Dissertation

Essays on Labor Economics and Inequality

Lukas Riedel

July 2023

Faculty of Economics and Social Sciences

University of Heidelberg

Supervised by Prof. Christina Gathmann, Ph.D.

Co-supervised by Prof. Dr. Sebastian Siegloch

Date of disputation: February 9, 2024

I thank Christina Gathmann, Kristina Hengen, Valentina Melentyeva, Ulla Mosmann-Riedel and Peter Riedel, Sebastian Siegloch, Michaela Slotwinski and Holger Stichnoth.

Contents

1 Introduction

In this dissertation, I collect four articles from the fields of labor economics and inequality and their intersection. As both fields are wide, I cover a subsection of topics that ranges from the measurement of gender inequality in the labor market over the evaluation of policies that address it to the broader question of assessing post-tax income inequality. Specifically, the first two chapters focus on female labor supply and gender inequality in the context of childbirth, the subsequent two move to inequality among women and general inequality. The dissertation evaluates public policies and methods to analyze them as well as it proposes methodological improvements.

Chapters 2 and 3 focus on childbirth which is both a major event in life but also an important origin of gender inequality. As still today, it is the female partner in a couple who takes on most of the burden of raising children, childbirth is often followed by prolonged labor market absence and subsequently reduced labor supply of women, though not men. The consequences are not transient. With—despite all achievements during the past decades—quite persistent gender roles, women often face differential treatment in the labor market even before they give birth as well as they select differently into occupations compared to men (Bertrand 2020). In the long term after birth, they have accumulated lower total earnings and pension claims. Hence, it is vital to understand the precise extents and roots of the career impacts of childbirth to provide meaningful advice to policymakers but also to individuals to make informed decisions.

Chapter 2 provides new insights on parental leave as one of the major family policy measures, both in Germany as well as in other industrialized countries. Parental leave typically consists of two policy instruments, job protected leave and parental benefit payments. Sebastian Findeisen, Jörg Heining, Sebastian Siegloch and I analyze the effects of extending parental leave on mothers' long-run labor market outcomes by exploiting variation provided by multiple reforms in Germany. Previous work by Schönberg and Ludsteck (2014) finds on average rather small effects of these policy changes. Adding to this, we empirically disentangle the causal effects of the two policy instruments. Holding constant the length of mothers' labor market breaks after the first childbirth, we show that parental leave significantly reduces the earnings losses of mothers who gave birth under the more generous parental leave policy and attribute this positive effect solely to job protection. Parental benefit payments, in comparison, have no substantial impact on mothers' labor market outcomes. Job protection does not only have a direct effect by enhancing employer continuity, its effect rather extends to mothers who switch the employer when they return to the labor market. Our results highlight that, even though career breaks come with substantial costs for women, job protection proves to be an effective measure to mitigate at least some of these costs while it also allows mothers to spend more time with their children.

Chapter 3 shifts the focus from analyzing a specific policy to the question how to analyze career trajectories after childbirth. Valentina Melentyeva and I study the earnings losses women incur after childbirth from a methodological perspective and provide new results on their extent. Quantifying and visualizing these losses using event studies to estimate so-called "child penalties" quickly became popular after the seminal contribution by Kleven, Landais, and Søgaard (2019). We build on the insights from the recent literature on static and dynamic difference-in-differences models following Goodman-Bacon (2021) to highlight that conventional event studies are prone to yield biased estimates, since the characteristics of women and treatment effects of childbirth are heterogeneous by the age at which mothers give their first birth. We address this issue by developing a new approach to estimate child penalties that takes the specifics of the setting into account and estimates the unbiased and causal effects of childbirth on mothers' labor market outcomes. Applying this approach, we show that there is substantial heterogeneity in child penalties by age at first childbirth. We, further, point out that the conventional approach yields a considerable underestimation of the average penalty. While we use our novel approach to estimate the average effect of having a child on earnings, it can also be extended to evaluate specific policy changes to asses their precise effects on women's careers.

Chapters 4 and 5 expand the focus from inequality that arises from the specific event of childbirth to, more generally, wage and income inequality. Chapter 4 retains the emphasis on women but widens the perspective from childbirth to an analysis of the impact of the public provision of childcare on wage inequality. While analyses of wage inequality among men and how it is related to labor market institutions such as unions or minimum wages are common (see, for instance, Dustmann, Ludsteck, and Schönberg 2009), the female part of the workforce often receives less attention. To some extent, this is driven by the differential labor supply of women due to the consequences of childbirth. However, this also suggests that family policies which address women's labor supply play the role of labor market institutions for women. I, therefore, analyze the relationship between a large-scale expansion of public care for children of kindergarten age and wage inequality among women as well as gender wage inequality in the German labor market. Exploiting the regional variation in how strongly childcare was expanded, I show that in regions with stronger increases in childcare wage inequality among women increased less strongly compared to regions with smaller increases. This is primarily driven by the lower half of the wage distribution and qualitatively similar for both full- and part-time workers. Larger expansions in childcare, however, do not contribute to a further reduction of the gender wage gap, suggesting they are associated with a more negative selection of women into work. While the labor supply effect of this reform has already been shown on the micro-level by Bauernschuster and Schlotter (2015), I extend the analysis of this childcare reform to show effects on the distribution of wages. My findings demonstrate that a policy change that benefits women (as more of them are in employment) and lower wage inequality among them does not necessarily lower gender wage inequality.

Finally, Chapter 5 widens the perspective again, from labor earnings to total post-tax income and from focusing on women to the entire population. Holger Stichnoth and I re-assess one of the key assumptions in the recent inequality literature on *Distributional National Accounts* (Dina, Piketty, Saez, and Zucman 2018). Dina aim to construct inequality statistics that are consistent with the national accounts. This requires to allocate the entirety of national income to individuals, which is a mostly straightforward task for cash income, however, in-kind income from collective expenditure (for instance education, defense or infrastructure) is typically not observable at the individual-level. The Dina literature, therefore, uses the assumption that collective expenditure is distributed proportionally to cash income among individuals. This has the disadvantage of ex-ante ruling out redistribution through collective expenditure. We show how public expenditure on education, as an important part of collective expenditure, is actually distributed. We find that the actual distribution of education expenditure clearly rejects assuming a proportional distribution, but rather favors an allocation as lump-sums. The consequence of our results is that the actual levels of post-tax inequality are lower than the current estimates in the Dina literature. We demonstrate a straightforward way to extend existing Dina analyses to provide more accurate representations of post-tax inequality.

The dissertation makes contributions beyond the topics discussed in the single chapters. It emphasizes the importance of public policies in shaping individual behavior and economic outcomes and it demonstrates the crucial role of correct and comprehensive measurement in studying such policies.

Chapter 3, where we reassess the estimation of child penalties, concludes that the earnings losses of women after childbirth are substantially larger than conventional methods have estimated so far. These losses are predominantly perceived as undesirable by the general public and lead to several other long-term costs and economic inefficiencies. Our findings, thus, provide additional reasoning for policy interventions. Chapters 2 and 4 then study parental leave and pubic childcare provision, i.e. two of the major polices that shape if and when women re-enter the labor market after childbirth and how successful they can purse their careers afterwards. For parental leave, we show the effectiveness of the policy and can specify which set of mothers benefits from it. Additionally, we highlight that the effectiveness of parental leave is solely due to its job protection pillar. Job protection is a policy instrument that more recent reforms in Germany have not addressed and that other countries offer to comparatively small extents, which stresses the relevance of our findings. As existing research has found, public childcare is a similarly effective policy. I add to this by showing that the public provision of childcare has effects beyond directly changing labor supply decisions. It alters when and which women select into different kinds of work such that childcare also leads to changes in wage inequality. My results underline further that a finding as the absence of a contribution of childcare to a lower gender wage gap, while it at first might seem counterintuitive and as an unwanted outcome, can stem from a successful policy. Since childcare changed the selection of women into work, some of those with otherwise zero earnings due to nonparticipation started to contribute to the gender gap as well. In a cross-sectional view, this reduces potential selection biases through selective participation such that gender inequality can decrease less than expected. In a life-cycle perspective, on the other hand, additional work by women is clearly beneficial. The chapter on Distributional National Accounts studies the inequality consequences of the public provision of education. Similarly to childcare, education is a public service that is provided in-kind rather than in-cash. Contrary to the analysis of childcare, in Chapter 5 we do not focus on indirect inequality effects through behavioral changes but on the distributional effect of the monetary value of education. This chapter highlights that the redistributive role of taxes comes not only from collecting and using them for cash transfers but also acts through the less visible, yet sizeable, channel of providing essential services which public debates tend to neglect.

With the chapter on the comprehensive measurement of redistribution via collective expenditure and the chapter on estimating child penalties, the dissertation, further, makes methodological contributions. Despite the fact that they study very settings, both chapters revisit methods that are already popular and regularly applied. We show their limitations and provide improvements that are straightforward to implement in most settings. Both chapters provide tools to more precisely measure and thus to more profoundly understand various dimensions of inequality. We, therefore, hope to also contribute to better informed and more relevant policy advice in the future.

There is a further point and at the same time a personal observation that deserves to be mentioned. All chapters build on preexisting work. That is, they either analyze policies and policy changes that have been subject to prior research or they revisit existing methods. Nevertheless, they extend on their respective origins and provide new insights on the effects of public policies and their assessment. At the start of this dissertation, this was not planned and I doubt that it could have been planned in this exact way. It is, in fact, the result of various research and policy consulting projects that I worked on and conversations I had over the past years and that encouraged further engagement with their respective subjects. Following up on certain aspects that developed from seemingly minor issues into problems that demanded increased attention helped enabling this dissertation. This point is worth stressing as it provides a good description of the process of doing research. This process is often characterized by incremental, at times unpredictable steps which, however, always contribute to an accumulation of knowledge on that further research can build. While the process itself might sometimes be slow and discouraging, this outcome is, after all, every time rewarding.

References

Bauernschuster, Stefan and Martin Schlotter (2015). "Public child care and mothers' labor supply – Evidence from two quasi-experiments". In: *Journal of Public Economics* 123, pp. 1–16.

Bertrand, Marianne (2020). "Gender in the Twenty-First Century". In: *American Economic Review: Papers & Proceedings* 110, pp. 1–24.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). "Revisiting the German Wage Structure". In: *The Quarterly Journal of Economics* 124 (2), pp. 843–881.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225 (2), pp. 254–277.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). "Children and Gender Inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11 (4), pp. 181–209.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2018). "Distributional National Accounts: Methods and Estimates for the United States". In: *Quarterly Journal of Economics* 133 (2), pp. 553–609.

