# Children and Their Effect on the Gender Wage Gap and Possible Heterogeneity in Germany

Bachelor Thesis in Business Analytics & Quantitative Marketing

*BSc$^2$ Econometrics and Economics*

Date: July 2 2021


Under Supervision of

Dr. Pilar García Gómez

Second Assessor

Dr. Mikhail Zhelonkin

Written By

Anouk Veltman (466547)

*Erasmus School of Economics*



**Abstract**

This research examines the presence of a child penalty in male and female earnings in Germany. To estimate the impact of children on gender inequality in the labor market, an event study is used. A long term child penalty of around 70% is found in earnings, driven by a wage rate gap. Furthermore, possible heterogeneous treatment effect is explored with the use of Bayesian analysis. This heterogeneity is explored based on psychiatric problem indicators and self-rated health scores. However, these results have been found to be inconclusive. Based on the earnings distributions of men and women it is concluded that women have not caught up to the income level of men, leading to harder to interpret results for women.



Key Words: Child Penalty - Bayesian Analysis - Heterogeneous Treatment Effects


Erasmus University Rotterdam

# Contents

# 1 Introduction

For many years researchers have been attempting to find a reason for the pay gap between men and women (Kunze, 2005; Oostendorp, 2004; Petersen & Morgan, 1995). It is well known that on average women make less money than their male coworkers. For instance, in the United States (US) a woman makes about 0.84 cents for every dollar a man earns (Cohen & Huffman, 2007). A reason for the wage gap between men and women in the past can be attributed to the fact that women had a low education level and less experience in the work field. Starting in around 1900 the female workforce participation rate in the US started to increase. In 1900 this percentage was only around 20% which grew to 34% in 1950 and up to 61% in 2000, which is relatively close to the male rate of 75% (O'Neill, 2003). In their paper, O'Neill and Polachek (1993) show that an increase in female's work experiences accounts for about a fourth of the narrowing of the wage gap in the 80s. The increase in level of education of women, or women obtaining a higher education at all also contributes to this convergence. This increase in work force participation does not just apply to women in general but also to mothers, where in 1975 only about 47% of mothers with children under the age of 18 were working that percentage rose to 72 in 2002 (Pugh, 2005).

With the increased availability of the same resources to women as men, it would be expected that women have the same opportunities as men. Indeed, papers look into the trend of the narrowing wage gap in the 70s and 80s and predict that the relative earnings of white women will continue to increase in the years after the late 1980s (Blau & Beller, 1988). However, despite the convergence of the wage gap, there is still a wage gap present in current days, a good proportion of which cannot be attributed to education level or workforce participation (Kunze, 2018). Actually, recent studies for Western countries even show an increase in most measures of wage and earnings inequality in, for example, West Germany (Antonczyk, Fitzenberger, & Sommerfeld, 2010).

The concept that gender inequalities are larger at the top of the hierarchy compared to the bottom is known as the 'glass ceiling effect' (Cotter, Hermsen, Ovadia, & Vanneman, 2001). This is another effect showing that for some reason women tend to earn less money for the same work as men, and that this disparity only grows the more people earn. This phenomenon is another example of historic gender inequality leading to a wage gap. Historically 'women jobs' do not prepare women to end up in high managerial functions (Acker, 2009). However, with women being able to attend college and get the same jobs as men it is quite remarkable to see that this glass ceiling has not declined. Yet again, leading to the question what causes this inequality, if it is not

access to education or historic discrimination?

Because the effects of education or workforce participation are not able to explain the full extend of the wage gap, it is of interest to look into other factors playing into this phenomenon. This paper focuses on the effects of having children on the earnings of men compared to women. The so called 'Child Penalty' is defined as the percentage by which women fall behind in earnings compared to men due to having children. Previous papers found that children do in fact have a negative effects on women's pay (Waldfogel, 1997). This paper will explore these effects on a German sample.

This leads to the following main research question:

> *What is the impact of children on the labor market trajectories of women relative to men?*

With a large portion of the wage gap still remaining unexplained it is interesting to explore the effects that children have on the wage gap. This paper explores the effects of children on the earnings gap between men and women for a German sample. The first part of this research looks into the size of the child penalty and explores if this is a reason for the large gender wage gap. With this the impact of children on labor market trajectories are explored, and hence the main research question is answered. For the main research question, an event study is used which is discussed in detail in Section 4.1. This research focuses on Germany using data from the German Socio-Economic Panel (SOEP), of which a more detailed description is given in Section 3.

Aside from solely comparing males and females as a whole, the event study will also be performed on subgroups conditional on several characteristics. As it has been shown that women differ in preferences regarding employment and having children (Andresen & Nix, 2019), it is interesting to see if the child penalty changes when conditioning on employment status. Furthermore, to explore possible gender role division different family sizes, defined by the number of children in a household, is another variable conditioned on in this research. Finally, Anderson, Binder, and Krause (2002) suggest a human capital hypothesis stating that women at the top of the income distribution suffer a higher child penalty because they give up more human capital when getting pregnant. Hence, the sample is divided up into subgroups where the top 25% and 50% earning males and females are excluded.

Finding ways to close the wage gap is important for a wide variety of reasons. For instance, Aizer (2010) finds that decreases in the wage gap can reduce violence against women. When women are able to earn the same amount as men they feel more independent and are able to stand up for themselves more. So aside from the financial equality, the closing of the wage gap can also bring large personal and social benefits. This also brings on the question of whether women who, for instance, feel better about themselves, and are happier in their life before the birth of their first child are able to stand more ground in their workplace and possibly suffer a lower child penalty. Therefore, this research also looks into possible heterogeneous effects of children among women compared to men. Other papers look into the possible causes for women experiencing the larger wage shock after child birth. A possible explanation for this would be gender norms, and differences in male and female's preferences regarding child care. Andresen and Nix (2019) show that biology plays a role in the initial drop in income for the person giving birth to the child, as well as women having a preference for spending time with their child over career.

This paper explores possible variation in the child penalty which can be explained by characteristics of the men and women in the sample. Research shows that mental health effects many aspects of a person's life, among which is income. de Lange, Taris, Kompier, Houtman, and Bongers (2005) find a relationship between work and mental health, where mental health effects labor market outcomes and vice versa. The second part of research looks into the effects of psychiatric problems and self-rated health scores on the child penalty. Furthermore, this research looks into heterogeneity and how this can help target resources better, and possibly lead to policy improvement.

Leading to the following sub research questions:

*Are there certain factors leading to a heterogeneous treatment effect?*
*And how can this information be used to develop policies to decrease the wage gap?*

The rest of this paper is structured as follows, first an overview of previous literature discussing the child penalty is presented in Section 2. After this, an overview of the used data is presented in Section 3, followed by a description of the chosen methods in Section 4. Subsequently, the results are presented in Section 5, followed by a discussion of these results. Finally in Section 7 the results are summarized followed by giving policy recommendations and suggestions for further research.

# 2 Literature Review

Various papers find that parenthood is much more consequential for mothers than it is for fathers, when it comes to earnings and employment (Baker, 2010; Viitanen, 2014). Such studies find that a possible reason for the earnings gap between mothers and fathers is related to access to socioeconomic resources and support (Luxton, 2006). However, this research is still relatively new and uncertainty remains. Hence, the continuation of research into the phenomenon remains important.

## 2.1 Previous Findings

Kleven, Landais, Posch, Steinhauer, and Zweimüller (2019) find child penalties ranging from 21% in Denmark to 61% in Germany. For this research, it is of interest to see if similar results are obtained for Germany using a slightly more recent version of the data set with a slight variation in the data selection. As found by Kleven, Landais, Posch, et al. (2019) Germany suffers the highest child penalty out of all the countries in their study. A possible explanation for this could be that Germany has been known for applying a caregiver-strategy in support of women's care, rather than in their employment. In earlier years, when Germany was still divided into East and West, West-Germany was known for being a male-breadwinner/ female-homemaker state Sainsbury et al. (1999). This research looks into if the child-penalty is still present in as large of a magnitude today, and whether this male/ female divide still exists.

de Quinto, Hospido, and Sanz (2020) look into the child penalty estimate for Spain, using the event-study specification proposed by Kleven, Landais, and Søgaard (2019), which is also used in this research. They find that in the long run ($t = 10$), where $t$ denotes years, the child penalty in Spain is around 28%. They find that labor market outcomes follow a similar trajectory for men and women before the birth of their first child. After the birth of the first born child women's income diverges considerably. In this example, the child penalty can be attributed to several factors. The first being that women reduce their hours worked after their first child, contributing to a 9.8% child penalty in the first year growing up to 23% in the 10 years after. Furthermore, the probability of men working part time decreases by 8% whereas for women this goes up by 30% after the firstborn. All these factors are also interesting to explore for the German sample used for this research. This research explores these possible explanations by running the event study conditional on different employment statuses.

