

Feldhoff, Charlotte H.

**Working Paper**

## The child penalty: Implications of parenthood on labour market outcomes for men and women in Germany

SOEPpapers on Multidisciplinary Panel Data Research, No. 1120

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Feldhoff, Charlotte H. (2021) : The child penalty: Implications of parenthood on labour market outcomes for men and women in Germany, SOEPpapers on Multidisciplinary Panel Data Research, No. 1120, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/231531>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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

1120 2021

**SOEP** papers  
on Multidisciplinary Panel Data Research

# The Child Penalty: Implications of Parenthood on Labour Market Outcomes for Men and Women in Germany

Charlotte H. Feldhoff

## **SOEPPapers on Multidisciplinary Panel Data Research** at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at <http://www.diw.de/soeppapers>

### **Editors:**

Jan **Goebel** (Spatial Economics)

Stefan **Liebig** (Sociology)

David **Richter** (Psychology)

Carsten **Schröder** (Public Economics)

Jürgen **Schupp** (Sociology)

Sabine **Zinn** (Statistics)

Conchita **D'Ambrosio** (Public Economics, DIW Research Fellow)

Denis **Gerstorff** (Psychology, DIW Research Fellow)

Katharina **Wrohlich** (Gender Economics)

Martin **Kroh** (Political Science, Survey Methodology)

Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Fellow)

Thomas **Siedler** (Empirical Economics, DIW Research Fellow)

C. Katharina **Spieß** (Education and Family Economics)

Gert G. **Wagner** (Social Sciences)

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)

DIW Berlin

Mohrenstrasse 58

10117 Berlin, Germany

Contact: [soeppapers@diw.de](mailto:soeppapers@diw.de)



# The Child Penalty: Implications of Parenthood on Labour Market Outcomes for Men and Women in Germany

Charlotte H. Feldhoff

January 2021

Whilst gender inequality has been falling in the developed world, child-related gender inequality in pay has stayed constant. In this paper I use German panel data spanning across 33 years from 1984 until 2017 including over 50,000 individuals. The main contribution of this paper is the analysis of the effect of parenthood on women's and men's earnings using propensity score matching. I estimate the annual average treatment effect of parenthood over the 20 years following the birth of the first child to be -10500€ for women and +6800€ for men. When comparing the percentage loss of potential earnings, I find that women suffer a long-run child penalty of 63% compared to men. I then examine the relationship between the treatment effect and gender norms, willingness to take on risk for your career and priorities regarding job characteristics. There exists evidence which suggests that all of these factors are associated with changes in the individual treatment effects.

JEL Codes: J13, J16, J21, J31, C12, C13

# 1

## Introduction

Over the last century gender inequality has been declining in all areas of life in the developed world. However, this decline has been slowing over time and has now come to a halt. The main components of gender-inequality used to be education-related inequality and child-related gender inequality. However, nowadays women and men attain similar levels of education, which has meant that education-related inequality has almost disappeared (*Share of Age Group with Specified Level of Educational Attainment by Gender* 2010). Whilst education-related inequality has been falling dramatically, child-related inequality has not only stayed the same but even increased (Avellar and Smock 2003; Kleven, Landais, and Søgaard 2019). This leaves child-related gender inequality as the single most important factor driving gender inequality in earnings today (Kleven, Landais, and Søgaard 2019).

This paper contributes to the literature on child-related gender inequality in the labour market. The immense literature on gender inequality in the labour market is summarized by for example Blau and Kahn (2017) and Olivetti and Petrongolo (2016). There is not one definition of gender inequality in income but rather many different definitions measuring different types of gender inequality in earnings. The most cited definition measures the difference in earnings between men and women who do an equal amount of work. Here gender equality is stipulated as equal pay for equal work. That gap tries to capture discrimination against women

on the labour market and is estimated to be around 16% in the EU (European commission 2017). The definition of gender inequality in earnings used in this paper, captures the idea that gender inequality is much more than the differences in wages for equal work. It takes into consideration that many women cannot carry out equal work as in the majority of cases they are still the ones responsible for household production. Not controlling for hours worked, this total gap in earnings is around 39% in the EU (European commission 2017). The latter type of gender inequality becomes especially interesting when studying the difference in the effect of childbirth on women's and men's earnings as having children makes it even more difficult for women to carry out equal work.

This paper focuses on estimating the overall effect of children on income for women compared to men in Germany whilst not controlling for labour market outcomes. This difference in effect for men and women is coined the child penalty. For this purpose, I define the child penalty to measure the total effect of all children as I do not condition on the number of children. There are two leading reasons why I believe that the child penalty is so important and requires further studying. First, Kleven, Landais, Posch, et al. (2019) found that the short-run child penalty i.e. the child penalty 3 years after the first child was born, is of large magnitude with it varying from 25% to up to 80% in the 6 countries<sup>1</sup> studied. Second, besides its magnitude the child penalty is also found to be extremely persistent. Kleven, Landais, and Sogaard (2019) and Kleven, Landais, Posch, et al. (2019) find the difference in income due to the child penalty persisting at large magnitudes even 20 years after the birth of the first child. There are several reasons for choosing Germany as the country of study. First, there exists rich administrative data which is required for the analysis. Second, Germany experiences one of the largest child penalties in the developed world making it an interesting case to study. Third, it allows the re-evaluation of Germany's child penalty as determined by Kleven, Landais, Posch, et al. (2019) using matching methods.

---

<sup>1</sup>UK, US, Germany, Austria, Denmark and Sweden

For the purpose of this study I use data from the German Socio-Economic Panel Study (SOEP 2019). In this paper propensity score matching is used to estimate the child penalty. Matching is used to pair parents with similar individuals of the same sex who do not have children. This then allows me to estimate the counterfactual income which is the hypothetical income the individual would have earned if the individual would have chosen not to have children. Matching has several advantages over the event study approach used by Kleven, Landais, and Søgaard (2019) and Kleven, Landais, Posch, et al. (2019). The key advantage of matching is that it introduces heterogeneity in the treatment effects. In particular this improves the estimation of the long-run child penalty which relies on a lot of extrapolation in the event study approach. Heterogeneity is also advantageous because it permits the further analysis of the relationship between individual treatment effects and risk preferences, gender norms and priorities regarding job characteristics. To the best of my knowledge, only one other paper by Simonsen and Skipper (2006) has used matching to estimate the effect of children on income. The key difference is that they estimate women's yearly reduction in earnings using cross-sectional data instead of panel data. Hence, contrary to my paper, they cannot analyse the development of the child penalty.

By and large, my results can be summarized into 5 main findings. The first is that the average effect of children on women's income is negative. Women lose on average 10500€ in income per year in the 20 years after childbirth whereas men gain almost 7000€ in income per year. Second, the long-run child penalty which is the average child penalty from event time 5 to 20, where the event is the birth of the first child, amounts to 63%. Third, I find that more conservative views on gender roles are associated with lower income for fathers relative to men without children. For women the reverse holds with mothers having a smaller loss in their income relative to their counterfactual income when they have more modern gender views. Fourth, fathers get punished in terms of their income relative to their counterfactual income for being more risk averse with regards to their career whereas this does not seem to be the case for women. Finally both men and women for

whom family friendliness is a prioritized job characteristic get penalised compared to their childless peers. The reverse is the true if income or career advancement are a priority. Evidently, gender norms have an effect on male and female preferences and hence should be considered in combination. These results are in line with what the large majority of the literature finds.

The outline of the paper is as follows. It starts off by giving a brief overview of the recent literature in section 2 which is followed by the methodology in section 3 of this paper. In chapter 4 I proceed by giving an overview of the data which is succeeded by my results in chapter 5. In chapter 6 I discuss the strengths and weaknesses of the methods used and interpret my results. Finally, I finish with a brief conclusion in chapter 7.



# 2

## Literature Review

The effect of children on women can be split up into pre- and post-fertility effects. In the literature, estimates of the child penalty usually only encompass post-fertility effects as they are easier to quantify whereas pre-effects need to be somehow proxied. Pre-fertility effects are defined as the effects anticipated fertility brings about, for example not pursuing further education or choosing a different career path (Budig and England 2001). One approach is to proxy pre-fertility effects using choices regarding education (Kleven, Landais, and Sogaard 2019; Kleven, Landais, Posch, et al. 2019). If one assumes that pre-effects can indeed be proxied by education choices, then recent data indicates that pre-effects have reduced over time as men and women complete similar levels of education in developed countries today. There still remain differences in the type of education completed with a lower percentage of women completing degrees in STEM subjects. Therefore, there seems to have been a shift from pre-effects to post-effects making up the majority of the child penalty. The post-fertility effect is the effect captured by most estimates and can be defined as the impact of childbirth on a woman's labour choices. Examples include working fewer hours or stopping working temporarily or permanently (Paull 2008; Fitzenberger, Sommerfeld, and Steffes 2013). In this paper I abstract from pre-fertility effects when estimating the treatment effect and the child penalty. Whilst I expect pre-fertility effects to be considerably smaller than post-fertility

effects, I conjecture that they exist and hence my estimate of the treatment effect is a lower bound of the total effect of fertility.

The impact of childbirth on earnings can be explained by changes in the extensive and/or the intensive margin and/or the wage rate. The extensive margin of labour supply captures whether the person is in employment or not whereas the intensive margin measures the numbers of hours worked.

In the seminal paper by Kleven, Landais, and Sogaard (2019), the authors conduct an event study which investigates the impact of children on earnings using data from Denmark. They determine the child penalty by running separate regressions for men and women on income with event dummies, age dummies and year dummies as independent variables. The paper has 3 main findings. The first contribution establishes that Denmark has a persistent long-run child penalty of 21%. In their follow up paper in 2019 they replicate this study for 5 other countries finding long-run child penalties ranging from 21% to 51% across the 6 countries (Kleven, Landais, Posch, et al. 2019). These results show that women and men across the developed world are substantially affected by child-related gender inequality. The second insight is that when decomposing gender inequality in earnings, it becomes clear that child-related gender inequality has gained in importance making up 80% of gender inequality in 2013 compared to only 40% in 1980 <sup>2</sup>. This highlights the importance of the child penalty for today's gender inequality. The third insight of the paper is that the persistence of the child penalty could be transmitted between generations. By looking at child penalties of grandmothers and grandfathers they show that parents pass on the child penalty to daughters but not to sons. The aforementioned phenomenon shows how childhood environment affects women's preferences over family and career which in turn impacts the child penalty.

In this paper I try to build on those findings by attempting to show the causal effect of children on income using propensity score matching. Furthermore, I attempt to get a clearer answer to how preferences of women and men are associated with the impact of childbirth on their income.

---

<sup>2</sup>See figure A.1 in the appendix

Matching has two main advantages over the event study approach used by Kleven, Landais, and Søggaard (2019) and Kleven, Landais, Posch, et al. (2019). The first advantage is that it allows the introduction of heterogeneity in the treatment effects (Simonsen and Skipper 2006). In the seminal paper by Kleven, Landais, and Søggaard (2019), identification of the event coefficients is achieved by variation in the age at first birth and in the year of birth of the parents. This means that the event coefficient is the same across individuals of the same gender in the same event time. Furthermore, the predicted wage is simply the sum of the coefficients for the age and the dummies meaning that individuals of the same sex, age and birth year have the same predicted wage. Thus, the impact of children on income is the same for individuals of the same sex who are of the same age and born in the same year. Matching gives rise to individual specific predicted counterfactual wages which vary with many demographic variables and schooling. The counterfactual wage has been determined by evaluating the income of a control group which makes the estimates based on actual counterfactual data. Moreover, this is especially useful when trying to understand what characteristics are associated with treatment effects in the long run. In the event study approach the long-run child penalty relies on heavy extrapolation which makes it more prone to bias. Hence, especially when calculating the long-run child penalty, the event study approach is inferior to matching.

The ability to control for background characteristics is the second advantage matching has over the event study approach. When matching on background characteristics, matches only occur between individuals who are comparable in their demographic background including their age, year of birth, country of birth, secondary schooling completed by the age of 20 etc. Therefore, it does not only take into account the effect of age and year of birth on income but also other pre-existing characteristics for example regional differences in pay and the returns to schooling. This allows one to determine a child penalty based on the individual's characteristics which play an immense role in future income and fertility. Furthermore, by matching on observed schooling one also partly controls for differences in unobservables, for example grit. Unobservables are likely to have a substantial impact on future income

and fertility, in particular those unobservables regarding work ethic. Matching allows me to at least partly control for those unobservables. By only considering the effects of age and year of birth on income, one misses out on capturing the effect of many of those factors affecting individual child penalties. Thus, matching allows me to determine a more accurate causal relationship of childbirth on earnings.

To the best of my knowledge there exists only one other paper which has used propensity score matching in the past to examine the effect of children on income (Simonsen and Skipper 2006). It finds the direct effect of motherhood to be less than 2%. There exist 5 key differences between their paper and this paper. First, Simonsen and Skipper (2006) only look at the motherhood penalty and do not consider the effect on men. Second, they use Danish cross-sectional data instead of German panel-data which leads to a smaller sample size studied and does not allow them to determine the difference in the effect of childbirth on earnings in each event time. Third, they focus on women who are fully employed which means that they focus on the pay gap for equal work rather than the total gap caused by childbirth. The fourth difference is that they use regression-adjusted propensity score matching and not propensity score matching by itself. Finally, the fifth difference is that they do not look into the relationship between treatment effects and other variables of interest including gender norms. Hence, even though similar methods are deployed in both papers, the paper focuses on a different part of the child-related gender inequality and only considers women in isolation instead of comparing them to men. I find that the convergence in income which is mainly caused by childbirth is driven not only by the effect of children on women's income but also their effect on men's income. This conclusion could only be drawn by looking at both women and men and allows for a more thorough analysis of the drivers of the child penalty.

Having looked at the evidence proving not only the existence of the child penalty but also its prevalence and magnitude, one can say that it is probably the most significant issue regarding gender inequality today. The next question one needs to ask oneself is why this phenomenon exists. The literature believes that there are 4 possible explanations. The first focuses on the biological differences between women

and men. The reasoning goes as follows: with women giving birth there is not only a short-run effect on the mother's health but also a long-run one as mothers might become more attached to their children than fathers. This effect on their health and their available time impacts their earning and productivity. The second possible explanation could be that as on average men earn more than women, the household decides it's best for the women to 'specialise' in childcare and the father to 'specialise' in working. The third potential reason is that societal views about gender norms automatically lead to women choosing childcare over work whilst the opposite holds true for men. The fourth explanation builds on that idea that women's preferences over family and work is different to those of men (Nix, Andresen, et al. 2019).

Papers which compare child penalties of heterosexual parents to those of same-sex female parents find that biological differences cannot explain the long-run penalty (Nix, Andresen, et al. 2019; Moberg 2016). They learn that whilst heterosexual parents experience a persistent long-run child penalty which is solely borne by the mother, same-sex couples only experience a child penalty for two years following childbirth which is borne by both. They further contrast couples with a similar comparative advantage differential and find that the comparative advantage does not affect the child penalty. Consequently, they discard the second possible explanation i.e. specialisation. Hence, both of those reasons cannot explain the long-run child penalty leaving societal views and preferences as possible explanation which is why this paper focuses on the effects of societal views and preferences on the child penalty.

There have been several studies trying to link gender roles to gender inequality. Kleven, Landais, and Sjøgaard (2019) shows how societal views on women's labour market participation change with childbirth. Before childbirth around 90% of sampled individuals believe women should work full-time which drops to 15% for women with children under school age. They then plot gender norms against the child penalty and find a strong positive correlation. In this paper I try to analyse this further by running regressions to determine the correlation between the individual treatment effects and gender norms.

Lastly, there is an argument of preferential differences between women and men driving the child penalty. Several studies have looked at intergenerational transmission of preferences (Fernández, Fogli, and Olivetti 2004; Kleven, Landais, and Søgaaard 2019). Kleven, Landais, and Søgaaard (2019) finds a positive correlation between a mother's child penalty and that of her daughter. In contrast, Fernández, Fogli, and Olivetti (2004) finds labour force participation of married women to be positively correlated with the labour force participation of their husbands' mothers. The question whether female child penalties are driven partly by female preferences or by male preferences formed during childhood remains unresolved. Later in this paper I try to build on this to give more evidence on how preferences and child penalties are correlated. A key driver of differential pay is the variability of earnings. If an individual's earnings fluctuate a lot, the individual gets rewarded for bearing this risk with a wage premium (Bertrand 2011). Hence, by examining differences in risk preferences I might be able to explain part of the child penalty through differences in risk aversion between men and women. Furthermore, I look at the relationship between job characteristics and the individual treatment effect as many papers have found that job/sector characteristics, for example family friendliness, have large effects on wages paid. Thus, men prioritising these characteristics could explain a piece of the child penalty (Gibb et al. 2014; Lundberg and Rose 2000; Petersen, Penner, and Høgsnes 2011).

# 3

## Methodology

### 3.1 Intuition

In this paper I try to estimate how the effect of having children on income differs between men and women. In an optimal scenario, fertility would be randomly assigned to the individual which would eliminate selection bias. Selection bias stems from individuals selecting themselves into either group which I expect would lead to individuals who choose to have children to differ from those who choose not to have children. Therefore, if fertility was randomly assigned there would be no selection bias. This would mean that differences in income would only be caused by differences in fertility and not by other intrinsic differences. However, in observational data, treatment is not assigned randomly and hence one needs to be concerned with the differences besides treatment between the treated and untreated individuals. In this study treatment is defined as having at least one child. There are various approaches to mimic a randomized control trial including for example difference-in-differences analysis. The reason I chose matching is because it is superior to most other approaches in that it allows me to analyse a much longer period of time. Matching tries to mimic a randomized trial by ensuring that the control and treatment groups are as similar as possible in selected variables i.e. trying to make treatment the only difference between the control and the treatment group.

To be able to proceed I need to introduce some notation. The variable  $t$  measures the event time where the event is childbirth. Thus, event time  $t$  equals to 0 at childbirth and states the time period relative to the birth of the first child. The actual income of a treated individual is denoted by  $Y_i(1)$  where  $i$  stands for individual  $i$ . Besides  $Y_i(1)$  I am also interested in  $Y_i(0)$  which is the income the treated individual would have earned if she or he had chosen not to have children. To get the treatment effect for treated individual  $i$  I simply need to take the difference between  $Y_i(1)$  and  $Y_i(0)$ . I now face the fundamental problem of causal inference which says that in practice one can never observe both outcomes  $Y_i(0)$  and  $Y_i(1)$  for one individual at one moment in time as the individual either has children or does not have children (Simonsen and Skipper 2006). Hence, I observe  $Y_i(1)$  but not  $Y_i(0)$  for a treated individual. This means I need to estimate the counterfactual wage  $Y_i(0)$  which is what matching achieves.

Matching overcomes the fundamental problem of causal inference by balancing the treatment and the control group in its covariates mimicking random treatment assignment. The idea behind matching is to assign childless individuals to individuals with children to whom they are most similar to. This allows me to predict the counterfactual for the treated  $Y_i(0)$  i.e. the earnings individual  $i$  would have earned if they had chosen not to have children. The individuals need to be as similar as possible in characteristics which matter for treatment assignment and the outcome. As I expect there to be a significant difference between the impact of children on men compared to women, I allow matches to only happen between individuals of the same sex. Thus, if matching achieves well balanced groups in all the relevant covariates, the average difference in income between the treatment and the control group can be given a causal interpretation. Hence, matching allows me to determine the effect of childbirth on individual  $i$  by calculating the difference between  $Y_i(1)$  and  $Y_i(0)$ .