Schönberg, Uta and Johannes Ludsteck (2014). "Expansions in Maternity Leave Coverage and Mother's Labor Market Outcomes after Childbirth". In: *Journal of Labor Economics* 32 (3), pp. 469–505.

2 Unpacking Parental Leave: The Role of Job Protection

Joint with Sebastian Findeisen (University of Constance), Jörg Heining (IAB – Institute for Employment Research, Nuremberg) and Sebastian Siegloch (University of Cologne)

Parental leave is one of the most important policies that shape the post-birth careers of women. We exploit a sequence of parental leave reforms in Germany that extended both the job protection period and the duration of parental leave benefits to different extents to study the effects of the two policy instruments parental leave consists of. Using administrative social security data, we first replicate the stylized facts that mothers respond to extensions in parental leave and that the average effect of longer leave-taking on their careers is negative. In a second step, we analyze the causal effects of job protection and parental benefit payments. Holding constant the length of mothers' post-birth labor market break, we show that extending job protection significantly reduces losses in long-run earnings while extensions of the benefit duration have no measurable impact. The positive effect of employment protection works both by enhancing employer continuity as well as by improving outside opportunities for mothers who change their employer.

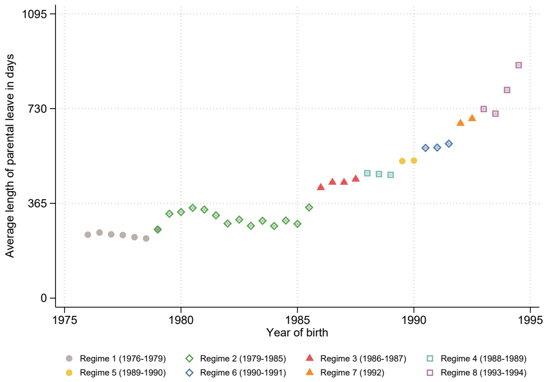
2.1 Introduction

It is a stylized fact that having children has large and long-lasting negative labor market consequences for women. Mothers across the world experience a persistent decline in earnings after childbirth that stem from both interruptions of labor market participation as well as from reductions in intensive margin of labor supply or from working in different, lower-paying jobs. This pattern is often coined as the child penalty (Kleven, Landais, and Søgaard 2019; Kleven, Landais, Posch, et al. 2019) and well-documented by a large body of literature (Angelov, Johansson, and Lindahl 2016; Lundborg, Plug, and Rassmussen 2017; Adda, Dustmann, and Stevens 2017). The losses in current earnings after childbirth have important implications as they translate to multiple disadvantages. For instance, relying on the partner's income during labor market breaks and thus forgoing to accumulate human capital makes it harder for women to change their labor supply in case of a divorce. In the long term, lower earnings further translate to lower contributions to the pension system which increases the risks of old-age poverty. These negative earnings effects cumulate to childbirth being one of the main drivers of the still persisting and sizeable gender earnings gap (Olivetti and Petrongolo 2016; Blau and Kahn 2017 and Bertrand 2020 provide overviews). Most economists and policymakers agree that these striking patterns call for policy interventions that reduce earnings losses after childbirth and thus further close the gender gap.

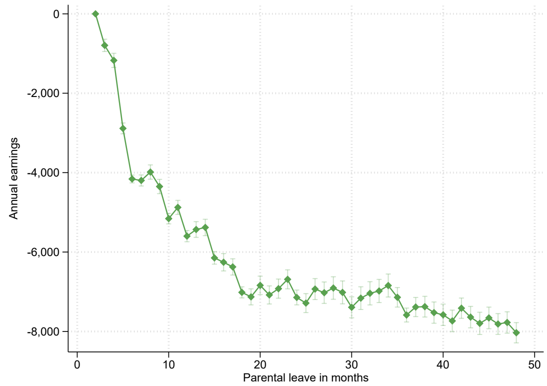
In this paper, we reassess the role of parental leave policies in shaping the labor market careers of women. Parental leave is a common policy in many industrialized countries that aims to balance the wish of mothers to spend time with their child while facilitating re-entry to the labor market. To this end, parental leave typically grants a combination of two policy instruments. Benefit payments that offset some of the income loss during the post-birth labor market break and a period of job protection during which mothers have the right to return to the pre-birth employer in their pre-birth job. Theoretically, the effects such a policy are ambiguous. On the one hand, job protection can help mothers to continue their career, for instance by retaining firm-specific human capital. On the other hand, prolonged absence from the labor market, promoted by benefit payments, can lead to depreciation of human capital or can be viewed as a signal of low productivity. Most existing empirical evidence on extending parental leave (Schönberg and Ludsteck 2014; Kleven, Landais, Posch, et al. 2021) points towards relatively small aggregate effects in the long-run.

To shed new light on parental leave policies, we empirically disentangle the career impacts of job protection and benefit payments. We exploit a sequence of seven parental leave reforms in Germany, taking place between the late 1970s and the mid 1990s. The reforms considerably extended either the job protection or the benefit payment period or both. Before 1979, mothers were eligible to 2 months of parental leave (with job protection and benefit payments). Until 1993, the job protection period was expanded to 36 months, benefit were paid for up to 24 months. In our analysis, we focus on reforms that either increased job protection alone or the benefit duration alone to isolate the causal effects of each policy instrument.

We use administrative social security data provided by the IAB in Germany from which we draw the universe of mothers. We exploit that parental leave reforms affected mothers based on the birth date of their children, which is arguably difficult to manipulate systematically. For each of the reforms, we focus on the two adjacent parental leave regimes around the reform. We compare mothers that gave their first birth in the old regime to mothers that gave the first birth in the new—more generous—parental leave regime. Holding constant the duration of parental leave, we then estimate the causal effect of the policy. We provide evidence that there is neither selection into parental leave regimes nor that selection of mothers into parental leave durations changes with a reform, i.e. that mothers' characteristics are smooth in parental leave taking.



(A) Length of the post-birth labor market break by parental leave regime.



(B) Relationship between length of parental leave taking and annual earnings seven years after birth.

FIGURE 2.1: Length of parental leave taking and earnings in the long run.

Notes: The left figure documents the increase in parental leave taking with each extension job protection and benefit payments. The right figure shows that with longer leave taking annual earnings decrease. Both figures show regression results of the dependent variable on birthday of the first child in half-year steps (left figure) and months of parental leave taking respectively (right figure) along with 95 percent confidence intervals. *Source:* Own estimations using the IEB data described in section 2.3.

We start with descriptive results. In Figure 2.1a we replicate the existing finding from the literature that the length of the post-birth labor market break responds to policy changes. We plot how the time mothers spend on leave after their first child evolves between the mid 1970s and the mid 1990s separating by parental leave regime. The figure shows that leave duration increases in discrete jumps, precisely at the implementation of policy changes. We also document the negative and long-lasting effect of parental leave on maternal earnings. Figure 2.1b plots the relationship between the earnings of mothers seven years after they gave birth to their first child and the time the spent on parental leave in months. It shows a negative and nonlinear relationship between earnings and the length of mothers' labor market breaks. Each additional month of parental leave is associated with lower earnings in the long run.

To reconcile the descriptive finding that mothers take prolonged labor market breaks even though they appear to come at high costs, we then turn to estimating the causal effects of extending parental leave. Holding constant the duration of parental leave, we show that women under the new, more generous policy regime experience significantly smaller earnings losses. The magnitude of this effect is substantial. Seven years after childbirth it amounts to almost 6 percent of annual earnings or to almost 10 percent of cumulative earnings. Strikingly, such policy effects are only observable for mothers who return after a parental leave period that is longer than the old regime's coverage period but shorter than the new regime's coverage period. The development of earnings for mothers who return after a parental leave period that is shorter than the old regime's coverage is virtually identical across regimes. This finding, first, suggests that selection of mothers does not differ across two adjacent policy regimes, largely ruling out selection as a driver of our results. Second, this pattern explains why previous studies that analyze the impact of parental leave reforms using different identification strategies tend to find rather small average effects. Utilizing a reform in 1992 that doubled the job protection period from 18 to 36 months while keeping the duration of benefit payments constant at 18 months allows us to attribute the positive effect of parental solely to job protection. For two reforms that extended only the duration of benefits, we find no effect of longer parental leave on earnings. Further, we provide detailed results on how job protection is able to reduce post-birth earnings losses. We document that mothers who return under job protection—again holding leave duration constant—have a higher probability to return to their pre-birth employer as well as to work for employers that pay higher average wages and have a higher establishment fixed effects (Card, Heining, and Kline 2013; Bellmann et al. 2020). In addition, we show the impact of job protection is not limited to enhancing employer continuity. Differentiating between mothers who stayed with their pre-birth employer and those who switched to a different employer, we find relative gains in annual and cumulative earnings, average firm wage and establishment fixed effect for mothers who return under job protection in both groups. This suggests that job protection, beyond enabling mothers to continue existing employment relationships, improves the outside options of mothers who change the employer thus leading to better matches and improved labor market outcomes.

Our study is related to a large literature evaluating the aggregate effects of parental leave reforms. Schönberg and Ludsteck (2014) provide the seminal analysis for the German context using those reforms that we use as identifying variation as well. Their main finding is that while extending parental does reduce short-run employment rates, effects on employment and earnings are small in the long-run. They also find evidence for employer continuity, though not quantify its impact. Lalive, Schlosser, et al. (2014) analyze medium-run impacts of parental leave reforms in Austria finding no negative effects as well. They further provide a simulation-based assessment of the importance of job protection and benefit payments that favors a combination of both instruments in terms of care-giving after birth and medium-run employment rates. We confirm the empirical findings of both papers and take them as a starting point to answer the question why substantially longer absence from the labor market does not do greater harm to mothers' long-run post-birth labor market outcomes. Our paper is the first to empirically disentangle the separate impacts of job protection and benefit payments on maternal earnings in the long-run. To this end, our analysis focuses on mothers with some minimum labor market attachment (i.e. who return after at most five years on leave) and, further, on reforms that extend the duration of either job protection or of benefit payments such that we compare mothers who are eligible to only one policy instrument while otherwise being similar. This allows us to conclude that small aggregate effects of parental leave extension on earnings are the result of job protection offsetting some of the negative consequences of longer leave taking.

Further evidence on how parental leave affects time away from work, labor supply, earnings, fertility and inequality is provided by Ruhm (1998), Baker and Milligan (2008), Lalive and Zweimüller (2009), Bailey et al. (2019), Bana, Bedard, and Rossin-Slater (2020), Ginja, Jans, and Karimi (2020), and Kleven, Landais, Posch, et al. (2021). Olivetti and Petrongolo (2017) and Rossin-Slater (2017) provide surveys of the existing literature. The general consensus from most studies is that longer parental leave eligibility increases the time mothers spend away from the labor market while the long-run consequences for labor supply and earnings are sometimes negative but most often small or zero. Dahl et al. (2016) focus on an expansion of benefit payments. Apart from increasing leave taking they do not find them to have significant effect on maternal labor market outcomes, children or marriage but rather adverse redistributive effects.

