ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Child Penalties in the United States: A difference-in-differences extension[1]

Bachelor Thesis

July 4, 2021

Author
Yoëlle Kilsdonk

Supervisor
Dr. P. García-Gómez

Second assessor
Dr. M. Zhelonkin

Abstract

This paper reproduces the child penalty estimates of Kleven et al. (2019a) in female and male earnings in the United States. The estimates are based on event studies around the birth of the first child. I find that women's earnings sharply diverge after birth, while male earnings seem unaffected. Moreover, I find a long-run child penalty in earnings of 36%. Furthermore, as a robustness check, I provide a difference-in-differences event study that uses childless individuals as controls over multiple treatment periods combined with Coarsened Exact Matching. The analysis confirms the key qualitative findings from the event study for women. However, it alludes to new insights in which fathers might be adversely affected by childbirth, contradicting the findings of Kleven et al. (2019a).

[1]The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Table of contents

1 Introduction

Many studies show that women's earnings are adversely affected by motherhood for at least ten years after the birth of the first child, while men's labor earnings seem to be virtually unaffected by parenthood. This gap between men and women is generally referred to as the 'child penalty'. The child penalty is researched by, amongst others, Kleven et al. (2019a), who seek to find evidence and explanations for child penalties across countries. Their main finding is that the child penalty is accountable for the remainder of the gender gap. They use the methodology of Kleven et al. (2019b) applied to an expanded dataset to examine child penalties in Denmark, Sweden, the United States, the United Kingdom, Austria, and Germany.

This paper researches the child penalty in the United States divided into two parts, the first being a replication of the results of Kleven et al. (2019a). Therefore, this paper researches how large and persistent the child penalties on women's earnings are compared to men, using event studies around the birth of the first child. Subsequently, this research provides a difference-in-differences extension, which serves as a robustness check for the results obtained in the replication part of this paper. This method uses childless men and women as a control group and those who have had children as the treatment group. To perform the difference-in-differences analysis, I first use k-to-k Coarsened Exact Matching to create two observably similar groups using age, education, marital status, relation to head, interview year and lagged income as matching variables.

I find a long-run child penalty of 36%, comparable to that of Kleven et al. (2019a). Moreover, this papers estimates for women closely follow the original estimates of Kleven et al. (2019a), with similar earnings before birth and a sharp drop in earnings for mothers after childbirth. On the other hand, I find some quite large differences within the percentage effects per event time for males, although this could be due to differences in obtaining the data and the magnitude of the original estimates. That said, the absolute differences between estimates are small. Furthermore, the results from the difference-in-differences approach seem to corroborate the convergence of mothers earnings, as well as the persistence of the penalty. A new insight that emerges from the difference-in-differences event study is that men might be affected by parenthood after all, albeit unclear in what periods exactly due to the randomness of the observation exclusion in the matching procedure.

The structure of this paper is as follows. First, I provide an overview of the main findings in existing literature on child penalties. Then, I describe the dataset and how to obtain it. Subsequently, I present a clear mathematical formulation of the used models, for both the replication and the extension, along with the necessary assumptions. Next, I present and discuss the results, followed by summarizing thoughts and a concise answer to the research question.

1.1 Literature

Many researchers of gender inequality believe that children are the critical factor behind the remainder of the gender wage gap (England, 2005; Waldfogel, 1998). Therefore, many researchers investigate this phenomenon. The existing body of literature can be subdivided into two categories. The first category focuses on the reason behind the child penalty. One of the reasons presented in the literature is that mothers might substitute income for advantageous, non-monetary job characteristics, i.e., mothers are confronted with a trade-off between higher earnings and mother-friendly jobs (Felfe, 2006; Budig and England, 2001; England, 2005; Becker, 1985), which is in line with Chung et al. (2017) who find that the child penalty is (partially) caused by a decrease in the earnings of the female spouse. Another explanation follows from a decline in hours worked (Waldfogel, 1997; Adema et al., 2020). Discrimination, categorized as either statistical or taste, is also found to be an important determinant of the child penalty (Budig and England, 2001; Phelps, 1972; Correll et al., 2007). Statistical discrimination is based on the perception that mothers are less productive, on average, and taste discrimination is based on the notion that employers prefer working with childless women (Gough and Noonan, 2013). Finally, many papers, such as Andresen and Nix (2019) and Kleven et al. (2019b), acknowledge that gender norms still account for a large portion of the penalty, despite gender convergence.[2]

The second category is primarily interested in which demographics are most affected, and in what policies are most effective to combat child penalties. For instance, Zhang (2009) uses surveys to compare and measure the differences in earnings. She finds that the earnings gap between women with and without children increase with age, education, work experience and number of children, the latter of which is corroborated by Kleven et al. (2019a). In regard to education, however, there does not seem to exist a general consensus. Adema et al. (2020) find a less steep decline in earnings for higher educated couples compared to Zhang (2009), whereas Anderson et al. (2002) only find a difference in earnings for skilled people, and no difference at all for people who did not finish high school. The latter result is in line with England et al. (2016), who find that white women with high skills and high wages experience the highest total penalties. In contrast, Killewald and Bearak (2014) find that the child penalty is larger for women in the middle of the wage distribution than for women below or above the median.

Clearly, study results vary widely, which might be caused by differences in the time period

[2] Olivetti and Petrongolo (2016) provide an overview of the factors driving gender convergence, and novel perspectives on remaining gender gaps.

covered, the sample, or the analytic technique used. However, typically the gross child penalty in earnings is estimated to be between 5 and 10 percent per child (Gough and Noonan, 2013). This paper contributes to the economic literature relating to the impact of children on the labor market outcomes of parents, while also serving as a robustness validation of the results obtained in Kleven et al. (2019a).

2 Data

Kleven et al. (2019a) use individual-level administrative data from Austria, Denmark and Sweden, and survey data for Germany, the United Kingdom and the United States.[3] This research uses the Panel Study of Income Dynamics (PSID) survey data of the United States to replicate the results of Kleven et al. (2019a) and to produce new results for this papers extension.

2.1 Variable descriptions

To estimate the impact of children on earnings of mothers and fathers using difference-in-differences and event studies around the birth of the first child, I use high-quality panel data with information on children and labor market outcomes provided by the PSID. Table 1 shows the variable names and corresponding description for the dataset.

The binary variable `sex` in Table 1 is equal to 0 for female individuals and is equal to 1 for men. The `event_time` variable captures the time between the current year s and the year of the first child-birth. Consequently, this variable is calculated as the difference between the current year and the calendar `year_first_child`. The `birthyear` of an individual is also manually imputed from the difference between year s and the age of individual i in year s. The last known `marital_status` of an individual ranges from married to never married and widowed, divorced or separated. The variable `education` is categorized as grade school (0-6), middle school (7-8), high school (9-12), college (13-16), and graduate (>16) based on years of education. The education level was not reported in the 1969 interview, and was therefore derived from the education level in 1970.[4] Moreover, the categorical variable `relation_to_head` takes on values 0 to 9 from 1969-1982 and 0 to 98 from 1983-2017, where 1 and 10 indicate head and

[3] An in depth explanation as to how to obtain the data for each of the six countries is given in their Online Appendix (Kleven et al., 2019a).

[4] This derivation is based on the education level in 1971. First, if an individual's education level in 1971 is one year higher than in 1970, then 1969 education is set to the education level in 1970 minus one. Second, if 1970 education is equal to 1971 education, then 1969 education is set equal to the education level in 1970. Note however that this might not be correct for all individuals, since an individual could start/stop education at any moment.

2 and 20 indicate spouse. For the final dataset, the relations are recoded from 10 to 1 and 20 to 2 for consistency. All other values for `relation_to_head` are excluded, since the PSID does not provide data on income for any other individuals than those who are head or spouse.

Table 1: Variable description

Variable name	**Description**
`age`	Age of individual i in year s
`birthyear`	Year of birth of individual i
`education`	Highest attained grade of an individual i in year s
`event_time`	Current event time t
`id`	Identification number
`income`	Total labor income of individual i in year s
`marital_status`	Last known marital status of individual i
	Categories: married = 1, never married = 2, widowed = 3, divorced = 4, NA > 8
`relation_to_head`	Relationship of individual i to head of the family in year s
	Categories: head = 1, spouse = 2, other > 3
`sex`	Binary variable for gender
	Categories: female = 0, male = 1
`year`	Indicator variable for each year s
`year_first_child`	Year in which the first child of individual i was born
	Categories: parent = $1969, \dots, 2017$, non-parent = 9999

Notes: This table shows the variable name with description for each variable in the dataset. The indices range from event time $t = -5, \dots, 10$ and year $s = 1969, \dots, 2017$.