While various papers found quantitative evidence for a wage gap caused by the birth of children, these results vary per country and per paper. For a large study done in the US it was found that there is a wage penalty of around 7% per child (Budig & England, 2001). However, results from the US are hard to generalize to the rest of the world, because rules and regulations vary so much compared to other countries. For instance, the American maternity leave policy is found to be among one of the worst in the world, with women returning to work very fast. Over a third of women in the US return to work three months after the birth of a child, where in countries such as Sweden, Germany, and the United Kingdom this is only around 5% for new mothers (Gustafsson, Wetzels, Vlasblom, & Dex, 1996; Klerman & Leibowitz, 1990).

A suspected reason for the wage gap for mothers is thought to be related to the types of jobs that attract women who later on in life want to have children. These so called 'mother friendly' jobs are expected to be easier to combine with parenting, but most of the time also require the mother to give up the possibility of a higher wage. However, the characteristics of these types of job are not able to explain the majority of the penalty (Budig & England, 2001). Furthermore, not only do women seek out these jobs to feel more secure about their occupation later, a similar problem is also apparent when women already have children. The reasoning for the child penalty does not seem to be related to the actual characteristics of the jobs themselves, but to the tendency of mothers to work more part-time compared to non-mothers (Budig & England, 2001). Aside from before the birth of children it has also been shown that companies actively discriminate against hiring mothers compared to fathers Correll, Benard, and Paik (2007).

Anderson et al. (2002) look into seeing if highly skilled workers experience a higher motherhood penalty because they give up more human capital when becoming pregnant. They find a total penalty of around 15% per child, again, also using an American sample. This research uses females without children as the counterfactual. They do find that women with a lower education level suffer a way lower, to almost non-existent child penalty consistent with this human capital hypothesis. For high school graduates, the years they loose out in the workforce due to pregnancy is not able to account for the wage loss compared to their counterfactual. Hence, more research into possible heterogeneity is needed, possibly on other subgroups as well. Which is something that is further explored in the next section.

## 2.2 Heterogeneous Treatment Effects

Because it is important to investigate what factors affect the penalty for having children it is of interest to see if there is a variability in subgroups of women. For example, Anderson et al. (2002) and de Quinto et al. (2020) find estimates for workforce participation and wage that differ across different levels of education for women. This paper will also look into possible groupings of women and the differences in their grouped average child penalties. This will be of interest particularly for possible policy recommendations focused on closing the wage penalty for motherhood. Being able to provide targeted policies will hopefully be able to combat the wage gap more successfully than current polices in place.

It is suggested by Anderson, Binder, and Krause (2003) that there is an unobserved heterogeneity present between mothers and non-mothers. They also consider heterogeneity among mothers in their timing of return to the labor force. They find little evidence to support the work-effort explanation of the gap, since the medium skilled mothers, in their paper defined as high-school graduates, suffer a larger penalty when compared to both low- or high-skilled mothers.

This paper focuses on heterogeneity based on psychiatric problem indicators and self-rated health scores. Mental health is something that effects a myriad of aspects of an individual's life. Much research has been done to research the relation between employment and an individuals' well-being. The relationship between mental health and work has been found to be bi-directional, meaning not only does work effect your mental health but this relation also works the other way around (de Lange et al., 2005). This brings on the question of could differences in the mental health of individuals also be playing into the wage gap as discussed in this paper. As found by Jayakody and Stauffer (2000) mental health can cause a lack in self-confidence, which can lead a person to be hesitant to take on new challenges or find employment. Hence, mental health can heavily impact employment and earnings of women. This paper builds on previous research into heterogeneity in income differences related to the birth of children by exploring the effects of mental health and overall health effects on the child penalty.

### 2.3 Policy

A variety of papers have evaluated policy reforms surrounding, for instance, maternity and paternity leave. In these papers the authors examine the effects of children during or after birth on the income of the mothers and father and examine possible differences (Andresen & Nix, 2019; Budig, Misra, & Boeckmann, 2012). Several policies regarding parental leave around child birth have been put into place to try and create equal opportunity for the mother and father. When it comes to parental leave, generally countries have longer maternity leave options compared to paternity leave (Baird, Frino, & Williamson, 2009; Paull, 2008). However, the child penalty is still present in, for example, Scandinavian countries, where child leave can be split evenly among parents (Kleven, Landais, & Søgaard, 2019).

Another example being, providing longer parental leave to allow for better recovery and adjustment period after the birth of a child. However, this also has negative effects, because in some cultures this can be interpreted as the mother being expected to provide care for their children or being a reason for the low employment of women (Budig et al., 2012; Schönberg & Ludsteck, 2014). These cultural differences also make it challenging to perform analysis regarding child policies and come up with some form of universal policies. Another reason for this being the household division between mother and fathers (Cooke, 2006). Fortunately, there have also been instances of successful policies being put into place as an attempt to combat the wage gap. For child penalty specifically, it has been found that in, for instance, in Finland the child penalty can be reduced with the help of tax and social security systems in place (Sieppi & Pehkonen, 2019).

From these widely varying parental leave policies it can be seen that it is still very unclear what the best way is to promote gender equality. As in some countries they do appear to be effective while in other countries the effects are small or non-existent. This paper adds to existing literature by providing evidence of a child penalty present in Germany, and hopefully bring to light underlying variables playing into differences in child penalty for certain subgroups.

This paper builds upon other research done on the child penalty by researching the child penalty in Germany. More detail on the data used for this will be given in Section 3. Some of the reasons for more of a traditional family role division in Germany is found to be related to their tax, education, and occupational system in combination with the lack of adequate daycare for children Aisenbrey, Evertsson, and Grunow (2009). If this research does prove to find a relatively high child penalty for the German sample, this could be interesting for policy recommendation targeting a reform in Germany's systems regarding daycare, education, tax and occupation.

## 2.4 Methods

A popular method for estimating the gender wage gap is an estimation through regressions, more specifically ones including a gender dummy (Weichselbaumer & Winter-Ebmer, 2005). In this specification wages are estimated separately for individuals of the two binary gender groups. In their paper Weichselbaumer and Winter-Ebmer (2005) find that, overall worldwide the wage differentials have fallen substantially from about 65% in the 1906s to only 30% in the 1990s.

Another method which has been used to look into the child penalty is by simply interviewing men and women on the integration of parenthood and work-life (Baker, 2010). The interviews were transcribed and afterwards analyzed thematically. However, this method does not allow for clear quantitative interpretation of the variables. All they can analyze is the answers provided to the questions, such as women stating they are scared of getting fired after getting pregnant, or quitting their job to take care of the children. The downside of this being that there is no reliable statistical interpretation of the variables in this case.

A study performed on German data used propensity score matching to estimate the effect of children on the earnings on men and women (Feldhoff, 2021). This paper makes use of the same data set I use for this research. Feldhoff (2021) finds a child penalty similar to that found by Kleven, Landais, Posch, et al. (2019), namely 63%. In this paper matching is chosen as the preferred method, because it balances out the treatment and control group by their covariates. This analysis using a different methodology finds similar results as when using the event study methodology as proposed by Kleven, Landais, and Søgaard (2019).

Another form of regression which has been used to estimate the difference in hourly earnings between mothers and childless women is a fixed effects regression (Budig & England, 2001). In this study Budig and England (2001) find that the penalty for motherhood is around 7% per child. Hence, this paper finds that women who have more children suffer a higher penalty. Which is in line with the idea that children negatively effect mother's incomes. This study was done for American women, with fixed effects for years and persons.

O'Neill (1985) looks into the effect of increasing experience of women in the work force over the past four decades on the narrowing of the wage gap. However, this is only able to explain a part of the narrowing of the wage gap, and hence there do seem to be other factors at play. This research contributes to the existing research on the pay gap by focusing on the child penalty and how it relates to income differences between men and women.

This paper focuses on an event study, using the methods of Kleven, Landais, Posch, et al. (2019) which are replicated in this paper. That paper focuses primarily on aggregate effects as opposed to individual or subgroup effects. However, when evaluating the effects of an event like child birth estimating overall average causal effect is of restricted value. Especially when this research is to be used for policies it is of interest to know how the child penalty differs for different individuals to create a better targeted policies. This research extends the research of Kleven, Landais, Posch, et al. (2019) by looking into possible heterogeneous treatment effects and possible groupings of women that are affected more by the birth of their first child.

# 3 Data

Data from Germany is used to perform the event study and calculate the child penalty. The data base is produced by the Deutsches Institut für Wirtschaftsforschung (DIW). The German Socio-Economic Panel (SOEP) is a longitudinal survey of approximately 11,000 private households in the Federal Republic of Germany from 1984 to 2019 and the eastern German länder from 1990 to 2019. The data set contains information on household composition, employment, occupation, earnings, health, and satisfaction indicators. A detailed description of how this data was obtained can be found in Appendix A.