## 3.2 Matching

### 3.2.1 Matching in theory

There are many different matching techniques one can use. I use propensity score matching which was first introduced by Rosenbaum and Rubin (1983). A propensity score is a conditional probability measure of treatment participation given those observables. In this paper it measures how likely a person is to have a child given those observables.

$$p(\mathbf{x}) = Pr[D = 1 | \mathbf{X} = \mathbf{x}] \quad (3.1)$$

In the above equation  $p(\mathbf{x})$  symbolises the propensity score for the observables  $\mathbf{x}$  e.g. age and year of birth.  $D$  represents a dummy variable which equals one for individuals with children and zero for childless individuals. One of the reasons why propensity scores are so widely used is that they overcome the problem of dimensionality by summarizing all of the covariates into one scalar. Therefore, matching has to balance similarity along only a single dimension instead of multiple dimensions. Hence, individuals can be compared using propensity scores which allow matching on the distributions of covariates without requiring exact matches for each covariate (Stuart 2010). This increases the overall number of matches. Furthermore, propensity score approaches reduces extrapolation and successive dependence on the outcome model specification which leads to more robust inferences (Ho et al. 2007). There are three main assumptions which need to be satisfied to allow for casual inference:

**Assumption I (Ignorable treatment assignment or hidden bias)**

$$(Y_0, Y_1) \perp D | X \quad (3.2)$$

This assumption originates from the conditional independence assumption stipulated by Rosenbaum and Rubin (1983). Assumption I states that treatment assignment ( $D$ ) is independent of the potential outcomes given the covariates ( $X$ ) (Rosenbaum and Rubin 1983). The assumption is rather strong as it assumes that

all variables relevant to treatment assignment and outcome are controlled for. If this assumption holds matching can recreate random treatment assignment and hence allows the estimation of an unbiased counterfactual and hence treatment effect. This assumption can be adapted to the propensity score specification where the focus is on the probability of parenthood and not on the set of covariates. Thus, I condition on the probability of having children instead of conditioning on all the covariates. As discussed previously, this overcomes the non-parametric issue of dimensionality with the propensity score being a scalar.

$$(Y_0, Y_1) \perp D \mid p(\mathbf{x}) \quad (3.3)$$

There are many possible ways to check whether propensity scores achieved adequate similarity in covariates between treatment and control groups. A common method to analyse how well matching has worked is to examine the balance of the covariates post matching. If the covariates include all information relevant to treatment assignment and outcome, then analysing the balance will test whether the hidden bias assumption holds. In an empirical setting, it is close to impossible to observe all of this information as some important variables e.g. work ethics are unobservables. For the time being I assume that all the right covariates have been picked and later, in section 3.2.2, discuss how likely this is.

If matching worked optimally the distributions of the covariates are exactly the same in both the control and the treatment groups. The optimal method to evaluate the quality of the matches is to compare the multidimensional histogram of the covariates. However, empirically it is best to inspect lower-dimensional summaries including the mean (Stuart 2010). A popular numerical balance diagnostic used to analyse the similarity of the means is the standardized bias. The standardised bias is calculated by dividing the difference of the means between the control and the treatment group by the standard deviation of the treatment group (Rosenbaum and Rubin 1985):

$$\frac{\bar{X}_t - \bar{X}_c}{\sigma_t} \quad (3.4)$$

In addition to the similarity in means, I further require the sample to have similar marginal distributions of covariates. One approach to examine whether this is the case is to look at the ratio of the variances of the treatment and control group for each covariate. The better the balance in marginal distributions, the closer the variance ratios are to one. After carrying out matching there should be no statistically significant differences between covariate means of the treated and comparison units. These mean comparisons can be contrasted with the unconditional means of the treatment and the control group before matching, which are likely to be statistically significant in most cases.

**Assumption II (Overlap or common support)**

$$X: 0 < Pr(D = 1 | X) < 1 \forall X \quad (3.5)$$

Assumption II specifies that for each value of  $X$  there is a positive probability of being both treated and untreated. To judge whether this assumption holds one should check whether there is common support by visually inspecting the densities of the propensity scores of the treatment and the control group. When assumption I and II hold, the treatment assignment is said to be strongly ignorable (Rosenbaum and Rubin 1983). However, for causal interpretation assumption III also needs to hold.

**Assumption III (Stable Unit Treatment Value Assumption)**

The stable unit treatment value assumption was first popularized by Angrist, Imbens, and Rubin (1996). It implies that the outcome of one individual is not affected by treatment assignment to any other individual i.e. no spillover effects. In this case it would mean that individual  $i$ 's outcome i.e. income is not affected by the probability of individual  $j$  having children. Furthermore, it assumes that treatment is the same for everyone meaning that there is only a single version of the treatment.

I believe that both components of this assumption are likely to be satisfied here. Arguably, an individual having children affects the people around them. For example, your sister having a child might mean that you occasionally help her to look after the child. However, it is very unlikely that due to those caring responsibilities you start earning less. Excluding step-fathers and step-mothers, treatment should not have

an effect on other people's income. Additionally, most individuals in the sample are unlikely to have ever met so the treatment of one individual should not directly affect another individual's income. The second component of the assumption should also be satisfied with treatment being the same for everyone. Of course one can say that treatment differs for individuals who give birth to twins compared to those who give birth to only one child or that some mother's health are affected more severely than others. However, on average treatment i.e. having children is clearly defined and its effect should be similar for everyone besides some rare exceptions.

How well these assumptions hold in my sample are shown in the results section. If these assumptions are satisfied, the average differences in outcomes between the matched treated and control individuals are attributable to the treatment.

Once matching on the observables is completed, I derive at the counterfactual outcome i.e. the hypothetical wage  $Y_i(0)$  for individuals with children. This gives me the possibility to determine the individual and the average treatment effect for an individual. Here  $g$  indicates the sex of the individual which is included to capture the belief that childbirth has different effects on income for men and women. The individual treatment effect is simply the difference between the income individual  $i$  actually earned ( $Y_{ist}^g$ ) and the income they would have earned ( $\hat{Y}_{ist}(0)^g$ ).

$$Treatment\ effect_{ist}^g = Y_{ist}(1)^g - \widehat{Y_{ist}(0)^g} \quad (3.6)$$

The average treatment effect measures how having children affects the average male's income and the average female's income.

$$ATE^g = \mathbb{E}[Y(1)^g | X] - \mathbb{E}[Y(0)^g | X] \quad (3.7)$$

To calculate the child penalty, one needs to find the percentage effect of having a child on women and men in each event time period. The first step is to determine the mean hypothetical income by averaging across individuals of the same sex at each event time. This leaves me with an average counterfactual wage for all mothers and an average counterfactual wage for all fathers for each event time. The next step is to find the percentage loss or gain each individual incurs from their

hypothetical wage by becoming a parent. The percentage loss for women at event time  $t$  is calculated by dividing the individual treatment effect at event time  $t$  by the average hypothetical income for women at event time  $t$ . For men you calculate the percentage gain by dividing the individual treatment effect at event time  $t$  by the average hypothetical income for men at event time  $t$ .

$$P_{ist}^g \equiv \frac{Y_{ist}(1)^g - \widehat{Y_{ist}(0)^g}}{\mathbb{E}[\widehat{Y_{ist}(0)^g} | t]} \quad (3.8)$$

The variable  $P_{ist}^g$  measures the percentage effect of having a child on labour income for individual  $i$  of sex  $g$  in time period  $s$  at event time  $t$ . The child penalty for each event time is then simply:

$$C_t = \mathbb{E}[P_{ist}^m | t] - \mathbb{E}[P_{ist}^f | t] \quad (3.9)$$

Here  $m$  stands for male and  $f$  for female i.e. the child penalty is the difference of the difference. To be able to analyse the effect of childbirth relative to the year prior childbirth i.e. event time  $-1$ , I set the percentage earned of potential income equal to zero for both men and women.

### 3.2.2 Matching in practice

#### General setting

In addition to choosing the type of matching which I chose to be propensity score matching, one needs to decide on many other components. I opt for a logit treatment model with 1:1 matching. The logit model is the standard matching model used in the literature. To ensure robustness, I also used a probit model which finds almost exactly the same results. This indicates that the model's specification has little effect on the treatment effect. When selecting the matching algorithm, one is faced with a trade-off between bias and efficiency. I opt for 1:1 nearest neighbour matching meaning that each treated individual gets matched to a maximum of one individual. The reason I chose 1:1 matching is the fact that it produces higher quality matches than  $k:1$  matching on average and hence is of higher statistical power (Snedcor and Cochran 1980). 1:1 matching minimises bias but ignores some of the information

which may lead to a reduction in efficiency. However, given the large sample size I am not as concerned about having fewer matches if it leads to higher quality matches.

Furthermore, I allow for replacement which means that an individual in the control group can be matched to more than one treated individual if he or she is also similar to another treated individual. Replacement is a useful tool to increase the size of the control group relative to the treated group which makes the estimator more stable. Thus, using replacement increases the total number of matches and by using a caliper one can ensure that the matches are of high quality (Stuart 2010). The caliper sets a maximum to the distance between the propensity score of the control individual and that of the treated individual. This method ensures that only close enough individuals in terms of propensity score are matched, where close enough is determined by the size of the caliper. Consequently, only individuals who have similar enough covariate distributions can be matched with each other. The literature proposes a caliper of around 0.25 of the standard deviation of the propensity scores which should be reduced in case the control group is smaller than the treated group (Rosenbaum and Rubin 1985; Stuart 2010). As this is the case here, I set the caliper to 0.2 of the standard deviation which ensures high quality matches without causing a large reduction in sample size.

Additionally, I check the robustness of the matching method by deploying two different commands in Stata where one command takes into account that the propensity score is estimated whereas the other one does not. I find that there is close to no difference in the average treatment effect and I hence settle for the more sophisticated method which takes into account that the propensity scores are estimated.

Lastly, I separated the design and the analysis of the study which means that the propensity score process is done without the use of the outcome variable. This reduces bias by ensuring that the design is unaffected by the outcome.

### Choice of variables

The selection of covariates is crucial as the control and the treatment sample need to be as similar as possible for variables that matter for treatment assignment and income whilst not introducing bias. There are two main requirements that the covariates should meet. First, it is vital to include all variables in the matching process which are related to both the treatment assignment and the outcome (Stuart 2010). Thus, factors which are believed to influence self-selection should be included in the set of covariates. When control individuals are compared to treated individuals one would want them to be similar as possible in characteristics which matter for labour income and fertility. If I match on variables which are irrelevant for either, matching will not give me a good prediction of the hypothetical income. For example, let's assume I matched the sample perfectly on individuals' preferences on having pets. Even if I match perfectly on those preferences it certainly does not mean I can argue for causal inference of the treatment effect. Preferences on pets are likely to not be related to labour income or treatment assignment and hence, making the two samples similar in terms of that covariate, does not help me in estimating the counterfactual income. Unless these preferences are somehow related to some other factor which matters for either income or fertility this would be a bad choice of covariate.

Second, variables which are affected by the treatment cannot be included as covariates and are better included in the outcome analysis (Stuart 2010). For example if I had chosen labour market experience as a covariate, I would have not had an accurate estimate of  $Y(0)$ . The reason being that childbirth affects labour market experience which matters for income. Therefore, if you match on experience the hypothetical income is not what that individual would have earned if they had chosen not to have children. A mother is more likely to drop out of the labour market, either temporarily or permanently, than a comparable woman without children. Hence, having children directly affects the amount of work experience an individual has. Childless women with low levels of experience will be intrinsically different to mothers with similar levels of work experiences as the reason for the low levels

of experience are likely to be different. Therefore, assuming that income increases with work experience, matching on experience will underestimate  $Y(0)$  for mothers.

The literature has found that it is optimal to include as many covariates as possible if they satisfy the second requirement even if they might not satisfy the first requirement. This practice has been found to minimise bias (Stuart 2010). The only time one should not follow that advice is if the covariate has a lot of missing data and hence significantly increases the amount of missing propensity scores.

Balancing on the right set of covariates is very important which is why I discuss my choices of covariates in the following paragraphs. As previously mentioned, I want to include as many covariates as possible which are expected to affect the outcome and treatment assignment but are itself not affected by the treatment. The first set of covariates I chose to match on are demographic variables including age, year of birth, region of birth, gender, racial background, disability and month of birth. They satisfy the conditions set out as they are deterministic i.e. not affected by treatment and are inclined to be relevant for the outcome and treatment assignment. The reason it is important to match on age is that it allows one to control for life-cycle events which are a key driver of income. Additionally, year of birth is included to control for time trends, for example business cycles and inflation. The reason I control for both country of birth and region of birth, if born in Germany, is to adjust for national and regional differences e.g. quality of schooling. I also match on variables capturing the racial background of the person as many studies have shown that it affects labour market outcomes including income, working conditions and interview offers (Bertrand and Mullainathan 2004; Lian and Matthews 1998; Solé and Parella 2003). Finally disability and severity of disability, the first being a dummy and the latter being measured in percentages, are also matched on. Undoubtedly, disability and its severity influences labour market outcomes. It is assumed here that childbirth does not affect either which holds for almost all pregnancies in Germany.

Apart from matching on demographic variables, I match on type of school certificates. According to the Mincerian wage regression and many other studies,



education is a key determinant of an individual's wage and is hence the reason for its inclusion (Mincer 1974; Breierova and Duflo 2004; Card 1993; Card 2001; T. C. Martin 1995; Psacharopoulos and Hinchliffe 1973; Psacharopoulos 1985; Patrinos and Psacharopoulos 2002). Germany has a three-tier secondary schooling system with three types of general certificates of education. At the age of 9 the average pupil gets advised to go into either of those systems depending on their grades. It is possible to change at a later stage, but this is an exception rather than the norm. The highest tier of education is completed after 12 to 13 years of education, whereas the second tier requires 10 years of schooling and the third requires 9 years of schooling. To go to university the highest certificate is required. By only including individuals who have their child after the age of 20 and with the majority of students choosing which school to go to at the age of 7, the assumption of fertility not affecting the type of certificate the individual achieves should hold.

If matching is done optimally there is no difference in observables and unobservables other than treatment between the matched individuals. This is very hard to achieve as there exist many unobservables which have large impacts on income. However, not including such variables does not imply that I have not controlled for them. If I match on a covariate which is correlated with good work ethics than I have indirectly also matched on work ethics. Hence, the only variables I have to worry about are the ones which are unobservable, relevant and are not correlated with any covariates included in the matching. It is difficult to say how well I have captured these unobservable variables which affect both treatment assignment and income. However, covariates capturing schooling decisions will certainly be correlated with grit, work ethics, aspirations and other important characteristics determining the treatment assignment and outcome. On the whole, there certainly will be some unobservable characteristics not captured here but it is probable that the most important drivers have been matched on, allowing for careful causal interpretation.

### Handling of missing data

In a theoretical setting data availability is always assumed to be perfect. However, this is almost never the case in an empirical setting. The handling of missing data is an important component which requires some consideration when judging the validity of the results. In my sample, missing data arises mainly through the construction of the propensity scores. Propensity scores can only be estimated for individuals who have no missing data for any of the 15 covariates. This leads to an increase in the frequency of missing data for propensity scores. Depending on the type of missing data, different approaches to deal with missing data are optimal. There are three different types of missing data. First, data can be missing completely at random when the reason it is missing is fully independent of observed and unobserved information. Second, it can be missing at random i.e. independent given observable information, or third it can be missing not at random. Stratification of missing data status and conditional-mean imputation are common methods to deal with missing data to increase sample size. However, the literature has found that they have questionable effects on data quality with many increasing bias (Greenland and Finkle 1995). Thus, if sample size is not a restriction, the literature believes that dropping all the individuals with missing data for any of the 15 covariates is one of the better approaches for dealing with missing data. Choi, Dekkers, and Cessie (2019) and many others have shown that this approach often outperforms many of the simple imputation methods. Another approach to deal with missing data could be to use more complex methods including multiple imputation methods which increase sample size. However, given my large sample size I believe that the small gain in sample size, and hence statistical power achieved by employing these sophisticated methods, is outweighed by the complexity of their application. Choi, Dekkers, and Cessie (2019) states that for all three types of missing data, dropping all individuals who have some missing data yielded valid non-biased casual treatment effects. Fortunately, my sample size is very large allowing me to opt for this complete case study approach which only considers the individuals with non-missing values for the propensity score. To summarize,

the method of handling missing data here does not introduce any bias and allows matching to yield a causal treatment effect.

### **Robustness checks**

As previously mentioned, propensity score matching only directly matches on observables and not on unobservables. This is a drawback if unobservables are important for treatment assignment and/or outcome and cannot indirectly be controlled for. The difference-in-differences approach (DID) overcomes this limitation by exploiting the time dimension of panel data. Many individuals in my sample were observed for periods before and after having children. By comparing the individual with themselves before and after treatment, DID controls for unobservable characteristics which do not change over time. Thus, DID reduces unobserved heterogeneity leaving only time-varying unobserved heterogeneity uncontrolled for. Therefore, it gives me a way to quantify unobserved differences between treated and untreated units, relaxing the ‘selection on observables’ assumption. However, it adds the parallel trend assumption which says that income of the control group and the treatment group have to follow the same trend if there would have been no treatment.

To increase the likelihood that the parallel trend assumption holds it is important to remove observed heterogeneity. The more similar the treatment and the control groups are in terms of relevant characteristics which matter for treatment assignment and outcome, the more similar their income growth rates are likely to be. This can be carried out using a DID matching estimator which evaluates the effect on the change in the outcome variable, before and after the intervention. The approach I follow is to match a treatment individual observed in event time -1 to a control individual which gives me a pre-treatment value for the treated and the untreated. I match on the exact same covariates used previously and deploy the same method as I did to find the ATE. Subsequently, I find the change for the treated individual by subtracting his or her pre-treatment income from her post-treatment income in event time one. The reason I chose event time one and not event time zero as the post-treatment period, is because the majority of the

treatment effect materialises the months and years after childbirth. In particular this is the case for children born in December of event time zero. For the control individual I use the observation 2 years after the matched pre-treatment observation to calculate this change. Once those two values are calculated for each individual one can find the DID by averaging this difference for each group and then subtracting this difference for the control group from the treatment group.

$$DID^g = \mathbb{E}[(Y_{post}^g | D = 1, p(x)) - (Y_{pre}^g | D = 1, p(x))] - \mathbb{E}[(Y_{post}^g | D = 0, p(x)) - (Y_{pre}^g | D = 0, p(x))] \quad (3.10)$$

To conclude, if the DID finds a significant effect of similar magnitude to what simple matching finds, unobserved time invariant heterogeneity cannot explain a substantial part of the treatment effect found. Thus, reducing the chance that unobserved heterogeneity can explain the treatment effects found.

### 3.2.3 Outcome analysis using linear regressions

Having calculated the individual treatment effect, I now move on to the second part of this paper which explores possible factors associated with the impact of childbirth on income. I use linear regressions to analyse the treatment effect further which allows me to control for some variables previously not matched on. It is important to note that I look at correlations which cannot be given a casual interpretation.

I run separate regressions for women and men to look for possible correlations between the individual treatment effect and variables of interest. First, I run a multiple linear regression with only the control variables as the independent variable which forms the baseline specification. Here  $i$  stands for individuals and  $s$  for the survey year. It is important to note that  $\pi$  is a vector which contains the coefficients for each of the  $k$  control variables.

$$Treatment\ effect_{is}^g = \alpha + \pi \sum Controls_{is} + v_{is}^g \quad (3.11)$$

Following the baseline regression, I run the same regression but also include the variables of interest. The variables of interest are gender norms, willingness

to take risks for your career and priorities regarding job characteristics. These variables are of interest to me because the literature has stipulated that they may be the underlying reasons of the child-related gender inequality in income (Kleven, Landais, and Sogaard 2019; Gjerdingen and Center 2005; Gibb et al. 2014; Lundberg and Rose 2000; Petersen, Penner, and Høgsnes 2011). I run separate multiple linear regression for each of these categories to overcome sample size restrictions<sup>3</sup>. For gender norms each individual was only surveyed once which means that I now only have cross-sectional data.

$$Treatmenteffect_i^g = \alpha + \delta Gendernorms_i + \pi \sum Controls_i + v_i^g \quad (3.12)$$

$$Treatmenteffect_{is}^g = \alpha + \delta Jobcharacteristics_{is} + \pi \sum Controls_{is} + v_{is}^g \quad (3.13)$$

$$Treatmenteffect_{is}^g = \alpha + \delta Riskpreferences_{is} + \pi \sum Controls_{is} + v_{is}^g \quad (3.14)$$

---

<sup>3</sup>For some of these variables data is only available for one survey year and only for a subset of the individuals. Hence, if all of the variables of interest were included at once there would be very few observations left.