There is a smaller but growing literature that focuses on the reactions of the employer side as parental leave policies that lead to increased absence of mothers from work have potentially negative effects on firms' outcomes. Results vary from little additional costs for firms (Brenøe et al. 2020) to more substantial ones that materialize either directly (Ginja, Karimi, and Xiao 2023) or through changes in the firms hiring or promotion policies (Thomas 2021; Huebener et al. 2021). Instead of directly looking at employers, we provide insights on how mothers re-allocate to firms when returning to the labor market. We extend on the finding that job protection enhances employer continuity by showing that both mothers who stay with their pre-birth employer and those who switch the firm gain from job protection.

The remainder of this paper is organized as follows. Section 2.2 summarizes the institutional setting of the German parental leave system and the corresponding reforms. Section 2.3 describes our administrative dataset. In Section 2.4, we present our empirical approach and discuss the causal identification of the effects of parental leave extension and of single policy instruments. Section 2.5 presents the results along with robustness checks. Section 2.6 concludes.

2.2 Institutional Background

The German parental leave legislation consists of two policy instruments. It, first, grants job protection, i.e. gives mothers the possibility to have a labor market break after childbirth with the right to return to the pre-birth employer in a similar job as before childbirth for a prespecified amount of time. Second, mothers are payed monthly parental benefit payments during their break. Parental leave is mainly organized country-wide by the federal government, but additionally, following Germany's federal structure, allows for extra benefit payments from the single states. In this paper, we exploit variation in both the federal legislation as well as in additions from two of the largest states to distinguish the effects of job protection and benefit payments.

Table 2.1: Length (in months) of job protection and benefit duration periods by parental leave regime (federal level).

Regime	Births since	Job Protection	Benefit Duration
1	1952	2	2
2	May 1, 1979	6	6
3	January 1, 1986	10	10
4	January 1, 1988	12	12
5	July 1, 1989	15	15
6	July 1, 1990	18	18
7	January 1, 1992	36	18
8	January 1, 1993	36	24
9	January 1, 2001	36	12/24
10	January 1, 2007	36	12+2

Notes: Representation based on the relevant laws at the federal level (*Mutterschutzgesetz* and *Bundeserziehungsgeldgesetz*). The benefit payments amounted to DM 750 between 1979 and 1985 and DM 600 from 1986 to 2000. The *Mutterschutzgesetz* 1952 does not specify a birth cutoff; it came into effect on February 6, i.e. applying at least to all those who had been pregnant at that time. Starting in 2001, parents were able to choose between 12 and 24 months of benefit payments. Since 2007, up to 14 months of benefits are granted if both parents take parental leave with one parent taking at least 2 months.

Federal Parental Leave Legislation Table 2.1 gives an overview of the different parental leave regimes at the federal level. Both the duration of job protection and benefit payments indicate maxima. Except for the period of two months after birth, mothers are free to return to work earlier. Benefits are only payed while on leave.

Time off work after childbirth was first introduced in 1952 as maternity protection (*Mutterschutzgesetz, MuSchG*) focusing on only a relatively short period after birth. Initially, it mandated two months of both job protection and benefit payments after childbirth during which mothers are not allowed to work. Mandated maternity protection is in place and unchanged until today which allows us to use it to identify mothers in the IEB data (see Section 2.3). The reform in 1979 introduced voluntary parental leave after the mandated protection period. It extended the duration of job protection and benefit payments to six months. Benefits amounted to DM 750 per month. In 1986, legislative focus shifted from only protecting mothers after birth to giving parents—i.e. from then onward, mothers and fathers—time to care for their newborns (*Gesetz zum Erziehungsgeld und zur Elternzeit, BErzGG*). At the same time the duration of job protection and benefits was increased to ten months. The benefit level was lowered to DM 600 where it remained until 2000. Subsequently, parental leave was extended to 12 months (1988), 15 months (1989) and 18 months (regime 6 in 1990). Our main identifying variation comes from the reform that marked the shift from parental leave regime 6 to regime 7 in 1992. This was the first time when only the job protection period was extended. Its length was doubled to 36 months while the duration of benefit payments remained at 18 months. Thus, this reform allows to observe the isolated effect of job protection for mothers who return to the labor market after between 19 and 36 months of leave.

In 1993 the benefit duration was extended to 24 months and in 2001 two forms of benefit payments were introduced. Parents could choose between receiving DM 900 per month for 12 months or DM 600 for 24 months. Starting in 2007, benefits amounts were calculated as a percentage of pre-birth income and their duration was changed to either 12 (if only one parent takes parental leave) or 14 months (if both parental go on leave and one parent takes at least two months).[1] Since they offer the clearest identification, we focus in this paper on parental leave regimes 2 to 7.

Additional Benefit Payments by (Some) Federal States To be able to study the isolated effect of benefit payments for mothers who return to the labor market while they are not under job protection but receive benefits, we use additional identifying variation at the level of federal states. Several federal states introduced additional benefit payments (*Landeserziehungsgeld*) that are designed similarly to the nationwide benefits. They grant monthly payments continuing the federal benefits without a break, but do not offer additional job protection. We use the introductions of these benefits in the second and third largest federal states of Germany, Bavaria and Baden-Württemberg.[2] In 1986, along with the reform switching to the federal parental leave regime 3, Baden-Württemberg started to pay *Landeserziehungsgeld* for additional 12 months after the federal payment (*Richtlinie für die Gewährung von Landeserziehungsgeld, RL-LErzG*). Thus, mothers in Baden-Württemberg could receive parental benefits for 22 instead of only 10 months. Three years later, in 1989, Bavaria introduced a similar payment for 6 months (*Bayerisches Landeserziehungsgeld, BayLErzGG*). Its duration was extended to 12 months in 1993.[3] In terms of amount the *Landeserziehungsgeld* was lower than the federal benefit (DM 400 in Baden-Württemberg and DM 500 in Bavaria) but in a similar range.

1 While the lump-sum benefit payments favored lower earning mothers, the benefit scheme introduced with the 2007 reform is more beneficial for higher earning women (see Raute 2019, who assesses the fertility effects of the reform). Aggregate labor market effects of this policy change are analyzed by Kluve and Schmitz (2018) and Bergemann and Riphahn (2023).

2 In the early 1990s, three East German states (Mecklenburg-Vorpommern, Saxony and Thuringia) introduced similar benefits. As we focus on West Germany we do not consider them here.

3 See also Table 2.7 in the Appendix.

2.3 Data: The Integrated Employment Biographies

We use data drawn from the *Integrated Employment Biographies* (IEB) provided by the *Institute for Employment Research* (IAB). The IEB consist of detailed spell data of the universe of employees subject to social security contributions in Germany. Starting in 1975, they include daily-level information on work, wages and benefit receipt spells that allows to trace an individual's entire employment biography. Further, they record wages and personal characteristics as education and occupation. In terms of content our data are comparable to the more widely used SIAB data (Frodermann et al. 2021) which are a two percent draw from the IEB and thus are much smaller.

There are two general shortcomings of the IEB. First, since they are constructed from social security records, they do not include individuals who are self-employed[4] or civil servants[5]. The self-employed are not a primary target group for our analyses since job protection regulations do not apply to them. Civil servants on the other hand are eligible to parental leave, however, given the generally more secure nature of their employment relationships, they likely face different constraints compared to regular employees when making decisions on work around childbirth. Further, civil servants represent a relatively small group within the German workforce of almost 39 m. in 1991.

Second, wages are censored at the social security contribution threshold compressing the observed distribution of wages. Card, Heining, and Kline (2013) measure 1 to 3 percent censored wages for women per year. Looking at our main specification we measure shares of 4 respectively 5 percent of censored wages in the last pre-birth spell (i.e. in a group that is positively selected compared to all observations of mothers that include the post-birth period), while seven years after childbirth when we measure outcomes in our main specification, fewer than 1 percent of mothers are recorded with a censored wage. Given that censoring is not especially prevalent among women we consider this issue in our case to be a minor one.

As their main purpose is to record information that is relevant to calculate social security benefits such as unemployment insurance payments, the IEB do not include information on childbirth. It is nevertheless possible to infer births from the data using some assumption. The first approach that gained attention was proposed by Schönberg (2009) who uses a certain type of employment interruption. Due to changes in the data, this method can no longer be applied. We therefore rely on the approach by Müller, Filser, and Frodermann (2022). For the group of mothers we focus on—those working prior to childbirth—they use de-registrations from employment due to the receipt of wage replacement payments from the statutory health

[4] 3.6 m. or 4.5 percent of the population in 1991 (`https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Tabellen/liste-bevoelkerung-erwerbstaetigkeit.html`).

[5] 1.8 m. in 1991 of whom less than one third were women (`https://www.destatis.de/DE/Themen/Staat/Oeffentlicher-Dienst/Tabellen/beschaeftigte-geschlecht.html`).

insurance that includes maternity allowances (*Mutterschaftsgeld*). These allowances are payed for the time of maternity protection during which mothers are not allowed to work. Typically, this period covers at least the six weeks before and the eight weeks after birth which allows to treat those women as mothers who are absent from the labor market for at least 98 days. Since wage replacement payments are recorded as well for those in long-term illness, the additional restriction only to treat de-registrations that appear until the age of 38 as indicating childbirth is imposed. Müller, Filser, and Frodermann provide a comparison of the number of mothers they identify with the official birth statistics which shows that for the years 1990 to 1992 around 60 percent of the total number of childbirths are identified. Note, that this share is likely to be larger for our target group. Müller, Filser, and Frodermann compare with the total number of births, while for us only births by women in employment that is subject to social security contributions are relevant. Further, we specifically analyze first births. For these, the identification strategy performs considerably better than for subsequent births. The latter can only be identified if mothers re-entered the labor market after their first birth.

For some of our analyses we merge characteristics of mothers' employers from the IAB's *Establishment History Panel* (Bhp). As for the Ieb, our version the of the Bhp includes information of the universe of German establishments that employ at least one employee at the cutoff date June 30. It has to be noted that this employer data refers specifically to establishments, not firms, since data collection happens at the place, not the legal entity, of work. Firms with multiple sites cannot be linked. While this might be a shortcoming when investigating questions that are centered around firms, it has little impact on our analysis as we focus primarily on the mothers' side.