For the main research question, individuals who have their first child between the ages of 20 and 45 are considered. Observations between 5 years before and 10 years after the birth of the first child are used. As income reported by the individuals is the gross labor income of the last 12 months, I report the income of individuals in event year $t-1$. For this reason, the event time runs from $t = [-6, 9]$. For this research there are no restrictions on the relationship status of the parents. Because indexes and scores used in the Bayesian analysis are obtained for $t = -1$, all individuals who are not surveyed the year before the birth of their first child are excluded. To obtain more accurate results, all individuals with less than 8 observations from $t = -6$ to $t = 9$ are excluded as well. Furthermore, since the heterogeneous groups are based on characteristics observed before child birth, individuals with less than two observations before the birth of the first child are also excluded. This data selection process is similar to that of Kleven, Landais, Posch, et al. (2019).

This results in a final data set of 35,269 observations of 2,529 individuals, hence an average of a little under 14 years observed per individual. The most important variables are *earnings*, *gender*, *birth information*, *event time*, and *survey year*. Earnings are defined as the reported total gross

labor earnings in the past 12 months, and gender can take on two values. Birth information is the year of birth reported for each child, from which the event time is obtained as the difference between the survey years and the year of birth. Descriptive statistics and relevant sample sizes for this research can be found in Tables 1 and 2. From these tables is is observed that the sample is relatively balances between males and females. However, it is apparent that about twice as many men are fully employed compared to women.

Table 1

*Descriptive Statistics*

| Variable | Min | Max |
|---|---|---|
| Age(Females) | 14 | 53 |
| Age(Males) | 14 | 56 |
| Survey Year | 1984 | 2019 |

Table 2

*Sample Sizes*

| | Females | Males |
|---|---|---|
| Total | 18,968 | 16,301 |
| Employed | 13,957 | 15,128 |
| Full-Employment | 6,531 | 12,991 |

To extend the research performed by Kleven, Landais, Posch, et al. (2019) a dummy variable and a self rated health score at $t = -1$ is added. All the variables are created using existing variables on psychiatric problems and self-rated health scores included in the SOEP data set. The dummy variable indicates if an individual was experiencing psychiatric problems in the year before the birth of the first child. Hence, for a women with psychiatric problems the years before the birth of their first child the dummy is equal to 1 and for all other women the dummy is equal to 0. The self rated health score is a person's self rated health score at that same time. The self rated health score is a scale from 1 (very good) to 5 (bad). These are used to create mutually exclusive subgroups of the sample to estimate heterogeneous treatment effects.

# 4 Methods

## 4.1 Replication

The method used is an event study, based on the specification in the paper of Kleven, Landais, Posch, et al. (2019), with the event being the birth of the first child. Using the panel data as described in Section 3. Individuals included in this study are people who have children between the ages of 20 and 45, with data recorded 5 years before the birth of the first child and 10 years following the birth.

Since the event in this study is the birth of a child, this is denoted by $t = 0$, where $t$ ranges from -6 to 9. The following regression is run for both men and women, separately:

$$Y_{ist}^{g} = \sum_{j \neq -1} \alpha_j^g \cdot \mathbf{I}[j = t] + \sum_k \beta_k^g \cdot \mathbf{I}[k = age_{is}] + \sum_y \gamma_y^g \cdot \mathbf{I}[y = s] + \nu_{ist}^g \tag{1}$$

Where $Y_{ist}^g$ is the variable of interest for every individual $i$ of gender $g$ in year $s$ at event time $t$. The regression includes three dummies, time of the event, age and year respectively. The time at $t = -1$ is excluded to allow for measuring the impact of birth a child relative to the year just before the birth. The age dummy is included to control for life-cycle trends and the year dummy to control for time trends such as wage inflation and business cycles. Since men tend to be older when having their first child compared to women, the inclusion of the age dummy is important. The identification of all three dummies is possible conditional on age and year, there is variation in event time driven by variation in the age at which individuals have their first child.

We wish to interpret the results of the regression in percentages, however due to non-participation Equation (1) is specified in levels rather than logs. Hence the level effects are converted into percentages by calculating:

$$P_t^g \equiv \frac{\hat{\alpha}_t^g}{E[\tilde{Y}_{ist}^g | t]} \tag{2}$$

Where $\tilde{Y}_{ist}^g$ is the predicted outcome when the event dummy is left out, hence, $P_t^g$ captures the year-$t$ effect of children as a percentage of the counterfactual outcome absent children. Finally, the child penalty on women relative to men at event time $t$ is defined as:

$$P_t \equiv \frac{\hat{\alpha_t^m} - \hat{\alpha_t^w}}{E[\tilde{Y_{ist}^w} | t]} \tag{3}$$

Which measures the percentage by which women are falling behind men due to children at even time $t$.

## 4.2 Bayesian Analysis

To extend the research of Kleven, Landais, Posch, et al. (2019), this paper looks into possible heterogeneity of mothers and fathers and how certain variables affect the income of groups of mothers differently. Since the paper by Kleven, Landais, Posch, et al. (2019) only shows the results for males versus females it does not explore differences within the group of females. Because they only show general results for the complete female group versus the complete male group no

heterogeneity between females in this sample is explored. Hence, their analysis is unable to present differences between subgroups of mothers. In this extension, I create groups of mothers to reveal underlying factors playing into different child penalties. The focus is on using certain health and psychiatric related characteristics of the females in the year before the birth of their first child and see if there is heterogeneity in the treatment effect from child birth. Kleven, Landais, and Søgaard (2019) did look at the possible heterogeneity in results based on generational data, I will extend this by looking at other possible factors causing heterogeneity and possibly grouping together women, based on self-rated health scores and psychiatric indicators as described in Section 3. This research will also look into heterogeneity in the male sample, as a way to compare the magnitude of the effect. However, since the women suffer the child penalty they are the main focus also for policy recommendation's presented in Section 7.3

As stated in Section 2.4 looking into heterogeneous treatment effects can be very informative about underlying mechanism driving the child penalty and also in the development of more tailored policy reforms. To look into the heterogeneous treatment effect a Bayesian analysis is performed. A reason for working with Bayesian analysis is that Bayesian approaches to subgroup analyses includes all the subgroup-level treatment effects into one joint model, allowing inferences per subgroup to be driven by the full data as opposed to just the subgroup. This contributes to more precise estimates of the subgroups, since the subgroup analysis are able to borrow information from each other. Hence, it provides more accurate estimates than simply dividing up the sample and running separate event studies using the specifications in Section 4.1.

Another advantage of using Bayesian analysis is that this approach uses the distribution on the subgroups directly. Whereas, other approaches focus on finding heterogeneous treatment effects or focus on hypothesis testing. Other papers look into heterogeneous treatment effects in medical research, where often the analysis is based on knowledge of epidemiology or other medical knowledge (Kent, Steyerberg, & van Klaveren, 2018). Because this research does not concern medical data, a Bayesian model is appropriate as it can take into account underlying effects without needing extra medical knowledge.

Individuals are divided up into groups based on characteristics obtained the year before the birth of the first child to prevent the effects of the birth of the child interfering with the outcome. As stated in Section 3 these are related to psychiatric problems and self-rated health scores.

For the Bayesian analysis of heterogeneous treatment effect the R package `beanz` created by Wang, Louis, Henderson, Weiss, and Varadhan (2018) is used. For the first analysis two groups per gender are created, one with psychiatric problems before birth and one without. For the second analysis five groups are created based on the five levels of self-rated health scores. Furthermore, the analysis is also run on just the two gender groups. For the remainder of this section general notation is used. The disjoint subgroups are denoted by $g$. Where $\theta_g$ denotes the treatment effect in subgroup $G = g$, and $\hat{\theta}_g$ is the estimated $\theta$ in subgroup $G = g$, with $\hat{\sigma}_g^2$ associated variance estimate. Because we have a continuous response variable, namely wage, we can define the treatment effect as follows:

$$\theta_g = E(Y|Z = 1, G = g) - E(Y|Z = 0, G = g) \tag{4}$$

Where $Z = 1$ after the birth of the first child. As suggested by Jones, Ohlssen, Neuenschwander, Racine, and Branson (2011) we assume that the maximum likelihood estimation treatment effect $\hat{\theta}$ can be approximated by:

$$\hat{\theta}_g|\theta_g \sim \mathcal{N}(\theta_g, \sigma_g^2) \tag{5}$$

The subgroup treatment effect $\theta_g$ is defined as the effect $Z$ has on the earnings per subgroup $g$. Hence, a positive treatment effect means that having children positively impacts the earnings of the defined subgroup. $\hat{\theta}_g$ with its associated standard errors $s_g$ are computed by fitting a generalized linear model with a Gaussian distribution to each of the $G$ subgroups. With the use of this analysis the possible heterogeneous treatment effects of subgroups for both men and women are estimated. With these estimations possible policy recommendations are suggested to aim to overcome the wage gap.

# 5 Results

This section presents the results of this research. First, I discuss the results of the event study, and explore the effects of the first born child on the income difference between men and women. Then I discuss the results of the event study performed on several groups conditioned on employment status, number of children, and income level. Finally, I discuss the results obtained from the Bayesian heterogeneous treatment effect analysis.