# 4

## Data

### 4.1 Data description

This paper uses data from the German Socio-Economic Panel Study (SOEP) to analyse the child penalty (SOEP 2019; Goebel et al. 2019). The data is not publicly available but can be requested from the Research Data Centre of the German SOEP which is located at the German Institute for Economic Research, DIW Berlin. The SOEP is a longitudinal study which first collected data in 1984 and now collects data on 30,000 individuals and 15,000 households annually (*Forschungsbasierte Infrastruktureinrichtung 'Sozio-oekonomisches Panel (SOEP)'* n.d.). The study gathers key information on households and on individuals over the age of 15 who have lived at some point in Germany. Since 1990 the study also comprises individuals from the new eastern counties previously in the German Democratic Republic.

The survey comprises information on core topics including population, demographics, education, labour market outcomes, income, housing health, preferences and overall satisfaction. This paper makes use of the information on individuals regarding education, demographics, labour market outcomes and data from the mother-child questionnaires. In the second section of this paper, information regarding gender norms, priorities when choosing your job and willingness to take risk are also made use of.

To study the effect childbirth has on parents' income I require detailed panel data on the date of birth of the first child, income and gender. The effect of childbirth for any one individual can only be calculated if the information on all of those three variables exist. Here income is being supplied as an annual figure, representing total gross income from employment. This figure is the sum of the individual's income from their primary and secondary employment and their income from self-employment. All of the income is measured in euros with income from years before the introduction of the euro being converted into euros<sup>4</sup>. Furthermore, I have converted current euro amount into constant euro amounts using consumer price indices provided by the Statistisches Bundesamt using a method proposed by Grabka (2017) (*Verbraucherpreisindex für Deutschland - Lange Reihen ab 1948 - April 2020* 2020). To convert household labour income from a previous period to a more recent period, one multiplies the original household labour income by the ratio of the new consumer price index to the old consumer price index (Grabka 2017).

In this paper I use income, instead of logged income, to allow the inclusion of individuals who report no annual income.

## 4.2 Sample selection

The data set used in this paper is a subset of the SOEP core data set and consists of individual core information and respondent status for each year in a long format. The panel is unbalanced and spans across 33 years starting in 1984 to 2017. The longest I follow a parent for is 26 years as I aim to estimate the long run child penalty which is calculated by tracking individuals from 5 years before to 20 years after the birth of the first child. Before undertaking matching, I consider 27,367 individuals who have a child and 26,495 individuals who do not have children. Here having children refers to having biological children and hence does not consider stepchildren. Therefore, the underlying assumption is that the existence of biological children is of higher importance with regard to its effects on income than the existence of stepchildren. Further the sample is restricted to individuals who are 15 to 65

---

<sup>4</sup>The conversion rate is 1 Euro = 1,95583 DM

years old. The reason for this restriction is that this is the age group which makes up the majority of the work force and hence is the one of interest in this paper. Additionally, only women and men who had their first child between the ages of 20 and 45 are part of the sample. The reasoning behind this is that this study tries to estimate the effect of childbirth on the average parent and hence does not include teenage pregnancies. Furthermore, individuals with children are excluded if they have not finished their core education i.e. secondary education by the time they have their first child. This restriction is in place to ensure that the decision regarding years of core education is less likely to be affected by childbirth which becomes relevant during matching. As this study intends to predict the impact of childbirth on earnings for the German population, only individuals who lived in Germany at the time of the survey are included. Furthermore, the individuals have to have lived at least one year before childbirth in Germany. Again, the reasoning behind this is to see how income earned in Germany is affected by childbirth.

After applying the restrictions discussed above, the sample is left with an average of over 6 data points per individual sampled. The average age at first childbirth is 27 for women and 30 for men. This is somewhat older than the full sample as there were more pregnancies before the age of 20 than after the age of 45 which were excluded. The average income for men with children is 34,500€ and for men without children is 18,500€. For women with children the average wage is 12,000€ compared to 13,500€ for women without children. It is not surprising that the average wage for men is significantly higher when one does not control for labour market outcomes.

The driving force behind this large difference in annual income between men and women is labour force participation with 61.4% of men working full-time in 2017 compared to only 29.1% of women. Furthermore, only 5.5% of men work part-time compared to 29.3% of women. The large gap between the income of men with and without children and the rather small gap for women with and without children might be surprising but can be explained by the demographic differences between the two groups. Even though key variables including the average age and full-time experience of the individuals with children are very similar to that of individuals



without children for individuals of the same sex, there are large differences in the distributions of those variables. When looking at the distribution of age, both the women and men in the control group have a much larger share of very young people who are likely to be in education and a much larger share of very old people who are likely to be in retirement explaining the lower income. For men, the percent of individuals in the control group in education in 2017 was 9.7% compared to less than 1.4% for men in the treatment group which is very similar for women. These large differences in distribution of key variables and also small but significant differences in means are the reason matching is required. Without using matching, one would believe that men without children earn only about half of what men with children earn which is not due to having children but due to being different in key variables. For women, one could believe that there is little difference in income for women with and without children which again is not due to treatment but due to differences in other characteristics which I prove by using matching.

Please refer to table A.1 in the appendix to have a more detailed overview of the sample means of men and women of the key variables considered in this paper.

### 4.3 Representativeness of the sample

This paper is aiming to make inference from the results of this study to the whole of Germany. To be able to make this inference, it is crucial that the sample considered is representative of Germany's population. For this condition to be satisfied it is important that, in all the key variables important for the results, the sample is similar to the rest of Germany.

The sample examines 25,000 men and 28,000 females over a time span of 33 years meaning that around 53.2% of the individuals are female. The national average in 1984 was 52.3% and 50.7% in 2017 which suggests that the sample is close to the population mean (*Bevölkerung nach Geschlecht* 2019). Another key statistic required to be similar to the national average is that of employment. In the sample 75.2% of the individuals surveyed after the year 2010 say that they are employed compared to the national average from 2011 to 2017 of 73.7% (*Erwerbstätigenquoten*

1991 bis 2018 2019). Breaking this down to men and women I find it to be 80.9% and 69.9% compared to the national average of 77.9% and 69.4% respectively. In addition to considering the extensive margin of labour market participation I also need to look at the intensive margin. Both of these variables are crucial drivers of the child penalty. Here, the intensive margin is best portrayed by the percentage of individuals in full- and part-time employment. In 2017 48% of women employed in Germany worked part-time compared to only 11% of men (*Blickpunkt Arbeitsmarkt* 2019). In the sample 50% of women and 8% of men in 2017 were part-time employed (*Blickpunkt Arbeitsmarkt* 2019). Another variable crucial for women's labour outcome is the average age at which they have their first child (S. P. Martin 2000). In 1980 the average age for women in Germany to have their first child was 24 which increased to 30 in 2018 (*Daten zum durchschnittlichen Alter der Mutter bei Geburt nach der Geburtenfolge für 1. Kind, 2. Kind, 3. Kind und insgesamt 2018* 2019). Hence, the average age of 27.35 observed in the sample is very close to the average age of first-time mothers over the 33 years in Germany. Finally, there are large differences in income between the western counties and eastern counties of Germany. According to the federal statistical office, around 19.5% of the German population lives in the "new" eastern states in 2018 which is similar to the mean of the sample which is 17.8% (*Bevoelkerung* 2020).

To conclude there seems to be strong evidence that this sample is representative of Germany as a whole. Therefore, one can reason that the results found in this paper can be generalized to the rest of Germany.

# 5

## Results

### 5.1 Matching

Matching aims to equate the distribution of the covariates between the treatment group and the control group. In the following sections I undertake tests to examine how well matching managed to balance variables between the treatment and control groups. The reason balancing matters is because the better the balance of the sample the more robust the results are going to be.

#### 5.1.1 Assumption I

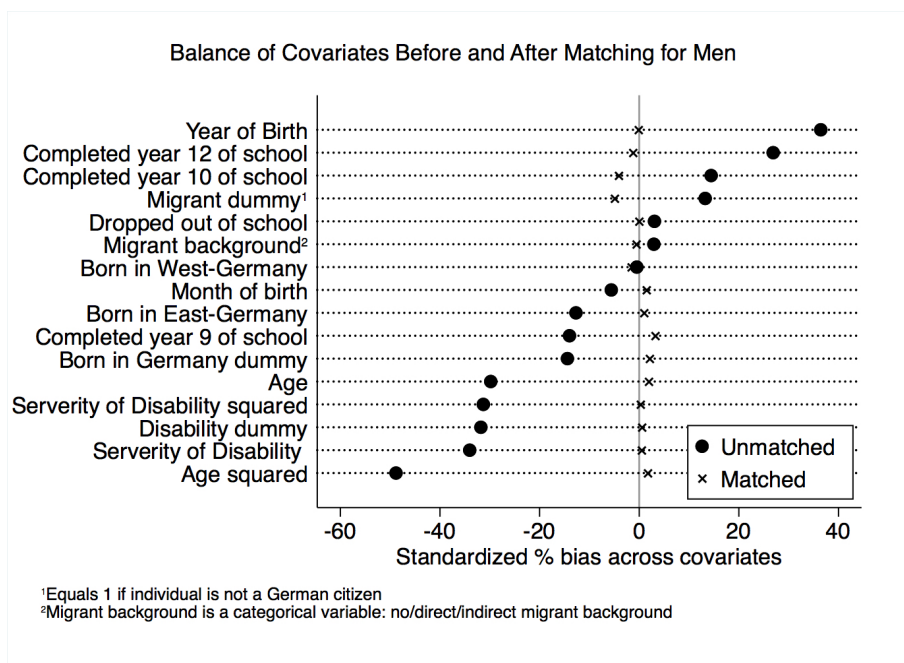
The hidden bias assumption (3.3) states that treatment assignment is independent of the potential outcomes given the set of covariates. This assumption holds if the treatment group is the same as the control group for all the relevant covariates. If this is the case one says that the sample is perfectly balanced.

One way to assess the balance of the sample is to compare the means of the covariates in the treatment group to the control group. A numerical method to test the similarity of the means is the B statistic which summarizes the overall standardized difference in the means of the covariates (Rubin 2001). The widely used benchmark for good matching is a value below 25% (Rubin 2001). Prior to matching, there are severe sample imbalances for both men and women with the B

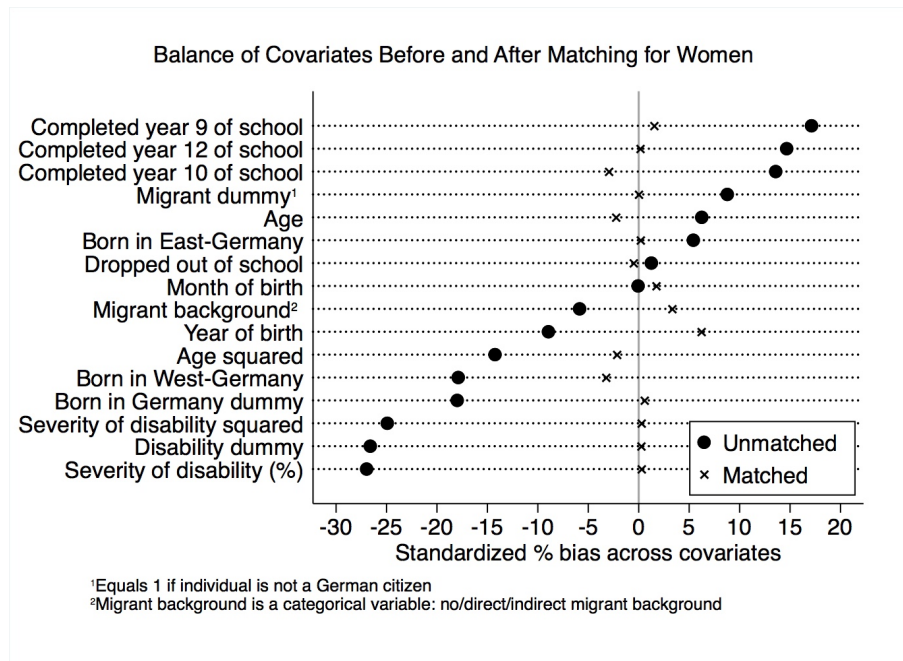
statistic being 125.8 for men and 130.4 for women. Through matching I reduce the B scores for men to 9 and for women to 11.3 indicating that matching has worked very well as it significantly reduced the differences in means between the control and the treatment group. One can also look at the means in more detail by examining the difference in means for each covariate for both groups. The statistic used to examine this is known as the standardized difference in percentage which is the mean difference as a percentage of the average standard deviation (D’Agostino Jr 1998).

$$Standardizedpercentagebias = 100 \times \frac{(X_t - X_c)}{\sqrt{\frac{(\sigma_t^2 + \sigma_c^2)}{2}}} \quad (5.1)$$

As depicted in figures 5.1 and 5.2, post matching for both male and female, the standardized percentage bias for all the covariates is around 5% or less which is below the critical value of 10% (Morgan 2018). This indicates that the sample is very well balanced after matching. When looking at the balance before matching one can see that matching was necessary with many covariates having a standardized percentage bias of over 10% with some having a bias of up to 50%.



**Figure 5.1:** Balance of covariates before and after matching for men



**Figure 5.2:** Balance of covariates before and after matching for women

Marginal distributions are another dimension in which covariates in the control and treatment group should be similar. One approach to check whether this is the case is to look at the R statistic (Rubin, 2001). The R statistic summarizes all of these ratios for each covariate into one number which should be between 0.5 and 2 post matching. For both men and women, this requirement is not met before matching. Prior to matching, the treatment and control groups had very different covariate distributions with the male sample having a R statistic of 0.19 and the female sample of 0.17. Post matching the R statistic is 1.21 for men and 1.01 for women displaying very good balance.

To conclude, given the covariates matched on, I have achieved good balance across the control and treatment groups for both men and women. A well-balanced sample is a necessary, however not sufficient, condition for causal inference. To allow for causal interpretation, I also require the covariates to capture all of the information regarding treatment assignment and fertility. As discussed previously, this is likely to hold to a certain extent but unfortunately cannot be tested.

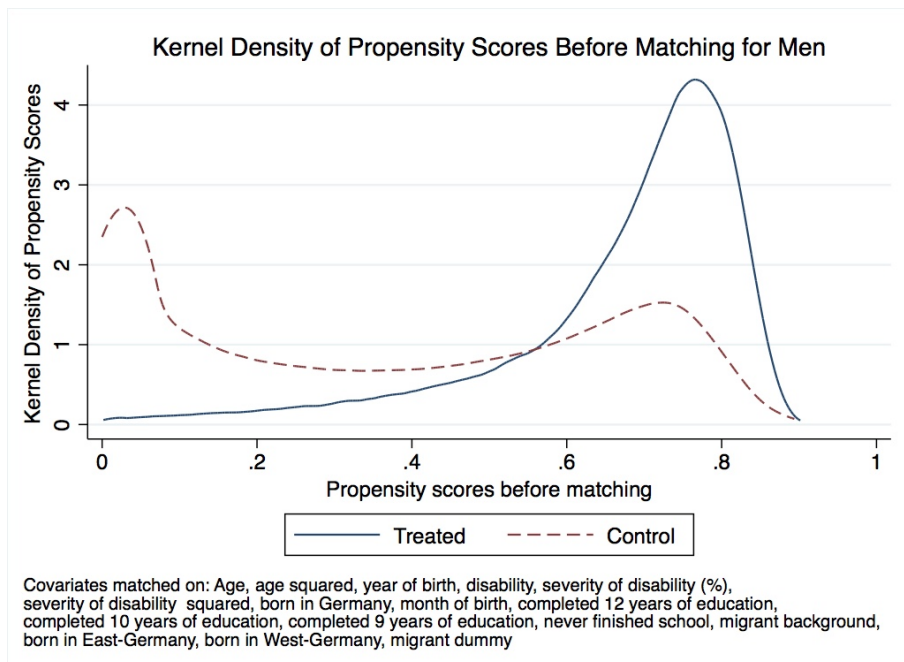
### 5.1.2 Assumption II

The second assumption (3.5) captures the idea that given any possible combination of the covariates observed, there should be some positive probability of having children and some positive probability of not having children. Hence, any individual in the sample has to have a propensity score above zero and below one. To check whether this assumption is satisfied one should look at the evidence on common support. I used the common support option which ensures that only individuals with propensity scores in the area of common support are included. This leads to the exclusion of 17 male observations and 35 female observations. Therefore, this assumption is satisfied which is shown in figures 5.3 – 5.6. The figures prove that there is always some probability of having a child or not having a child.

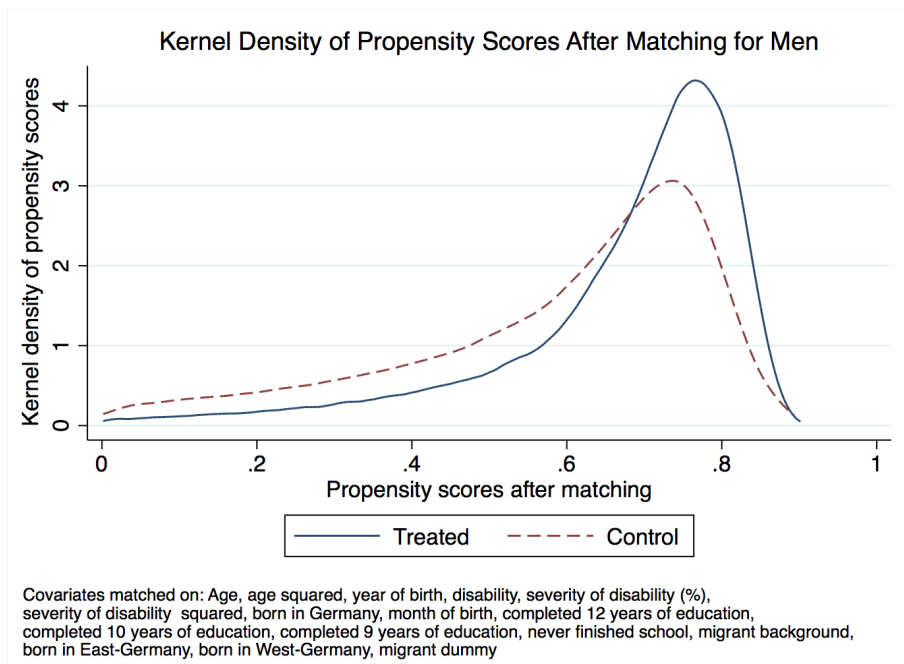
The assumption is also known as the overlap assumption as additional to common support it assumes that there is significant overlap between the distribution of covariates i.e. the propensity scores of the control and the treatment group. The reason insufficient overlap is problematic is that it leads to the treatment effect to depend strongly on extrapolation. This in turn makes the model's prediction less reliable (Stuart 2010). As shown in figures below, matching leads to a much better overlap for both women and men which means that at least one very similar control individual exists for every treated individual who can form a high-quality match. To conclude assumption II is satisfied as the kernel densities appear to have a large overlap and by excluding some observations only individuals with common support are included.