We combine the data from the Ieb and the Bhp to a panel data set that records all relevant information for five years prior to the first childbirth until up to 15 years after the first birth. Using the daily-level precision of our data, we measure years relative to the first birth always as full years, not calendar years. Therefore, in year seven after childbirth which we use in our main specification, we observe the outcomes of mothers exactly seven years after birth, regardless of their month of childbirth. This ensures that results are not biased if birthdays are not uniformly distributed across the year such that for some, one would mechanically measure larger cumulative earnings in a panel in calendar time.

For each year, we observe between around 150,000 and 250,000 births. In our estimations we restrict the sample to mothers who give their first birth and show sufficient labor market attachment, i.e. we only consider those who eventually return to work after a break of at most five years. In our main specification, we focus an parental leave regimes 6 and 7. Table 2.2 gives an overview of the characteristics of mothers who gave birth under these regimes. Mothers who gave birth under regime 6 are on average 26.8 years old and have a median parental leave time of 18 months. Prior to birth, 7 percent of them owns a college degree (75 percent a vocational

Table 2.2: Summary statistics for mothers with their first child born under parental leave regimes 6 and 7.

	Regime 6				Regime 7			
	Mean	SD	p25	p75	Mean	SD	p25	p75
Year of first childbirth								
Age	26.79	4.43	24	30	27.08	4.52	24	30
Education								
Share no vocat. degree	0.17	0.38	0	0	0.17	0.38	0	0
Share vocat. degree	0.75	0.43	1	1	0.75	0.43	0	1
Share university degree	0.07	0.25	0	0	0.07	0.26	0	0
Average annual earnings	23,979	11,235	16,115	30,859	24,906	11,511	16,834	32,052
Work experience (years)	5.92	4.01	1.39	6.33	5.89	4.11	2.54	8.44
Seven years after birth								
Parental leave duration	16.94	11.17	7	21	19.80	13.45	7	33
Annual earnings	19,731	13,134	10,674	27,179	19,027	13,887	7,972	27,055
Cumulative earnings	105,118	74,257	50,436	141,663	97,702	74,701	43,199	133,106
Share in different firm	0.33	0.47	0	1	0.32	0.47	0	1
Share in part-time work	0.54	0.50	0	1	0.58	0.49	0	1
Share censored wage	0.009	0.092	0	0	0.009	0.095	0	0
Observations	114,840				83,284			

Notes: Summary statistics for mothers who gave birth under parental leave regimes 6 and 7. The upper panel collects statistics for the year of first childbirth, the lower panel for year seven after childbirth. Average annual earnings are calculated over the three years before the first childbirth. Cumulative earnings are collected for the time since childbirth. All earnings variables are given in Euro, inflation adjusted to 2015 as base year. Parental leave duration is measured in months. *Source:* Own calculations using the IEB data described in this section.

degree, the remaining ones do not have a professional degree); a majority of 83 percent works in occupations where they perform skilled tasks, whereas 11 percent carry out more complex tasks, the remainder works in unskilled occupations. They have on average 5.9 years of labor market experience. Mothers in regime 7 are in terms of their average age, education and experience almost identical. Across parental leave regimes 6 and 7 the median time spent on parental leave (18 months) is constant, though the mean increases from 17 to 19 months. Despite having higher average earnings in the three years prior to childbirth (Euro 24.9 K vs. Euro 23.9 K), mothers from regime 7 have somewhat lower annual earnings seven years after they gave birth (Euro 19 K vs. Euro 19.7 K) and as well have lower cumulative earnings over the seven year period after birth (Euro 98 K vs. Euro 105 K). This is consistent with the negative relationship between longer leave taking and earnings we document in Figure 2.1b. It further is consistent with the finding, that mothers from regime 7 have a by 4 percentage points larger probability to work in part-time after birth.

2.4 Empirical Approach

In this paper we aim to go beyond assessing the aggregate effect of a parental leave extension. We, first, differentiate between the causal impacts of the two main policy instruments parental leave consists of: job protection and benefit payments. Second, we focus on the job protection pillar of parental leave and explore how it impacts multiple dimensions of maternal labor market outcomes, especially the selection of mothers to employers. This section explains how we approach these questions empirically.

2.4.1 The Impact of Parental Leave on Mothers' Careers

In most countries that implement them, parental leave policies are composed of job protection and parental benefit payments, both of varying length (Rossin-Slater 2017). The overarching political objective is that mothers are enabled to enjoy time with their newborns while being able to return to their previous job and have some replacement for their foregone earnings. Both policy instruments give different incentives to mothers.

While spending more time with the child is often perceived in a positive way, the career consequences are generally negative. As Figure 2.1b documents, mothers incur sizeable and long lasting earnings losses from increasing their time away from the labor market after the birth of the first child. These losses can be the result of a number of factors. First, during their time at home mothers' human capital depreciates; for those who return to a different employer, firm-specific human capital is lost; similarly, occupation-specific human capital can only be retained if a mother continues to work in her pre-birth occupation. Second, employers can perceive the choice of a longer parental leave period as a indicator that a mother is less willing to put effort into her career. As a consequence, such mothers might be less likely to be considered for promotions. Third, as losses for those who choose to re-enter the labor market increase in the length of the post-birth break, the relative gains from re-entering decrease such that re-entering is discouraged which adds further costs.

A policy that only grants benefit payments incentivizes mothers to take prolonged labor market breaks or it discourages commitment to work (Blau and Kahn 2013). The strength of such incentives depends on the benefit amount and duration. Mothers' awareness that longer absence from the labor market can have negative effects in the longer run introduces further variation in the strength of the incentive. Contrary to job protection, benefit payments do not offer a mechanism that could offset negative consequences of prolonged leave. On average, we therefore do not expect positive effects of extending parental benefit payments on subsequent labor market outcomes (in line with the findings by Dahl et al. 2016).

Job protection, on the other hand, incentivizes longer labor market breaks as well, but at the same time eases re-entering and thus offers a mechanism that—at least partly—can help to compensate for the costs of longer work interruptions. While costs of depreciation of general human capital still apply to mothers who return under job protection, the right to return to the pre-birth employer in a similar occupation helps retaining firm- or occupation-specific human capital, thus reduces the costs of pausing work. A job protection policy, therefore, can decrease the likelihood of permanent absence from the labor market. For mothers who re-enter while under job protection, it increases the likelihood of continuity in terms of both employer and occupation. In addition to mothers who benefit from returning to their pre-birth employer, job protection can also benefit mothers who change their firm when returning to the labor market (that could, for instance, be necessary when the need to care for the child limits the ability to commute). Job protection grants these mothers a period during which they retain their pre-birth employer as an outside option such that they have more time to search for a new job that better fits their preferences and constraints. When comparing mothers with similar lengths of parental leave, we thus expect that job protection policies reduced the earnings losses for all mothers who return to the labor market in general and specifically, both for mothers who return to their pre-birth employer as well as for those who switch to a different employer. Existing studies that analyze reforms in both job protection and benefit payments, tend to put more emphasis on the effects of job protection (Lalive and Zweimüller 2009; Schönberg and Ludsteck 2014) or on a combination of both policy instruments (Lalive, Schlosser, et al. 2014).

2.4.2 Empirical Implementation

An ideal experiment to distinguish the effects of benefit payments and job protection on maternal careers would be to randomize mothers who take post-birth labor market breaks of similar length into groups that are either treated with job protection, benefit payments, both, or with a placebo. To establish a real-world analog, we proceed as follows.

To estimate the causal effects of job protection and benefit payments on the labor market outcomes of mothers, we exploit variation in the parental leave policy that is driven by the reforms described in Section 2.2. Over time and eight reforms, German parental leave policy became more and more generous. Figure 2.2 gives a stylized illustration of such a reform that extends parental leave. It plots how parental leave coverage depends on the duration of the post-birth labor market break for mothers who gave birth before and after the reform. Policy coverage for mothers under two subsequent parental leave regimes depends on the time they spend on leave (on the x-axis). A reform replaces the old parental leave regime j (depicted in blue) with a new regime k (depicted in orange). Coverage is represented by darker colored horizontal bars. Returning to the labor market after the end of coverage is represented by lighter colors. The new regime grants mothers extended parental leave coverage. For the old regime, the vertical blue line marks the end of coverage, in the new regime, this is represented by a vertical orange line. This situation offers several possible comparisons in an empirical analysis. A before-after comparison of *all* mothers who gave birth under the old regime j with *all* mothers who gave birth under the new regime k gives the average treatment effect of the reform as an aggregate of different factors. Extending parental leave coverage incentivizes mothers to extend their post-birth labor market break. If mothers react accordingly (as shown in Figure 2.1a, this is the case), human capital depreciation is expected to be larger. Comparing the outcomes of mothers across regimes then contains a mechanical effect of increases in leave duration (given the results in Figure 2.1b, we expect it to be negative). It does not allow to distinguish this duration effect from other and potentially counteracting effects due to changes in the policy instruments, because it includes mothers who are subject to a variety of policy constraints when returning.

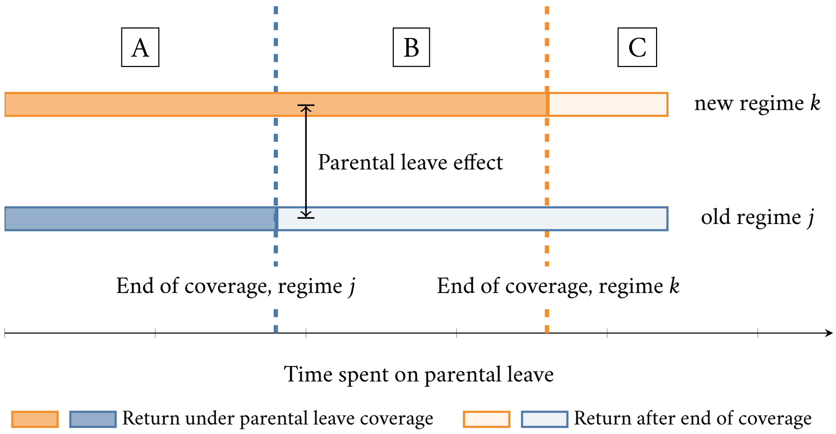


FIGURE 2.2: Stylized representation of a parental leave reforms and how coverage differs depending on the duration of the post-birth labor market break.

Notes: The figures illustrates different groups of mothers that are constituted by each parental leave reform. In section A, to the left of the vertical blue line, there are only mothers who return to the labor market while under parental leave coverage. Between the blue and the orange vertical lines, in section B, mothers who gave birth under regime j return after parental leave coverage has ended while those from regime k are still covered. In section C, to the right of the vertical orange line, no mother returns under parental leave coverage. Holding time on parental leave constant, comparing mothers from both regimes in section B allows to assess the effects of parental leave. *Source:* Own representation.

