costs irrespective of gender and age	directly paid by firms, firms' costs a direct function of workers' gender and age
From 1 Jan 2006	contribution per employee (Umlage U2), firms' costs irrespective of gender and age	contribution per employee (Umlage U2), firms' costs irrespective of gender and age

Note: When determining firm size "employee" means full-time equivalent employee.

Table 2: Summary statistics: firms by size and sector

	All firms		Manufacturing		White collar		Public sector	
Number of employees	10-30	> 30	10-30	> 30	10-30	> 30	10-30	> 30
Share female	0.41	0.33	0.29	0.20	0.50	0.45	0.69	0.64
	(0.29)	(0.25)	(0.25)	(0.15)	(0.23)	(0.20)	(0.26)	(0.18)
Share female under 35	0.40	0.35	0.27	0.22	0.51	0.48	0.66	0.66
	(0.35)	(0.25)	(0.30)	(0.15)	(0.31)	(0.19)	(0.33)	(0.19)
Mean age of employees	43.16	42.49	42.78	42.30	42.79	41.36	44.40	43.30
	(4.58)	(2.74)	(3.97)	(2.12)	(5.17)	(2.96)	(4.69)	(3.52)
Monthly full-time wage, women under 35	1858.58	2773.41	1714.22	2947.09	2224.39	2790.01	2029.98	2397.45
	(707.84)	(674.92)	(640.39)	(669.83)	(798.66)	(721.51)	(643.94)	(474.63)
Monthly full-time wage, men under 35	2104.75	2963.99	2049.39	3110.26	2537.27	3084.55	2226.35	2630.82
	(731.21)	(684.00)	(591.64)	(629.61)	(1025.74)	(804.47)	(774.09)	(626.45)
Share of women working full-time	0.64	0.74	0.67	0.79	0.69	0.67	0.54	0.63
	(0.31)	(0.19)	(0.32)	(0.15)	(0.30)	(0.19)	(0.30)	(0.21)
Share of men working full-time	0.86	0.94	0.92	0.95	0.84	0.91	0.76	0.87
	(0.22)	(0.12)	(0.16)	(0.22)	(0.25)	(0.17)	(0.27)	(0.16)
Share female of FT employees	0.35	0.28	0.244	0.17	0.43	0.37	0.57	0.56
	(0.29)	(0.23)	(0.24)	(0.13)	(0.24)	(0.17)	(0.29)	(0.20)
Number of new hires	4.15	32.24	4.95	423.16	25.89	197.18	6.37	143.81
	(14.96)	(94.53)	(14.10)	(909.80)	(143.56)	(264.26)	(8.21)	(242.34)
Number of newly hired women	1.96	14.35	1.41	110.04	16.05	79.83	3.92	85.22
	(10.31)	(37.03)	(5.47)	(207.42)	(103.65)	(96.52)	(5.02)	(124.10)
At least one new hire (=1)	0.84	0.96	0.89	0.99	0.86	1	0.89	0.99
	(0.37)	(0.17)	(0.35)	(0.07)	(0.324	-	(0.31)	(0.07)
At least one woman hired (=1)	0.56	0.85	0.48	0.94	0.69	0.97	0.81	0.98)
	(0.49)	(0.36)	(0.50)	(0.24)	(0.46)	(0.16)	(0.39)	(0.12)
Founding year of the firm	1987.72	1985.55	1988.72	1981.5	1988.61	1983.16	1986.50	1984,87
	(8.51)	(8.90)	(8.20)	(8.92)	(9.30)	(9.85)	(8.79)	(1744.97)
Individual-year observations	251,325	5,348,020	116,773	2,966,185	35.783	603.512	70,839	1,498,442
Individual-year observations under 35	64,411	1,359,753	31,830	746,965	9,970	183,849	14,250	364,158
Firm-year observations	10,658	17,822	5,428	8,679	1,458	2,209	2.539	5,648

Notes: Number of employees denote full-time equivalent employees. Table entries are calculated at the individual-year and firm-year level, pooled over the years 2001-2010. The three main sectors depicted in the table do not cover all workers in the sample, i.e. the numbers of observations by sector do not sum horizontally to the numbers given for all firms in columns 1 and 2.

Table 3: Within large firm gender wage gap

	All employees				New hires			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-period × female	0.0259***	0.0113***			0.0033	0.0149		
	(0.0082)	(0.0040)			(0.0207)	(0.0091)		
Time trend			0.0027	-0.0010			-0.0049	0.0002
			(0.0025)	(0.0015)			(0.0084)	(0.0032)
Post-2006 trend			0.0053	0.0065***			0.0134	0.0051
			(0.0037)	(0.0023)			(0.0129)	(0.0047)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	-	Y	-	Y	-	Y	-	Y
Controls	-	Y	-	Y	-	Y	-	Y
Clusters	2,008	2,007	2,008	2,007	1,998	1,975	1,998	1,975
N	1139325	1139324	1139325	1139324	217,225	217,202	217,225	217,202

Notes: Coefficients show the interaction term of equations (2) and (3). Controls: Age of firm, region at NUTS 1 level (federal states) and sector (German one-digit classification for sectors from 2003). Standard errors are clustered at the firm level, observations denote individual by year observations. */**/*** denote statistical significance at the 10/5/1% level respectively.