### 5.1.3 Treatment assignment

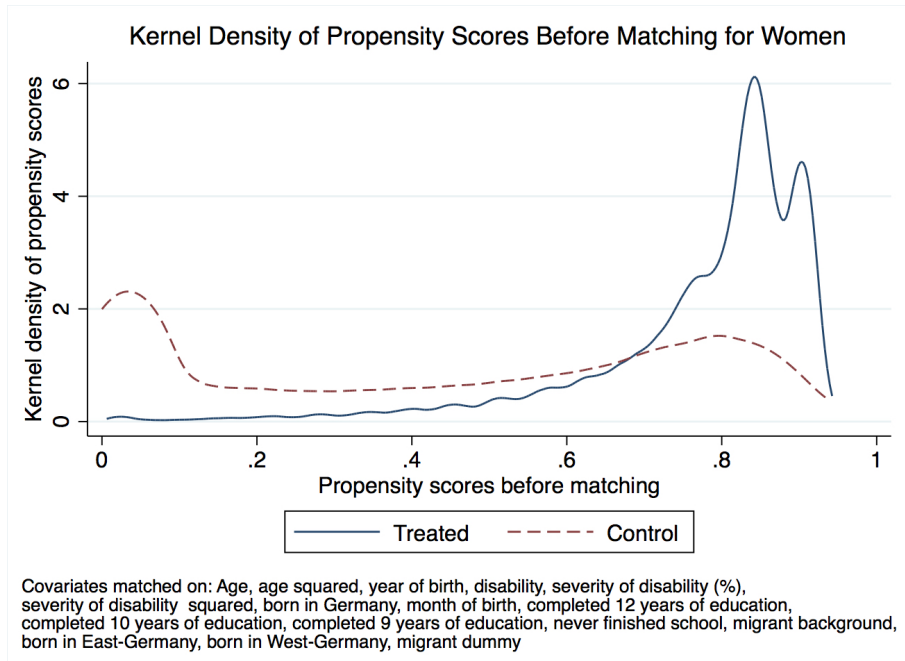
The likelihood of an individual to have children depends on many different covariates. The next paragraph examines whether these covariates have the expected effect on individuals' likelihood of being treated. If they match the trends seen in the population the sample is more likely to be representative of Germany as a whole. Before matching I find that both men and women who immigrated to Germany are more likely to have more children than individuals born in Germany which is in line



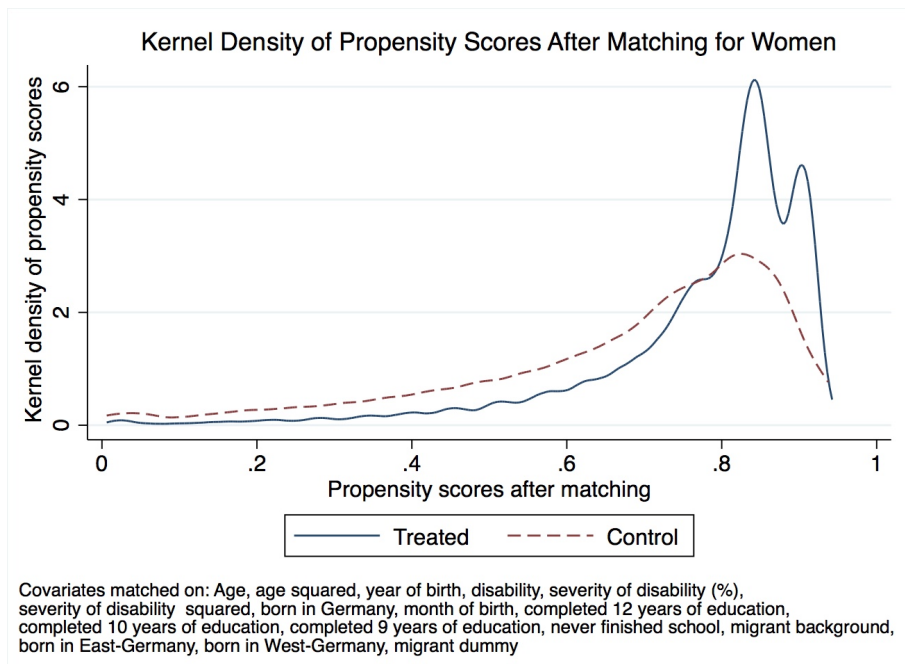
**Figure 5.3:** Density of propensity scores before matching for men



**Figure 5.4:** Density of propensity scores after matching for men



**Figure 5.5:** Density of propensity scores before matching for women



**Figure 5.6:** Density of propensity scores after matching for women



with what the statistics show for the whole of Germany (Frankfurter Allgemeine Zeitung 2018). Unsurprisingly, I find that the probability of having children increases with age. Furthermore, the probability of having children for women decreases with years in education which is also the case in Germany (Schmitt and Wagner 2006). An interesting trend is that women who are born in east Germany have higher fertility rates compared to women born in west Germany whereas men have a lower probability of having children. This can be explained by the fact that in east Germany 53% of the population is male compared to 49% in west Germany (Geis and Orth 2020). Furthermore, over the years, women in east Germany have had higher fertility rates than women in west Germany (*Zusammengefasste Geburtenziffer nach Kalenderjahren* 2019). In a nutshell, the likelihood of being treated based on your covariates before carrying out matching is in line with what can be observed in the rest of Germany.

#### 5.1.4 The treatment effect

After completing matching, which is carried out separately for men and women, I find the average treatment effects for women and for men. The average effect on men's earnings over the 20-year long period following the birth of the first child, is equal to 6,810€ per year which is significant at the 1% level. A total number over 95,000 observations, split into control and treatment group, were used to calculate this effect. For women I find the average treatment effect to be -10,526€ per year which is significant at the 1% level. A total number of around 67,000 observations split into treatment and control group were used to estimate this average treatment effect. Thus, having children affects women's earnings negatively and men's earnings positively.

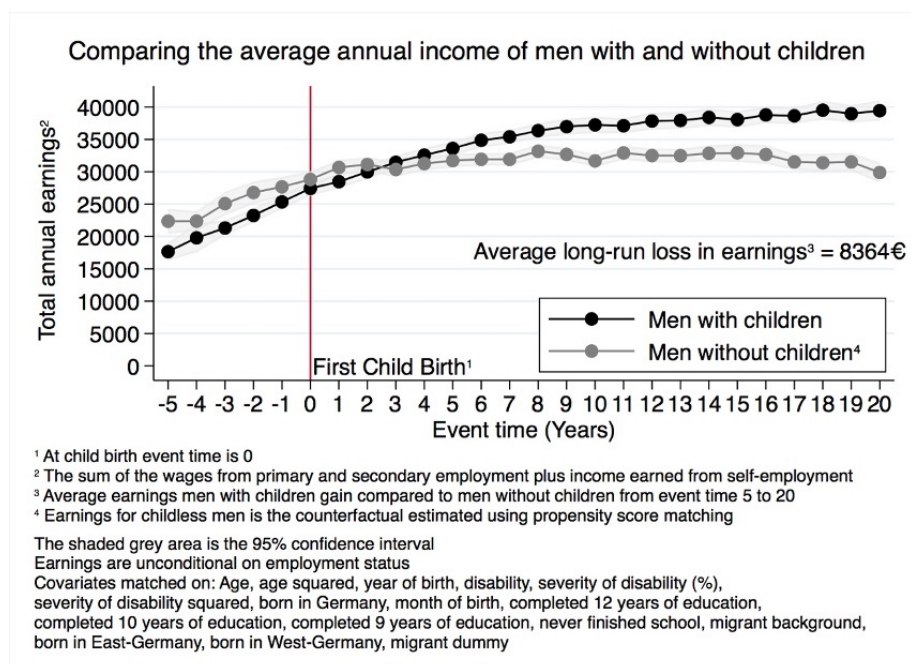
#### 5.1.5 The child penalty

The child penalty builds on the previously identified treatment effect by comparing the difference in impact of children on fathers' and mothers' earnings relative to

their counterfactual income for each event time  $t$  from 5 years before up to 20 years after the birth of the first child.

Before comparing men and women, one should look at differences between individuals of the same sex with and without children. Figure 5.7 depicts the earnings of men with children and their estimated counterfactual wage. It is important to emphasize that the light grey line is not something I observe i.e. it is not simply the labour income of the male control group but the predicted hypothetical outcome I estimated using the control group. For the 5 years before childbirth and the 2 years after childbirth men's counterfactual income is slightly higher than their actual income but for most years this difference is not significant with the standard errors overlapping. From event time 3 onward, fathers' income steadily grows whereas the counterfactual income levels off. Thus, from event time 6 onward men with children earn significantly more than if they had chosen not to have children. This gap continues to grow with its maximum occurring 20 years after childbirth when men with children earn annually 10,000 euros more than their childless counterparts. These differences in annual gross earnings are driven by differences in income growth, with childless men experiencing no income growth after event time 1 compared to fathers' income growing continuously over the whole 26 years studied.

When comparing mothers' actual earnings to their potential incomes one can see that throughout the whole period studied women without children have significantly higher incomes than women with children. This suggests that there are not only post-fertility effects but also pre-fertility effects. The gap between the two is constant until childbirth with both their income following a similar pattern of growth. The reason the impact of childbirth on income is not visible until event time 1 is because in event time 0 women are giving birth hence for some of the year they might still be working and/or receive job-protected parental leave i.e. still receive similar amounts of income. Following the year of childbirth this gap in earnings doubles within a year and then even triples compared to event time -1. This divergence is almost fully driven by the reduction in actual earnings rather than growth in potential earnings of mothers. Starting from event time 2 until event time 20, the difference



**Figure 5.7:** Comparing the income of men with and without children

in mothers' actual income and their potential earnings continuously shrinks. From event time 2 until event time 8, the convergence in incomes is induced entirely by mother's recuperation of their income earned before childbirth. From event time 9 onward until event time 15 the convergence is driven equally by mothers' income growth and childless women's income decline. In the final 5 years considered, the gap in income shrinks only very slightly which is caused by very moderate growth in mothers' income. Twenty years after the birth of the first child, women with children still earn over 7,000 euros less per year than their childless counterpart which is similar to the gap in income before childbirth.

I can now calculate the long run child penalty using equations (3.6) – (3.9). As set out in the methodology section using equation (3.9), I take the difference in the percentage effect of childbirth on women and men. To arrive at this percentage effect, I divide the individual's difference between their actual income and their counterfactual income by the average counterfactual income of that event time for their sex (3.8). The result is shown in figure 5.9.

The long-run child penalty, which is defined as the child penalty averaged from 5 to 20 years after childbirth, is found to be 63.4%. In brief, the interpretation

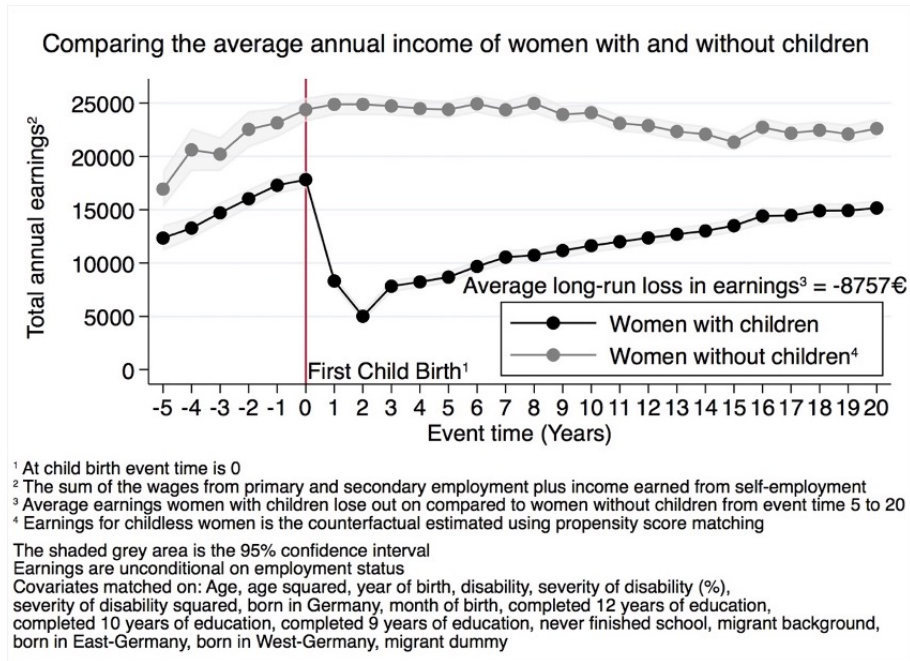


Figure 5.8: Comparing the income of women with and without children

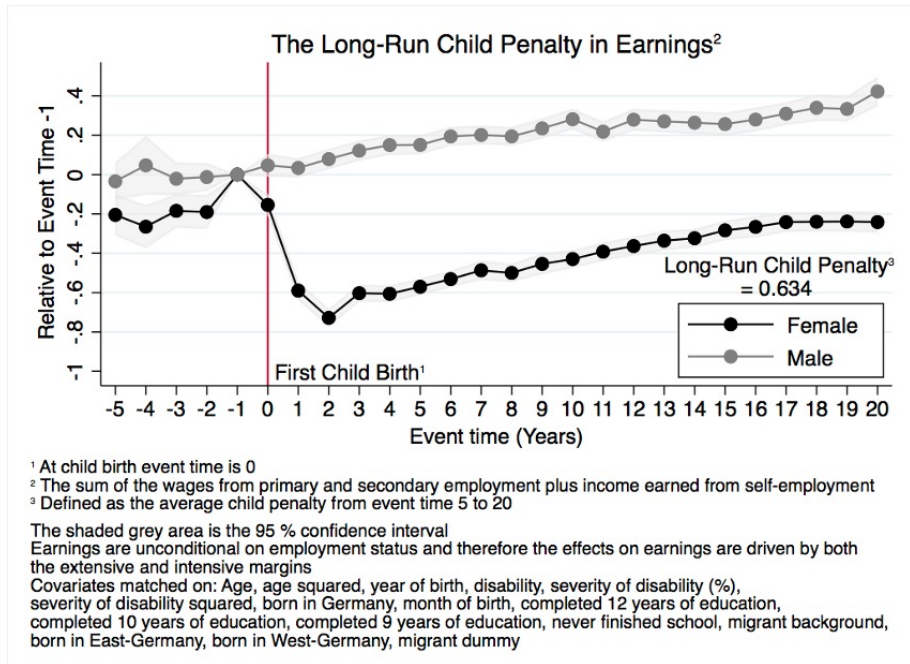


Figure 5.9: The long-run child penalty in earnings

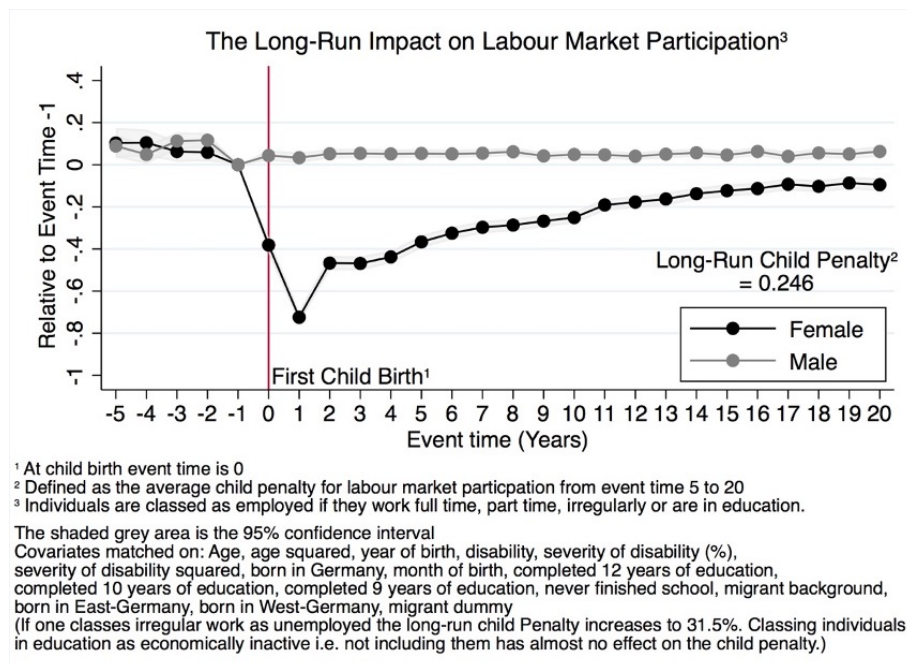
of this results is that women earn 63.4% less of their potential earnings than men earn of their potential earnings. Evidently, the average mother's potential income is significantly lower than the average father's potential income at each point in time. Therefore, the child penalty demonstrates not the difference in actual earnings earned by women and men but how the impact of children on actual income relative to potential earning differs between women and men. Before childbirth, men's and women's actual earnings earned as a percentage of their counterfactual earnings follows a fairly similar trend. In event time -5 and -3 their standard errors are overlapping implying that the difference is not significant. However, once the first child is born the percentage of actual income earned is significantly different at the 1% level between men and women for the whole 20 years after the birth of the first child. These findings support Kleven, Landais, and Sogaard (2019) and Kleven, Landais, Posch, et al. (2019) findings that children are the key driver of gender inequality in earnings.

The child penalty is highest in event time 2 at over 80% and from then onward shrinks and amasses to around 50% at event time 17 which is driven by women's actual income increasing quicker relative to their counterfactual income than men's actual income. From event time 17 onward the child penalty increases again to about 60%. It is vital to look at the composition of the child penalty to understand what drives it. During the 16 years following childbirth, the child penalty is primarily driven by the effect of childbirth on women's earnings, especially in the first 2 years following birth when women experience a 70% loss of their potential income. From event time 3 onward, the child penalty continuously falls due to women recuperating some of their lost income. However, during the whole time father's income grows relative to their counterfactual income which reduces this effect. During the final years studied the child penalty increases again to around 60%, which is fully driven by men's income rising relative to their counterfactual income. On the other hand women's loss in potential earnings levels off during those final years. The child penalty is therefore principally driven by the response of women's earnings rather than of men's earnings to children in the first 15 years after the birth of the child.

This switches for the years thereafter with the impact on women’s earnings making up only 20% of the 60% child penalty in event time 20.

To understand which margins drive this child penalty I look at the effect of parenthood on labour market participation, labour market hours and the wage rate. To do so I utilise the same matching methods for each of those dependent variables.

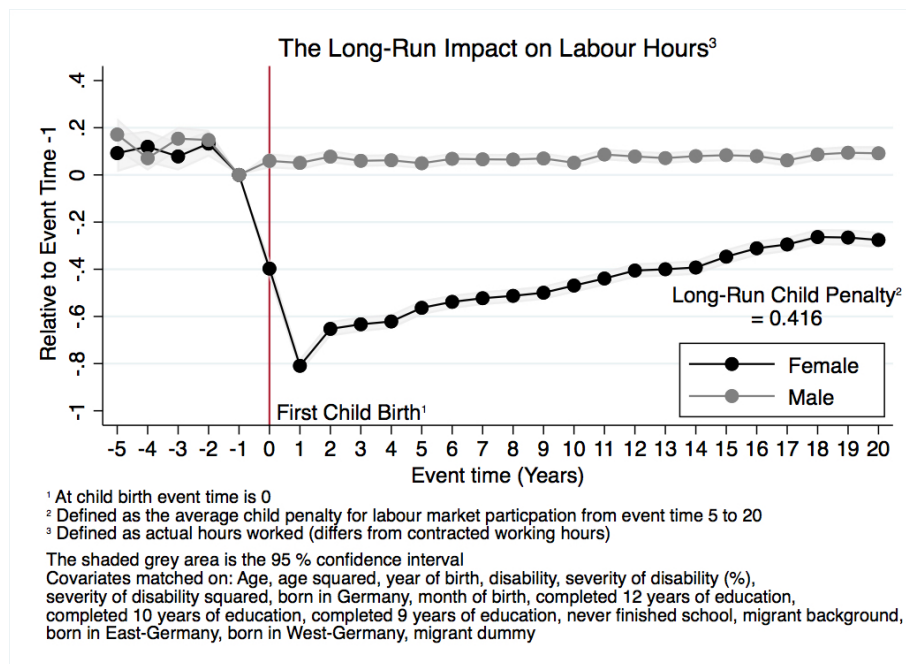
First, I analyse the effect of parenthood on labour market participation. As shown in figure 5.10 men’s labour market participation seems to be almost unaffected by childbirth, with fathers being slightly more likely to participate in the labour market compared to childless men. For women this cannot be said with women’s participation rate dropping by over 70% compared to their childless counterpart 2 years after the birth of the first child. Over time most mothers re-enter the labour market but even after 20 years around 10% of mothers have yet to re-enter the labour market showing just how persistent the effect of childbirth is on women’s labour market outcomes.