To be able to distinguish the effects of policy instruments, it is necessary to compare mothers who differ in their eligibility for these instruments while accounting for leave duration. The parental leave reforms allow to define groups of mothers with similar leave durations, but who are exposed to different policies. Figure 2.2 distinguishes between the three sections A, B and C. In section A, we observe only mothers who return before the end of parental leave coverage as defined by the old regime j, i.e. to the left of the vertical blue line in the figure. In section C, there are mothers who return after the end of parental leave coverage as defined by the new regime k, i.e. to the right of the vertical orange line. For mothers from both sections A and C, policy instruments do not differ across parental leave regimes. If they choose to take the same amount of parental leave in either regime, they are unaffected by the reform, thus untreated. In section B, there are mothers who return later than the old regime's coverage period but before the new regime's coverage is exhausted (i.e. between the blue and the orange line). Hence, mothers giving birth under the new regime are still under parental leave coverage, but mothers in the old regime are not. Therefore, by focusing on comparing mothers in section B across regimes while controlling for their leave duration allows to differentiate the effect of single policies from the mechanical effect of longer leave taking due to the reform. Within section B, mothers from the new regime k are treated with job protection, benefit payments or both (depending on the specific reform) while mothers from the previous regime j have experienced the same amount of human capital depreciation until their return but do not enjoy the potential benefits of either policy.

Which policies can by analyzed is given by the changes that each reform introduces. Most of them extended both job protection and benefit payments in parallel, thus comparing mothers in section B across reforms yields a joint effect of job protection and benefit payments. As our main identifying variation we use the reform in 1992 which extended only the length of job protection coverage, allowing us to isolate its effect. To isolate the effects of benefit payments, we use reforms in two large federal states.

To bring this approach to the data, we estimate two specifications. The first one leverages the large number of observations in our administrative dataset to obtain the effect of each additional month of parental leave on the change in post-birth labor market outcomes. Formally, we estimate the following specification separately for each regime r:

$$Y_{it}^{r} = \sum_{l=3}^{48} \beta_l \mathbb{1}\,(\text{leave duration months}_i = l) + \chi \mathbf{X}_{t=-1} + \varepsilon_{it}, \tag{2.1}$$

where β_l measures the effect of returning to work after l months of parental leave on outcome Y_{it} of mother i at time t away from her first childbirth. The reference group for β_l are mothers who return after two months, the mandated maternity protection period. The vector $\mathbf{X}$ holds controls for individual characteristics, i.e. average earnings in the three years prior to childbirth as well as pre-birth indicators for education, occupation, federal state and age. We control for these variables as they can have an impact both on when mothers return to they labor market as well as on their long-run outcome trajectories. This specification gives the precise effect of returning to work after l months of parental leave. Comparing the coefficient estimates from two adjacent regimes for a given value of l yields the difference in the changes in the outcome conditional on taking l months of parental leave. Focusing this comparison on the subset of coefficient estimates that fall in the range of parental leave durations that we define as section B in Figure 2.2 allows to asses the effect of the policy (either job protection, parental benefit payments or both) net of human capital depreciation. This approach, further, allows to plot outcome trajectories by parental leave duration for each regime for a graphically assessment.

We complement the detailed month-by-month comparisons of mothers with different parental leave lengths with a simplified pooled estimation that replaces the indicators for each month of parental leave with indicators for the sections A, B and C defined above (see Figure 2.2). To ensure that the sections are clearly distinguishable from each other, we do not include observations from mothers who return right at the cutoff between two sections. For regime pair 6/7, for instance, this applies to mothers who return exactly after 18 or 36 months. We estimate the following specification for each regime pair $\{j, k\}$ around a reform:

$$Y_{it}^{r\in\{j,k\}} = \sum_{r\in\{j,k\}} \gamma_r \mathbb{1}\,(\text{regime} = r) \times \sum_{g\in\{A,B,C\}} \delta_g \mathbb{1}\left(\text{leave duration group}_i = g\right) + \chi \mathbf{X}_{t=-1} + \varepsilon_{it}, \tag{2.2}$$

where r indexes parental leave regimes and g indexes groups of mothers depending on in which sections of leave duration they return. This model reduces the 45 indicator variables from equation (2.1) to three indicators δ_g for each section $g \in \{A, B, C\}$ interacted with indicators γ_r for being either in the previous regime j or in the new one k. In this specification, the difference between the two regime indicators for section B gives the average effect of the policy that applies solely to mothers from the new regime in section B.

To establish that the difference between outcomes of mothers from two regimes, while holding constant parental leave duration as described above, represents the causal impact of being exposed to either policy instrument, our estimation approach requires two main underlying assumptions. First, that mothers do not strategically plan when to have their child with respect to the changing policy environment. Precisely, that they do not time childbirth to select themselves into a more generous parental leave regime. Given the close succession of most reforms, this assumption is reasonable ex ante. The reform in 1992 that marks the change from regime 6 to regime 7 and provides our main identifying variation happens only 18 months after the previous and 12 months before the next reform. As different sources of uncertainty play a role when someone tries to precisely time conception and childbirth, it is unlikely that a substantial number of women has the opportunity to select into the new regime and successfully does so. Further, Schönberg and Ludsteck (2014) argue based on checking the media coverage on parental leave reforms that reporting usually did not start earlier than two months before a policy change. That again limits the possibility for women to adjust their decision making. The second assumption is that mothers are indeed comparable across parental leave regimes conditional on their individual leave duration. In other words, that reforms in the parental leave legislation do not introduce changes in the characteristics of mothers who choose specific parental leave durations. This could, for instance, be that especially well educated and better earning mothers, who are sufficiently well informed, rational and forward looking enough to anticipate the benefits parental leave and especially job protection can have, select into longer leave durations as long as they are job protected. Then, in a comparison of mothers across regimes, even when holding parental leave duration constant, it could happen that observed differences are driven by selection. From a theoretical perspective, this is something that cannot be fully ruled out. However, such considerations are unlikely to be a main driver of mothers' decision making. It, first, requires mothers to have access to information on the negative consequences of longer leave taking as well as on the potentially offsetting effects of parental leave policies. Second, it requires that mothers give enough weight to such information such that they impact their decisions when choosing on parental leave durations. Existing results speak against this. Expected labor market consequences are typically not found to be an important driver of the decisions of women as they are perceived as less important or smaller than they in fact are. Kuziemko et al. (2018) show on data from the US and UK that women rather underestimate the impacts of having children on subsequent labor market outcomes. Kiessling et al. (2019) analyze wage expectations of German college students, i.e. a group for that an above-average knowledge on earnings penalties of children can be assumed. They find that young females do anticipate earnings penalties for having children, but that gender gaps in wage expectations are mostly driven by sorting into college majors and occupations and by anticipated differences in working hours (after having children), though not by child-related career breaks. In Section 2.5.2 we discuss based on our findings on parental leave durations and how they are related to earnings if there are indications for selection of specific types of mothers into certain parental leave durations. In Section 2.5.6 we provide further empirical assessments of both identifying assumptions.

2.5 Results

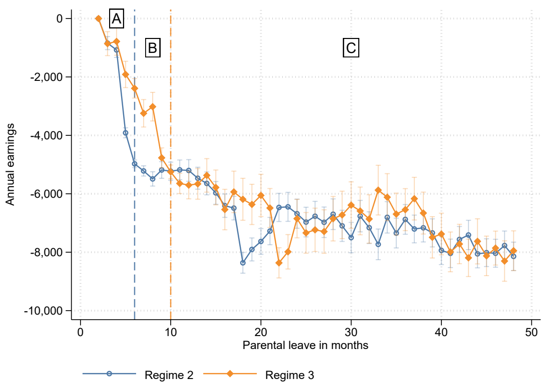
In this section we present our main results. We start by showing that extending parental leave is clearly successful in reducing mothers' long-run earnings losses. We further present evidence that this positive effect can be solely attributed to job protection coverage while benefit payments have no such effect. We then expand on the value of job protection by showing that returning under job protection significantly enhances employer continuity and enables mothers to work for higher paying employers. In addition, we show that the beneficial effect of job protection is not solely driven by the possibility to continue an existing employment relationship but rather that both mothers who stay with their pre-birth employer as well as those who switch to a new employer upon return to the labor market profit. Throughout this section we focus on outcomes measured seven years after the birth of the first child. For this time span, we expect that the most crucial labor supply decisions have already been made. It is after the child has entered kindergarten and (in a majority of cases) primary school which both can affect the childcare constraints mothers face. In the last part of this section, we present robustness checks.

2.5.1 The Effect of More Generous Parental Leave Policies

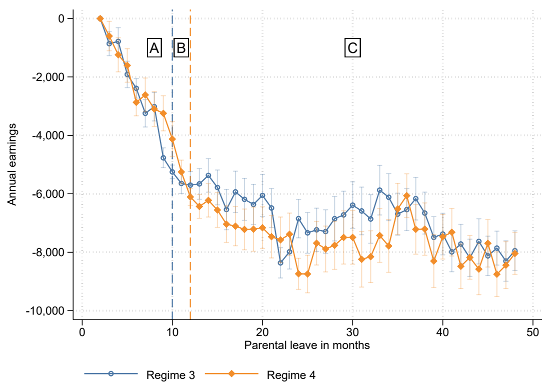
As shown in Figure 2.1 in the introduction, there is a clear negative association between the length of parental leave and earnings in the long run. Going beyond this stylized fact we aim to assess the causal effect of extending parental leave coverage on mothers' long-run earnings losses. To this end, we employ the empirical strategy described in Section 2.4. Here, we focus on the parental leave reforms that move from regime 2 to regime 6 and extend both job protection and benefit payments in a parallel way. We therefore estimate the combined effect of both policy instruments.

Graphical Evidence for Annual Earnings In Figure 2.3, we plot the detailed month-by-month effects of longer parental leave on annual earnings seven years after childbirth. All results are obtained using regression as described in Equation (2.1) that estimate the loss in annual earnings for each additional month of parental leave relative to returning right after the end of the mandatory maternity protection period of two months. In the plots we always compare the earnings of mothers returning after the same duration of parental leave in two adjacent regimes around a reform (the old regime is plotted in blue, the new one is plotted in orange). As parental leave policies became more generous over time, the new regime is always the more generous one. The dashed vertical lines indicate the maximum coverage of parental leave. The blue vertical line marks the end of coverage in the old regime, the orange vertical line for the new regime. The vertical lines also distinguish the three groups, A, B and C, that we define in Figure 2.2 to differentiate mothers with respect to their exposure to different policy constraints.