Finally, the main variable of interest is the `income` of the head and spouse of the family. This income variable includes wages, bonuses, overtime, commissions, professional practice, the labor part of farm income, business income, and income from roomers and boarders for both genders. From 1993 onward, the measure given by the PSID as total labor income no longer includes the labor part of farm income and business income. Consequently, I manually add the labor part of income from unincorporated business. However, the PSID only includes the labor and asset part of farmer income combined, rather than separately. The inclusion of asset income could lead to negative earnings and is inconsistent with the definition of labor earnings used in the previous years. Therefore, from 1993 onward, farm income is excluded. Lastly, the PSID conducts bi-yearly interviews from 1999-2017 in which the reported income in interview year s is actually the income of year $s + 1$. For example, the labor income in 2012 is reported in 2011. Since the income of an individual is determined by their relation to the head, which is a yearly status that is unavailable for even years, there is no way to know how to properly link the income of year $s + 1$ to individual i. Therefore, to maximize the number of datapoints, I treat the income in year $s + 1$ as the income in year s from 1999 onward.

2.2 Dataset

The original PSID dataset consists of 34,326 individuals over 39 years ranging from 1969-2017. For both the extension and replication part of this paper, I split the dataset into two parts: one with individuals who have children and one with individuals who do not have children, which I will refer to as the 'children dataset' and 'non-children dataset', respectively. The replication part of this paper is solely based on the children dataset.

The remainder of this section is structured as follows. First, I focus on the children dataset, and I provide descriptive statistics along with an explanation of the cleaning process. Next, I focus on the the non-children dataset before and after applying Coarsened Exact Matching (CEM). I end the section with descriptive statistics for the non-children dataset.

2.2.1 Children dataset

To abide by the constraints set in Kleven et al. (2019a), I ensure that each individual in the dataset has (1) had their first child between the ages of 20 and 45, and that (2) each individual is observed at least eight times total, of which (3) at least once before and once after birth, within a 15-year time window, ranging from 5 years prior to the birth of first child to 10 years post-birth.

To this end, I first delete all observations with event times smaller than -5 and larger than 10 to comply with the second part of constraint (3). Next, I remove all individuals for whom the minimum event time is larger than -1 or the maximum event time is smaller than 1, to ensure the first part of constraint (3). Subsequently, to obtain the birthyear, I calculate the difference between the interview year and the age in that respective year. I then subtract the birthyear of the first child to obtain the age at the birth of the first child, and I delete all observations who do not suffice constraint (1). Thereafter, I make sure that the unique ID number of each individual is observed at least eight times total, which ensures the satisfaction of constraint (2). Finally, I account for outliers.[5] This yields a final children dataset consisting of 3,814 individuals over 39 years, with a total of 46,404 observations.

Table 2 shows descriptive statistics for the children dataset. The dataset consists of 1,941 males, all of which are head, and 1,873 women of which 806 are head and 1,799 are spouse. Additionally, in comparison to Kleven et al. (2019a), male individuals in my dataset are 2.2 years older on average when they get their first child. This number is approximately equal to 1.6 years for women. The difference in average age could be due to the fact that Kleven

[5] For an in depth explanation of the data cleaning process with respect to outliers, see Appendix A.

et al. (2019a) follow individuals from 1968-2019, while I only follow them from 1969-2017, which naturally leads to data loss. Lastly, the average number of years spent on education is equal to 14 for women, which is 0.2 years more than men.

Table 2: Descriptive statistics children dataset

	# of individuals		**Year of FC***		**Age at FC**	**Birthyear**	**Income**	**Marriage**	**Education**
	Head	Spouse	Range	Mean	Mean	Mean	Mean	Median	Mean
Male	1,941	0	1970-2008	1986	28.0	1959	$31,182	Married	13.8
Female	806	1,799	1970-2008	1986	26.5	1958	$13,215	Married	14.0

Notes: This table shows descriptive statistics for the children dataset. An individual can be classified as both head and spouse, since the classification is based on the total time period of 1969-2017. The mean of education is based on the mean of the maximum attained years of education per individual. *FC stands for first child.

Figure 1 and 2 show the distribution of the highest attained grade and the last known marital status of men and women. It holds for both sexes that most individuals have completed college, followed by high school, graduate middle school and finally grade school, although more women than men have completed higher education. Furthermore, most individuals are married, or were married at one point during the sample period.

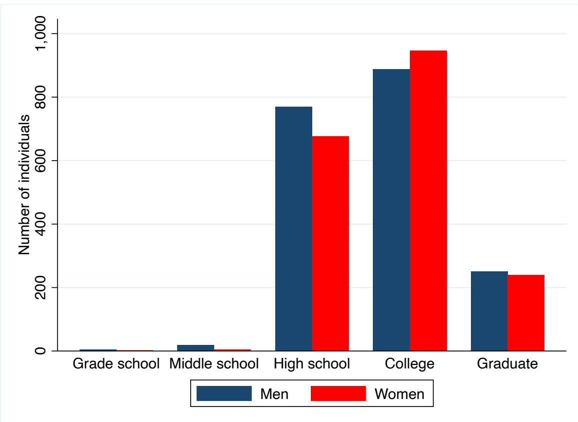


Figure 1: Education distribution

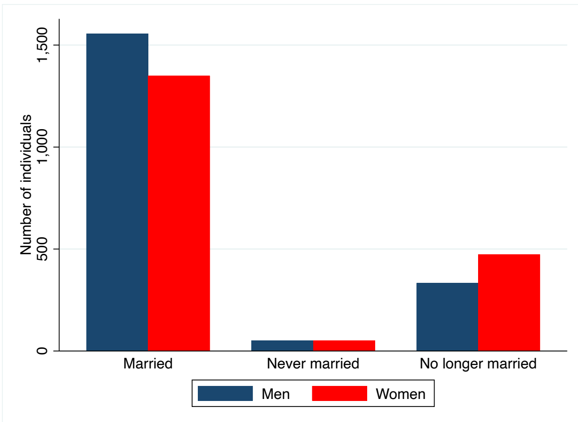


Figure 2: Marital status distribution

Notes: These figures show the distribution of the maximum completed years of education per individual and the last known marital status for men and women in the children dataset.

2.2.2 Non-children dataset

I use the non-children dataset as a control group in order to perform a difference-in-differences event study for the extension of this paper. To this end, I first perform CEM to the non-children dataset using the final children dataset to make the treatment group (individuals with children) and the control group (individuals without children) observably similar. Subsequently, I assign placebo births to each individual in the non-children control group (see Section 3.2.1 for details).

Table 3 shows the descriptive statistics for the non-children dataset before and after matching. The non-children dataset before matching consists of 10,500 untreated individuals, of which 4,956 women and 6,387 men. Thus, approximately 56% of individuals is male, which is 6% more than in the children dataset. The mean income of women is \$12,693, similar to women with children. In contrast, the mean income of men is nearly 35% lower than the mean income for men in the children dataset at \$20,324. Furthermore, most women are born in 1951, while most male individuals are born in 1954. Thus, men and women in the non-children dataset are 5 and 7 years older than men and women in the children dataset, respectively.

Table 3: Descriptive statistics non-children dataset

		# of individuals	**Birthyear**	**Income**	**Marital status**	**Education**
Pre-matching	Women	4,956	1951	\$12,693	No longer married	11.0
	Men	6,387	1954	\$20,324	Never married	10.9
Post-matching	*C*.Women	267	1956	\$14,953	Married	13.9
	C.Men	336	1956	\$24,426	Married	13.7
	T.Women	208	1957	\$14,009	Married	13.6
	T.Men	352	1958	\$30,803	Married	13.3
Standardised differences	Women		-0.10	0.05	0.03	0.11
	Men		-0.14	-0.18	-0.09	0.13

Notes: This table shows unweighted mean values for birthyear, income and maximum attained years of education, along with the median of marital status. The standardised differences are calculated as $\frac{\bar{X}_C - \bar{X}_T}{\sqrt{\sigma_T^2 + \sigma_C^2}}$, where $\bar{X}_C$ refers to the mean value of variable X of the control group, and σ_C^2 denotes its variance. The post-matching row shows descriptive statistics for both the control (C.Sex) and treatment sample (T.Sex).

After performing CEM, the non-children dataset consists of 688 men and 475 women, as shown in Table 3. The mean income of men in the treatment group is equal to \$30,803, which is only 1% lower than in the children dataset. The mean income of women is equal to \$14,009, which differs just as much from the children dataset as the pre-matching income differs from the children dataset. Furthermore, the mean birthyear of men and women is more comparable to the individuals in the children dataset after matching, with the mean birthyear for men and women equal to 1958 and 1957. Thus, treated individuals are only one year older compared to individuals in the children dataset, on average.