**Figure 5.10:** The long-run impact of parenthood on labour market participation

Second, I examine the effect of parenthood on the intensive margin. A similar pattern emerges for both men and women. Over the whole 25 years considered, except year -1 when I set the percentage effect to zero, men with children work

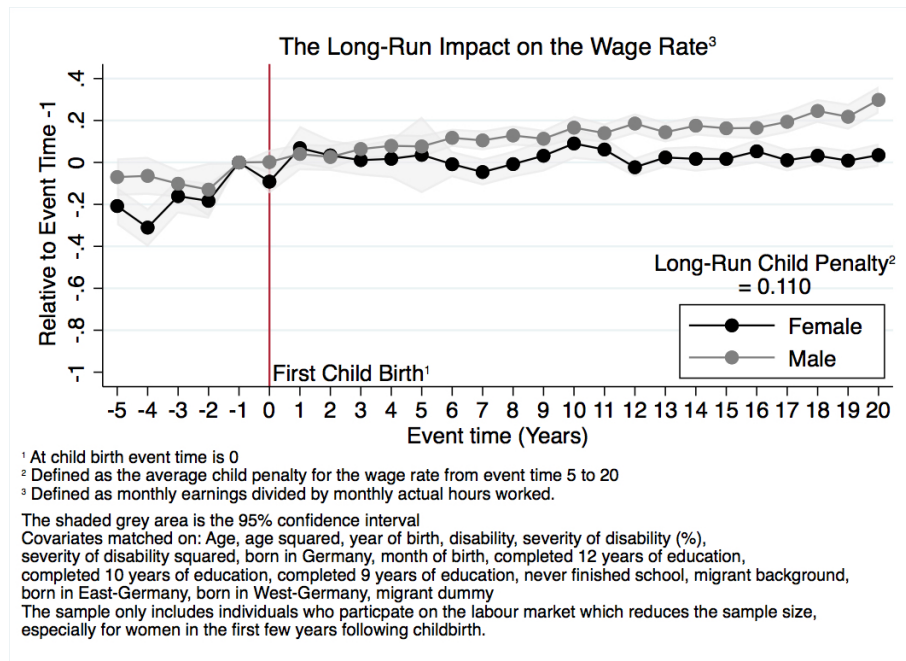
significantly longer hours than their childless counterpart. For women the effect of childbirth on labour hours is even stronger than its effect on labour market participation. Apart from the larger initial drop in labour hours, hours worked also recover at a slower rate than participation does. Thus, leading to a remaining reduction in labour hours worked of over 25% 20 years after childbirth.



**Figure 5.11:** The long-run impact of parenthood on labour hours

Third, I assess the impact of parenthood on the wage rate. The wage rate is calculated using actual hours worked, which is self-reported, and monthly income. Over a quarter of women in the sample do not work and hence do not have an hourly wage rate. This causes a significant reduction in sample size leading to larger standard errors especially for the first few years after childbirth when 70% of mothers have dropped out of the labour market. As shown in Figure 5.12, one can see that the wage rate of women participating in the labour market is not affected by parenthood. This is a rather surprising result which could be the result of a combination of factors. First, the result is less robust due a large reduction in sample size, dropping from around 4,000 women to 800 in event time 1. Additionally, there might be inaccuracies in the data with hours worked being self-reported and hence potentially differing to the contracted hours leading to the wage rate being





**Figure 5.12:** The long-run impact of parenthood on the wage rate

inaccurate. Furthermore, mothers might include parental subsidies in their income earned which causes an increase in their hourly wage rate. When looking at men it becomes clear that fathers earn substantially more per hour than men without children. This difference increases significantly over time with fathers earning 30% more than their counterfactual income in even time 20.

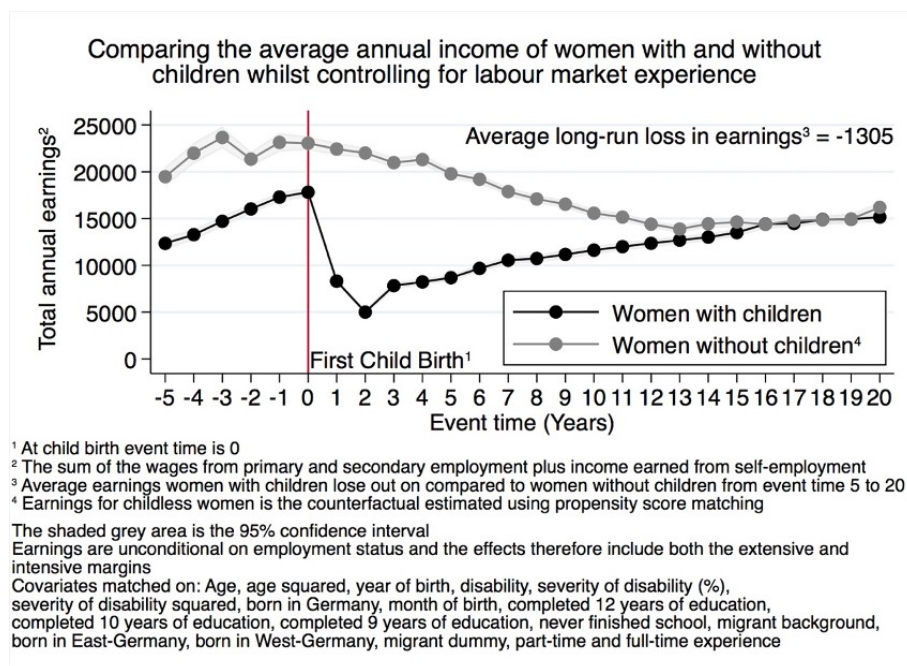
To summarize, women's reduction in earnings following childbirth is mainly driven by the intensive and the extensive margin whereas men's rise in earnings following childbirth is mainly driven by differences in the wage rate. In the long run the child penalty is driven by mothers working 30% fewer hours than women without children and by fathers' wage rate being 30% higher than childless men's wage rate.

### The child penalty – matching on experience

To examine what role labour market experience plays in Germany, I now include both part- and full-time experience in the work place to the set of covariates. Of course now the potential income can no longer be interpreted as their real  $Y_i(0)$ , as childbirth often reduces mothers years of experience. What I find is that work experience seems to be one of the main reasons why women with children earn



less than women without children in the long run. When controlling for labour market experience women earn their counterfactual income 16 years after the birth of the first child. It is worth re-emphasizing that matching no longer determines the causal effect of having children. The potential income plotted in figure 5.13 is no longer their real potential income but the income they would have earned if they had chosen not to have children given their level of experience. In this scenario I compare mothers with certain years of experience to childless women who are similar in their demographic background and with a similar number of years of experience and schooling. Therefore, the long-term impact of motherhood on earnings is almost entirely driven by the loss in work experience which lowers human capital. In the short-run the large gap in earnings between mothers' actual income and their counterfactual income remains, which is due the fact that it is almost entirely driven by mothers working fewer hours compared to their childless peers.

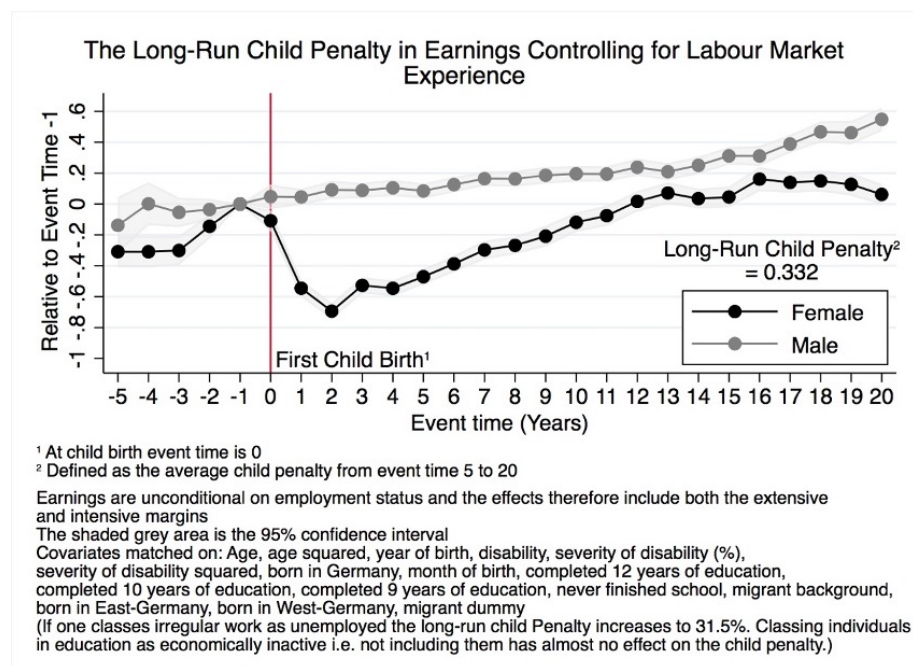


**Figure 5.13:** Comparing the average annual income of women with and without children when controlling for labour market outcomes

When replicating this for men almost no difference in earnings is found, suggesting that fathers' work experience is almost completely unaffected by childbirth. The

only difference being that there is no longer a gap between actual and counterfactual income in the years before childbirth.

These findings can be explained by women still being the primary carer for children which leads to mothers having to temporarily step back from the labour market to focus on household production. The data reflects this hypothesis with mothers at the age of 50 having 1 and a half fewer years of work experience<sup>5</sup> compared to childless women of the same age. For men this difference is much smaller with fathers having half a year of work experience less than men without children.



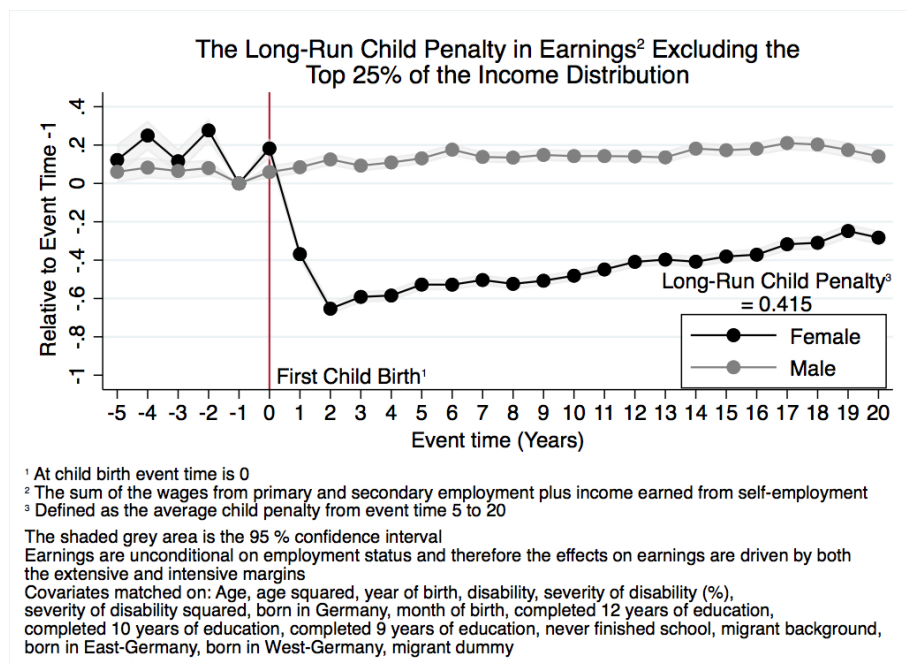
**Figure 5.14:** The long-run child penalty in earnings when controlling for labour market experience

These trends are reflected in the child penalty which shrinks dramatically to 33% when controlling for labour market experience. The short-run loss in earnings cannot be explained by loss in experience and is mostly driven by the intensive and the extensive margin.

<sup>5</sup>This is not taking into consideration whether this is part-time or full-time experience.

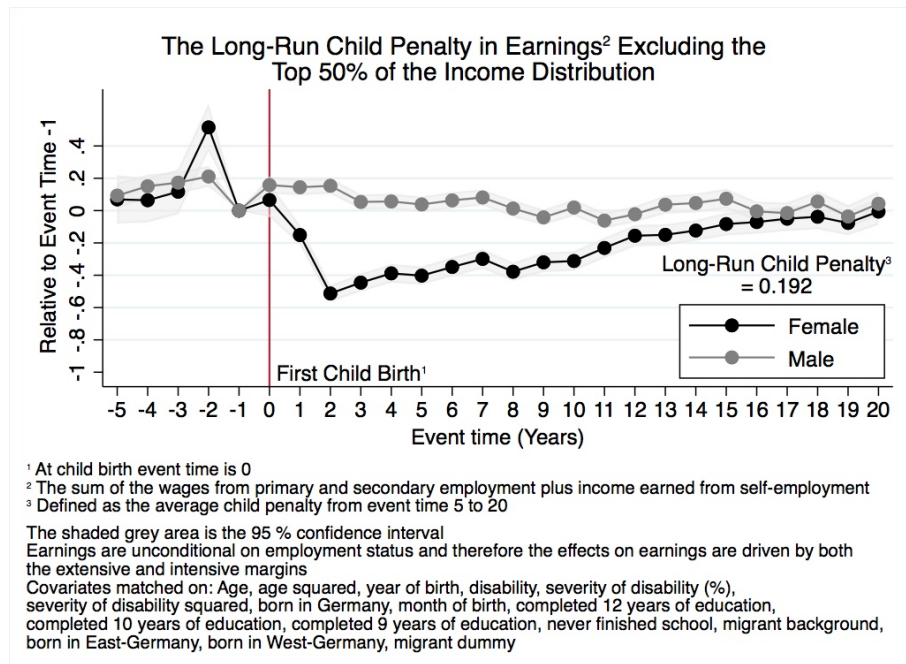
### The child penalty across the income distribution

A concern raised in the literature is that the child penalty is an upper class phenomenon which does not exist for middle and low income individuals Kleven, Landais, and Sogaard (2019). To assess this claim, I first remove the top 25% and then the top 50% of female and male earners and then calculate the child penalty for those individuals.



**Figure 5.15:** The Long-Run Child Penalty in Earning when excluding the top 25% of the income distribution

The child penalty drops to 46% when excluding the top 25% and to 20% when excluding the top 50% as shown in figures 5.15 and 5.16 respectively. There appears to be a positive correlation with the child penalty increasing along the income distribution. This is mainly driven by the reduction in father's actual income relative to their counterfactual income. As one moves down the income distribution men's actual income starts to get closer to their counterfactual income. By excluding the top 25% of the income distribution more fathers than childless men are excluded as they earn significantly more on average than men without children. For women this trend seems to exist as well but in a much weaker form. Excluding the top quartile of the income distribution leaves the effect of children on



**Figure 5.16:** The Long-Run Child Penalty in Earning when excluding the top 50% of the income distribution

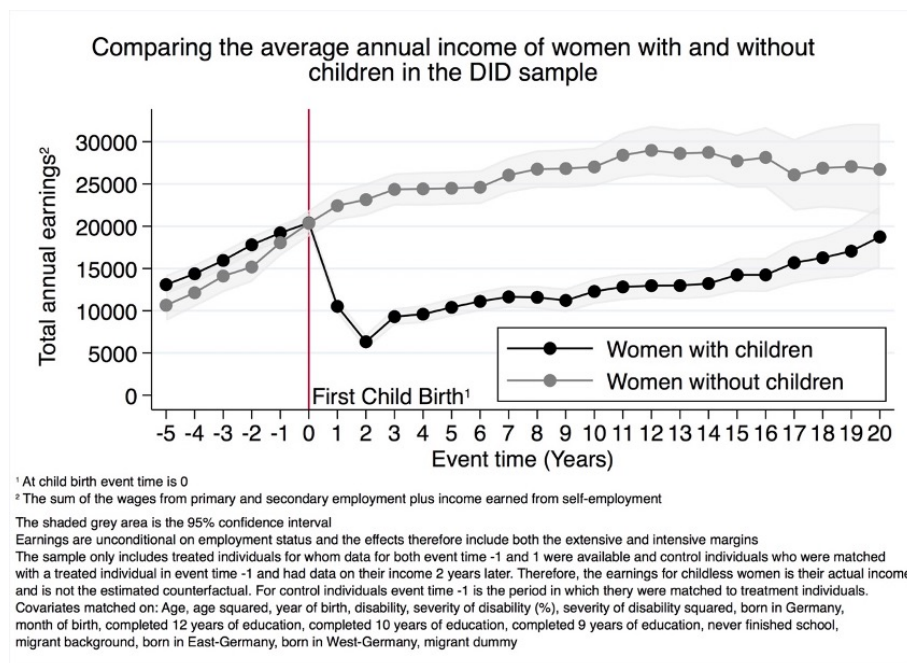
women's income relative to their counterfactual income almost entirely unaffected. Only when excluding the top half of the income distribution do mothers' earnings rise compared to their counterfactual income.

The reason for not looking at the bottom 25% of the income distribution is because all women with and without children in the quartile earn no income at all, therefore both their actual and counterfactual income is zero. To conclude, the child penalty does seem to be larger for upper class and middle class individuals. This is mainly driven by father's actual income decreasing relative to their counterfactual income.

### 5.1.6 Robustness check: Difference-in-differences matching

I employ DID matching to check how robust my findings are. The first step is to assess how robust my DID matching results are by checking whether the common trend assumption holds. One way of checking whether the assumption holds is to look at the trend of income earned the years before event time -1. In the graphs below, one can see that the level of income earned differs between fathers and men

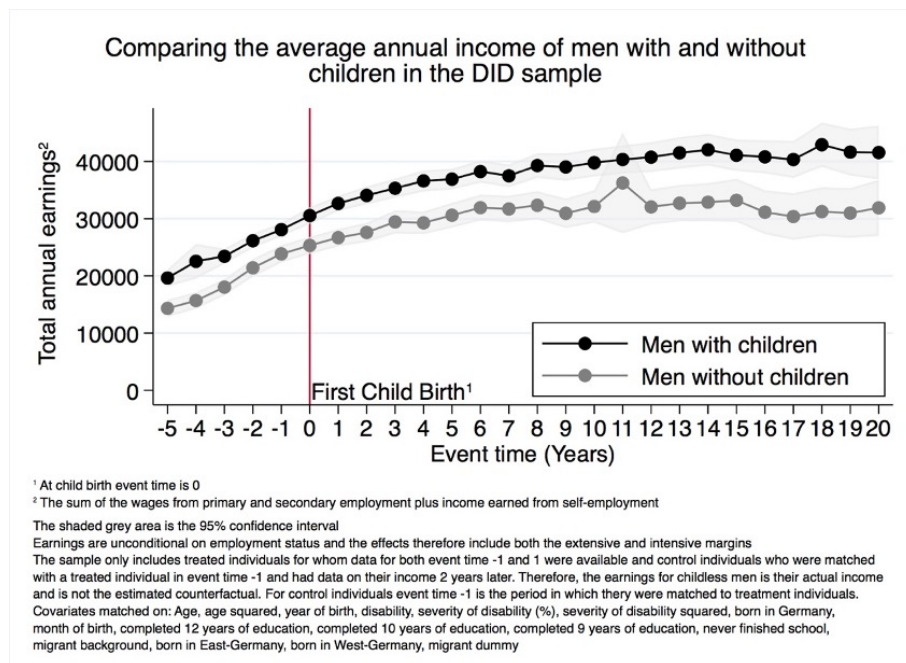
without children, and mothers and women without children. However, DID only requires the trends to be the same which is the case in the 5 years prior to childbirth as shown in figures 5.18 and A.1. This result suggests that if it would not be for treatment, they would continue following the same trend. Of course there is still a small possibility that they would follow different trends even without treatment but the likelihood of this is much smaller given the fact that they followed the same trend for 4 years prior event time -1. Furthermore, there exists no obvious reason why these trends would change without treatment.