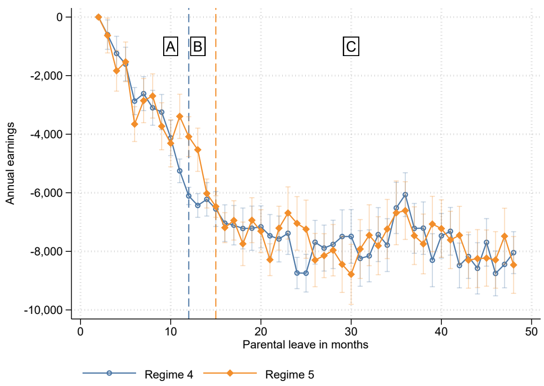
Across all regimes we observe a similar pattern. As in Figure 2.1, the relationship between the length of parental leave taking and earnings is negative and nonlinear. Mothers from both regimes experience significant losses compared to mothers who return to the labor market immediately after two months of maternity protection. For those who take the longest leave durations, the losses amount to around Euro 8 K or to around one third of average pre-birth earnings. In the following we discuss the earnings effects conditional on holding constant the duration of parental leave for the three sections A, B and C of the plots where mothers face different policy constraints.



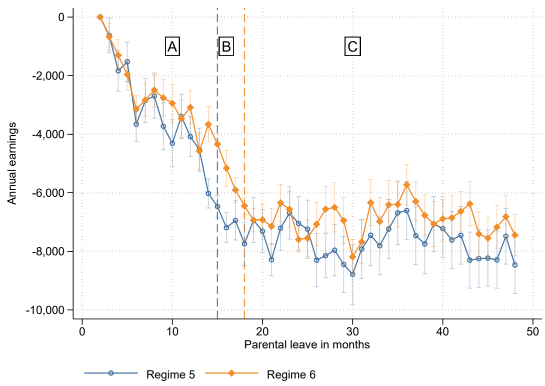
(A) Regime pair 2/3



(B) Regime pair 3/4



(C) Regime pair 4/5



(D) Regime pair 5/6

FIGURE 2.3: Effect of extending job protection *and* parental benefit payments on annual earnings seven years after birth.

Notes: The figure plots the results of regressions of annual earnings seven years after childbirth on the length of the post-birth labor market break in months along with 95 percent confidence bands. The regressions (see also Equation (2.1)) are carried out separately for parental leave regimes and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as indicators for age and year. *Source:* Own estimations using the IEB data described in Section 2.3.

Section A: Mothers in both regimes return under parental leave coverage. To the left of the vertical blue line, we compare mothers from both regimes who returned to the labor market before parental leave coverage is exhausted. All of them enjoy job protection as well as they receive maternity benefit payments. For most reforms, maternal earnings follow a similar path. For mothers in both regimes, we observe a clearly negative relationship between the time they stay away from the labor market and their earnings seven years after birth. For this group, the negative effect of an additional month of parental leave is the largest one; their losses are the steepest. We detect no relevant difference between mothers who return in the old or new regime. Table 2.8 in the Appendix lists pooled estimations results of the differences between regimes. For mothers returning in group A we estimate average differences between around Euro −160 and Euro 360, that are not significantly different from zero throughout.

Section B: Only mothers in the new regime return under parental leave coverage. Between the vertical blue and orange line, in section B of the figure, we compare mothers who return after the old regime's coverage period has ended but before the new regime's coverage is exhausted. Hence, mothers giving birth under the new regime are still under job protection and receive maternity benefits, while mothers in the old regime are not. Contrary to the mothers we plot in section A, we find substantial differences between the adjacent parental leave regimes. Across all regime pairs, we estimate significantly lower earnings losses for mothers under the new regime at a given length of parental leave. The differences range between Euro 488 (regime pair 3/4) and Euro 1,407 (regime pair 5/6) (see also Table 2.8 in the Appendix).

Section C: Mothers from both regimes return after the end of parental leave coverage. To the right of the vertical orange line in section C of each panel, we plot the earnings of mothers who return after parental leave policies have expired, even under the more generous new regime. Similar to those in section A, there is again no difference in policy constraints at return. Again similar to mothers in section A, we find no strong differences in earnings between mothers from the two adjacent regimes. We also observe that earnings continue to stabilize such that there is no more marked negative association with the length of parental leave.

It is a striking result that the earnings trajectories in adjacent regimes are virtually identical in sections A and C of the respective figures, i.e. where the policy constraints are identical for mothers from both regimes. The only strong deviation we observe throughout all reforms is detected in section B where the policies that apply to mothers differ across regimes. We interpret this pattern as a strong evidence that being covered by parental leave successfully offsets a substantial fraction of the earnings loss incurred due to longer parental leave duration. The pattern further offers evidence that selection into specific parental leave durations is similar across adjacent regimes. If, otherwise, mothers' selection patterns into groups A, B and C would change with a parental leave reform we would rather expect earnings differences in more than only one group.

Table 2.3: Pooled estimation results of length of parental leave taking on cumulative earnings for the seven years after birth, regimes 2–7. Differences between more recent and previous regime (p-values are given in brackets).

	Section			
Regime Pair	A	B	C	Observations
Extension of job protection and benefit payments				
2/3	1,474	11,290	1,472	263,311
	[0.192]	[0.000]	[0.060]	
3/4	-608	6,036	-1,270	143,727
	[0.620]	[0.000]	[0.240]	
4/5	-1,439	7,373	5,110	111,314
	[0.335]	[0.000]	[0.000]	
5/6	936	8,422	3,113	127,638
	[0.512]	[0.000]	[0.018]	
Extension of job protection alone				
6/7	2,583	9,504	4,554	147,346
	[0.059]	[0.000]	[0.001]	

Notes: The table shows the differences in the losses in cumulative earnings between the more recent and the previous parental leave regime. Positive values indicate mothers in the new regime are relatively better off. The groups are defined as illustrated in Figure 2.2, i.e. in group A mothers in both regimes return while having job protection, in group B only mothers in the more recent regime have job protection and in group C no mother returns under job protection. Differences are obtained from regressions as specified in equation (2.2) controlling for average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) and age. *Source:* Own estimations using the IEB data described in Section 2.3.

Pooled Estimations for Cumulative Earnings Moving from graphical to numerical results, the upper part of Table 2.3 collects estimation results for regime pairs 2/3 to 5/6 where we pool observations over the groups A, B and C (as illustrated in Figure 2.2). We estimate the specification as defined in equation (2.2), i.e. we interact indicators for returning in group A, B or C with indicators for having given birth under either the old or the new parental leave regime. The table lists the differences in losses between the new and the old regime where positive values indicate that mothers in the new regime are better off as their relative losses are smaller compared to mothers in the previous regime. To give a complete picture of the full earnings trajectory since returning to the labor market, the table shows results for all post-birth earnings cumulated until year seven after birth (Table 2.8 in the Appendix shows analog results for annual earnings).

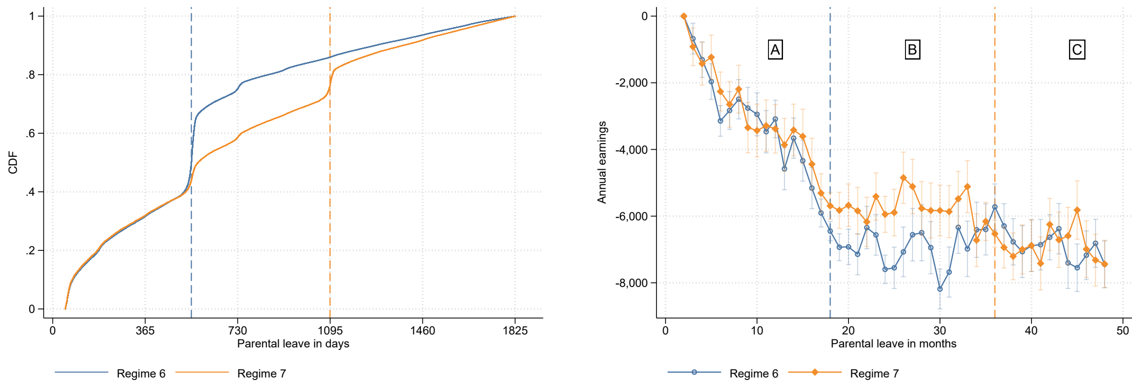
Overall, our pooled estimation results on cumulative earnings confirm the conclusions we have drawn from the previous graphical assessment. Comparing mothers from regime pairs 2/3 to 5/6 who return in section A we find no statistically significant differences between regimes. In comparison, estimates for section B show that mothers who gave birth under the new regime and return to the labor market before job protection and benefit payments in the new regime expire but after expiration of both in the old regime suffer from significantly smaller losses in cumulative earnings during the seven years after birth. Their advantage in cumulative earnings is sizeable at between Euro 6 K (regimes 3/4) and Euro 11.3 K (regimes 2/3) in absolute terms or between 5 and 9 percent of total cumulative earnings. Comparing within regime pair, the advantage in earnings losses for mothers in the new regime in section B is up to ten times larger than in section A. In section C, mothers from both regimes return while neither being under job protection nor receiving benefit payments. This group consists of those mothers who have the longest leave durations and are likely more heterogeneous. They show to be less attached to the labor market, for instance because they put a larger weight on household work and caring for their children or their partners provide sufficient financial stability that they can afford to re-enter the labor market rather late and with potentially higher reservation wages. We further cannot rule out that they gave birth to additional children[6] or that they have made specific arrangements with their employers that allow them to have longer breaks. Consistent with the assumption of a more heterogeneous group, differences in earnings losses for mothers from section C show some variation. While we measure significant differences between mothers from the new and the old regime these are small in magnitude compared to section B. The pattern we document here for cumulative earnings is confirmed by analog results for annual earnings that we list in Table 2.8 in the Appendix. There, we find that mothers from section B in the new regime have between about 500 (regimes 3/4) and Euro 1.4 K (regimes 5/6) lower earnings losses than their counterparts from the previous regime.

2.5.2 The Effect of Job Protection

So far, we have studied the combined effect of job protection and parental benefit payments which were extended in parallel across regimes 2 to 6. To disentangle the separate causal effects of job protection and benefit payments on long-run maternal labor market outcomes, we start by isolating the effect of an increase in the job protection period. To this end, we restrict the analysis to the reform in January 1992 that moved from parental leave regime 6 (births from July 1, 1990 to December 31, 1991) to regime 7 (births from January 1, 1992 to December 31, 1992) and increased the job protection period from 18 to 36 months while leaving the benefit payment period unchanged at 18 months. This reform introduced the longest increase of the job protection period and is also the only one to extend job protection alone. Consistent with the results by Schönberg and Ludsteck (2014), the aggregate effect of this reform is small. Comparing the post-birth earnings from mothers in both regimes (see Table 2.2) yields that in

[6] For the method by Müller, Filser, and Frodermann (2022) to reliably identify a subsequent birth it is necessary that a mother re-enters the labor market before having the second child.

new regime mothers earn on average Euro 704 (or 3.5 percent of earnings of mothers in the old regime) less than in the old regime. In the following, we discuss how leave taking and earnings losses conditional on leave duration are affected by the reform.