3 Methodology

3.1 Replication

I use the event study specification of Kleven et al. (2019a) and Kleven et al. (2019b) to replicate the results for the child penalties for men and women in the United States. For each parent in the dataset, the event time t is indexed relative to the year of the first childbirth, with t ranging from 5 years prior to birth of the firstborn to 10 years after. I run the following regression separately for men and women

$$Y_{ist}^{g} = \sum_{j \neq -1} \alpha_j^g \cdot 1[j = t] + \sum_k \beta_k^g \cdot 1[k = \text{age}_{is}] + \sum_y \gamma_y^g \cdot 1[y = s] + \varepsilon_{ist}^g, \tag{1}$$

where Y_{ist}^{g} equals gross labor earnings, excluding taxes or transfers, specified in levels for individual i of sex g in year s at event time t.[6] The first term on the right-hand side includes event time dummies, the second includes age dummies, and the final term includes year dummies. The year prior to the birth of the first child is taken as the benchmark, therefore I exclude $t = -1$ from the event time dummies. Thus, α_j^g measures, and can be interpreted as, the impact of the arrival of the first child relative to the year before the first childbirth. Lastly, I include age to account for life cycle trends and I include year dummies to account for time trends, such as wage inflation or business cycles.

Next, I use the approach of Kleven et al. (2019a) to convert the estimated level effects into percentages. Therefore, the percentage effects of parenthood on earnings across event time t for each gender g are calculated as

$$P_t^g \equiv \frac{\hat{\alpha}_t^g}{\mathbf{E}\left[\sum_k \hat{\beta}_k^g \cdot 1[k = \text{age}_{is}] + \sum_y \hat{\gamma}_y^g \cdot 1[y = s] \middle| t\right]} \equiv \frac{\hat{\alpha}_t^g}{\mathbf{E}\left[\tilde{Y}_{ist}^g \middle| t\right]},$$

where $\tilde{Y}_{ist}^g$ is the predicted outcome when omitting the contribution of the event time dummies, and therefore children, using the estimates from Equation (1). Hence, P_t^g captures the year-t effect of children as a percentage of the outcome absent children. Therefore, in line with Kleven et al. (2019a), I define the child penalty as

$$P_t = \frac{\hat{\alpha}_t^m - \hat{\alpha}_t^w}{\mathbf{E}\left[\tilde{Y}_{ist}^w \middle| t\right]}, \tag{2}$$

where P_t represents the percentage by which women are falling behind men due to having children at event time t. Finally, I calculate the long-run average penalty P_t from event time

[6] Taking the log would result in the omission of observations for non-participating individuals, therefore Equation (1) is specified in levels.

five to ten. The reason I report the long-run penalty is that, while the approach described above is based on the event of having a first child, the long-run child penalty also includes the impact of children born after to the first one. Therefore, the long-run penalty is able to capture the total effect of children on gender inequality (Kleven et al., 2019b). However, while determining short-term child penalties (e.g. $t = 1$ or 2) primarily depends on a smoothness assumption common to all event studies, determining long-term penalties necessitates stronger assumptions and may require the use of a control group (Kleven et al., 2019b). Thus, to consider the causal identification of this analysis, I provide a difference-in-differences event study.

3.2 Extension

Based on the event study around the birth of the first child, arguments can be made for the causal influence of the birth of a first child on men and women's earnings. However, in order to establish a causal relationship, it is important to know whether similar outcomes would have occurred if the individuals had not had a child, *ceteris paribus*. As mentioned in Section 3.1, the identification of long-term effects requires stronger assumptions, because the penalties are implicitly based on men as a 'control' group for women, since it measures how far women fall behind men due to children. The presence of parallel pre-birth trends after controlling for life cycle and time trends would provide support for this assumption, but the pre-birth trends tend to be less informative for event times further from the birth of the first child. Additionally, it can be problematic to compare women to men if men are also subjected to the treatment, even though the event time analysis might suggest that men are unaffected.

Consequently, I consider a difference-in-differences extension of the event study as an identification check, to evaluate the effect of a first born on the earnings of an individual. This approach uses childless individuals as controls over multiple treatment periods combined with k-to-k Coarsened Exact Matching (CEM). The resulting average treatment effect provides a robust difference-in-difference estimate of treatment effects.

3.2.1 Coarsened Exact Matching

CEM is a monotonic imbalance-reducing matching method, which allows bounding of higher level imbalance in some characteristics of the distribution through an *ex-ante* choice of coarsening (Bertoni et al., 2020). The main goal of matching in general is to prune observations such that the remaining dataset has a better balance between the control and treatment group. That is, the empirical distributions of covariates $\mathbf{X}$ are more similar between groups (Iacus et al., 2012). Using a matching method leads to less model dependence and reduced statistical bias. Moreover, the use of CEM never increases the imbalance between the control and treatment group.[7]

The CEM algorithm creates a set of strata with the same coarsened values of matching variables. I use a k-to-k transformation of CEM, which produces a matching result with equal number of treated and control individuals in each matched stratum. This result is obtained by randomly dropping observations[8]. CEM sets the matching weight equal to 1 for individuals who can be matched, and equal to 0 for those who cannot. Therefore, the estimates are to be interpreted as the average treatment effects.

For this study, male and female individuals in the treatment and control samples are matched using the panel data described in Section 2.2.1 and 2.2.2. The matching variables include age, education (coarsened to five categories), marital status (coarsened to four categories), interview year, the relation to head and a lagged income variable to account for the earnings trajectory of individuals. This matching results in l strata, each containing $k1$ treated and $k2$ control individuals, with $k1 = k2 = k$. Once the matching is complete, I calculate the rounded median year of first child per stratum $r \in l$ using the $k1$ treated individuals. Subsequently, I assign the rounded median to all $k2$ control individuals in stratum r as their *temporary* placebo birthyear. Now, since an individual is observed over multiple years and can therefore be matched to multiple individuals, it is possible that a control individual $i \in k2$ has been assigned multiple placebo birthyears. To solve this problem, I generate a random variable with uniformly distributed numbers assigned to each observation, and I determine the minimum of these random numbers per individual i. Then, I assign the *final* placebo birthyear to individual i by selecting the *temporary* birthyear that is associated with the minimum random value of that individual. As a result, each control individual has one unique placebo birthyear in the dataset. Finally, I ensure that conditions (1) to (3), as described in Section 2.2.1, are met.

3.2.2 Difference-in-differences

Difference-in-differences is a quasi-experimental approach that compares the changes in outcomes over time between a treatment and a control group. The control group consists of those who never have children, but have been assigned a placebo birth and are observed in a 15-year window around the placebo, and the treatment group consists of individuals who are observed in a 15-year window around the birth of the first child.

The treatment variable D_i is binary and is equal to one if individual i has had a child at any point in the PSID dataset, and equal to 0 if not. The variable Y^g_{ist} denotes the outcome, gross

[7]See (Iacus et al., 2011, 2012) for a detailed discussion of CEM properties and a comparison with other matching methods.

[8]I use random exclusion of observations over other exclusion methods in order to reduce the chance for bias.

labor earnings in this case, that would be realized for individual i of sex g in year s at event time t. The estimation of Y^g_{ist} is similar to that of Equation (1), but with an additional difference-in-differences term and using CEM weights least squares instead of ordinary least squares. Same as before, I run the following regression separately for men and women

$$\begin{aligned} Y^g_{ist} = &\sum_{j \neq -1} \alpha^g_j \cdot 1[j = t] + \sum_k \beta^g_k \cdot 1[k = age_{is}] + \sum_y \gamma^g_y \cdot 1[y = s] \\ &+ \sum_{j \neq -1} \delta^g_j D_i \cdot 1[j = t] + \eta^g_{ist}. \end{aligned} \tag{3}$$

3.3 Assumptions

3.3.1 Event study

Since panel data includes repeated observations of the same individual, it is likely that the error terms are correlated. This is confirmed by the Breusch-Pagan test for heteroskedasticity, which strongly rejects the null hypothesis of constant variance. As a consequence OLS is no longer efficient and the OLS standard errors are incorrect. However, OLS is still consistent and unbiased. To account for the inconsistency of the covariance matrix of the estimated regression coefficients, I use robust standard errors clustered on the individual level. These standard errors are also used for the difference-in-differences regressions.

3.3.2 Difference-in-differences

The difference-in-differences approach rests on two main assumptions: parallel trends and common shock (Angrist and Pischke, 2008; Dimick and Ryan, 2014).

The parallel trends assumption states that the trend in the outcome variable of the treatment and control group in the absence of treatment are the same, or similar at least. To asses whether the parallel trends assumption holds in practice, one can empirically examine graphs of the outcome variable over time, or one could assess the significance of the interaction term between time and treatment in the pre-intervention period (Dimick and Ryan, 2014). That is, if the interaction between event time and the dummmy for the treated is zero before the first child birth, it would be reasonable to assume that the trends would have continued in parallel, had the treatment never occurred. If the assumption does not hold, a difference-in-differences analysis would be biased.