**Figure 5.17:** Comparing the income of women with and without children in the DID sample

These results suggest that the DID matching results are robust. For women the DID is -13,000€ as the income of women with children drops around 8,700€ from event time -1 to 1, whereas for women without children income rises by around 4,300€. This suggests that there is clearly a strong negative effect of treatment on women's earnings which is significant at the 1% level. This is slightly larger than the average female treatment effect previously calculated. The reason is due to the fact that the impact of children on income is strongest in the first few years after childbirth and reduces over time. The treatment effect, previously calculated,





**Figure 5.18:** Comparing the income of men with and without children in the DID sample

estimated the effect of childbirth over the 20-year long period after childbirth and hence is expected to be smaller. Therefore, the results suggest that simple matching accounted for most of the time non-varying unobservable heterogeneity. For men the DID is estimated to be around 1,700€ which is significant at the 5% level. For childless men income increases by 2,800€ whereas for fathers it increases by 4,500€ from event time -1 to +1. The average treatment effect, calculated previously, estimated it to be larger than that which is driven by the difference in income between men with and without children increasing over time. When looking at figure 5.7 there seems to be no difference in income between men with and without children for the first 5 years. This suggests that the child penalty previously estimated is a lower bound and might be even higher when controlling for unobserved heterogeneity. To conclude, the child penalty cannot be explained by unobserved time invariant heterogeneity. Furthermore, when controlling for unobserved time invariant heterogeneity the size of the child penalty seems to be unchanged.

## 5.2 Regression analysis

Once I have determined the individual treatment effect, I can examine potential factors driving the impact of childbirth on income. I continue to look at men and women separately as childbirth has a very different effect on their income and hence this impact on earnings is expected to have different drivers. For the following analysis the dependent variable is the individual treatment effect i.e. the difference between the actual income and the hypothetical income. The first step is to establish a baseline which only includes the control variables. The baseline is then followed by regressions which besides the control variables add in further variables of interest. When adding in the variables of interest to the baseline regression, the sample size is substantially reduced which increases standard errors.

The baseline was constructed by building on the existing literature in labour economics. The Mincer earnings function is the most influential labour income regression in economics which explains labour income as a function of total years of schooling, labour market experience and labour market experience squared. (Mincer 1974; Lemieux 2006). In almost all settings, the returns have been found to be positive for schooling and labour market experience and negative for experience squared. This supports the idea that increased schooling and increased work experience enhance human capital and hence productivity which in turn leads to an increase in labour income. The negative return on work experience squared implies that experience displays decreasing marginal returns. I previously controlled for education undertaken at school level but not higher education. Thus, I now control for university degree completed as well.

Besides those factors, I include a dummy for east Germany to control for regional differences in labour income and potential income. Since the reunification of Germany in 1990, the former Federal Republic of Germany has persistently outperformed the former German Democratic Republic in most of the key labour market outcomes with wages still being 16% lower in the east of Germany (*Die Arbeitsmarktsituation von Frauen und Männern* 2019; Bundesministerium für Wirtschaft und Energie 2019). Therefore, it is very important to control for those differences. Furthermore,

I control for marital status. Marriage affects not only individuals' financial security but also the income tax rates they face and hence labour market decisions. Lastly, I control for the age at which the individual had their first child as many studies have shown that timing of fertility matters for future income (S. P. Martin 2000; Waldfogel 1998; Amuedo-Dorantes and Kimmel 2005; Miller 2011).

A negative coefficient implies that the independent variable is correlated with lower actual earnings relative to the hypothetical earnings. For women this means that their loss in income caused by childbirth is even larger whereas for men it implies that their gain caused by childbirth is smaller.

It becomes apparent that some associations appear to be more robust than others. The coefficients which are robust are the ones which display a clear positive or negative correlation with treatment across all the regressions and hence are the ones which potentially have an underlying causal interpretation.

I find that for 2 out of the 4 regressions married mothers display significantly lower earnings relative to their potential incomes than unmarried mothers. The reverse seems to be true for men with this phenomenon being known as the marriage premium. Furthermore, there is a substantial and significant negative association between living in East-Germany and the dependent variable for men, which holds throughout and does not exist for women. This implies that the child penalty should be lower in East-Germany due to lower actual earnings for men relative to their potential incomes. Another persistent association is the negative relationship between the dependent variable and the age at which women had their first child. This seems to be counter-intuitive and the opposite to what the literature finds, indicating that it is a spurious correlation (Amuedo-Dorantes and Kimmel 2005; S. P. Martin 2000).

The coefficients for work experience and schooling from the Mincer regression are strongly robust. For women an increase in part-time experience is always positively correlated with the dependent variable which is significant at the 1% level. Those results show how important any form of labour market experience is for women if they want to decrease their amount of potential income lost. The magnitude of



**Table 5.1:** Linear regression: Baseline

VARIABLES	Men	Women
Married dummy	2,609*** (590)	-2,998*** (307.6)
East-Germany dummy	-11,687*** (522.4)	-1,258*** (360.2)
Part-time experience <sup>1</sup>	-1,490*** (225.2)	859.9*** (69.27)
Full-time experience <sup>1</sup>	1,521*** (74.63)	1,147*** (53.16)
Part-time experience squared	73.67*** (20.09)	-14.91*** (4.47)
Full-time experience squared	-23.02*** (1.991)	2.41 (2.429)
Completed university dummy	27,512*** (487.4)	10,078*** (306.9)
Age had first child	125.4*** (41.96)	-412.0*** (32.85)
Constant	-25,364*** (1,296)	-11,362*** (858.1)
Difference <sup>2</sup>	7,272.80	-9,542.60
Observations	42,016	39,441
R-squared	0.129	0.109

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Dependent variable: Difference = actual earnings – hypothetical earnings where hypothetical earnings are the earnings the individual would have earned if had chosen not to have children Hence, a negative coefficient means that actual earnings have decreased relative to hypothetical earnings

<sup>1</sup>Number of years spent in part/full-time employment

<sup>2</sup>Mean of the dependent variable for the observations considered in the regression

this association is also rather large with one year of part-time experience being associated with a decrease in the gap of around 10%. For men the coefficient for part-time work experience is always found to be negative but not always significant. This suggests that part-time experience is associated with lower actual income relative to their hypothetical income. When looking at the correlation between full-time experience on the labour market and the dependent variable I find for both men and women an increase in actual income relative to their potential income. These coefficients are always found to be significant at the 1% level and seem to be robust to the model's specification. There seems to be no strong association between experience squared and the dependent variable apart from the significant negative association for full-time experience squared for men. Both men and women seem to have positive returns to completing a university degree with the coefficient always being significant at the 1% level. There seems to be evidence that the association between the individual treatment effect and education is of much larger magnitude for men than for women. These findings imply that more educated men and women have higher actual earnings relative to their potential incomes where the coefficient is substantially larger for men. This implies that university educated mothers earn a larger percentage of their potential income compared to mothers without a degree. However, due to that increase being smaller for women compared to equally educated men, their child penalty actually increases.

The three main areas I would like to explore in the following section are how gender norms, willingness to take risks for career and priorities when choosing their job are correlated with the individual treatment effects. The last two areas could provide me with evidence that preferences play an important role for the effect of childbirth on earnings.

First, I examine the relationship between gender norms and individual treatment effects. When I refer to traditional gender roles, I mean that women are seen to be the ones who are responsible for household production and men are responsible for the financial security of the family.

Mothers with modern gender roles have smaller treatment effects compared to mothers with more traditional views on gender roles. The reverse is true for men. Looking at the variables more closely men that believe they should help with domestic work, experience lower actual earnings relative to their hypothetical earnings. The same applies to men who believe that men and women are equal in the household and at work. This association is significant at the 5% level and at the 10% level respectively. For women there is a negative correlation between believing that men are more in charge of family's financial security than women which is significant at the 10% level. Furthermore, women who believe that women and men are equal in the household and at work have significantly more actual income compared to their hypothetical income.

To summarize, there seems to be supporting evidence for gender norms explaining part of the individual treatment effects. However, these results only show correlations and cannot by itself be interpreted as causal effects.

The third regression (table 5.3) tries to examine the relationship between risk preferences regarding career and the individual treatment effect. The results are that risk preferences with regard to career are more important for men with the coefficient for low risk preferences to be significantly negative at the 5% and significantly positive at the 5% level for high risk preferences. For women the coefficients have the same sign but are not found to be positive. Thus, fathers get penalised for being risk averse and rewarded for being risk tolerant.

Lastly, I look at the relationship between individuals' treatment effects and individuals' preferences regarding the importance of certain job characteristics. As shown in table 5.4, an apparent feature for all of the significant coefficients is that most of them are significant at the 1% level and are of large magnitude suggesting that preferences regarding job characteristics are very relevant for individual treatment effects. Unsurprisingly one of the results is that individuals, for whom income is a priority when choosing their job, have significantly higher levels of actual income than their hypothetical income. What might come as a surprise to some is that the coefficient with the largest magnitude is preferences

**Table 5.2:** Linear regression: Gender norms

VARIABLES	(1) Men	(2) Women
Married dummy	7,642*** (2,777)	229.6 (1,344)
East-Germany dummy	-11,045*** (2,698)	-1,536 (1,687)
Part-time experience <sup>1</sup>	1,823* (968)	1,289*** (381.9)
Full-time experience <sup>1</sup>	1,797*** (449.6)	1,239*** (296.7)
Part-time experience squared	43.36 (60.36)	-40.34 (24.72)
Full-time experience squared	-27.29** (13.19)	-0.366 (14.13)
Completed university dummy	30,845*** (3,018)	8,717*** (1,604)
Age had first child	158.1 (266.8)	-486.0*** (181.9)
Women care more for family than career <sup>2</sup>	-915.2 (2,800)	-2,253 (1,502)
Men more in charge of family's financial security <sup>2</sup>	438.5 (2,786)	-2,375* (1,391)
Men help with domestic work <sup>2</sup>	-7,284** (3,553)	803.9 (2,222)
Men and women equal in hh/family and at work <sup>2</sup>	-4,954* (2,738)	3,427** (1,427)
Children suffer if father concentrates on work <sup>2</sup>	-3,464 (2,790)	-2,102 (1,441)
Working mother as emotionally available <sup>2</sup>	6,109** (2,941)	1,868 (1,648)
Constant	-26,827*** (8,943)	-12,652** (5,836)
Difference <sup>3</sup>	8,615.40	-9534.2
Observations	1,259	1,352
R-squared	0.186	0.137

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Dependent variable: Difference = actual earnings – hypothetical earnings where hypothetical earnings are the earnings the individual would have earned if had chosen not to have children. Hence, a negative coefficient means that actual earnings have decreased relative to hypothetical earnings

<sup>1</sup>Number of years spent in part/full-time employment

<sup>2</sup>Dummy variable is equal to 1 if agree with the statement and 0 if disagree

<sup>3</sup>Mean of the dependent variable for the observations considered in the regression

**Table 5.3:** Linear regression: Willingness to take on risk for career

VARIABLES	(1) Men	(2) Women
Married dummy	3,898** (1,536)	-1,249 (938)
East-Germany dummy	-10,465*** (1,626)	-108.9 (1,088)
Part-time experience <sup>1</sup>	-2,706*** (784.5)	759.8*** (249.6)
Full-time experience <sup>1</sup>	1,707*** (229.1)	1,318*** (186.8)
Part-time experience squared	180.6** (82.57)	-22.83 (16.68)
Full-time experience squared	-26.12*** (6.33)	5.587 (7.925)
Completed university dummy	28,040*** (1,608)	9,791*** (988.2)
Age had first child	206.6 (135.4)	-419.2*** (114)
Willing to take low risk for career <sup>2</sup>	-3,597** (1,471)	-1,436 (935.1)
Willing to take high risk for career <sup>2</sup>	3,627** (1,674)	793.2 (1,448)
Constant	-28,366*** (4,199)	-11,270*** (3,008)
Difference <sup>3</sup>	10,855	-7299.9
Observations	4,866	4,437
R-squared	0.116	0.098

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Dependent variable: Difference = actual earnings – hypothetical earnings where hypothetical earnings are the earnings the individual would have earned if had chosen not to have children Hence, a negative coefficient means that actual earnings have decreased relative to hypothetical earnings

<sup>1</sup>Number of years spent in part/full-time employment

<sup>2</sup>This is a variable recoded in dummies capturing an individual's willingness to take risk: scale of 0-10 where 0 means that you are not willing to take any risks. Here low risk for career captures those who answer with 0-3, mid risk (excluded here due to collinearity) 4-6 high risk 7-10

<sup>3</sup>Mean of the dependent variable for the observations considered in the regression

regarding family friendliness of the job. For men and women, the coefficient is significant at the 1% and 5% respectively. It's of such large magnitude that men's positive treatment effect turns into a negative one if they cite family friendliness as a priority. For women citing family friendliness as a priority almost doubles their negative treatment effect. Hence, making family friendliness a priority for your job means for men and women with children that they lose out on a lot of their potential income. Finally, there is a positive and significant correlation between career advancement being a priority and the treatment effect. This association is still significant at the 1% level for men and at the 10% level for women but is of smaller magnitude. Additionally, there seems to be no correlation between importance for recognition and the treatment effect.

**Table 5.4:** Linear regression: Priorities when choosing job

VARIABLES	(1) Men	(2) Women
Married dummy	3,511 (2,327)	-1,777** (767.6)
East-Germany dummy	-11,728*** (1,465)	199.4 (985.5)
Part-time experience <sup>1</sup>	-1,425*** (552.6)	1,448*** (203.6)
Full-time experience <sup>1</sup>	1,897*** (259.6)	1,275*** (164.9)
Part-time experience squared	47.86 (35.01)	-56.92*** (12.41)
Full-time experience squared	-28.53*** (7.672)	-0.279 (7.123)
Completed university dummy	34,833*** (1,633)	9,690*** (905)
Age had first child	35.3 (151.9)	-521.5*** (106.8)
Income is a priority when choose job <sup>2</sup>	8,583*** (1,826)	3,713*** (859)
Family friendliness is a priority when choose job <sup>2</sup>	-11,790*** (3,483)	-4,852** (2,308)
Opportunity for career advancement is priority <sup>2</sup>	4,691*** (1,326)	1,464* (775.9)
Recognition is a priority when choose job <sup>2</sup>	-1,492 (1,378)	590.1 (824.9)
Constant	-27,025*** (6,200)	-9,221*** (3,499)
Difference <sup>3</sup>	10,627.20	-9,338.30
Observations	3,827	3,971
R-squared	0.197	0.121

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Dependent variable: Difference = actual earnings – hypothetical earnings where hypothetical earnings are the earnings the individual would have earned if had chosen not to have children  
Hence, a negative coefficient means that actual earnings have decreased relative to hypothetical earnings

<sup>1</sup>Number of years spent in part/full-time employment<sup>2</sup>Dummy equals 1 if agree with the statement<sup>3</sup>Mean of the dependent variable for the observations considered in the regression

# 6

## Discussion

### 6.1 The treatment effect

The average treatment effect for men is found to be an annual gain of 6,810€ and an annual loss of 10,526€ for women. The interpretation of these figures is that the average income for men with children is 6,810€ higher on average each year than comparable men without children. The men are comparable in their demographics and in their years of education in high school. Therefore, fathers'  $Y(1)$  exceeds their  $Y(0)$  as fathers earn more than comparable men without children. For women the result implies that women with children earn on average 10,526€ less per year than comparable women without children over the relevant time period of 20 years after childbirth. Therefore, they miss out on a large portion of their hypothetical income by having chosen to have children.

There seems to be clear evidence that childbirth has statistically significant effects on women and men which, when tested, were also significantly different from each other. However, even though the effects are significant they are only economically meaningful if they are of a relevant magnitude. Even if the impact of childbirth on income is found to be significant it might not be meaningful if it only leads to a trivial difference between individuals with and without children. To assess how substantial the effect is one needs to compare the average treatment effect to



the relevant labour income means. For men with children I find the average labour income to be 34,500€ a year compared to 12,000€ for women. This means that men with children earn around 20% more income compared to what they would be earning if they had decided not to have children. Women with children earn on average over 80% less than women without children. It is needless to say that treatment therefore has a substantial effect on income. This is especially important as the effect is calculated for a 20 year-long period, which is the majority of the average individual's time spent in employment (Eurostat 2019; Hytti and Valaste 2009).

All things considered, I believe that the treatment effect is robust as it does not change substantially when changing different parts of the specification. For example, changing the number of nearest neighbours matched on has no effect on the significance of the treatment effect and just a small effect on the magnitude. The same holds for changing the caliper, not allowing for replacement or not considering that the propensity scores are estimated. As long as I only include covariates which are not affected by treatment, no obvious change to any part of the setting can change the key results. Of course, there still exists a number of unobserved factors for which one cannot control that affect the treatment effect. As shown previously, controlling for time invariant unobserved heterogeneity does not change the sign or the significance of the treatment effect with the magnitude only changing very slightly. Therefore, assuming that time varying unobserved heterogeneity is not of immense magnitude, the significance and sign of the treatment effect is robust. I believe that this assumption is not very strong as it is unlikely that childbirth would coincide with a universal change in unobserved time varying heterogeneity which is of large enough to significantly affect the treatment effect.

These findings are in strong agreement with the majority of the published literature on the topic, with almost all papers finding a negative effect of children on female's income and the majority of them finding a positive effect of children on male's income (Daniel, Lacuesta, and Rodriguez-Planas 2013; Petersen, Penner, and Høgsnes 2014; Waldfogel 1997; Waldfogel 1995). A key driver is the effect of childbirth on the average working hours of men and women.

Whereas men's working hours increase upon childbirth women's working hours decrease, leading to this divergence in income (Lundberg and Rose 2000). Kleven, Landais, Posch, et al. (2019) supports my findings as they find that the intensive margin is more important than the extensive margin in Germany which holds true even more so in the long-run. In addition to different working hours resulting into differences in labour market experiences, breaks from the labour market have also been shown to reduce income (Gangl and Ziefle 2009). Especially for women with more than one child this is of relevance as they are likely to take some parental leave for each child, in turn leading to more than one break from the labour market. Furthermore, mothers sort themselves into lower paying occupations compared to women without children, whereas fathers sort themselves into higher paying occupations compared to men without children (Gibb et al. 2014; Lundberg and Rose 2000; Petersen, Penner, and Hogsnes 2007).

## 6.2 The child penalty

Because of childbirth affecting men's and women's earnings in the opposite direction I find that the child penalty is driven not only by women's negative treatment effect but also by men's positive treatment effect. When contrasting father's income to childless men's income I find that fathers earn significantly more from event time 6 onward until the end of the period studied. These results are no anomaly in the literature and have been found in many other studies (Gibb et al. 2014; Lundberg and Rose 2000; Simonsen and Skipper 2008). The reason for this is that men with children earn higher wages whilst also working longer hours plus being less likely to be out of work (Gibb et al. 2014). There can be many explanations for these results. Men with children face higher expenses as they have to support a family and hence choose to work longer hours. Additionally, it could be that men with children are more educated than men without children and hence have a larger amount of human capital accumulated through schooling. I only matched on the type of school degree and hence did not control for education completed thereafter. My sample and the literature support this hypothesis with the average father having

completed one further year of education compared to the average childless man at the age of 50. However, Lundberg and Rose (2002) finds that even controlling for education, men with children earn more than men without children. Additionally, the literature finds that fathers sort themselves into higher paying occupations than men without children (Gibb et al. 2014; Lundberg and Rose 2000; Petersen, Penner, and Høgsnes 2011). Hence, it seems to be a combination of factors which cause this positive treatment effect for men with children.

I have found that the contribution of women's treatment effect to the child penalty is largely caused by most mothers working fewer hours with many dropping out of the labour market altogether. As shown in figure 5.12 I do not find a significant effect of childbirth on the hourly pay of mothers. This result might be slightly less robust due to a significantly smaller sample size. It could be that the reason why the hourly wage is not impacted by motherhood is that many mothers do not re-enter the labour market because the wage they get offered does not cover their opportunity cost. Therefore, if mothers' opportunity cost is a lot higher compared to similar women without children then this difference in opportunity cost might explain why motherhood does not affect the hourly wage. This is not an unreasonable assumption as mothers have to miss out on time with their children to work and have to pay for childcare which might offset the income earned through working.