(A) Empirical cumulative distribution functions of length of parental leave.

(B) Effect of extending job protection from 18 to 36 months on annual earnings seven years after birth.

FIGURE 2.4: Effect of extending job protection from 18 to 36 months on parental leave duration and annual earnings seven years after birth (regime pair 6/7).

Notes: The left-hand panel plots the empirical CDFs of parental leave duration in days. The right-hand panel plots the results of regressions of annual earnings seven years after childbirth on the length of the post-birth labor market break in months along with 95 percent confidence bands. The regressions (see also equation (2.1)) are carried out separately for regimes 6 and 7 and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as indicators for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

Changes in Leave Duration and Long-run Earnings We start in Figure 2.4a by taking a closer look at mothers' decisions how long to stay out of the labor market after their first childbirth. The figure plots the empirical cumulative distribution functions (CDF) of mothers' leave durations in days by parental leave regime (regime 6 is depicted in blue, regime 7 in orange). To ensure consistency with our other analyses, the sample is restricted to those mothers who return to labor market after at most five years on leave. On average, mothers in regime 7 take around 3 months longer than their counterparts in regime 6 to return to work (19.8 vs. 16.9 months on average). The upper part of the distribution of the leave duration accounts for almost all changes; the 75th percentile increases from 21 to 33 months (see also Table 2.2). For leave durations up to 18 months (the end of job protection in regime 6, indicated as a dashed blue line) the CDFs in both regimes are virtually identical. Similarly, for leave durations longer than 36 months (the end of job in regime 7, indicated as a dashed orange line), the CDFs are again close together (though showing that some additional mothers in regime 7 choose to return after breaks longer than 36 months). Between the two regime cutoffs where job protection coverage ends, the CDFs differ substantially. In regime 6, returning after 18 months is the most common choice. After the reform, a large share of mothers in regime 7 who, without the reform, would have returned to work after 18 months now chooses to return later. There is still a clear uptick in regime 7's CDF at 18 months, though substantially smaller. The CDF's slope to the left of the 18 months cutoff is slightly steeper compared to the regime before, showing that more mothers now return after between 18 and 36 months. Consequently, the jump in the CDF at 36 months when mothers return who use the full extension due to the reform is less pronounced in regime 7.

Figure 2.4b plots the month-by-month effects of longer parental leave on annual earnings seven years after childbirth. As in Figure 2.3 in the previous section, all results are obtained using regressions that estimate the loss in annual earnings for each additional month of parental leave relative to returning right after the end of the mandatory maternity protection period of two months. Regime 6 is plotted in blue, regime 7 in orange. Consistent to our previous findings for regimes 2 to 6, we observe a negative and nonlinear relationship between the time spent on parental leave and earnings in the long run. Mothers in both regimes experience significant earnings losses the longer they stay away from the labor market. Contrary however to our previous results that showed effects of extending job protection and benefit payments for the same amount of time, the results in this section indicate the isolated effects of an extension in job protection. In more detail, we observe the following.

Section A: Mothers in both regimes return under job protection coverage. To the left of the vertical blue line we compare mothers from both regimes who took up to 18 months of parental leave and returned to the labor market while still under the coverage of job protection and parental benefit payments. Similar to our previous findings, mothers from this section of the figure show the largest negative effect of an additional month they spend on parental leave. As their decisions were made under the same policy constraints, earnings of these mothers do not exhibit significant differences across parental leave regimes. In Table 2.8 in the Appendix we document that the average difference between mothers from regime 6 and regime 7 is numerically small (Euro 205) and not significantly different from zero.

Section B: Only mothers in regime 7 return under job protection. Between the two vertical lines, in section B of the figure, we compare mothers who take 19 to 36 months of parental leave. They allow us to identify the effect of job protection since only those who gave birth under regime 7 return to the labor market while still being under job protection coverage. We find that earnings losses for mothers from both regimes do not continue to increase in leave duration, but stabilize. Comparing the earnings losses across regimes for mothers who take parental leave of the same duration, however, reveals that for (almost) all durations, mothers from regime 7 exhibit clearly smaller earnings losses. On average, the difference in annual earnings between the regimes is Euro 1,123 (see also Table 2.8 in the Appendix). This amount is of substantial size as it corresponds to around 5.9 percent of annual earnings (respectively 6.3 percent of the median and 8.1 percent of a standard deviation) of mothers in regime 7. While we do find a similar pattern for the other reforms, this one stands out. In the reforms before, both policy instruments—job protection and benefit payments—were extended. The reform analyzed here, extends solely the job protection period. It thus allows us to attribute the difference in earnings losses solely to the fact that mothers who gave birth under regime 7 have the benefit of additional job protection coverage. It is further of note that all the previous reforms only introduced relatively short extensions of parental leave—by up to four months. The extension when moving from regime 6 to 7, though, is more than four times longer. It doubles the previous maximum from 18 to 36 months. This underlines that the beneficial effect of job protection is not restricted to short increases but allows mothers to have twice the time with their children, without having to suffer substantial additional earnings losses. Consistent to our previous findings, this effect of job protection is not specific for year seven after childbirth but persistent such that it adds up over time. The difference in cumulative earnings losses between mothers from regime 6 and 7 amounts to Euro 9,504 (see Table 2.3) which corresponds to 9.7 percent of total cumulative earnings (respectively 11.7 percent of the median and 12.7 percent of a standard deviation).[7]

Section C: Mothers in both regimes return after the end of job protection coverage. To the right of the vertical orange line, we plot the earnings of mothers who take 37 months or more of parental leave which means that in both regimes they return to the labor market without a right to return to their pre-birth job, i.e. policy constraints across regimes do not differ. Similar to our findings for other reforms as well as for mothers who return in section A, the graphical analysis shows there are no significant differences in earnings losses between mothers in regimes 6 and 7. On average, earnings losses in regime 6 are smaller by Euro 48 which is not significantly different from zero (see Table 2.8 in the Appendix).

Selection Into Different Parental Leave Durations An important assumption that underlies our approach to identify an isolated effect of job protection is that parental leave reforms do not change how mothers of similar characteristics select themselves into parental leave durations. This means comparability conditional on their leave duration across regimes is given. We address this assumption in Section 2.5.6 by showing that the relationship between observable characteristics and leave duration does not change between regimes after a reform. In the following, we build on the results in Figure 2.4 to address the potential impacts of unobserved factors.

[7] In Figure 2.10 in the Appendix we provide graphical evidence for the relationship between cumulative earnings and parental leave duration which confirms the pattern that we find for annual earnings.

For our estimations of the effect of job protection to be biased it would be necessary that for those mothers who return in section B of Figure 2.4b and conditioning on leave duration to account for human capital depreciation, there is an omitted variable driving selection into leave durations that also has an impact on post-birth earnings.

There are two cases that have to be distinguished. The first one considers mothers who gave birth under regime 7. It could be that the reform especially incentivizes mothers who are positively selected in terms of unobserved ability to take longer parental leave such that observations from regime 7 in section B are biased upwards. This is unlikely to be the case. In absence of the reform such positively selected mothers would had returned earlier, i.e. in section A. However, when comparing the CDFs from both regimes in Figure 2.4a, we find no indication of missing distributional mass. Instead, leave taking to the left of the cutoff for regime 6 (up to 18 months) is almost identical across regimes. For the two CDFs to be identical while at the same time positively selected mothers take longer leave durations due to the reform, it would be necessary that the reform also incentivized mothers who, in absence of the reform, would have taken more than 18 months of leave, to reduce their time on leave. Borrowing from the LATE terminology, that means we need to rule out the presence of defiers. Such monotonicity has to be assumed rather than it can be tested for directly (Angrist and Imbens 1994).[8] Given the incentives provided by the reforms, this assumption is nevertheless plausible. Further, there is no sign of excess mass to the left of the end of coverage for regime 6 as well as there is no missing mass to the right of this cutoff. In addition, we observe a continuously declining trend in post-birth earnings for longer leave durations, whereas positively selected mothers who take longer leave would rather result in an increase of earnings for longer leave durations. Defiers, on the other hand, would rather lead to a downward bias of earnings of mothers from regime 7 to the left of the cutoff for regime 6, which we do not observe. There is the further possibility that from those mothers who, without the reform, would have returned right after 18 months, a positively selected fraction takes longer parental leave. This would bias earnings observed in section B under regime 7 upwards. However, it would also bias earnings observed in regime 7 right at 18 months downwards. We do not observe such a pattern in Figure 2.4b. Instead, we find that for mothers taking 18 months of leave, earnings of those in regime 7 are slightly higher compared to regime 6 while there are no different trajectories across regimes.

The second case to consider is that mothers who gave birth under regime 6 and who return after more than 18 months of leave are particularly bad selected in terms of their unobserved ability. This would mean that observations from regime 6 in section B are biased downwards. Again, we find no indication for this. Mothers in regime 6 who take between 19 and 36 months

[8] Early work by Kowalski (2020) attempts to provide a framework for inference on defiers.

of leave have earnings that are on average similar to the earnings of those who return right after 18 months, i.e. while they are still covered by job protection. In case of selection, one would instead expect a significant and lasting earnings drop for mothers who take 19 and more months of parental leave and thus willingly opt out of returning under job protection. Overall, we do not find indication that unobserved factors would play a substantial and biasing role in our results.

Is Parental Leave Duration Used as a Signaling Device? Beyond the role of single policy instruments other papers have suggested parental leave reforms give mothers the opportunity to signal high productivity to the labor market (Kleven, Landais, Posch, et al. 2021; Tô 2021) which could explain (at least parts of) the overall small reform effects found by Schönberg and Ludsteck (2014) and others.

Given our results in Figure 2.4 we find no evidence that speaks in favor of signaling. Signaling would require that mothers who are eligible to more parental leave after a reform return to the labor market right before the end of parental leave that applied for the old regime. By doing so, they would be able to separate themselves both from mothers who are in the same regime as well as form those in the old regime. For our case of the reform moving from regime 6 to 7, that would mean that mothers who gave birth under regime 7 and could take up to 36 months of leave return already after strictly less than 18 months. Here, a similar rationale as for assessing the presence of defiers applies. The CDF for regime 7 in Figure 2.4a exhibits jumps at both 18 and 36 months but, importantly, there is no excess distributional mass before 18 months. Further, we observe that earnings as plotted in Figure 2.4b are smooth around and before 18 months leave duration. Thus, there is no sign that signaling is used nor that it would be a successful strategy to improve long-run earnings.