The second assumption is that of common shocks. In economics, a shock is an unexpected or unpredictable event (unrelated to the treatment) that affects a system. The common shocks assumption states that any event that occurs during, or following the intervention, should equally affect each group (Dimick and Ryan, 2014). There is no analytical way to confirm the validity of this assumption, but if the parallel trends assumption holds – which implies that there are no major differences in the labor market participation absent the treatment – one can assume that other relevant shocks (e.g. a liquidity crisis, culture shocks in the labor market, or family related shocks) would have affected both groups in the same way in the absence of having a child. Naturally, the preparation that goes into conceiving a child might cause individuals to enter different life trajectories, allowing shocks to be received differently, but this is likely reflected in the rejection of the parallel trends assumptions, thus invalidating the common trends assumption regardless.

3.3.3 Coarsened Exact Matching

The main assumption for CEM is much more difficult to validate than those of difference-in-differences. To use CEM, one must assume that treatment assignment is ignorable conditional on covariates $\mathbf{X}$. This assumption is often referred to as 'no unmeasured confounders' or 'no omitted variables'. In practice, this means that assignment of the treatment is independent of the potential outcomes (Iacus et al., 2012). Unfortunately, the assumption of conditional ignorability is untestable.

4 Results

First, I present the results of the replication part of this paper, followed by a comparison between the estimates in this paper and those of Kleven et al. (2019a) for women and men separately. I then conclude this section with the results and implications of the difference-in-differences extension.

4.1 Event study

Figure 3 shows the estimated effects of parenthood on earnings in the United States along with the estimated effects of Kleven et al. (2019a). The earnings of men and women evolve similarly before parenthood – taking life cycle and time trends into account – but diverge sharply after the first childbirth. Moreover, women experience an immediate drop in earnings of almost 30% as a result of motherhood, which further increases in magnitude to approximately 40%, and persists for at least 10 years. In contrast, men seem to be nearly unaffected by birth.

The most apparent difference between this papers female estimates and those of Kleven et al. (2019a), is in the long run P_t^w levels from event time 5 to 10. Specifically, I find that female earnings have plateaued around -40%, whereas Kleven et al. (2019a) find a slightly more positive plateau level of -36%. Although, fortunately, the absolute differences between estimates are small in magnitude, as shown in Figure 5.

Additionally, Figure 4 shows that all but two event time estimates are within a range of -30% to 15%, with exception of -69% and 189% for event times -3 and -2, respectively. However, since the original estimates at these event times are smallest (0.0637 and 0.0315), it seams reasonable that the percentage deviation is largest. In addition, Figure 5 does not show exceptional differences in magnitude at these event times.

Overall, the mean difference between estimates is equal to -0.021, which implies that I negatively overestimate the percentage effects of parenthood on earnings by 2.1% on average, relative to Kleven et al. (2019a). The mean absolute percentage difference is relatively low, equal to 27%. Furthermore, Table 4 in Appendix B shows that all event time estimates for women are significant, with positive and relatively constant estimates before the first childbirth, and large negative estimates after birth. This further corroborates the findings of Kleven et al. (2019a), and provides an indication for the causal relation between lower earnings and childbirth for women. This causal relationship is further investigated using difference-in-differences in Subsection 4.2.

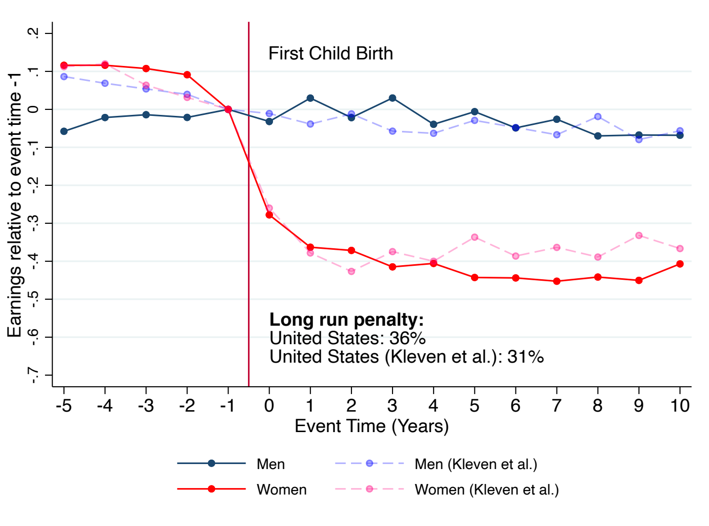


Figure 3: Percentage effects of parenthood on earnings

Notes: This figure shows this papers estimated percentage effects of parenthood on earnings across event time t for each gender g, i.e., P_t^g along with the estimates of Kleven et al. (2019a). The figure also displays long-run child penalties, defined as the average penalty P_t from event time five to ten.

Based on Figure 3, male estimates seem to deviate most from Kleven et al. (2019a) in the pre-birth period and at event time 1 and 3. These differences are corroborated by Figure 4, which shows vast percentage differences between the estimates for male individuals of this paper and the estimates of Kleven et al. (2019a), resulting in a total mean absolute percentage difference of 104%. However, the estimates of Kleven et al. (2019a) are small in magnitude. Therefore, a slight difference in estimates can cause a high percentage difference. For example, this papers estimates for men and women at event time 8 are both -0.05 lower in magnitude relative to Kleven et al. (2019a), but it results in a -260% difference for men, compared to a mere 20% difference for women. In fact, Figure 5 shows highly comparable deviations for men and women both in magnitude, and in mean (absolute) differences.

With this in mind, the least similar male estimate is that of event time -5, which differs 150% from the estimates of Kleven et al. (2019a), and differs almost -0.15 in magnitude – whilst all other estimates deviate less than 0.1 in magnitude from the original estimates. Additionally, Table 5 in Appendix B shows that the -5 event time estimate is significant and negative, as is the estimate for event time nine. This implies that the income five years before and nine years after birth are respectively $-\$1271$ and $-\$3144$ lower relative to the income one year before the first childbirth. However, given that none of the other pre-birth estimates nor the remaining post-birth estimates are significant, this does not seem to allude to a causal relation between earnings and fatherhood.

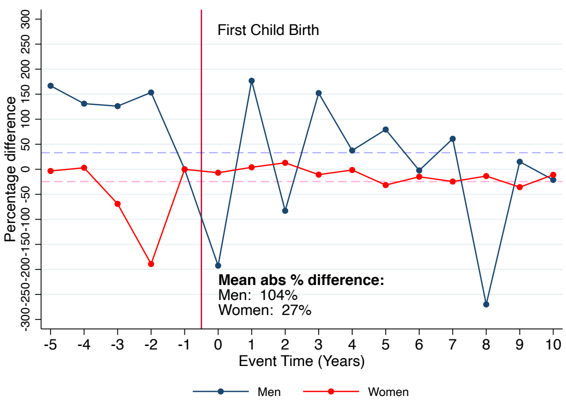


Figure 4: Percentage differences

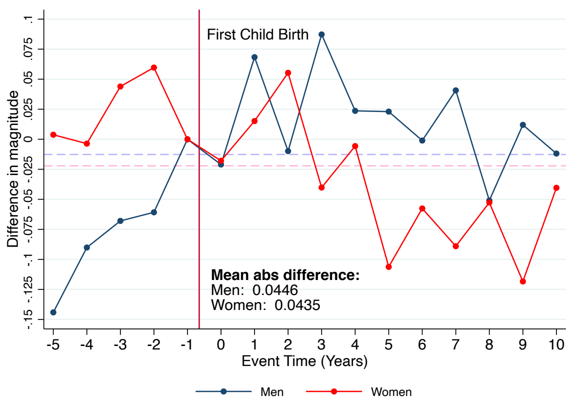


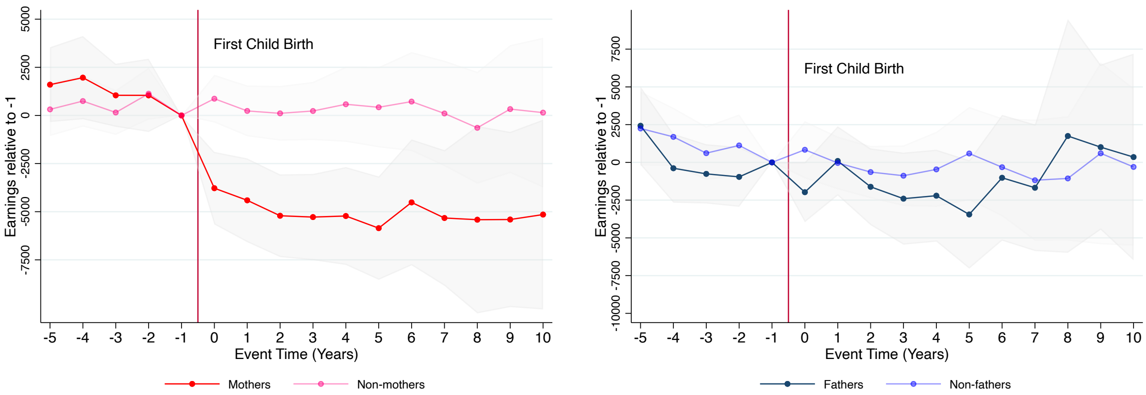
Figure 5: Differences in magnitude

Notes: These figures show the percentage differences and absolute difference between this papers estimates and those of Kleven et al. (2019a) for men and women. The dotted lines represent the mean values. The absolute mean percentage differences are reported on the bottom.