## 5.1 Replication

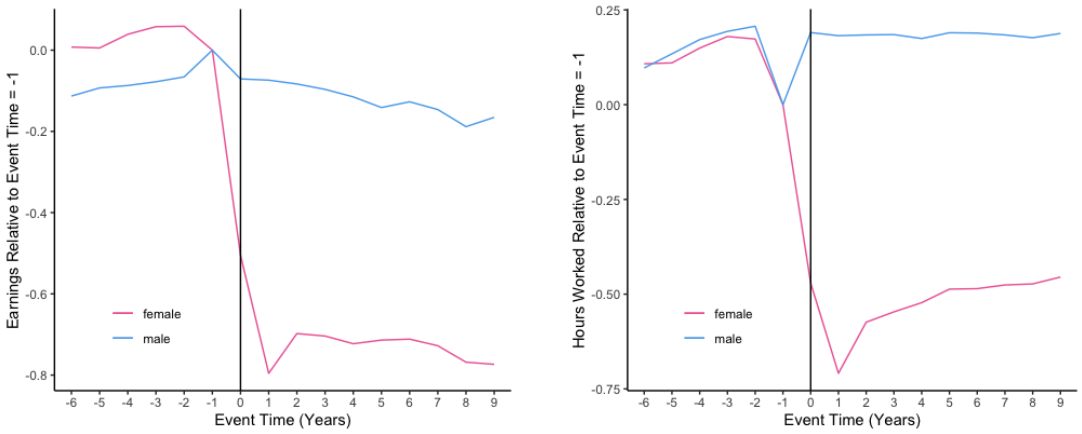
Figure 1 displays the earnings and hours worked relative to event time $t = -1$, in percentages as defined in Equation (2). Both plots show the gender specific impacts of the first child $P_t^m$ and $P_t^w$, these results are obtained using the specification as in Equation (2), with non-parametric controls for age and time trends.

In the years before the birth of the first child it is observed that the relative earnings of men and women evolve in roughly the same pattern. However, right after the birth of their first child women experience a large drop in their earnings. This initial drop is the largest at $t = 1$, where the difference is close to 80%. In the years following child birth this gap does close a little. However, as can be observed from this figure even years after child birth the child penalty is still present. For women around 9 years after child birth, their income has leveled out to around 77% less than their income the year before the birth of their first child. For the males in this sample their income has only dropped by around 17% in comparison to time $t = -1$.

A factor playing into labor earnings of individuals is the number of hours worked, hence, aside from total earnings, the effect on the number of hours worked by males and females in this sample is also looked at. This is shown in Figure 1 *(b)*, from this figure it can again be seen that before the birth of the first child that the number of hours worked by men and women evolve among the same trajectory. Right after child birth women experience a significant drop while men do not really seem to be impacted by the birth of their first child. Here, women's number of hours worked drops by almost 70% the first year after child birth. Whereas for men no negative change is experienced over the time period from $t = -6$ to $t = 9$ at all. 9 years after child birth the hours worked by females does differs a lot less relative to $t = -1$, relative to the long term effect on earnings, namely close to 45%. However, women do not show any long term recovery in this labor market variable either. For exact values of $P_t^m$ and $P_t^w$, for earnings and hours worked, at each time $t$ from -6 to 9, see Appendix B.

Overall, it is apparent that women suffer a significant drop in labor market outcomes whereas men seem to be relatively unaffected. This shows how continual the effects of children are in this sample. An interesting observation here is that the earnings gap seems to stay rather consistent over the years and does not show any narrowing pattern. Whereas the hours worked shows to be narrowing, which can be observed from the upward trajectory in Figure 1$(b)$. The combination of both of these factors suggests an hourly wage rate gap between mothers and fathers. Indicating

that women do in fact feel an urge to work more after child birth in terms of hours, but do not earn as much as men do for the same number of hours worked. Just like Kleven, Landais, Posch, et al. (2019) I find that in Germany a large part of the earnings penalty can be attributed to the wage rate effect.




*(a)* Earnings *(b)* Hours Worked

*Figure 1.* Impacts of Children on earnings*(a)* and hours worked*(b)*

The child penalty as defined in Equation 3 over the 6 years before birth till 9 years after can be found in Figure 2. The long run child penalty in Germany is found to be around 70%, which is very high when compared to other countries as found by Kleven, Landais, Posch, et al. (2019), where the child penalty ranges from 21% in Denmark, to 61% in Germany. Hence, the estimation in this paper is also slightly higher, namely 9%, than Kleven, Landais, Posch, et al. (2019) found in their paper 'Child Penalties: Across Countries'. This difference can potentially be caused by the slightly more recent data set used in this research and the somewhat different variable selection used.

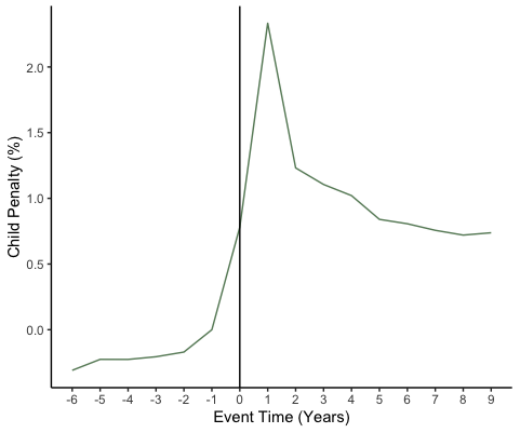


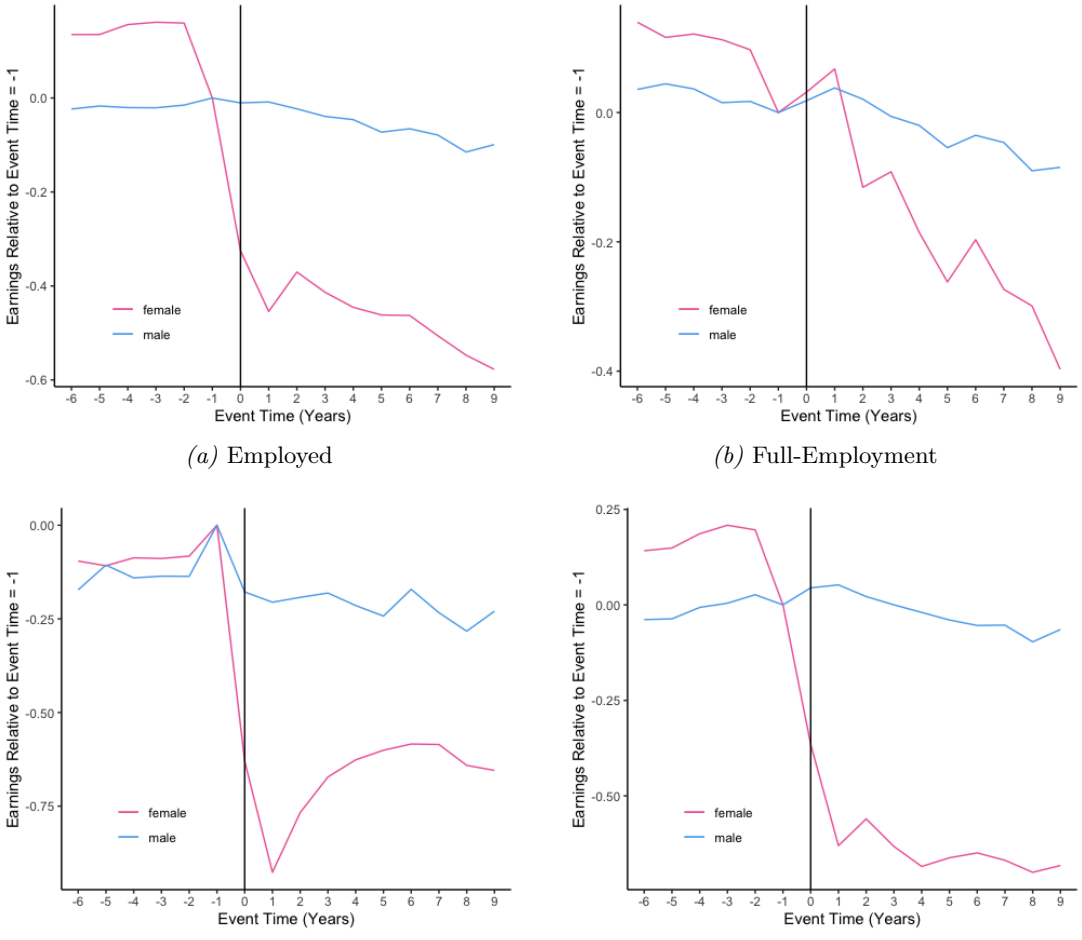

*Figure 2.* Child Penalty

### 5.1.1 Extension

To examine if certain variables effect the size of the child penalty, I create subgroups to see if there is a different impact on certain groups conditional on a variety of variables. All of these conditional subgroups analysis will be performed using the specification of Kleven, Landais, and Søgaard (2019). One of the groups explored is conditioning on employment status to check if women who are employed all throughout the survey years suffer a lower child penalty. The child penalty estimate, as defined by equation (3), in the employed group is found to be around 50% which is much smaller than the one found for the full sample. Aside from employment status the event study is also run conditional on the fact that individuals are working full time, which results in an even lower child penalty estimation of around 25%. Showing that preferences for working to work or even work full time possibly play a role in the magnitude of the child penalty. These results are displayed in Figures 3 $(a)$ and $(b)$. From this figure it is apparent that relative to their earnings at $t = -1$ women earn more before and this starts to decline quite rapidly after. Whereas for men their relative incomes do not seem effected much at all, especially in families with only one child, which is a variable explored in the next paragraph.