Table 4: Within large firm gender wage gap - above median wages

	All employees				New hires			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-period × female	0.0145**	0.0096*			0.0225	0.0203		
	(0.0067)	(0.0054)			(0.0166)	(0.0156)		
Time trend			-0.0019	-0.0028			-0.0043	-0.0063
			(0.0022)	(0.0019)			(0.0055)	(0.0052)
Post-2006 trend			0.0090***	0.0091***			0.0135*	0.0170**
			(0.0033)	(0.0028)			(0.0079)	(0.0085)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	-	Y	-	Y	-	Y	-	Y
Controls	-	Y	-	Y	-	Y	-	Y
Clusters	528	528	528	528	524	521	524	521
N	622,968	622,968	622,968	622,968	99,955	99,952	99,955	99,952

Notes: See Table 3

Table 5: Within large firm gender wage gap - below median wages

	All employees				New hires			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-period × female	0.0100	0.0143**			-0.0059	0.0117		
	(0.0142)	(0.0063)			(0.0354)	(0.0099)		
Time trend			-0.0028	0.0025			-0.0141	0.0035
			(0.0046)	(0.0020)			(0.0128)	(0.0041)
Post-2006 trend			0.0099	0.0016			0.0286	-0.0007
			(0.0061)	(0.0029)			(0.0179)	(0.0059)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	-	Y	-	Y	-	Y	-	Y
Controls	-	Y	-	Y	-	Y	-	Y
Clusters	1,478	1,477	1,478	1,477	1,472	1,452	1,472	1,452
N	516,302	516,301	516,302	516,301	117,255	117,235	117,255	117,235

Notes: See Table 3

Table 6: Within large firm gender wage gap - by sectors

	Manufacturing				White collar				Public sector			
	All		New hires		All		New hires		All		New hires	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post-period × female	0.0113***		0.0149		0.0207***		0.0201		0.0175**		0.0341*	
	(0.0040)		(0.0091)		(0.0048)		(0.0131)		(0.0075)		(0.0180)	
Time trend		-0.0010		0.0002		0.0016		-0.0113*		-0.0011		0.0054
		(0.0015)		(0.0032)		(0.0021)		(0.0065)		(0.0027)		(0.0051)
Post-2006 trend		0.0065***		0.0051		0.0042		0.0246**		0.0073*		0.0019
		(0.0023)		(0.0047)		(0.0032)		(0.0115)		(0.0040)		(0.0070)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters	2,007	2,007	1,975	1,975	1,004	1,004	994	994	259	259	254	254
N	1139324	1139324	217,202	217,202	700,156	700,156	109,276	109,276	147,259	147,259	35,626	35,626

Notes: See Table 3

Table 7: Full-time wages of female employees - small vs. large firms

	All employees				New hires			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Interaction term								
Post-period × large firm	0.0323	0.0418***			0.0788	0.0326		
	(0.0247)	(0.0113)			(0.0568)	(0.0226)		
Time trend			0.0194**	0.0181***			0.0242	0.0102
			(0.009)	(0.0042)			(0.0213)	(0.0104)
Post-2006 trend			-0.0198*	-0.0126*			-0.0119	-0.0026
			(0.0118)	(0.007)			(0.0287)	(0.017)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	-	Y	-	Y	-	Y	-	Y
Controls	-	Y	-	Y	-	Y	-	Y
Clusters	2,960	2,899	2,960	2,899	2,731	2,340	2,731	2,340
N	379,777	379,672	379,777	379,672	83,283	82,892	83,283	82,892

Notes: See Table 3

Table 8: Share of female employees - large vs. small firms

	All employees				New hires			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Interaction term								
Post-period × large firm	0.0138	0.0132*			0.0305**	0.0258***		
	(0.0101)	(0.0069)			(0.0126)	(0.0095)		
Time trend			0.0003	0.0015			0.0143**	0.0110**
			(0.0041)	(0.0025)			(0.0061)	(0.0045)
Post-2006 trend			0.0043	0.0020			-0.0160*	-0.0107
			(0.0054)	(0.0036)			(0.0088)	(0.0072)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	-	Y	-	Y	-	Y	-	Y
Controls	-	Y	-	Y	-	Y	-	Y
Clusters	3,316	3,272	3,316	3,272	3,271	3,144	3,271	3,144
N	27,328	27,284	27,328	27,284	22,737	22,610	22,737	22,610

Notes: Coefficients show the interaction term of equations (2) and (3). Controls: Age of firm, region at NUTS 1 level (federal states) and sector (German one-digit classification for sectors from 2003). Standard errors are clustered at the firm level, observations denote firm by year observations and are weighted by the number of employees. */**/*** denote statistical significance at the 10/5/1% level respectively.

Table 9: Share of female employees - small vs. large firms - by sectors

	Manufacturing				White collar				Public sector			
	All		New hires		All		New hires		All		New hires	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post-period × large firm	0.0132*		0.0258***		0.0158*		0.0275**		0.0191		0.0369	
	(0.0069)		(0.0095)		(0.0083)		(0.0129)		(0.0261)		(0.0315)	
Time trend		0.0015		0.0110**		0.0019		0.0169***		-0.0022		-0.0012
		(0.0025)		(0.0045)		(0.0029)		(0.0060)		(0.0085)		(0.0117)
Post-2006 trend		0.0020		-0.0107		0.0030		-0.0197**		0.0080		0.0092
		(0.0036)		(0.0072)		(0.0042)		(0.0094)		(0.0121)		(0.0176)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters	3,272	3,272	3,144	3,144	1,662	1,662	1,592	1,592	439	439	423	423
N	27,284	27,284	22,610	22,610	13,667	13,667	11,173	11,173	3,499	3,499	2,898	2,898

Notes: see Table 8