4.2 Difference-in-differences

The effect of the CEM matching on the unweighted pre-treatment descriptive statistics for men and women can be seen in Table 3. This table also includes standardised differences for birthyear, income, marital status, and education. According to Imbens and Wooldridge (2009), the standardised difference should be lower than 0.25 to ensure that the linear regression methods are not sensitive to the model specification. Table 3 shows that all standardised differences are well below this threshold, thus the treatment and the control group seem to be balanced. Moreover, it follows from the discussion in Section 2.2.2 that the matching increases the comparability of the descriptive statistics between those who have had a child, and those who have not.

Now, as an identification check, I compare the event study approach to a difference-in-differences approach for women and men separately. Figure 6 shows the estimated effects of children on earnings in this difference-in-differences design. These statistics are obtained from Equation (3). The details of how I construct the control groups for men and women, including how placebo births are assigned to those who never have children, are described in Section 3.2.



(a) Women who have children vs women without children

(b) Men who have children vs men without children

Figure 6: Estimated effects of children in a difference-in-differences event study design

Notes: The graphs show the estimated event time effects on earnings of women (Panel A) and men (Panel B). The figure reports difference-in-differences estimates between those who have children and those who do not have children (as opposed to previous penalty measures based on comparing men and women who *have* had children). The shaded 95% confidence intervals are based on clustered standard errors.

Panel A of Figure 6 shows the difference-in-differences analysis for women. As seen from the graph, mothers and non-mothers show a near identical pre-birth trend, with none of the estimates significantly different from zero. This is an indication for the validity of the parallel trends assumption discussed in Section 3.3.2. Moreover, the course of the line for non-parents is flat, which is to be expected since the line should be the change in earnings after controlling for age and wave fixed effects.

Table 6 in Appendix C shows that once mothers give birth, their earnings are immediately significantly affected with a decrease of approximately $-\$4650$. This negative effect persists for at least ten years, with a long run interaction estimate equal to approximately −\$5400. Earnings of childless women do not seem to be affected, as the event time estimates are insignificant for all event times.

Moreover, it is clear from Panel B of Figure 6 and the insignificance of the event time estimates in Table 7 in Appendix C, that the earnings for childless men are unaffected across event times. Combined with the insignificance of the interaction estimates in the pre-birth period, the parallel trends assumption seems valid.

A new insight from the difference-in-difference approach is that men's earnings might be significantly affected by parenthood after all. This follows from the significant interaction estimate of approximately −\$2810 in the year of childbirth.[9] However, the difference-in-difference approach required a k-to-k matching procedure, which randomly drops control observations to ensure an equal number of individuals in each stratum (see Section 3.2.1 for details). Since this observation exclusion happens at random, the regression output differs each time it is performed. As a consequence, other interaction estimates than those reported in Table 7 in Appendix C could be significant. Generally, one to three interaction estimates are significant after birth, while the event time estimates are nearly always insignificant. This provides some indication that the significant estimates from the event study discussed in Section 4.1 could be caused by childbirth after all.

With that in mind, if men *are* adversely affected by fatherhood (which follows from a significantly negative interaction estimate), the impact on women's earnings relative to men's as captured by the child penalties from the event study are smaller than the numbers reported in Figure 3 for those event times.

5 Conclusion

This thesis is a replication and extension of Kleven et al. (2019a). Thus, this paper researches how large and persistent child penalties are on women's earnings compared to men. I use data provided by the Panel Study of Income Dynamics, which includes precise and detailed information on the labour market outcomes of household heads and their spouses for a nationally representative sample of the population in the United States. I follow all individuals over a 15-year time window within 1969 to 2017, in which each individual is observed at least eight times.

Based on the event study, I find that even though men and women's earnings are similar

[9] This can be seen in Table 7 in Appendix C, and the confidence interval around the estimate for fathers at $t = 0$ in Panel B of Figure 6.

before birth, only women experience a negative percentage effect of parenthood on earnings across event times. The drop in earnings persists for (at least) ten years after the first childbirth. More specifically, I find that the long-run child penalty in earnings is equal to 36%, which is 5% higher than reported in Kleven et al. (2019a). The estimates I find for women generally do not differ much from those of the reference paper, while male estimates do differ a lot percentage wise. However, this is most likely due to differences in data collection and the small magnitude of the estimates. Fortunately, the absolute differences between the estimates of this paper and those of Kleven et al. (2019a) are less than 0.125 for each estimate.

Moreover, I provide a difference-in-differences extension as an identification check, using individuals without children as a control group. To construct this control group I use k-to-k Coarsened Exact Matching, and I assign placebo births accordingly. The results of this approach show that the earnings of childless women remain unchanged across event times, while mothers experience a persistent drop in earnings after childbirth, in line with the results of the event study. These results provide a strong indication for the causal effect of childbirth on mother's earnings. Furthermore, in contrast to the findings of Kleven et al. (2019a), the difference-in-differences analysis for men gives rise to the possibility of a causal relation between birth and a decrease in gross labor earnings for fathers. Although, it is unclear at which event times earnings are significantly affected, due to the randomness of the k-to-k matching. That said, if men *are* adversely affected by fatherhood, the impact on women's earnings relative to men's as captured by the child penalties from the event study in the replication are smaller than reported.

Finally, further research should be conducted on the most efficient way of matching and assigning placebo births, in order to accurately assess the robustness of the event study of Kleven et al. (2019a). Moreover, the significance in the interaction estimates after birth for fathers are quite dependent on what observations are used exactly. And since the k-to-k matching drops individuals at random, this poses a problem for the robustness of the method. This paper uses random exclusion of observations in order to reduce the chance for bias, but it would be of interest to explore other possibilities such as using nearest neighbor selection using, e.g., Minkowski or Euclidean distances. These results can provide additional information on the causal relation between fathers earnings and childbirth.

References

Y. Adema, S. Rabaté, and S. Rellstab. Inkomen moeders halveert bijna na komst kinderen, Dec 2020. URL https://www.cpb.nl/sites/default/files/omnidownload/ESB-artikel-Inkomen-moeders-halveert-bijna-na-komst-kinderen.pdf.

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

M. E. Andresen and E. Nix. What causes the child penalty? Evidence from same sex couples and policy reforms. Technical report, Discussion Papers, 2019.

J. D. Angrist and J.-S. Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

G. S. Becker. Human capital, effort, and the sexual division of labor. *Journal of labor economics*, 3(1, Part 2):S33–S58, 1985.

D. Bertoni, D. Curzi, G. Aletti, and A. Olper. Estimating the effects of agri-environmental measures using difference-in-difference coarsened exact matching. *Food Policy*, 90:101790, 2020.

M. J. Budig and P. England. The wage penalty for motherhood. *American Sociological Review*, 66(2):204–225, 2001. ISSN 00031224. URL http://www.jstor.org/stable/2657415.

Y. Chung, B. Downs, D. H. Sandler, R. Sienkiewicz, et al. The parental gender earnings gap in the United States. Technical report, 2017.

S. J. Correll, S. Benard, and I. Paik. Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, 112(5):1297–1338, 2007. ISSN 00029602, 15375390. URL http://www.jstor.org/stable/10.1086/511799.

J. B. Dimick and A. M. Ryan. Methods for evaluating changes in health care policy: the difference-in-differences approach. *Jama*, 312(22):2401–2402, 2014.

P. England. Gender inequality in labor markets: The role of motherhood and segregation. *Social Politics: International Studies in Gender, State & Society*, 12(2):264–288, 2005.

P. England, J. Bearak, M. J. Budig, and M. J. Hodges. Do highly paid, highly skilled women experience the largest motherhood penalty? *American Sociological Review*, 81(6):1161–1189, 2016.

C. Felfe. *The Child Penalty–A Compensating Wage Differential?* CEPS, 2006.

M. Gough and M. Noonan. A review of the motherhood wage penalty in the United States. *Sociology Compass*, 7(4):328–342, 2013.

S. M. Iacus, G. King, and G. Porro. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493):345–361, 2011.

S. M. Iacus, G. King, and G. Porro. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1):1–24, 2012. doi: 10.1093/pan/mpr013.

G. W. Imbens and J. M. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86, 2009.

A. Killewald and J. Bearak. Is the motherhood penalty larger for low-wage women? A comment on quantile regression. *American Sociological Review*, 79(2):350–357, 2014.

H. Kleven, C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller. Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*, 109:122–26, May 2019a. doi: 10.1257/pandp.20191078. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20191078.