Another interesting variable to condition on is the number of children in a household, to see if the effects differ for families with one child or two children. The results of conditioning on the number of children are displayed in Figures 3 $(c)$ and $(d)$. From this figure, once again, a similar pattern as before is observed, where relative income of women drops steadily after the birth wheres men are relatively unaffected. When it comes to number of children in a family it is interesting if certain families anticipate having more children and possibly divide up the gender roles in anticipation,

which could result in a much higher child penalty in the families with more children. For families who have one child in total the long term child penalty is found to be around 34%, which is close to half the observed child penalty of the full sample. Whereas this is over 88% for mothers of families who end up having two children. Hence it is quite apparent that having more children negatively effects women in a harsher way. Giving reason to believe that families anticipate having more children, then have a more traditional gender role division. However, this could also be related to the fact that in the two children household the second child is still younger and hence requires more care from the mother. Ideally, this research would examine the effects over a longer period of time. However, with this current size of the data set this would lead to a too small sample size.




*(a)* Employed

*(b)* Full-Employment

*(c)* One Child

*(d)* Two Children

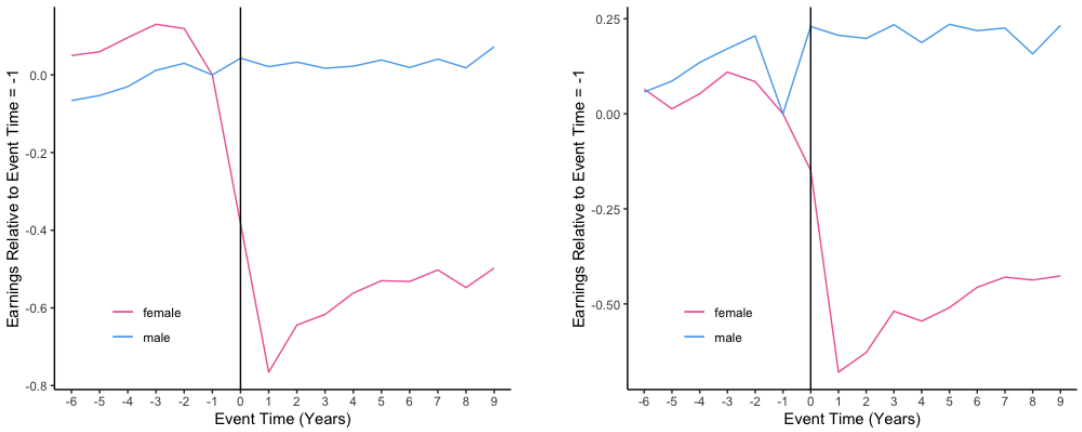
*Figure 3.* Impacts of Children on Individuals Conditional on Employment *(a)*, Full Employment *(b)*, having One Child *(c)*, and Two Children *(d)*

In literature it has been suggested that income effects caused by children is a more prevalent issue for the upper class (Anderson et al., 2002). To investigate this idea the top 25% and 50% earning males and females are excluded from the regression. This is done by calculating the top first and second quantile of the income distributions for males and females separately and then dropping all observations per gender that are above these income levels. Then the $P_9^g$ is calculated, these results are presented in Table 3, and the $P_t^g$ for all event times are presented in Figure 4.

Table 3

*$P_9^g$ with top earning individuals excluded*

| | $P_9^w$ | $P_9^m$ | $n^w$ | $n^m$ |
|---|---|---|---|---|
| Full data set | -77 | -16 | 18,968 | 16,301 |
| Top 25% Excluded | -50 | 0.07 | 14,226 | 12,223 |
| Top 50% Excluded | -43 | 23 | 9,463 | 8,147 |




*(a)* Top 25% Excluded

*(b)* Top 50% Excluded

*Figure 4.* Impacts of Children on Individuals when the top earning individuals are excluded (25%*(a)*, 50%*(b)*)

When excluding the top 25% female earnings the $P_9^w$ drops from almost -80% to -50%, showing that the drop in income is indeed more severe for the top earning mothers. When excluding the top 50% female earners $P_9^w$ drops even more to be 43%, but this drop is significantly less severe compared to going for the full sample to excluding the top quantile. For the males in this sample a positive change in $P_9^m$ is also observed, when excluding the top 25% and 50% earning males. All of the estimates presented here are with respect to the counterfactual income at $t = -1$.

The exact numerical results of each conditional group for every event time can be found in Appendix B.1.

## 5.2 Bayesian Analysis

A variety of subgroups were formed to run the Bayesian heterogeneous treatment effect analysis in an attempt to find out what subgroups' income levels are hurt the most by the birth of a first child. Something which is very apparent even from first glance at the results presented in Table 4, is that for all subgroups it can be observed that regardless of the other variables defining the subgroup, the treatment effect on the female subgroup is almost always negative with the exception of the psychiatric problem indicator.

Table 4

*Subgroup Results*

| Female | | | | Male | | | |
|---|---|---|---|---|---|---|---|
| ppi | srhs | Average Effect | $n$ | ppi | srhs | Average Effect | $n$ |
| - | - | -5280.439 | 18,968 | - | - | 13738.951 | 16,301 |
| 0 | - | -6173.400 | 1,043 | 0 | - | 16765.420 | 962 |
| 1 | - | 2157.628 | 72 | 1 | - | 5406.050 | 73 |
| - | 1 | -5626.773 | 2,034 | - | 1 | 17378.777 | 1,562 |
| - | 2 | -6072.458 | 6,565 | - | 2 | 13964.781 | 5,465 |
| - | 3 | -3201.480 | 2,189 | - | 3 | 12575.396 | 1,999 |
| - | 4 | -3509.885 | 517 | - | 4 | 8277.872 | 400 |
| - | 5 | -6034.917 | 42 | - | 5 | 7907.002 | 106 |

[1] ppi and srhs refer to the psychiatric problem indicator and self-rated health score respectively.

From Table 4 it can be seen that for the male sample the effects of the psychiatric indicator at $t = -1$ seems to result in a much lower effect on their estimated income after birth. Which could be an indication of psychiatric problems negatively effecting their income, since we observe a drop of around -65%. However, when looking at the females this effects seems to surprisingly be positive and even higher, namely around 85%. Unfortunately, the sample size of people with a psychiatric problems indicator is rather small, which could be a possible explanation for the surprising results obtained. In both the male and female sample only around 70 individuals were observed with

psychiatric problems the year before child birth whereas the sample of people without is around a thousand for each gender. Due to this small sample size it is unfortunately rather hard to draw any conclusions from these obtained results.

For the self-rate health score we do however obtain a way more leveled out sample for each of the scores per gender. Here only self-rate health score of 5 has a relatively low number of observations and the other scores each have between 400-7500 observations. For males, for all the self-rate health score there is a clear decrease in the effect in relation to a lower self-rated health score. Which is displayed in Figure 5. For women on the other hand this effect does not follow such a pattern. Here, it is observed that the estimations first increase and later on decrease again. For this reason it is once again tough to draw any clear conclusions from the obtained results for women.