Many studies have found that there is indeed a negative effect of motherhood on hourly wages (Waldfogel 1997; Budig and England 2001; Livermore, Rodgers, and Siminski 2011). Previous literature has argued that the effect of childbirth on labour market experience is one of the key reasons why women earn lower wages when returning to the labour market (Taniguchi 1999). Human capital is a key driver of income which is accumulated through schooling and labour market experience (Becker 1985; Harmon, Oosterbeek, and Walker 2003; Patrinos and Psacharopoulos 2002). Therefore, if motherhood reduces labour market experience then experience could be a key driver of the child penalty (Becker 1985; Harmon, Oosterbeek, and Walker 2003; Patrinos and Psacharopoulos 2002). Budig and England (2001) estimated that a third of the motherhood penalty can be explained

by a loss in human capital through fewer years of schooling and work experience. Waldfogel (1997) finds similar results with the majority of the motherhood penalty being explained by the loss in experience. In my sample women with children have actually completed one more year of education at the age of 50 than childless women. Hence, the amount of education received cannot be the main driver of the child penalty. This leads me to match on experience as shown in figure 5.13. I find that experience does play a major part in the treatment effect especially a few years after the birth of the first child. As I use panel data, I can examine how the importance of the loss in experience changes relative to the time of the first birth. Unsurprisingly, the importance of experience increases with time relative to event time 0. With the birth of the first child mothers will start losing years of experience compared to women without children. This gap in years of experience increases with time and could therefore explain more of the child penalty later on. This is also a slightly puzzling result as it seems to suggest that motherhood also has an impact on the hourly wage which is not what I found previously. This could be due to the fact that by matching on part- and full-time experience, I also make women more similar in terms of their employment status. Therefore, I do not only control for work experience but also indirectly control for labour hours worked.

In conclusion, the estimated child penalty is supported by the existing literature and economic intuition. A slightly puzzling result remains the substantial short-run penalty which remains even after matching on labour market outcomes.

### 6.3 Regressions

The following paragraphs attempt to discuss the robustness of the regressions. I try to do so by assessing whether the associations found are supported by economic intuition and the literature. Furthermore, I attempt to shine light on how the regression results can potentially explain parts of the difference in signs of the average treatment effect for men and women.

The Mincer earnings equation uses education, work experience and work experience squared as the independent regressors to predict income. Theoretically

and empirically the literature has found substantial and statistically significant positive returns to education with men mostly having larger returns to education than women (Harmon, Oosterbeek, and Walker 2003; Boockmann and Steiner 2006). In all of the regressions for both men and women I find this to be the case with actual income rising relative to their counterfactual income. It is important to point out that during matching I only controlled for education completed at school and not higher education which allowed for this positive association to occur. A similar picture can be seen when looking at full-time experience which has a significant positive association with the dependent variable. For women this also holds for part-time work whereas for men it does not. The reason for this could be that men who choose to work part-time are therefore also more likely to choose a family friendly job which pays worse. Therefore, if I do not control for family friendliness of the job, part-time experience might indirectly pick up that association. The significance of this negative association falls from 1% to 10% once I control for these preferences indicating that it is not part-time experience itself but other factors which are included in the error term at that point (table 5.2). The regressions do not find a strong negative association between the dependent variable and experience squared, which is not what the theory predicts. However, the association for full-time experience squared for men and part-time experience squared for women is always negative but not always significant.

Furthermore, I find that marriage seems to be negatively correlated with treatment for women and positively for men. There can be many reasons for this, including gender norms and the effect of marriage on the financial security for women (Persson 2020). An additional reason could be the German joint taxation of married couples with full income splitting (Ehegattensplitting)<sup>6</sup> which divides the total income earned by the couple by two before calculating tax rates (Wersig n.d.). If you declare to have a primary earner and a secondary earner, the primary earner's income gets taxed at a relatively lower rate and the secondary earner's income at a relatively higher tax rate. Evidently, women are in most cases the secondary

---

<sup>6</sup>Legal basis §32a Abs. 5. EStG : Paragraph 32 section 5 in the income tax act

earner which leads to the "splittingtariff" to increase men's participation rates and decreases women's participation rates (Bach et al. 2011; Bick and Fuchs-Schündeln 2017; Bundesministerium der Finanzen 2020; Fehr and Ujhelyiova 2013; Steiner and Wrohlich 2004). This effect is driven by men's supply of work being relatively inelastic to changes in income tax rates whereas women's labour supply is found to be more elastic with respect to income tax changes, especially on the participation margin (Keane 2011). The wage premium for married men is an established fact in the literature and is explained by higher human capital, sorting into higher paying occupations, wage differential, higher productivity and the selection of married men (Antonovics and Town 2004; K. Daniel 1995; Ginther and Zavodny 2001).

The negative association between living in east Germany and men's treatment effect might be surprising, especially as women do not seem to display this effect. The explanation for this stems from large cultural and economic differences especially before the reunification of East and West Germany. In the East, women's labour participation was 85% compared to 55% for women in the west (Maier 1993; Trappe and Rosenfeld 1998; Trappe and Sørensen 2006). There were many reasons for those stark difference. Key differences came from a larger need for women to work due to worse economic conditions, better childcare, more liberal abortion laws and different attitudes towards gender roles in the East (Konietzka and Kreyenfeld 2002). There has been convergence in the last 30 years with some of these differences fading, however many differences still persist today. These differences I find to be confirmed with 28% of women in the West working full-time compared to 37% in the East. Women in the eastern regions of Germany are more attached to the labour market which reduces their time for household production. Therefore, men have to help out more with home production which could explain this negative correlation between men living in the east and the treatment effect. Additionally, wages are in general lower in the East potentially reducing the wage gap between fathers and men without children which is the key driver of the positive treatment effect for fathers.

Many studies have found that mothers who have children at a young age experience a larger motherhood penalty which is why many women, especially

highly educated ones, choose to delay having children (S. P. Martin 2000; Waldfogel 1998; Amuedo-Dorantes and Kimmel 2005; Miller 2011). Given those findings you expect women and potentially men as well to have a positive association between the age at which they have their first child and the treatment. Surprisingly, this is only the case for men with women experiencing a significant negative association. This seems to suggest that there is some misspecification in the model.

In summary, I believe that most significant associations found are in line with the economic intuition. The key exception remains the negative relationship between age of childbirth and the dependent variable for women. In conclusion, one should be careful interpreting these results with some appearing to be spurious correlations.

Lastly I try to assess the robustness of the regression results, when adding in the variables of interest, by analysing whether the associations found are in line with what the literature finds. The first regression focuses on the relationship between gender norms and the effect of children on income. For the majority of households women are in charge of most of the home production and men are in charge of labour market production (Layte 2018; Sullivan 2000). One of the key reasons why this is the case is gender norms and preferences. Kleven, Landais, and Søgaaard (2019) discovered that countries with more traditional views on gender roles experience much larger child penalties. Here, I try to show that this same effect exists on an individual level. In my regression I observe that men with more modern views on gender roles have smaller positive treatment effects than fathers who have more traditional views. The reverse is true for women. Therefore, it is probable that men who are of the opinion that they have the same responsibility as women in home production, are also more likely to help with home production. This will lead to them spending more time on house production and less time working on the labour market which explains the lower actual earnings relative to the potential earnings. Thus, this may translate into women splitting home production more equally with their partner if possible and to increase their focus on production in the labour market. Women with children spend on average significantly more time on home production than women without children. The time spent on domestic work increases

substantially less upon childbirth for men than it does for women (Gjerdingen and Center 2005). Hence, the belief of being equal in the household and at work allows them to reduce this difference between the time they spend on house production compared to comparable women without children. Consequently, they increase their actual income which causes their individual treatment effect to fall (Nix, Andresen, et al. 2019; Kleven, Landais, Posch, et al. 2019). An interesting extension here would be to look at couples and see how their gender views affect each other.

The second and third regressions I look at try to explain the treatment effects through differences in preferences regarding risk aversion and job characteristics. The regressions show that the willingness of women to take on risk for their career seems to not be correlated with their treatment effects. For fathers however being risk averse regarding their career, reduced their actual income relative to their counterfactual income. The reverse is true for fathers who have a higher risk tolerance. In my sample the mean risk women reported to be willing to take, on a scale of  $10^7$ , was 3.8 compared to 4.4 for men which is significantly different at the 1% level. This finding is supported by the literature with Croson and Gneezy (2009) reporting that women are, on average, more risk averse than men. The reason these differences in risk aversion can explain wage differentials is that for two occupations which are the exact same apart from the variability of earnings, one will be paid a compensating wage differential to induce workers to accept these higher levels of risk (Bertrand 2011). If men take these jobs more often than women due to different levels of risk aversion, than risk aversion could explain part of the difference in the treatment effect. However, one would then expect there to be a positive association between risk tolerance and the treatment effect for women, as higher risk tolerance should mean that they are more likely to take these higher paying jobs. Given the correlations men and women, who are both willing to take on a lot of risk for their job, will experience a higher child penalty because the men's treatment effect increased whereas the women's treatment effect stayed the same.

---

<sup>7</sup>With 10 being the highest willingness to take risk for career



For women and men job characteristics are of huge importance regarding their associations with the treatment effect. Thus, if preferences for job characteristics differ significantly between men and women, they can partly explain the large difference in treatment effects between men and women. Women seem to agree more often with the statement that family friendliness is a priority, with 64% of women saying family friendliness is very important compared to 42% of men. As shown in section 5 there is evidence to believe that family friendly jobs pay significantly less than family unfriendly jobs as they are correlated with lower actual earnings relative to hypothetical earnings. This has been supported by many other studies which find that family friendly occupations pay less (Gangl and Ziefle 2009). This difference in preference explains partly why a higher percentage of women work in the public sector, which is known to be more accommodating to family needs than the private sector (Kleven, Landais, and Sogaard 2019). Other studies have found similar results with Schumann, Ahlburg, and Mahoney (1994) detecting a decrease in the pay differential between men and women when controlling for job characteristics. Furthermore, 10% more men than women say that income is very important and 15% more men than women say career advancement is very important to them. This shows that women put less emphasis on payment and job advancement and more emphasis on family friendliness which explains parts of the child penalty. These difference in preferences are likely fostered by societal views on gender roles.

In addition to these differences existing between men and women, they also exist within the sexes between individuals with and without children. The subsequent paragraphs attempt to explain where these differences in preferences and views in gender roles lie and what they could mean. Having a look at the difference in gender norms I find an interesting trend with women without children having more modern views on gender roles compared to women with children. The largest difference was detected for the variable measuring whether the individual believed that women and men were equal at home and at work, with women without children agreeing 17% more often with this statement than women with children. However, the sample size for women without children was small with around 280 sampled answers leading to

large confidence intervals meaning that only two were significantly different at the 5% level. Interestingly the trend was always the same which leads me to believe that with a bigger sample one would have the same result. For men there does not seem to be as clear pattern with the only robust finding being that men with and without children believing in more traditional gender roles than women which is significant at the 1% level. These gender roles could partly explain why women without children chose to not have children. This implies that there is self-selection into treatment. It is also possible that having children affects these views with to-be-mothers having similar views on gender roles as childless women but then changing those views when becoming mothers. The sample size here is too small to carry out this sort of analysis and hence I cannot suggest what the cause is. However, when doing a similar analysis in the subsequent paragraph I find that other preferences change significantly after childbirth. This suggests that the same might be the case for gender norms implying that it is not due to self-selection.

Looking at the willingness to take on risk in their career, I observe that men with children report to be willing to take on more risk which is significant at the 10% level but of small magnitude. For women the difference is also significant at the 1% level but here women with children are less willing to take on risk for their career, with their willingness to take on risk being 3.7 out of 10 compared to 4.1 for women without children. Again, one needs to ask oneself if this is because women with and without children are intrinsically different or if parenthood affected their preferences. Evidence supports the latter. When looking at women with children before childbirth they seem to be willing to take on the same amount of risk, suggesting that it is age and childbirth which shift these preferences and not general differences in preferences between women with and without children. For men I find that there is also no difference between men with and without children before childbirth. To conclude, parenthood is associated with risk preferences in the opposite direction for both men and women which gives me a possible explanation for the differences in treatment effects.

Unsurprisingly, for both women and men with children, family friendliness is significantly more important than for men and women without children with the difference being significant but of small magnitude of around 7%. Similar to risk preferences I find that for women without children, income and career advancement is significantly more important than it is for women with children. However, these differences are brought about through childbirth and age and are not due to systematic differences in characteristics and preferences. Therefore, these changes in preferences can explain self-selection of mothers into more family friendly occupations which leads to the negative treatment effect.

It's essential to point out that the control group for gender norms is rather small with only 300 women and 300 men. This should still be large enough for the tests to be of power but might not allow for generalisation to the rest of Germany. For willingness to take on risks, sample size of the control increases to almost 10,000 for men and 7,000 for women. For priorities of job characteristics, the sample size for the control group is raised to 900 for men and 800 for women. Hence, these results are likely to be more reliable.

To conclude, the negative impact of childbirth on women's earning can partly be explained by gender norms and differences in preferences regarding job characteristics and the level of risk an individual is willing take for their career. Therefore, one reason why mothers earn less is due to women self-selecting themselves into careers and jobs which pay lower wages. This self-selection is enforced by childbirth which leads to an even bigger difference in preferences regarding job characteristics. It is important to consider that these preferences are most likely formed by gender norms themselves which make women more likely to be in charge of home production.

# 7

## Conclusion

In summary, I used observational data on almost 55,000 German residents to estimate the effect of having children on women's and men's annual gross earnings. The key contribution of this paper to the literature is the use of propensity score matching to determine the treatment effect of childbirth on earnings and the child penalty. My analysis proves the existence of a persistent long-run child penalty which is of similar magnitude to what has previously been found in the literature (Kleven, Landais, and Søgaard 2019; Kleven, Landais, Posch, et al. 2019). To conclude, the paper has five main findings.

The first result, is that there is a negative effect of childbirth on women's earnings of around 10,500€. Men, however, experience a large positive effect on their income of almost 7,000€. This is the average annual effect of having children on income not controlling for labour market outcomes when considering the first 20 years after the first child is born. When matching on work experience to explain parts of this loss in earnings, I find that the loss in experience caused by childbirth is a key driver of mothers' loss in future earnings.

The second finding is the effect childbirth has on women's earnings compared to men. The result is that there is an average long-run child penalty of 63% which averages the child penalty over event time 5 to 20. Hence, women earn 63% less of their potential income compared to men.

These two findings support the claim by Kleven, Landais, and Sjøgaard (2019) and Kleven, Landais, Posch, et al. (2019) that parenthood is the key driver of gender inequality in earnings.

The third finding is that parenthood drives women to choose employment in family friendly environments and to put less emphasis on payment and job advancement. Both priorities are positively correlated with lower incomes. For men, parenthood has the opposite effect. In a nutshell: Parenthood for men is a motivator to earn more money whereas for women it is a demotivating factor. This is in line with the traditional gender role models: "A father hunts, a mother feeds."

This leads me to my fourth finding which finds a significant positive correlation between actual income relative to counterfactual income and more modern views of gender norms for women. For men the opposite holds. Hence, gender roles seem to affect not only women's but also men's treatment effect. This suggests that more modern gender roles reduce the child penalty by reducing the positive treatment effect for men and the negative treatment effect for women. This complements the finding by Kleven, Landais, and Sjøgaard (2019) that national views on gender roles affect a country's child penalty.

The fifth finding is that fathers get punished in terms of income relative to childless men for being more risk averse regarding their career and rewarded if they are less risk averse. For women I find similar patterns which are however found to be insignificant. Therefore, the contribution of men to the child penalty through experiencing a positive treatment effect is partly due to fathers being willing to take on more risk for their career than childless men.

This paper has the following policy implications for governments who wish to reduce child-related gender inequality in pay. One of the central findings has been that women lose 63% more of their potential incomes than men due to childbirth. The reason for this is that most women cannot carry out the same amount of work as men which is due to the fact that, in the majority of households, women remain responsible for home production. This means that the child penalty is actually much larger than the differential in pay for equal work. Hence, to even get to

an equal pay for equal work point, society needs to make it easier for women to carry out equal work. A possible way of allowing women to stay attached to the labour market participation after having children is to reduce the cost of childcare. This subsequently affects the choice whether going back to work is feasible or not. Previous research has found negative estimates of the elasticity of female labour supply to childcare costs suggesting that making childcare cheaper increases female labour supply (D. M. Blau and Robins 1988; Gong, Breunig, and King 2010; Andresen 2013; Bettendorf, Jongen, and Paul 2015; Bauernschuster and Schlotter 2015; Akgunduz and Plantenga 2018). However, there does exist some evidence that subsidised childcare simply crowds out informal childcare leading to little causal effect of subsidised childcare on women's labour supply (Havnes and Mogstad 2011).

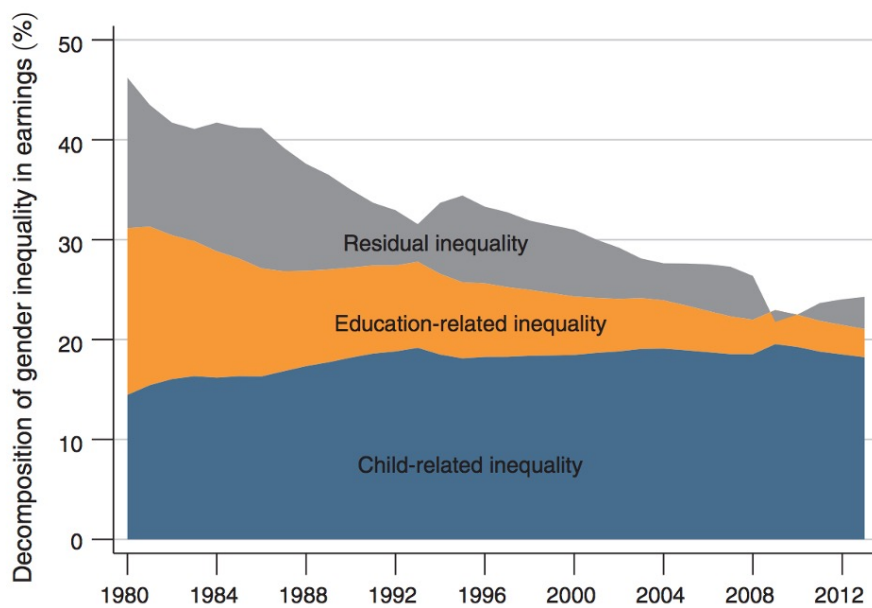
From a policy point of view, it would also be of interest to continue the work on the linkage between policy reforms and their effects on gender norms (Unterhofer and Wrohlich 2017). If certain family policies are found to significantly impact gender norms especially regarding children, then this could help to reduce child-related gender inequality.