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

C. Olivetti and B. Petrongolo. The evolution of gender gaps in industrialized countries. *Annual review of Economics*, 8:405–434, 2016.

E. S. Phelps. The statistical theory of racism and sexism. *The American Economic Review*, 62 (4):659–661, 1972. ISSN 00028282. URL http://www.jstor.org/stable/1806107.

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

J. Waldfogel. Understanding the "Family Gap" in Pay for Women with Children. *Journal of Economic Perspectives*, 12(1):137–156, March 1998. doi: 10.1257/jep.12.1.137. URL https://www.aeaweb.org/articles?id=10.1257/jep.12.1.137.

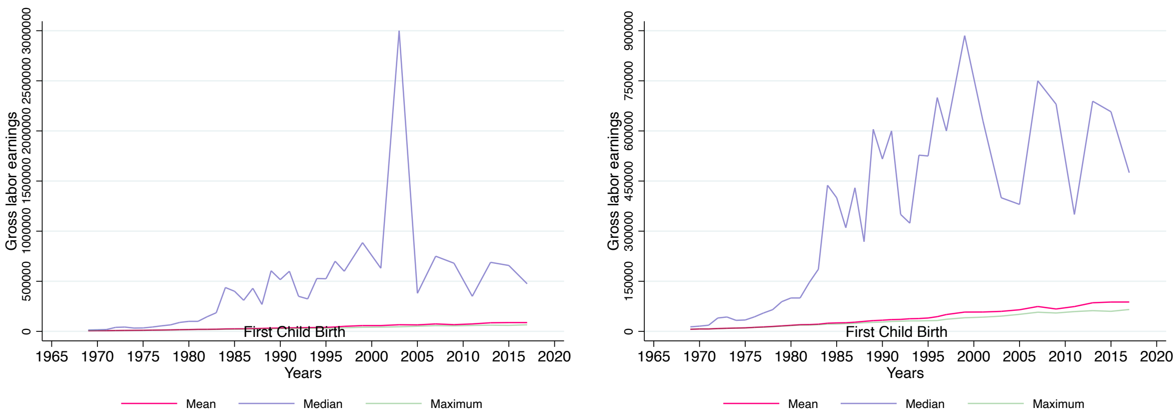
X. Zhang. Earnings of women with and without children. *Perspectives on Labour and Income*, 21(2):5, 2009.

Appendices

A Outliers children dataset

In order to maximize the accuracy of the estimates in this analysis, I investigate the data for outliers. I evaluate the mean, median and maximum earnings per interview year, for men and women separately.

Figure 7 shows the earnings for men before accounting for outliers in Panel A. Clearly, there is an isolated peak in 2013 at $3,000,000. This value is nearly 5 times larger than the highest income in 2011, and nearly 8 times larger than the maximum income in 2015. Upon further investigation, I find that the second to highest income in 2013 is $700,000, and the second highest income for this individual is equal to $260,000 in 2005. As a result, this observation appears to be an outlier for both the sample and the individual, and is therefore deleted from the dataset. After accounting for the outlier, the three highest incomes are equal to 1.5, 1.4, and 1 million in 2017, 2015 and 2011, respectively, as shown in Panel B of Figure 7. There do not appear to be any additional aberrant observations in the maximum earnings, and the mean and median evolve consistently.



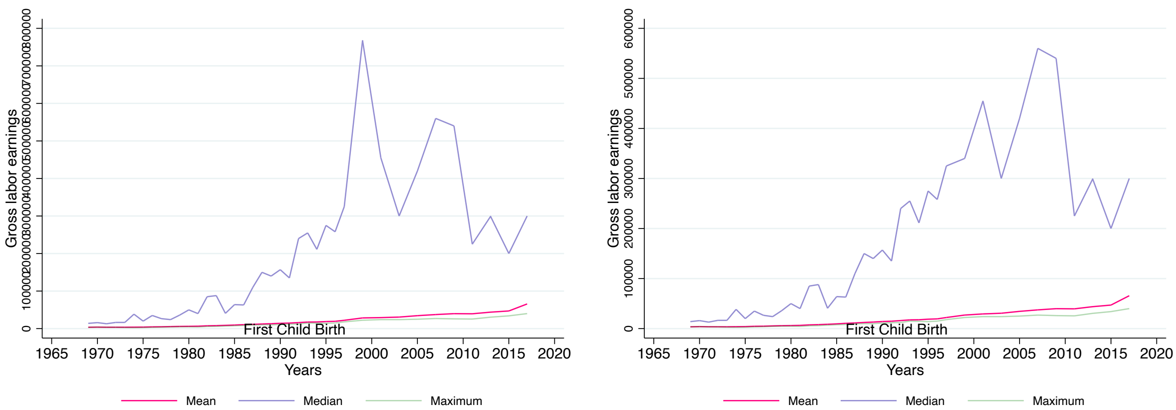
(a) Before accounting for outliers

(b) After accounting for outliers

Figure 7: Gross labor earnings of men over time

Notes: The graphs show the mean, median an maximum gross labor earnings over time for the children dataset before accounting for the outlier (Panel A) and after accounting for the outlier (Panel B). The minimum income is always equal to zero.

Similarly, Figure 8 shows the mean, median and maximum gross labor earnings for women. Again, there is one clear deviation in 1999 at $768,000. The year prior to and after 1999 show a maximum income of $325,000 and $455,000, respectively. Furthermore, the second highest income for the individual is $325,000, less than half the size. The second highest value in 1999 equals 340,000. Thus, analogous to the analysis of men, the observation is an outlier for the sample and for the individual, and is therefore deleted from the sample. The resulting graph is presented in Panel B of Figure 8. No additional values stand out after this deletion.



(a) Before accounting for outliers

(b) After accounting for outliers

Figure 8: Gross labor earnings of women over time

Notes: See the notes to Figure 7.

Finally, after identifying and accounting for each outlier, I reassure that the dataset meets every restriction described in Section 2.2.1.

B Event study

Table 4: Event study regression women

Year		Age		Event time	
γ_{1969}	0	β_{16}	0	α_{-5}	1503.4***
	$(\cdot)$		$(\cdot)$		(3.62)
γ_{1970}	1051.9**	β_{17}	3462.2*	α_{-4}	1533.8**
	(2.96)		(2.49)		(3.16)
γ_{1971}	1340.9***	β_{18}	4326.0***	α_{-3}	1396.0**
	(3.78)		(3.64)		(2.84)
γ_{1972}	1710.1***	β_{19}	5386.8***	α_{-2}	1205.6***
	(4.44)		(4.47)		(3.50)
γ_{1973}	2086.0***	β_{20}	6315.4***	α_{-1}	0
	(5.31)		(5.47)		$(\cdot)$
γ_{1974}	2398.9***	β_{21}	7185.3***	α_{0}	-4168.9***
	(5.92)		(6.22)		(-12.09)
γ_{1975}	2659.8***	β_{22}	8224.4***	α_{1}	-5944.5***
	(6.47)		(7.11)		(-18.23)
γ_{1976}	3396.8***	β_{23}	9470.5***	α_{2}	-6677.0***
	(7.99)		(8.19)		(-14.88)
γ_{1977}	3755.7***	β_{24}	10617.9***	α_{3}	-7957.1***
	(8.82)		(8.99)		(-18.26)
γ_{1978}	4502.4***	β_{25}	11227.2***	α_{4}	-8469.5***
	(10.26)		(9.64)		(-15.60)
γ_{1979}	5001.9***	β_{26}	12083.0***	α_{5}	-9793.8***
	(11.09)		(10.15)		(-17.42)
γ_{1980}	5355.3***	β_{27}	13327.4***	α_{6}	-10551.0***
	(11.77)		(11.21)		(-14.95)
γ_{1981}	5582.4***	β_{28}	14009.3***	α_{7}	-11379.7***
	(11.79)		(11.49)		(-16.30)
γ_{1982}	6525.1***	β_{29}	15140.4***	α_{8}	-11925.4***
	(13.02)		(11.86)		(-12.78)
γ_{1983}	7074.0***	β_{30}	15507.3***	α_{9}	-12775.0***
	(13.81)		(12.06)		(-12.37)
γ_{1984}	7584.1***	β_{31}	15506.4***	α_{10}	-12366.5***
	(14.78)		(11.64)		(-10.73)
γ_{1985}	8147.0***	β_{32}	17041.8***		
	(15.38)		(12.41)		
γ_{1986}	9073.5***	β_{33}	17382.6***		
	(16.88)		(12.01)		
γ_{1987}	9708.6***	β_{34}	18679.9***		
	(17.19)		(12.39)		
γ_{1988}	10598.5***	β_{35}	18031.1***		
	(18.02)		(11.61)		