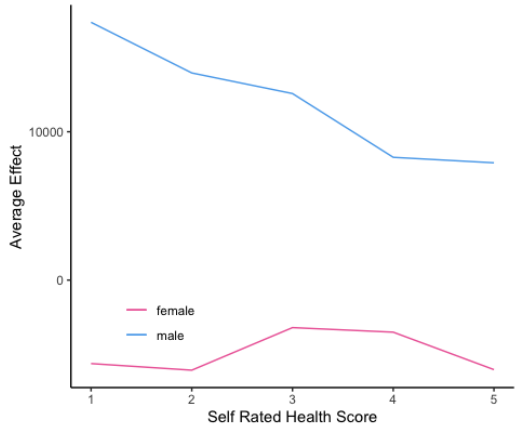
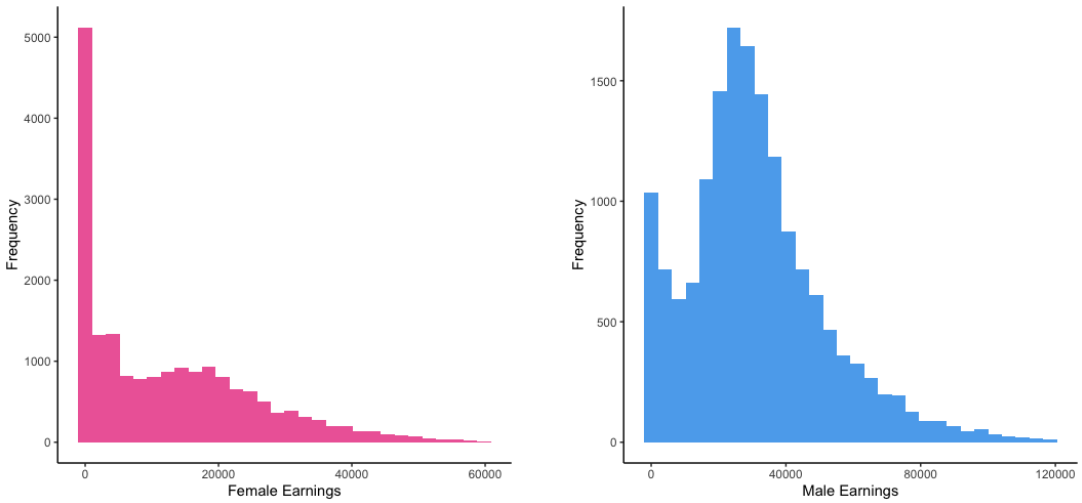



*Figure 5.* Self Rated Health Score Subgroup Effects

Even when comparing the female subgroups to the full female sample group it is tough to draw conclusions about the relative difference. It is especially interesting to observe that women with the psychiatric indicator have a 140% higher effect, which is a very high difference and also in the opposite direction than expected. Furthermore, the variance for all the estimates are also over 10 times larger than the average effects estimates. Making the reliability of the results also questionable.

Overall, it has been proven to be tough to draw conclusions based on the female estimates obtain from this research. A possible explanation for this could be related to the traditional gender roles where the male is the traditionally seen as the breadwinner of the family. Since men tend to be the ones with a more stable income it is easier to estimate effects of outside variables on their income. From Figure 6 you can see that the male earnings distribution appears to be more

normally distributed, whereas the female sample is strongly left skewed. This confirms the idea that it is hard to run similar analysis on earnings of males and females since the inherit structure of the income distribution for both genders is so different. Furthermore, this stability allows for a smaller variance and hence more accurate estimations. Unfortunately, as found in Section 5.1 women still suffer a significant child penalty. This does leave the question of how accurate these finding are when including all women unconditional on work status. This will be further discussed in the discussion.




*(a)* Female *(b)* Male

*Figure 6.* Earnings Distributions

# 6 Discussion

This section discusses the results obtained in this research. Firstly, the obtained results of the replication of the paper by Kleven, Landais, Posch, et al. (2019) are discussed. Then the extension of that replication conditional on employment status, number of children and income level are discussed. Finally, the results of the Bayesian analysis are discussed. Which is followed by a discussion of the limitations of this research.

## 6.1 Main Findings

### 6.1.1 Replication

Overall, it is clear that also in Germany the long run child penalty is still largely present. Figure 1 shows that the divide between mothers and fathers is not present before child birth but diverges quickly after the birth of the first child. This divide is present for both earnings and hours worked. The fact that the gap in hours worked does seems to be converging, as the hours worked by mothers after the birth of their first child shows an upward trend. While overall labor earnings does not seem to show such a trend, gives reason to believe that women suffer a wage rate gap after the birth of the first child. This supports the notion of a gender wage rate gap, which has been found by many other researchers (Antonczyk et al., 2010; Waldfogel, 1997). As stated in the introduction for many years researchers have been attempting to find a reason for the pay gap, this research shows that a large portion of this can be attributed to the effects that children have on labor market outcomes.

The data set used for this research includes data up to 2019. Hence, even with very recent data, the child penalty is still present for women. Confirming the idea presented by Kunze (2018) of a wage gap still being present, which can largely not be attributed to things such as education level. Furthermore, when comparing to other countries, Germany suffers a significantly higher child penalty. In fact it is found that the child penalty in Germany is over three times higher than the one found in Denmark by Kleven, Landais, and Søgaard (2019).

Aside from purely focusing on earnings or hours worked subgroups were created conditioning on variables such as employment status and number of children. From this it was observed that conditioned on full time employment reduces the child penalty to about 25%. Suggesting that women with preferences for working full time even with the birth of a child suffer less of a child penalty. Confirming the idea that the child penalty is partially caused by the fact that women have different preferences when it comes to employment and being a mothers. On top of which, it is also shown that conditional on employment in general women also suffer a lower child penalty. Further, confirming the idea that labor preferences play a role in the gender wage divide surrounding child birth.

As stated in the literature review it has been found in the past that Germany was characterized as being a male-breadwinner/ female- homemaker state, which would explain the highly child penalty found in this research. This traditional gender role division in the presence of children

is confirmed by the event study conditional on the number of children, where it is found that an increase in the number of children in a household is related to a higher child penalty.

Regarding the top earning individuals it is hard to identify the exact timing of each child and pinpoint the exact effect they have on earnings. Because the event study specification uses identification at $t = -1$, denoting the birth of the first child the exact effect of the second child is tough to pinpoint per mother. Another issue regarding the timing of children is the timing of maternity leave of mothers. Unfortunately, this data set does not show when exactly women go on maternity leave and return to the workforce. However, since women should get paid when on maternity leave it is hard to estimate these effects in general.

### 6.1.2 Bayesian Analysis

Overall, it is hard to draw a clear conclusion from the results obtained from the Bayesian analysis. However, the fact that the results are unclear for women specifically does confirm the idea that women's income has not progressed at the same level as men's income has over the years. Because the estimates obtained for the male sample show a much clearer pattern, as depicted in Figure 5 shows that historically men's income has stabilized to a level allowing for good research and estimation. The fact that this analysis is much harder for women goes to show that women's income has not caught up with the growth of that of their male peers. This is also depicted in Figure 6 where a clearly left skew distribution is seen for women whereas for men a seemingly more normally distributed earnings pattern is visible. Hence, just as Kunze (2018) found in his research a large wage gap remains present in current days. With the results found in this paper also the increase in earnings inequality found by Antonczyk et al. (2010) is not surprising.

The fact that overall the female income distribution shows a large spike at zero shows that overall women have a lower workforce participation rate. There are several possible explanations for this. Firstly, biologically there is a tendency for women to feel more of an urge to stay home with their kids, which is why there are relatively more women without a job than men (Andresen & Nix, 2019). Secondly, gender roles pressed on women by society but also by companies discriminating against mothers who work can cause them to work less, or decide to not work at all (Budig & England, 2001; Correll et al., 2007). Finally, since this research has confirmed the existence of a child penalty, when parents decide to have one person be a stay at home parent it makes sense to have the person with a higher wage to remain employment. This in combination with the hourly wage gap that has been found gives reason to have the woman work less or not at all, since the

father is able to make more money. A big problem that arises with this conclusion is that when women earn less and decide to also work less because of this, this also does not allow the wage gap to close since women are not encouraged to work and show off how successful they can be at this. Hence, giving in to this downwards spiral for mothers in the workforce.

### 6.2 Limitations

One of the main limitation of this research relates to the data set, and the observations present per variable of interest. Especially for mental health related variables after the data preparation needed for the event study little observations without missing values were left. Due to the low number of observations for the psychiatric problems indicator it is hard to draw conclusions about these results, and use these in policy recommendations.

Another limitation relates to the variables used for the extension, specifically the self-rated health score used. The reliance on self reported health instead of measures of health introduces a bias to the results. Since individuals have a varying perceived baseline of general health, one person's good health score could be a bad health score for someone else. Hence, these scores always need to be treated with caution. This could be overcome in future research by making use of actual observed health scores as opposed to self reported ones.

A variable that has been explored a lot in previous research related to an income inequality has been race. An intention for this research was to also explore possible heterogeneous effects based on the ethnic background of individuals in this sample. The SOEP data set comes with a race variable, however, after the data selection process it was found that this variable only contained missing values for each individual. For this reason, unfortunately, this subgroup analysis could not be explored any further.

## 7 Conclusion

This section concludes the research by answering the main research question and the two sub-questions, with the use of the obtained results. Finally, I present suggestions for further research based on the limitations of this research. This is followed by policy recommendations based on previous findings and the results from this research.

## 7.1 Child Penalty and Wage Gap

In this study I examine the presence of the child penalty as defined in Equation (3) in Germany. With the aim to answer the following main research question:

> *What is the impact of children on the labor market trajectories of women relative to men?*

Which can be answered with the results obtained from the event study performed. I find that in the long run women in Germany suffer a child penalty in earnings of around 70%. This child penalty in earnings also does not show a converging pattern but rather seems to level out at this level. For hours worked however, there is a converging pattern apparent in the long. This leads to the conclusion that women suffer a wage gap which can be attributed to the effects of having children. Hence, the impact of children on labor market trajectory is mainly present in the form of a wage rate gap between men and women.