# Appendices

# A

## Appendix

### A.1 Supporting material



**Figure A.1:** The composition of gender inequality in earnings

Source: Kleven *et al.* (2018)



Table A.1: Sample means

VARIABLES	Men			Women		
	No children	Children	Difference	No children	Children	Difference
Year of Birth	1966 (19.39)	1966 (10.09)	0.27*** (0.076)	1972 (16.74)	1968 (10.21)	4.033*** (0.066)
East-Germany dummy	0.216 (0.411)	0.194 (0.395)	0.022*** (0.002)	0.153 (0.36)	0.188 (0.391)	-0.035*** (0.001)
Year of immigration	1987 (17.39)	1984 (13.16)	2.275*** (0.196)	1989 (15.93)	1986 (13.25)	2.932*** (0.199)
Full time experience <sup>1</sup>	14.64 (14.38)	15.65 (8.421)	-1.013*** (0.058)	8.308 (11.12)	7.237 (6.074)	-1.809*** (0.022)
Part time experience <sup>1</sup>	0.631 (2.012)	0.582 (1.834)	0.05*** (0.009)	1.551 (3.898)	3.36 (4.365)	1.071*** (0.042)
Years in education	11.91 (2.613)	12.45 (2.869)	-0.546*** (0.014)	12.18 (2.767)	12.37 (2.674)	-0.196*** (0.014)
Age	37.79 (15.26)	39.05 (7.978)	-1.259*** (0.06)	32.33 (13.7)	36.48 (7.607)	-4.155*** (0.052)
Total number of children	0	1.988 (0.919)	-1.989*** (0.003)	0	1.97 (0.904)	-1.97*** (0.004)
Birth year of 1st child	—	1996 (9.811)	—	—	1995 (10.71)	—
Annual earnings <sup>2</sup>	18,482 (23,265)	34,532 (28,239)	16050*** (125.902)	13,568 (17,953)	12,071 (15,021)	1496.14*** (81.853)
Married dummy	0.35 (0.477)	0.819 (0.385)	-0.468*** (0.002)	0.251 (0.434)	0.739 (0.439)	-0.488*** (0.002)
Age had first child	—	30.17 (5.293)	—	—	27.35 (4.834)	—
Born outside Germany dummy	0.13 (0.336)	0.173 (0.379)	-0.043*** (0.002)	0.123 (0.328)	0.159 (0.366)	-0.036*** (0.001)

\*\*\*p&lt;0.01, \*\*p&lt;0.05, \*p&lt;0.1

Robust standard errors in parentheses for difference columns and standard deviations for all other columns

<sup>1</sup>Number of years spent in part/full-time employment<sup>2</sup>From primary, secondary employment and self-employment

# Bibliography

- Akgunduz, Yusuf Emre and Janneke Plantenga (2018). “Child care prices and maternal employment: A meta-analysis”. In: *Journal of Economic Surveys* 32.1, pp. 118–133.
- Amuedo-Dorantes, Catalina and Jean Kimmel (2005). “The motherhood wage gap for women in the United States: The importance of college and fertility delay”. In: *Review of Economics of the Household* 3.1, pp. 17–48.
- Andresen, Martin Eckhoff (2013). “Women and Children First?: Labor market effects of universal child care for toddlers”. MA thesis.
- Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996). “Identification of causal effects using instrumental variables”. In: *Journal of the American statistical Association* 91.434, pp. 444–455.
- Antonovics, Kate and Robert Town (2004). “Are all the good men married? Uncovering the sources of the marital wage premium”. In: *American Economic Review* 94.2, pp. 317–321.
- Avellar, Sarah and Pamela J Smock (2003). “Has the price of motherhood declined over time? A cross-cohort comparison of the motherhood wage penalty”. In: *Journal of marriage and family* 65.3, pp. 597–607.
- Bach, Stefan et al. (2011). “Reform of income splitting for married couples: only individual taxation significantly increases working incentives”. In: *DIW Economic Bulletin* 1.5, pp. 13–19.
- Bauernschuster, Stefan and Martin Schlotter (2015). “Public child care and mothers’ labor supply—Evidence from two quasi-experiments”. In: *Journal of Public Economics* 123, pp. 1–16.
- Becker, Gary S (1985). “Human capital, effort, and the sexual division of labor”. In: *Journal of labor economics* 3.1, Part 2, S33–S58.
- Bertrand, Marianne (2011). “Chapter 17-new perspectives on gender”. In: *Handbook of Labor Economics* 4 part B.
- Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination”. In: *American economic review* 94.4, pp. 991–1013.
- Bettendorf, Leon JH, Egbert LW Jongen, and Muller Paul (2015). “Childcare subsidies and labour supply—Evidence from a large Dutch reform”. In: *Labour Economics* 36, pp. 112–123.
- Bevoelkerung (2020). <https://www.statistikportal.de/de/bevoelkerung/flaeche-und-bevoelkerung>. Accessed: 2020-01-08.
- Bevölkerung nach Geschlecht (2019). [shorturl.at/pqCGV](http://shorturl.at/pqCGV). Accessed: 2020-02-27.

- Bick, Alexander and Nicola Fuchs-Schündeln (2017). “Quantifying the disincentive effects of joint taxation on married women’s labor supply”. In: *American Economic Review* 107.5, pp. 100–104.
- Blau, David M and Philip K Robins (1988). “Child-care costs and family labor supply”. In: *The Review of Economics and Statistics*, pp. 374–381.
- Blau and Lawrence M Kahn (2017). “The gender wage gap: Extent, trends, and explanations”. In: *Journal of Economic Literature* 55.3, pp. 789–865.
- Blickpunkt Arbeitsmarkt (2019). <https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Tabellen/erwerbstaetigenquoten-gebietsstand-geschlecht-altergruppe-mikrozensus.html>. Accessed: 2020-02-24.
- Boockmann, Bernhard and Viktor Steiner (2006). “Cohort effects and the returns to education in West Germany”. In: *Applied Economics* 38.10, pp. 1135–1152.
- Breierova, Lucia and Esther Duflo (2004). *The impact of education on fertility and child mortality: Do fathers really matter less than mothers?* Tech. rep. National bureau of economic research.
- Budig, Michelle J and Paula England (2001). “The wage penalty for motherhood”. In: *American sociological review*, pp. 204–225.
- Bundesministerium der Finanzen (2020). *Zur Reform der Besteuerung der Ehegatten*. [https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Ministerium/Geschaeftsbereich/Wissenschaftlicher\\_Beirat/Gutachten\\_und\\_Stellungnahmen/Ausgewaehlte\\_Texte/2018-09-27-Gutachten-Besteuerung-von-Ehegatten-anlage.pdf?\\_\\_blob=publicationFile&v=2](https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Ministerium/Geschaeftsbereich/Wissenschaftlicher_Beirat/Gutachten_und_Stellungnahmen/Ausgewaehlte_Texte/2018-09-27-Gutachten-Besteuerung-von-Ehegatten-anlage.pdf?__blob=publicationFile&v=2). Accessed: 2020-01-04.
- Bundesministerium für Wirtschaft und Energie (2019). *Jahresbericht der Bundesregierung zum Stand der Deutschen Einheit*. [https://www.bmwi.de/Redaktion/DE/Publikationen/Neue-Laender/jahresbericht-zum-stand-der-deutschen-einheit-2019.pdf?\\_\\_blob=publicationFile&v=26](https://www.bmwi.de/Redaktion/DE/Publikationen/Neue-Laender/jahresbericht-zum-stand-der-deutschen-einheit-2019.pdf?__blob=publicationFile&v=26). Accessed: 2020-02-24.
- Card, David (1993). *Using geographic variation in college proximity to estimate the return to schooling*. Tech. rep. National Bureau of Economic Research.
- (2001). “Estimating the return to schooling: Progress on some persistent econometric problems”. In: *Econometrica* 69.5, pp. 1127–1160.
- Choi, Jungyeon, Olaf M Dekkers, and Saskia le Cessie (2019). “A comparison of different methods to handle missing data in the context of propensity score analysis”. In: *European journal of epidemiology* 34.1, pp. 23–36.
- Crosen, Rachel and Uri Gneezy (2009). “Gender differences in preferences”. In: *Journal of Economic literature* 47.2, pp. 448–74.
- D’Agostino Jr, Ralph B (1998). “Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group”. In: *Statistics in medicine* 17.19, pp. 2265–2281.
- Daniel, Kermit (1995). “The marriage premium”. In: *The new economics of human behavior*, pp. 113–25.
- Daniel, Aitor Lacuesta, and Núria Rodríguez-Planas (2013). “The motherhood earnings dip: Evidence from administrative records”. In: *Journal of Human Resources* 48.1, pp. 169–197.

- Daten zum durchschnittlichen Alter der Mutter bei Geburt nach der Geburtenfolge für 1. Kind, 2. Kind, 3. Kind und insgesamt 2018* (2019). <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Geburten/Tabellen/geburten-mutter-biologischesalter.html>. Accessed: 2020-03-28.
- Die Arbeitsmarktsituation von Frauen und Männern* (2019). <https://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Regionen/Politische-Gebietsstruktur/Ost-West-Nav.html>. Accessed: 2020-04-14.
- Erwerbstätigenquoten 1991 bis 2018* (2019). <https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Tabellen/erwerbstaetigenquoten-gebietsstand-geschlecht-altergruppe-mikrozensus.html>. Accessed: 2020-03-28.
- European commission (2017). *The gender pay gap situation in the EU*. [https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu\\_en](https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en). Accessed: 2020-04-25.
- Eurostat (2019). *Duration of working life – statistics*. [https://ec.europa.eu/eurostat/statistics-explained/index.php/Duration\\_of\\_working\\_life\\_-\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php/Duration_of_working_life_-_statistics). Accessed: 2020-05-15.
- Fehr, Hans and Daniela Ujhelyiova (2013). “Fertility, female labor supply, and family policy”. In: *German Economic Review* 14.2, pp. 138–165.
- Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti (2004). “Mothers and sons: Preference formation and female labor force dynamics”. In: *The Quarterly Journal of Economics* 119.4, pp. 1249–1299.
- Fitzenberger, Bernd, Katrin Sommerfeld, and Susanne Steffes (2013). “Causal effects on employment after first birth—A dynamic treatment approach”. In: *Labour Economics* 25, pp. 49–62.
- Forschungsbasierte Infrastruktureinrichtung 'Sozio-oekonomisches Panel (SOEP)'* (n.d.). [https://www.diw.de/de/diw\\_01.c.615551.de/forschungsbasierte\\_infrastruktureinrichtung\\_\\_sozio-oekonomisches\\_panel\\_\\_soep.html](https://www.diw.de/de/diw_01.c.615551.de/forschungsbasierte_infrastruktureinrichtung__sozio-oekonomisches_panel__soep.html). Accessed: 2020-05-03.
- Frankfurter Allgemeine Zeitung (2018). *Zahl der Geburten steigt weiter an*. <https://www.faz.net/aktuell/gesellschaft/menschen/deutschland-geburtenrate-steigt-weiter-deutlich-an-15516425.html>. Accessed: 2020-01-28.
- Gangl, Markus and Andrea 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”. In: *Demography* 46.2, pp. 341–369.
- Geis, Wido and Anja K Orth (2020). *Abwanderung Aus Ostdeutschland: Weniger Frauen Gehen, Männerüberschuss Bleibt Bestehen*. [www.iwkoeln.de/studien/iw-kurzberichte/beitrag/wido-geis-anja-katrin-orth-weniger-frauen-gehen-maennerueberschuss-bleibt-bestehen-302444.html](http://www.iwkoeln.de/studien/iw-kurzberichte/beitrag/wido-geis-anja-katrin-orth-weniger-frauen-gehen-maennerueberschuss-bleibt-bestehen-302444.html). Accessed: 2020-03-28.
- Gibb, Sheree J et al. (2014). “The effects of parenthood on workforce participation and income for men and women”. In: *Journal of Family and Economic Issues* 35.1, pp. 14–26.

- Ginther, Donna K and Madeline Zavodny (2001). "Is the male marriage premium due to selection? The effect of shotgun weddings on the return to marriage". In: *Journal of Population Economics* 14.2, pp. 313–328.
- Gjerdingen, Dwenda K and Bruce A Center (2005). "First-time parents' postpartum changes in employment, childcare, and housework responsibilities". In: *Social Science Research* 34.1, pp. 103–116.
- Goebel, Jan et al. (2019). "The German Socio-Economic Panel (SOEP)". In: *Jahrbücher für Nationalökonomie und Statistik* 239.2, pp. 345–360.
- Gong, Xiaodong, Robert V Breunig, and Anthony King (2010). "How responsive is female labour supply to child care costs: New Australian estimates". In: *SOEP 2016-Codebook for the \$ PEQUIV file 1984-2016: CNEF variables with extended income information for the SOEP*. Tech. rep. SOEP Survey Papers.
- Greenland, Sander and William D Finkle (1995). "A critical look at methods for handling missing covariates in epidemiologic regression analyses". In: *American journal of epidemiology* 142.12, pp. 1255–1264.
- Harmon, Colm, Hessel Oosterbeek, and Ian Walker (2003). "The returns to education: Microeconomics". In: *Journal of economic surveys* 17.2, pp. 115–156.
- Havnes, Tarjei and Magne Mogstad (2011). "Money for nothing? Universal child care and maternal employment". In: *Journal of Public Economics* 95.11-12, pp. 1455–1465.
- Ho, Daniel E et al. (2007). "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference". In: *Political analysis* 15.3, pp. 199–236.
- Hytti, Helka and Maria Valaste (2009). "The average length of working life in the European Union". In: *The Social Insurance Institution; Online working papers 1/2009*.
- Keane, Michael P (2011). "Labor supply and taxes: A survey". In: *Journal of Economic Literature* 49.4, pp. 961–1075.
- Kleven, Henrik, Camille Landais, Johanna Posch, et al. (2019). "Child penalties across countries: Evidence and explanations". In: *AEA Papers and Proceedings*. Vol. 109, pp. 122–26.
- Kleven, Henrik, Camille Landais, and Jakob Egholt Søggaard (2019). "Children and gender inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11.4, pp. 181–209.
- Konietzka, Dirk and Michaela Kreyenfeld (2002). "Women's employment and non-marital childbearing: A comparison between East and West Germany in the 1990s". In: *Population* 57.2, pp. 331–357.
- Layte, Richard (2018). *Divided time: gender, paid employment and domestic labour*. Routledge.
- Lemieux, Thomas (2006). "The "Mincer equation" thirty years after schooling, experience, and earnings". In: *Jacob Mincer a pioneer of modern labor economics*. Springer, pp. 127–145.

- Lian, Jason Z and David Ralph Matthews (1998). “Does the vertical mosaic still exist? Ethnicity and income in Canada, 1991”. In: *Canadian Review of Sociology/Revue canadienne de sociologie* 35.4, pp. 461–481.
- Livermore, Tanya, Joan Rodgers, and Peter Siminski (2011). “The effect of motherhood on wages and wage growth: evidence for Australia”. In: *Economic Record* 87, pp. 80–91.
- Lundberg, Shelly and Elaina Rose (2000). “Parenthood and the earnings of married men and women”. In: *Labour Economics* 7.6, pp. 689–710.
- (2002). “The effects of sons and daughters on men’s labor supply and wages”. In: *Review of Economics and Statistics* 84.2, pp. 251–268.
- Maier, Friederike (1993). “The labour market for women and employment perspectives in the aftermath of German unification”. In: *Cambridge Journal of Economics* 17.3, pp. 267–280.
- Martin, Steven P (2000). “Diverging fertility among US women who delay child-bearing past age 30”. In: *Demography* 37.4, pp. 523–533.
- Martin, Teresa Castro (1995). “Women’s education and fertility: results from 26 Demographic and Health Surveys”. In: *Studies in family planning*, pp. 187–202.
- Miller, Amalia R (2011). “The effects of motherhood timing on career path”. In: *Journal of population economics* 24.3, pp. 1071–1100.
- Mincer, Jacob (1974). “Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.” In:
- Moberg, Ylva (2016). *Does the gender composition in couples matter for the division of labor after childbirth?* Tech. rep. Working Paper.
- Morgan, Charity J (2018). *Reducing bias using propensity score matching*.
- Nix, Emily, Martin Eckhoff Andresen, et al. (2019). *What causes the child penalty? evidence from same sex couples and policy reforms*. Tech. rep. Discussion Papers 902.
- Olivetti, Claudia and Barbara Petrongolo (2016). “The evolution of gender gaps in industrialized countries”. In: *Annual review of Economics* 8, pp. 405–434.
- Patrinos, Anthony Harry and George Psacharopoulos (2002). *Returns to investment in education: a further update*. The World Bank.
- Paull, Gillian (2008). “Children and women’s hours of work”. In: *The Economic Journal* 118.526, F8–F27.
- Persson, Petra (2020). “Social insurance and the marriage market”. In: *Journal of Political Economy* 128.1, pp. 000–000.
- Petersen, Trond, Penner, and Geir Hogsnes (2007). “The motherhood wage penalty: Sorting versus differential pay”. In:
- Petersen, Trond, Penner, and Geir Høgsnes (2011). “The male marital wage premium: Sorting vs. differential pay”. In: *ILR Review* 64.2, pp. 283–304.
- (2014). “From motherhood penalties to husband premia: The new challenge for gender equality and family policy, lessons from Norway”. In: *American Journal of Sociology* 119.5, pp. 1434–1472.
- Psacharopoulos, George (1985). “Returns to education: a further international update and implications”. In: *Journal of Human resources*, pp. 583–604.

- Psacharopoulos, George and Keith Hinchliffe (1973). *Returns to education: An international comparison*. Elsevier Amsterdam.
- Rosenbaum, Paul R and Donald B Rubin (1983). “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70.1, pp. 41–55.
- (1985). “Constructing a control group using multivariate matched sampling methods that incorporate the propensity score”. In: *The American Statistician* 39.1, pp. 33–38.
- Rubin, Donald B (2001). “Using propensity scores to help design observational studies: application to the tobacco litigation”. In: *Health Services and Outcomes Research Methodology* 2.3-4, pp. 169–188.
- Schmitt, Christian and Gert G Wagner (2006). “Kinderlosigkeit von Akademikerinnen überbewertet”. In: *DIW Wochenbericht* 73.21, pp. 313–317.
- Schumann, Paul L, Dennis A Ahlburg, and Christine Brown Mahoney (1994). “The effects of human capital and job characteristics on pay”. In: *Journal of Human Resources*, pp. 481–503.
- Share of Age Group with Specified Level of Educational Attainment by Gender* (2010). <http://datatopics.worldbank.org/education/wDashboard/dqattainmentage>. Accessed: 2020-01-29.
- Simonsen, Marianne and Lars Skipper (2006). “The costs of motherhood: an analysis using matching estimators”. In: *Journal of Applied Econometrics* 21.7, pp. 919–934.
- (2008). “An empirical assessment of effects of parenthood on wages”. In: *Advances in Econometrics* 21, pp. 359–380.
- Snedcor, George W and William G Cochran (1980). *Statistical Methods*. 7th. Iowa State Univ. Press.
- SOEP (2019). “Socio-Economic Panel (SOEP), Data from 1984-2017”. Version 34. In: DOI: 10.5684/soep.v34.
- Solé, Carlota and Sonia Parella (2003). “The labour market and racial discrimination in Spain”. In: *Journal of Ethnic and Migration Studies* 29.1, pp. 121–140.
- Steiner, Viktor and Katharina Wrohlich (2004). “Household taxation, income splitting and labor supply incentives—a microsimulation study for Germany”. In: *CESifo Economic Studies* 50.3, pp. 541–568.
- Stuart, Elizabeth A (2010). “Matching methods for causal inference: A review and a look forward”. In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 25.1, p. 1.
- Sullivan, Oriel (2000). “The division of domestic labour: twenty years of change?” In: *Sociology* 34.3, pp. 437–456.
- Taniguchi, Hiromi (1999). “The timing of childbearing and women’s wages”. In: *Journal of Marriage and the Family*, pp. 1008–1019.
- Trappe, Heike and Rachel A Rosenfeld (1998). “A comparison of job-shifting patterns in the former East Germany and the former West Germany”. In: *European Sociological Review* 14.4, pp. 343–368.
- Trappe, Heike and Annemette Sørensen (2006). “Economic relations between women and their partners: an East and West German comparison after reunification”. In: *Feminist Economics* 12.4, pp. 643–665.

- Unterhofer, Ulrike and Katharina Wrohlich (2017). "Fathers, parental leave and gender norms". In:  
*Verbraucherpreisindex für Deutschland - Lange Reihen ab 1948 - April 2020* (2020). <https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/Publikationen/Downloads-Verbraucherpreise/verbraucherpreisindex-lange-reihen-pdf-5611103.pdf>. Accessed: 2020-03-28.
- Waldfogel, Jane (1995). "Women working for less: Family status and women's pay in the United States and United Kingdom." In:  
— (1997). "The effect of children on women's wages". In: *American sociological review*, pp. 209–217.  
— (1998). "Understanding the "family gap" in pay for women with children". In: *Journal of economic Perspectives* 12.1, pp. 137–156.
- Wersig, Maria (n.d.). "Overcoming Joint Taxation: The German Case". In: *Challenging Gender Inequality in Tax Policy Making* ().  
*Zusammengefasste Geburtenziffer nach Kalenderjahren* (2019). <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Geburten/Tabellen/geburtenziffer.html>. Accessed: 2020-03-28.