Table 4 continued from previous page

Year		Age		Event time
γ_{1989}	11144.1***	β_{36}	21234.7***	
	(18.31)		(12.50)	
γ_{1990}	12143.4***	β_{37}	20923.8***	
	(19.33)		(10.68)	
γ_{1991}	12669.3***	β_{38}	22197.6***	
	(18.94)		(10.61)	
γ_{1992}	13788.7***	β_{39}	25423.4***	
	(18.90)		(10.14)	
γ_{1993}	14996.8***	β_{40}	26254.5***	
	(17.74)		(8.00)	
γ_{1994}	15020.6***	β_{41}	20360.0***	
	(18.61)		(7.49)	
γ_{1995}	16183.4***	β_{42}	29183.7***	
	(17.20)		(6.65)	
γ_{1996}	16718.8***	β_{43}	19041.3***	
	(16.23)		(5.64)	
γ_{1997}	19175.5***	β_{44}	28985.3***	
	(17.19)		(4.55)	
γ_{1999}	23819.7***	β_{45}	20106.4***	
	(16.05)		(4.97)	
γ_{2001}	25624.9***	β_{46}	22728.6***	
	(14.12)		(3.30)	
γ_{2003}	26663.0***	β_{47}	17226.6***	
	(13.59)		(3.68)	
γ_{2005}	30505.6***	β_{48}	35320.9**	
	(12.56)		(3.22)	
γ_{2007}	33446.8***	β_{49}	34708.5	
	(9.66)		(1.80)	
γ_{2009}	35996.2***	β_{50}	8127.7	
	(8.60)		(0.83)	
γ_{2011}	35771.2***	β_{51}	31749.6	
	(8.66)		(1.74)	
γ_{2013}	40216.9***	β_{52}	28143.3***	
	(7.25)		(9.43)	
γ_{2015}	42716.8***			
	(6.71)			
γ_{2017}	61872.5***			
	(4.46)			
N	22819			

Notes: This table reports estimates of the model specification reported in Equation (1) for women. Clustered standard errors are reported in parentheses. An asterisk denotes a significance level of ***0.001, **0.01, and **0.05.

Table 5: Event study regression men

Year		Age		Event time	
γ_{1969}	0	β_{17}	0	α_{-5}	-1270.9*
	(·)		(·)		(-2.19)
γ_{1970}	1418.6**	β_{18}	-5140.7**	α_{-4}	-470.8
	(2.63)		(-3.14)		(-0.76)
γ_{1971}	1052.5	β_{19}	-4863.5***	α_{-3}	-313.0
	(1.78)		(-3.80)		(-0.49)
γ_{1972}	2306.4***	β_{20}	-3794.2**	α_{-2}	-465.1
	(3.68)		(-3.00)		(-1.08)
γ_{1973}	2495.9***	β_{21}	-2363.8	α_{-1}	0
	(3.82)		(-1.96)		(·)
γ_{1974}	2993.2***	β_{22}	-2103.1	α_{0}	-774.2
	(4.45)		(-1.72)		(-1.71)
γ_{1975}	2832.5***	β_{23}	-639.2	α_{1}	788.6
	(4.12)		(-0.53)		(1.81)
γ_{1976}	3928.4***	β_{24}	-33.34	α_{2}	-625.4
	(5.71)		(-0.03)		(-1.05)
γ_{1977}	4763.7***	β_{25}	631.0	α_{3}	932.4
	(6.56)		(0.52)		(1.39)
γ_{1978}	5968.1***	β_{26}	1876.5	α_{4}	-1318.6
	(7.98)		(1.51)		(-1.76)
γ_{1979}	6813.2***	β_{27}	3569.6**	α_{5}	-215.1
	(9.07)		(2.60)		(-0.23)
γ_{1980}	7879.5***	β_{28}	4575.8***	α_{6}	-1888.3
	(10.26)		(3.48)		(-1.60)
γ_{1981}	9126.0***	β_{29}	6219.6***	α_{7}	-1075.3
	(11.14)		(4.68)		(-0.83)
γ_{1982}	9191.9***	β_{30}	7053.7***	α_{8}	-3073.8
	(10.95)		(5.11)		(-1.95)
γ_{1983}	9854.5***	β_{31}	8752.2***	α_{9}	-3144.2*
	(11.40)		(6.03)		(-2.01)
γ_{1984}	12361.2***	β_{32}	10423.7***	α_{10}	-3357.5
	(12.73)		(7.00)		(-1.66)
γ_{1985}	12840.8***	β_{33}	12649.0***		
	(13.06)		(7.82)		
γ_{1986}	12818.3***	β_{34}	14436.3***		
	(14.15)		(8.29)		
γ_{1987}	14238.4***	β_{35}	14987.4***		
	(14.87)		(8.07)		
γ_{1988}	16080.5***	β_{36}	22280.1***		
	(16.33)		(9.38)		
γ_{1989}	17699.8***	β_{37}	16785.8***		
	(15.24)		(7.68)		

Table 5 continued from previous page

Year		Age		Event time
γ_{1990}	18795.5***	β_{38}	21567.5***	
	(15.72)		(8.36)	
γ_{1991}	20075.2***	β_{39}	22114.6***	
	(15.55)		(6.69)	
γ_{1992}	20344.6***	β_{40}	28107.1***	
	(17.25)		(6.86)	
γ_{1993}	22012.9***	β_{41}	21690.1***	
	(17.40)		(6.10)	
γ_{1994}	22474.6***	β_{42}	28109.9***	
	(16.30)		(5.77)	
γ_{1995}	23994.0***	β_{43}	23401.3***	
	(15.99)		(4.67)	
γ_{1996}	28022.4***	β_{44}	18107.8***	
	(13.76)		(4.15)	
γ_{1997}	34448.9***	β_{45}	26501.3**	
	(16.46)		(2.96)	
γ_{1999}	41268.4***	β_{46}	19475.7**	
	(15.24)		(3.24)	
γ_{2001}	40850.4***	β_{47}	15144.0*	
	(14.83)		(2.31)	
γ_{2003}	43039.5***	β_{48}	11203.6	
	(17.29)		(1.74)	
γ_{2005}	46843.2***	β_{49}	20001.8	
	(16.37)		(1.64)	
γ_{2007}	55000.4***	β_{50}	8840.7	
	(12.30)		(1.16)	
γ_{2009}	46439.7***	β_{51}	-1656.5	
	(10.54)		(-0.22)	
γ_{2011}	52681.0***	β_{52}	7937.8	
	(11.68)		(1.54)	
γ_{2013}	62935.5***	β_{53}	-2174.8	
	(8.13)		(-0.14)	
γ_{2015}	65366.0***	β_{54}	-376.1	
	(6.49)		(-0.03)	
γ_{2017}	64959.5***	β_{55}	-19263.5***	
	(5.68)		(-7.59)	
N	23585			

Notes: See the notes to Table 4, but applied to men.

C Difference-in-differences

Table 6: Difference-in-differences regression women

Year		Age		Event time		Interaction	
γ_{1969}	0	β_{17}	0	α_{-5}	314.8	$\delta_{-5}D$	1284.6
	$(\cdot)$		$(\cdot)$		(0.45)		(1.10)
γ_{1970}	1424.8*	β_{18}	-881.4	α_{-4}	746.7	$\delta_{-4}D$	1216.1
	(1.97)		(-0.57)		(1.10)		(1.00)
γ_{1971}	1023.0	β_{19}	-1672.0	α_{-3}	152.9	$\delta_{-3}D$	893.0
	(1.39)		(-1.25)		(0.26)		(0.88)
γ_{1972}	1385.3	β_{20}	-1278.5	α_{-2}	1130.0	$\delta_{-2}D$	-86.00
	(1.82)		(-1.14)		(1.64)		(-0.07)
γ_{1973}	2334.6**	β_{21}	1259.8	α_{-1}	0	$\delta_{-1}D$	0
	(2.77)		(1.27)		$(\cdot)$		$(\cdot)$
γ_{1974}	2073.3*	β_{22}	1427.6	α_{0}	869.9	$\delta_{0}D$	-4649.5***
	(2.44)		(1.49)		(1.38)		(-3.93)
γ_{1975}	2076.2*	β_{23}	1515.5	α_{1}	235.1	$\delta_{1}D$	-4643.2***
	(2.45)		(1.71)		(0.35)		(-3.52)
γ_{1976}	3245.7***	β_{24}	3309.1***	α_{2}	110.1	$\delta_{2}D$	-5320.2***
	(3.61)		(3.66)		(0.15)		(-4.14)
γ_{1977}	3214.4***	β_{25}	3672.0***	α_{3}	231.3	$\delta_{3}D$	-5510.7***
	(3.50)		(4.19)		(0.30)		(-4.36)
γ_{1978}	4397.9***	β_{26}	3593.1***	α_{4}	581.6	$\delta_{4}D$	-5805.2***
	(4.73)		(4.42)		(0.58)		(-3.79)
γ_{1979}	5073.8***	β_{27}	4419.6***	α_{5}	425.1	$\delta_{5}D$	-6280.2***
	(5.39)		(5.12)		(0.40)		(-4.00)
γ_{1980}	6383.9***	β_{28}	5537.6***	α_{6}	719.5	$\delta_{6}D$	-5235.3**
	(6.57)		(6.16)		(0.55)		(-2.81)
γ_{1981}	6646.6***	β_{29}	5143.3***	α_{7}	103.4	$\delta_{7}D$	-5430.7**
	(6.92)		(5.71)		(0.07)		(-2.73)
γ_{1982}	7696.6***	β_{30}	7341.6***	α_{8}	-642.7	$\delta_{8}D$	-4772.8*
	(6.99)		(7.25)		(-0.43)		(-2.51)
γ_{1983}	8049.3***	β_{31}	6484.8***	α_{9}	330.3	$\delta_{9}D$	-5736.4*
	(8.19)		(6.37)		(0.19)		(-2.43)
γ_{1984}	9097.3***	β_{32}	7543.9***	α_{10}	146.1	$\delta_{10}D$	-5295.2*
	(8.52)		(6.54)		(0.07)		(-2.02)
γ_{1985}	9784.9***	β_{33}	9056.0***				
	(9.59)		(6.63)				
γ_{1986}	11084.7***	β_{34}	7401.7***				
	(9.93)		(5.83)				
γ_{1987}	11618.9***	β_{35}	7169.1***				
	(10.68)		(4.73)				
γ_{1988}	13684.5***	β_{36}	8385.3***				
	(11.40)		(5.44)				