Aside from the performed event study to examine the child penalty this research also aimed to find heterogeneous treatment effects. This was explored with the goal of answering the following sub research questions:

> *Are there certain factors leading to a heterogeneous treatment effect?*
> *And how can this information be used to develop policies to decrease the wage gap?*

These prove to be a little tougher to answer, as there is no one clear answer from the analysis. However, a factor that does seem to lead to the difficulty of estimating the treatment effects for women is related to the way their incomes are hit, and the fact that their income is a lot less stable than males.

The biggest factor playing into the difficulty of estimating heterogeneous treatment effects for women seems to be related to the underlying structure of they income. Historically speaking men had more of chance to obtain education earlier, work more, and also earn more. This had lead to them having a more stable and developed income level. This research has shown once again that women have not been able to catch up to their male peers. This also leads to inconclusive results and no clear conclusion on if there are certain factors leading to a heterogeneous treatment effect for this women in this sample.

The fact that women still experience such inequality in the work force, does allow for possible policy recommendations. Not only related to child penalty but also just labor force participation and labor force outcomes in general. These are presented in Section 7.3.

## 7.2 Suggestions for Further Research

Since in recent studies it has been shown that mental health can have a large effect on many things effecting labor market outcomes it is of huge interest to policy makers and government employees in general to be aware of these effects. This in combination with still a large part of the wage gap remaining unexplained leads to demand for more research into this. A suggestion for further research in this mainly pertains to the fact that not a lot of questionnaires such as the SOEP include variables relating to mental health effect, hence building models including such variables is challenging. The SOEP data is collected by the German Institute for Economic Research (DIW Berlin) and the Kantar Group. A suggestion for them would be to focus on collecting more variables related to mental health effects. This would allow for better research into the effects of these variables on not only income inequality but also many other types of studies.

Another interesting addition to this research is to perform a meta-analysis where your compare a wide variety of countries. Specifically, spanning across continents to see if there are differences which can be observed across not only country but continental borders. This is of interest when exploring possible cultural differences playing into the child penalty. Furthermore, since in certain countries it is shown that the child penalty is a lot lower this could help other countries learn from those and possibly help with policy reforms.

## 7.3 Policy Recommendations

The high child penalty found for Germany in this research in addition to many other papers confirms the idea presented in Section 2.3 stating that Germany still has a traditional family role tradition. Possible reasons for this are their tax, education, and occupational system. Therefore, to combat the effects that the traditional family role has on the child penalty in Germany as found in this research would target these systems.

Section 2.3 also discussed the issue that in various other countries policies targeted at lowering the child penalty were unsuccessful so far. However, in Finland with a reform in the tax and social security system they were able to successfully reduce the child penalty (Sieppi & Pehkonen, 2019). Examples of these systems are tax-free child benefits and child-care subsidies. Implementing such policies in Germany could also help reduce the child penalty there. Having a higher work force participation rate will also be beneficial in tax returns for the government, hence could make these reforms well worth their investment.

Another thing found in this research is related to the historic pattern of labor earnings for men compared to women. Over the years women are slowly given the same opportunities as men with regards to education level and employment. However, this research has shown that women's income, specifically that of mothers has not caught up to the level of the fathers. To combat this issue the German government could form policies to encourage women to work after giving birth to a child. For instance, by lowering the cost of child care, or provide free child care for a couple of days a week. When child care becomes more affordable women will feel less pressured to stay at home with the kids, hence allowing them to work more and gain more work experience.

## References


Acker, J. (2009). From glass ceiling to inequality regimes. *Sociologie du travail*, *51*(2), 199–217.

Aisenbrey, S., Evertsson, M., & Grunow, D. (2009). Is there a career penalty for mothers' time out? a comparison of germany, sweden and the united states. *Social Forces*, *88*(2), 573–605. Retrieved from http://www.jstor.org/stable/40645817

Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review*, *100*(4), 1847–59.

Anderson, D. J., Binder, M., & Krause, K. (2002). The motherhood wage penalty: Which mothers pay it and why? *American economic review*, *92*(2), 354–358.

Anderson, D. J., Binder, M., & Krause, K. (2003). The motherhood wage penalty revisited: Experience, heterogeneity, work effort, and work-schedule flexibility. *ILR Review*, *56*(2), 273–294.

Andresen, M. E., & Nix, E. (2019). *What causes the child penalty? evidence from same sex couples and policy reforms* (Tech. Rep.). Discussion Papers.

Antonczyk, D., Fitzenberger, B., & Sommerfeld, K. (2010). Rising wage inequality, the decline of collective bargaining, and the gender wage gap. *Labour economics*, *17*(5), 835–847.

Baird, M., Frino, B., & Williamson, S. (2009). Paid maternity and paternity leave and the emergence of'equality bargaining'in australia: an analysis of enterprise agreements, 2003-2007. *Australian Bulletin of Labour*, *35*(4), 671–691.

Baker, M. (2010). Motherhood, employment and the "child penalty". *Women's Studies International Forum*, *33*(3), 215–224. Retrieved from https://www.sciencedirect.com/science/article/pii/S0277539510000063 doi: https://doi.org/10.1016/j.wsif.2010.01.004

Blau, F. D., & Beller, A. H. (1988). Trends in earnings differentials by gender, 1971–1981. *ILR Review*, *41*(4), 513–529.

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American sociological review*, 204–225.

Budig, M. J., Misra, J., & Boeckmann, I. (2012). The motherhood penalty in cross-national perspective: The importance of work–family policies and cultural attitudes. *Social Politics*, *19*(2), 163–193.

Cohen, P. N., & Huffman, M. L. (2007). Working for the woman? female managers and the gender wage gap. *American sociological review*, *72*(5), 681–704.

Cooke, L. P. (2006). Policy, preferences, and patriarchy: the division of domestic labor in east germany, west germany, and the united states. *Social Politics: International Studies in Gender, State & Society*, *13*(1), 117–143.

Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American journal of sociology*, *112*(5), 1297–1338.

Cotter, D. A., Hermsen, J. M., Ovadia, S., & Vanneman, R. (2001). The glass ceiling effect. *Social forces*, *80*(2), 655–681.

de Lange, A. H., Taris, T. W., Kompier, M. A., Houtman, I. L., & Bongers, P. M. (2005). Different mechanisms to explain the reversed effects of mental health on work characteristics. *Scandinavian journal of work, environment & health*, 3–14.

de Quinto, A., Hospido, L., & Sanz, C. (2020). The child penalty in spain. *Banco de Espana Occasional Paper*(2017).

Feldhoff, C. H. (2021). *The child penalty: Implications of parenthood on labour market outcomes for men and women in germany* (Tech. Rep.). SOEPpapers on Multidisciplinary Panel Data Research.

Gustafsson, S. S., Wetzels, C. M., Vlasblom, J. D., & Dex, S. (1996). Women's labor force transitions in connection with childbirth: A panel data comparison between germany, sweden and great britain. *Journal of population economics*, *9*(3), 223–246.

Jayakody, R., & Stauffer, D. (2000). Mental health problems among single mothers: Implications for work and welfare reform. *Journal of Social Issues*, *56*(4), 617–634.

Jones, H. E., Ohlssen, D. I., Neuenschwander, B., Racine, A., & Branson, M. (2011). Bayesian models for subgroup analysis in clinical trials. *Clinical Trials*, *8*(2), 129–143.

Kent, D. M., Steyerberg, E., & van Klaveren, D. (2018). Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *Bmj*, *363*.

Klerman, J. A., & Leibowitz, A. (1990). Child care and women's return to work after childbirth. *The American Economic Review*, *80*(2), 284–288.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child penalties across countries: Evidence and explanations. In *Aea papers and proceedings* (Vol. 109, pp. 122–26).

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209.

Kunze, A. (2005). The evolution of the gender wage gap. *Labour Economics*, *12*(1), 73–97.

Kunze, A. (2018). The gender wage gap in developed countries. *The Oxford Handbook of Women*

*and the Economy*, 369.

Luxton, M. (2006). Feminist political economy in canada and the politics of social reproduction. *Social reproduction: Feminist political economy challenges neo-liberalism*, 11–44.

O'Neill, J. (1985). The trend in the male-female wage gap in the united states. *Journal of Labor Economics*, *3*(1, Part 2), S91–S116.

O'Neill, J. (2003). The gender gap in wages, circa 2000. *American Economic Review*, *93*(2), 309–314.

O'Neill, J., & Polachek, S. (1993). Why the gender gap in wages narrowed in the 1980s. *Journal of Labor economics*, *11*(1, Part 1), 205–228.

Oostendorp, R. (2004). *Globalization and the gender wage gap*. The World Bank.

Paull, G. (2008). Children and women's hours of work. *The Economic Journal*, *118*(526), F8–F27.

Petersen, T., & Morgan, L. A. (1995). Separate and unequal: Occupation-establishment sex segregation and the gender wage gap. *American Journal of Sociology*, *101*(2), 329–365.

Pugh, A. J. (2005). Selling compromise: Toys, motherhood, and the cultural deal. *Gender & Society*, *19*(6), 729–749.

Sainsbury, D., et al. (1999). *Gender and welfare state regimes*. Oxford University Press.

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

Sieppi, A., & Pehkonen, J. (2019). Parenthood and gender inequality: Population-based evidence on the child penalty in finland. *Economics letters*, *182*, 5–9.

Viitanen, T. (2014). The motherhood wage gap in the uk over the life cycle. *Review of Economics of the Household*, *12*(2), 259–276.

Waldfogel, J. (1997). The effect of children on women's wages. *American sociological review*, 209–217.

Wang, C., Louis, T. A., Henderson, N. C., Weiss, C. O., & Varadhan, R. (2018). beanz: An R package for Bayesian analysis of heterogeneous treatment effects with a graphical user interface. *Journal of Statistical Software*, *85*(7), 1–31. doi: 10.18637/jss.v085.i07

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys*, *19*(3), 479–511.

# Appendices

## A Data Retrieval

### A.1 Source

The data is obtained from the German Socio-Economic Panel (SOEP), which provides rich information on labor supply dynamics and allows to observe the date of birth of children to adult members of the sample. It was conducted between 1984 and 2016 in West Germany and between 1990 and 2016 in East Germany. The data is obtained from the SOEP website [1], using the contract from the supervisor of this paper, Dr. García-Gómez. The annual surveys are conducted by the Kantar Group and the German Institute for Economic Research (DIW Berlin).

### A.2 Variable Definitions

`Earnings` When surveyed, individuals are asked to report the total of their gross labor income in the last 12 months. If an individual is surveyed in event year $t$, we therefore define their reported gross labor income in the survey as their earnings for event year $t-1$.

`Employment` Employment status of the parent is defined based on wages and hours worked. More precisely, an individual is considered employed if he had positive wages and worked at least 52 hours in the previous year.

`Birth Information` Year of birth is reported for each child. Event time is defined as the difference between the survey year and the year of birth.

[1] https://www.diw.de/en/diw_01.c.678568.en/research_data_center_soep.html accessed on May $7^{th}$ 2021

# B Replication Results

Table 5

*Full Replication Results*

| | | Earnings | | Hours Worked | |
|---|---|---|---|---|---|
| Time | Child Penalty | $P_t^w$ | $P_t^m$ | $P_t^w$ | $P_t^m$ |
| -6 | -0.3091371 | 0.007574518 | -0.11305221 | 0.1079487 | 0.09691493 |
| -5 | -0.2257177 | 0.005364342 | -0.09296282 | 0.1102534 | 0.13413777 |
| -4 | -0.2261603 | 0.039054134 | -0.08700263 | 0.1493078 | 0.17170352 |
| -3 | -0.2057148 | 0.057819176 | -0.07782463 | 0.1798536 | 0.19346560 |
| -2 | -0.1700269 | 0.058870601 | -0.06591366 | 0.1732371 | 0.20695688 |
| -1 | 0.0000000 | 0.000000000 | 0.00000000 | 0.0000000 | 0.00000000 |
| 0 | 0.7802476 | -0.504958935 | -0.07079316 | -0.4711986 | 0.19025979 |
| 1 | 2.3329070 | -0.795622625 | -0.07387867 | -0.7089121 | 0.18204793 |
| 2 | 1.2317170 | -0.697766643 | -0.08322759 | -0.5739703 | 0.18407266 |
| 3 | 1.1056180 | -0.703812722 | -0.09651581 | -0.5467108 | 0.18487910 |
| 4 | 1.0208615 | -0.722780704 | -0.11501486 | -0.5224701 | 0.17449510 |
| 5 | 0.8400113 | -0.713805719 | -0.14130673 | -0.4867436 | 0.18980275 |
| 6 | 0.8065353 | -0.711486831 | -0.12712206 | -0.4854743 | 0.18892131 |
| 7 | 0.7566236 | -0.727713966 | -0.14625778 | -0.4757881 | 0.18405003 |
| 8 | 0.7202819 | -0.768198931 | -0.18835681 | -0.4729716 | 0.17666851 |
| 9 | 0.7380874 | -0.773478195 | -0.16551612 | -0.4549341 | 0.18788253 |

## B.1 Conditionals

Table 6

*Results Conditional on Variables*

| | Employed | Full-Employment | Total Kids = 1 | Total Kids = 2 |
|---|---|---|---|---|
| Time | | Child Penalty | | |
| -6 | -0.2399193 | -0.12502887 | -0.24930025 | -0.3636511 |
| -5 | -0.2139342 | -0.07447149 | -0.07282205 | -0.3204793 |
| -4 | -0.2275815 | -0.08821543 | -0.15541998 | -0.2641588 |
| -3 | -0.2179095 | -0.10614009 | -0.13246381 | -0.2389609 |
| -2 | -0.1929961 | -0.08042492 | -0.12821432 | -0.1729091 |
| -1 | 0.0000000 | 0.00000000 | 0.0000000 | 0.0000000 |
| 0 | 0.520580 | -0.00677262 | 0.68078031 | 0.8667094 |
| 1 | 0.7578614 | NA | 2.01244140 | 2.3796088 |
| 2 | 0.4635439 | 0.13573648 | 0.89056635 | 1.3432948 |
| 3 | 0.5008404 | 0.07786010 | 0.59206592 | 1.5175112 |
| 4 | 0.5080456 | 0.14610859 | 0.39811444 | 1.5525693 |
| 5 | 0.4617129 | 0.17645895 | 0.28824713 | 1.2176722 |
| 6 | 0.4532360 | 0.12390102 | 0.36816755 | 1.0273174 |
| 7 | 0.4815966 | 0.17973888 | 0.25890857 | 1.0153026 |
| 8 | 0.4700198 | 0.14454823 | 0.24613839 | 0.9275901 |
| 9 | 0.5272351 | 0.23963304 | 0.34529449 | 0.8858335 |

Table 7

*Top Earning Individuals Excluded*

| | Top 25% Excluded | | Top 50% Excluded | |
|---|---|---|---|---|
| Time | $P_t^w$ | $P_t^m$ | $P_t^w$ | $P_t^m$ |
| -6 | 0.04986348 | -0.06612929 | 0.06505802 | 0.05731211 |
| -5 | 0.05961352 | -0.05305193 | 0.01317463 | 0.08615689 |
| -4 | 0.09607474 | -0.03017675 | 0.05281755 | 0.13512126 |
| -3 | 0.13036763 | 0.01169640 | 0.10982944 | 0.17122949 |
| -2 | 0.11946708 | 0.03001845 | 0.08468561 | 0.20510330 |
| -1 | 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 |
| 0 | -0.38222786 | 0.04306766 | -0.14865590 | 0.22947025 |
| 1 | -0.76552813 | 0.02131251 | -0.67952078 | 0.20652861 |
| 2 | -0.64418815 | 0.03298933 | -0.62774830 | 0.19822871 |
| 3 | -0.61675171 | 0.01727208 | -0.51918515 | 0.23443863 |
| 4 | -0.56183785 | 0.02245466 | -0.54523782 | 0.18731655 |
| 5 | -0.52991124 | 0.03846495 | -0.50974843 | 0.23544539 |
| 6 | -0.53205686 | 0.01922361 | -0.45633323 | 0.21866744 |
| 7 | -0.50224156 | 0.04081230 | -0.42960265 | 0.22558656 |
| 8 | -0.54755986 | 0.01836255 | -0.43701944 | 0.15758486 |
| 9 | -0.49750826 | 0.07257244 | -0.42642980 | 0.23258359 |

# C Code Explanation

## C.1 Sample Selection

`sampleselection_rep.R, sampleselection_extension.R`

Here, a matrix is created which stores all the unique person ID's and each observed event time. From this matrix the data selection is made. In the first selection all individuals with missing event time at $t = -1$ are excluded, then everyone with less than 8 observations, and finally everyone with less than 2 observations before $t = -1$. At the end of this file new survey year, age, and event time variables are created to move the earnings of the individual's to the right year, since income is reported for $t - 1$.

## C.2 Event Study

`eventstudy_rep.R, eventstudy_hoursworked.R, eventstudy_conditional.R`

The event study files, performs the event study for the replication with earnings and hours worked, and on the conditionals presented in Section 5.1.1. The sample is split into a males and females group, after which the full regression using factors is run on the variable of interest. Then in a loop the reduced regressions are run to obtain the estimations needed for $\tilde{Y}^g_{ist}$. At the end of each files vectors are create to create the plots presented in the results section.

## C.3 Extension

`extension.R`

In this file first an empty censor variable is created since `beanz` requires such a variable. Then the Bayesian analysis is run on each of the subgroups of interest. At the end of this file earnings distribution plots are create to visualize the income of both genders where the top 1% earning individuals are exclude for each gender to obtain a more insightful plot.