Table 6 continued from previous page

Year		Age		Event time	Interaction
γ_{1989}	15610.0***	β_{37}	7720.7***		
	(12.22)		(4.75)		
γ_{1990}	15416.1***	β_{38}	7038.8***		
	(11.83)		(4.01)		
γ_{1991}	15760.8***	β_{39}	8226.5***		
	(12.41)		(4.21)		
γ_{1992}	16494.4***	β_{40}	5445.3**		
	(11.88)		(2.64)		
γ_{1993}	16264.5***	β_{41}	7282.0**		
	(10.83)		(3.09)		
γ_{1994}	16684.7***	β_{42}	8474.5**		
	(11.05)		(3.31)		
γ_{1995}	16240.9***	β_{43}	6905.5*		
	(10.49)		(2.30)		
γ_{1996}	14633.9***	β_{44}	7818.5		
	(8.31)		(1.89)		
γ_{1997}	17504.5***	β_{45}	5683.1		
	(8.86)		(1.62)		
γ_{1999}	20101.0***	β_{46}	609.0		
	(10.17)		(0.13)		
γ_{2001}	19236.9***	β_{47}	13808.1**		
	(8.29)		(2.81)		
γ_{2003}	16903.7***	β_{48}	17053.2		
	(7.29)		(1.37)		
γ_{2005}	21296.7***	β_{49}	16807.6*		
	(5.63)		(2.46)		
γ_{2007}	20453.5***	β_{50}	12669.0		
	(4.76)		(0.90)		
γ_{2009}	17894.0***	β_{51}	5825.8		
	(3.44)		(0.34)		
γ_{2011}	19312.7***	β_{52}	25211.5***		
	(3.89)		(5.21)		
γ_{2013}	22733.3***				
	(3.92)				
γ_{2015}	22064.7**				
	(3.00)				
γ_{2017}	31413.1**				
	(2.61)				
n	5273				

Notes: This table reports estimates of the model specification reported in Equation (3) for women. Clustered standard errors are reported in parentheses. An asterisk denotes a significance level of ***0.001, **0.01, and *0.05.

Table 7: Difference-in-differences regression men

Year		Age		Event time		Interaction	
γ_{1969}	0 (·)	β_{18}	0 (·)	α_{-5}	2247.1 (1.79)	$\delta_{-5}D$	178.6 (0.10)
γ_{1970}	-60.58 (-0.07)	β_{19}	173.6 (0.09)	α_{-4}	1690.7 (1.72)	$\delta_{-4}D$	-2078.3 (-1.45)
γ_{1971}	285.5 (0.39)	β_{20}	1424.4 (0.81)	α_{-3}	611.4 (0.68)	$\delta_{-3}D$	-1371.9 (-1.05)
γ_{1972}	361.4 (0.46)	β_{21}	741.6 (0.45)	α_{-2}	1124.5 (1.08)	$\delta_{-2}D$	-2084.3 (-1.46)
γ_{1973}	1439.3 (1.86)	β_{22}	2571.2 (1.51)	α_{-2}	0 (·)	$\delta_{-1}D$	0 (·)
γ_{1974}	2594.8** (3.12)	β_{23}	2947.1 (1.70)	α_0	835.4 (0.86)	δ_0D	-2810.2* (-1.97)
γ_{1975}	3257.5*** (3.89)	β_{24}	3881.6* (2.46)	α_1	-36.34 (-0.04)	δ_1D	130.7 (0.08)
γ_{1976}	3400.8*** (3.84)	β_{25}	5095.8** (2.94)	α_2	-641.2 (-0.73)	δ_2D	-976.0 (-0.64)
γ_{1977}	3853.4*** (4.40)	β_{26}	6304.5*** (3.73)	α_3	-883.5 (-0.86)	δ_3D	-1519.9 (-0.84)
γ_{1978}	5379.4*** (5.70)	β_{27}	8926.1*** (4.99)	α_4	-463.9 (-0.37)	δ_4D	-1742.3 (-0.93)
γ_{1979}	5832.4*** (5.35)	β_{28}	9463.2*** (5.23)	α_5	587.7 (0.37)	δ_5D	-4035.6 (-1.74)
γ_{1980}	7831.1*** (7.55)	β_{29}	9822.2*** (5.50)	α_6	-320.0 (-0.19)	δ_6D	-698.9 (-0.27)
γ_{1981}	8863.7*** (7.76)	β_{30}	11808.4*** (5.89)	α_7	-1182.7 (-0.58)	δ_7D	-496.8 (-0.18)
γ_{1982}	9756.5*** (8.77)	β_{31}	11723.5*** (6.28)	α_8	-1062.7 (-0.51)	δ_8D	2810.9 (0.85)
γ_{1983}	10292.1*** (9.71)	β_{32}	13915.2*** (6.79)	α_9	600.8 (0.20)	δ_9D	408.3 (0.11)
γ_{1984}	11297.5*** (9.77)	β_{33}	13728.1*** (6.74)	α_{10}	-301.4 (-0.11)	$\delta_{10}D$	652.8 (0.17)
γ_{1985}	12521.5*** (8.86)	β_{34}	15183.0*** (6.96)				
γ_{1986}	12849.7*** (9.74)	β_{35}	15152.3*** (6.85)				
γ_{1987}	13216.4*** (10.28)	β_{36}	19990.2*** (8.20)				
γ_{1988}	14805.6*** (9.70)	β_{37}	16681.2*** (7.20)				
γ_{1989}	15882.2*** (11.70)	β_{38}	18298.5*** (7.07)				

Table 7 continued from previous page

Year		Age		Event time	Interaction
γ_{1990}	16965.0***	β_{39}	18013.8***		
	(11.62)		(6.76)		
γ_{1991}	18143.4***	β_{40}	21776.9***		
	(11.46)		(7.45)		
γ_{1992}	20250.4***	β_{41}	22466.7***		
	(12.45)		(6.96)		
γ_{1993}	18264.4***	β_{42}	25265.5***		
	(11.43)		(6.81)		
γ_{1994}	19136.6***	β_{43}	28266.8***		
	(11.69)		(6.84)		
γ_{1995}	20368.6***	β_{44}	23514.3***		
	(11.75)		(6.06)		
γ_{1996}	22747.0***	β_{45}	31226.1***		
	(11.93)		(5.74)		
γ_{1997}	26594.0***	β_{46}	24846.0***		
	(12.15)		(4.98)		
γ_{1999}	28437.0***	β_{47}	24845.7***		
	(11.17)		(4.06)		
γ_{2001}	29653.9***	β_{48}	19463.0**		
	(9.60)		(2.72)		
γ_{2003}	34214.1***	β_{49}	19003.0*		
	(10.02)		(2.21)		
γ_{2005}	35330.2***	β_{50}	17532.4**		
	(7.76)		(2.60)		
γ_{2007}	37290.8***	β_{51}	14188.6		
	(9.45)		(1.59)		
γ_{2009}	34062.4***	β_{52}	18482.9**		
	(7.25)		(2.72)		
γ_{2011}	43384.6***	β_{53}	2316.7		
	(7.31)		(0.17)		
γ_{2013}	49771.8***	β_{54}	7813.2		
	(5.72)		(0.60)		
γ_{2015}	39612.9***	β_{55}	-9092.5		
	(4.75)		(-1.82)		
γ_{2017}	50340.5***				
	(5.54)				
n	7521				

Notes: See the notes to Table 6, but applied to men.