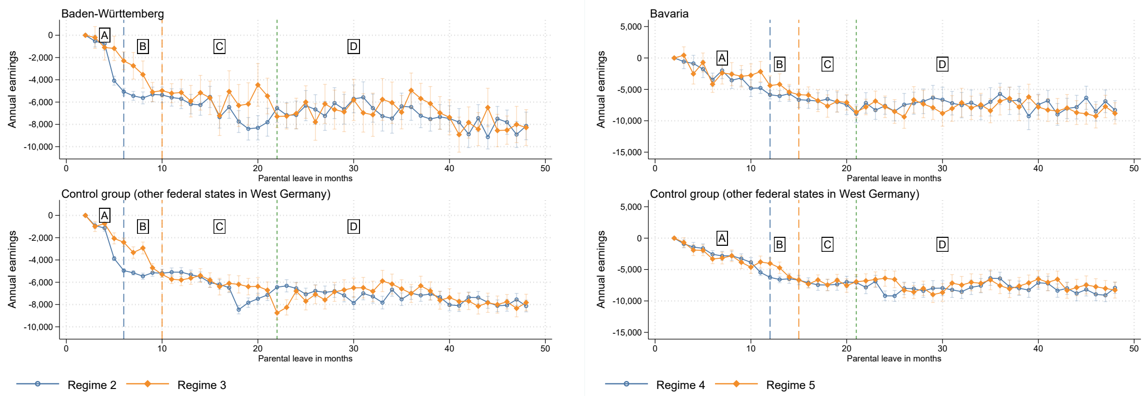
2.5.3 The Role of Benefit Payments

In the previous section, we have shown that job protection alone is a sufficient measure to substantially decrease mothers' post-birth earnings losses. Nevertheless, Lalive, Schlosser, et al. (2014) argue that the combination of job protection and benefit payments is preferable over only one policy instrument in allowing for time with the child and maintaining a career. Dahl et al. (2016), on the other hand, highlight potentially adverse redistributive effects of benefit payments while they do not yield gains in terms of labor market outcomes. In this section, we therefore turn our attention to the impact of parental benefit payments on long-run earnings losses.

There are no nation-wide policy changes that would allow to assess the isolated effect of benefit payments. Therefore, we make use of two reforms in Baden-Württemberg and Bavaria that extend the benefit duration beyond the job protection period. They introduce additional benefit payments paid immediately following the federal benefits.[9] Some points are noteworthy about this particular state-level benefits. First, they are introduced in 1986 (Baden-Württemberg) and in 1989 (Bavaria), i.e. close to the federal-level reform moving from regime 6 to 7 in 1992 that we use to isolate the effect of job protection. This lowers the risk that substantial changes in mothers' preferences or social norms over time threat the comparability of our analyses. Second, the additional benefits are introduced close to the nationwide introduction of parental leave regimes 3 and 5 (in Baden-Württemberg right at the same date, in Bavaria six months earlier) which allows us to employ a similar strategy as before. That is, we compare mothers from two adjacent policy regimes who return to the labor market either (a) while covered by both job protection and (federal-level) benefit payments, or (b) while only mothers from the new regime are covered by job protection and (federal-level) benefit payments, or (c) while only mothers from the new regime are covered by (state-level) additional benefit payments or (d) while they are not covered by any policy instrument. Third, the reforms happen while job protection is still at lower levels compared to the increase from 18 to 36 months in 1992. Fourth, they are implemented in the two states that are commonly viewed as the most conservative ones in terms of the populations' gender norms. The last two points underline the appeal of these reforms as they rather increase the incentives for mothers to take up the additional benefit payments and stay away from the labor market for a longer time. Hence, we expect that our estimates on the effects of benefit payments constitute an upper bound.

In Figure 2.5, we use the same methodology as before. We plot regression results that compare mothers who gave birth under a new policy regime (plotted in orange) with mothers from the old policy regime (plotted in blue) holding their amount of parental leave taking constant. The results in the top row are for Baden-Württemberg and Bavaria, i.e. the treated federal states where additional benefits are introduced for mothers in the more recent regime. The vertical blue line indicates the amount of leave taking where the nationwide parental leave coverage for the old regimes ends. The orange line indicates this for the new regime. The vertical green line (for Baden-Württemberg at 22 months and for Bavaria at 21 months) indicates until which leave duration mothers in the new regime can receive additional benefits. The relevant section in the figure where mothers between the two adjacent regimes differ only with respect to their eligibility to the additional state-level benefits is marked as C. Along with this treated group we plot the results from similar estimations for the other West German federal states which serve as a control group, because there mothers who return in section C are not eligible to additional benefit payments. Differences between the treated and the control group with respect to the

[9] For a detailed description see Section 2.2.



(A) Introduction of additional benefits in Baden-Württemberg (1986).

(B) Introduction of additional benefits in Bavaria (1989).

FIGURE 2.5: Effects of extending parental benefit payments in Baden-Württemberg and Bavaria on annual earnings seven years after birth.

Notes: The figure plots the results of regressions of annual earnings seven years after childbirth on the length of the post-birth labor market break in months along with 95 percent confidence bands. The regressions are carried out separately for mothers who gave birth under the more recent regime (orange line), i.e. those who receive additional benefit payments, and mothers from the previous regimes who do not receive these benefits (blue line). Graphs in the top row show estimations for Baden-Württemberg (left) and Bavaria (right), i.e. the federal states that pay additional benefits; the bottom row plots results from the other West German states as a control group. All estimations control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as indicator for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

differences in earnings losses between the new and the old regime would indicate that there is an impact of benefit payments on post-birth earnings.

In both treated states and the control group (though for the former measured with less precision), we observe the, by now familiar pattern, that mothers from the new regime who return in section B while still being under job protection receive higher earnings compared to those who take similar amounts of parental leave but gave birth under the old, less generous regime. This confirms that the previous findings hold for Baden-Württemberg and Bavaria as well. Turning to mothers in the treated states who return in section C, we find no evidence of labor market effects of benefit payments. Those mothers who are eligible to additional benefit payments show almost no differences compared to their non-eligible counterparts from the old regime. For Baden-Württemberg, there are differences between the regimes that are statistically significant and indicate lower earnings losses for the more recent regime at leave durations of 19 to 21 months. They, however, follow a pattern that we observe as well for the control group of other federal states that do not pay additional benefits. Hence, this cannot be attributed to parental benefit payments. In Bavaria, earnings losses are almost identical across regimes. Overall, we have to reject the hypothesis that benefit payments have a significant effects on mothers' long-run earnings. Thus, they do not play a role as a labor market policy but rather as a social policy.

2.5.4 Employer Continuity

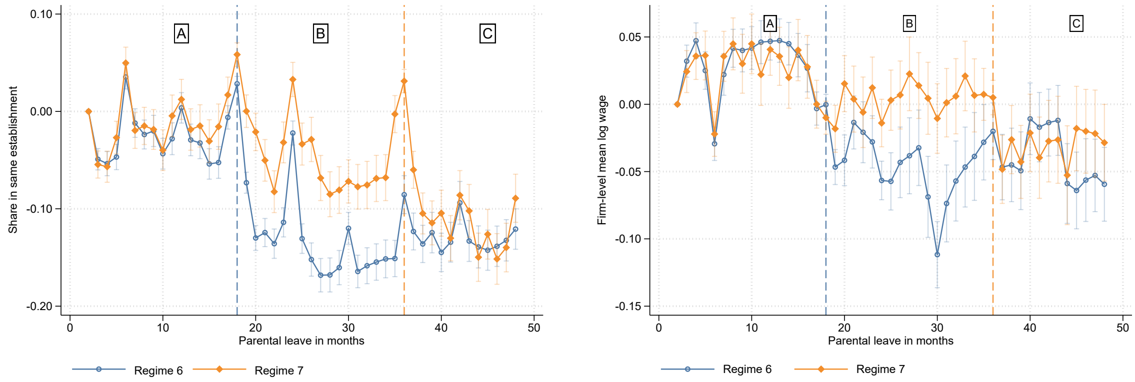
Having established the crucial role of job protection for mothers' post-birth careers, this and the following section explore its role in more detail. Upon re-entry to the labor market, a mother faces the choice to either continue her existing employment relationship with the pre-birth employer or to switch to a different employer. Job protection is often argued to be beneficial because it makes it easier to continue to work for the pre-birth employer. Mothers then maintain their firm-specific human capital, do not have to search for a new employer and their labor market attachment is thus expected to increase.

Focusing again on regime pair 6/7 that extends the job protection period from 18 to 36 months while keeping benefit payments constant at 18 months we, first, confirm the finding by Schönberg and Ludsteck (2014) that the reform had only a small effect on employer continuity. As shown in Table 2.2, seven years after childbirth around 1 percentage point fewer mothers in regime 7 have switched their employer. When we zoom in to differentiate by the duration of parental leave (Figure 2.6a), we find that increasing employer continuity is almost entirely driven by mothers who take longer leave under the new regime but still return under job protection coverage. Seven years after childbirth, job protection accounts for a up to 10 percentage points larger probability of still working with the pre-birth employer.

Even though we have no measure for maintenance of firm-specific human capital, our data allow to shed light on the firm side. In Figure 2.6b we plot the development of the establishment-level mean log wages by parental leave duration for mothers in regimes 6 and 7. As for most other outcomes there is no difference for mothers in sections A and C of the figure. In section B where only mothers in regime 7 benefit from job protection, we find that they work at establishments that pay around 5 percentage points higher average wages. This is driven by the pattern that mothers in regime 6 who take longer parental leave incur losses relative to returning early while for those in regime 7 there is almost no change up until a leave duration of 36 months. In both regimes, returning right after the end of job protection (i.e. after 19, respectively 37 months) results in a sharp drop in establishment wages by 5 percentage points. Again similar for both regimes, establishment wages for mothers who take less than 18 months of leave are higher than for mothers who return immediately indicating a more positive selection.

We find a qualitatively similar but less pronounced pattern for establishment fixed effects (Card, Heining, and Kline 2013; Bellmann et al. 2020) but no significant effects on establishment size (see Figure 2.11 in the Appendix).

Overall, these findings indicate that maintaining firm-specific human capital is not the sole driver of the beneficial effect of job protection. Besides enabling mothers to stay with their pre-birth employers it enables them to stay with employers who, on average, pay the higher wages.



(A) Relative probability of working for the pre-birth employer.

(B) Establishment-level mean log wage.

FIGURE 2.6: Effect of extending job protection from 18 to 36 months on working at the pre-birth employer and mean log establishment wages seven years after birth, regime pair 6/7.

Notes: The figure plots the results of regressions of the respective outcomes on the length of the post-birth labor market break in months. The regressions are carried out separately for regimes 6 and 7 and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as for indicators for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

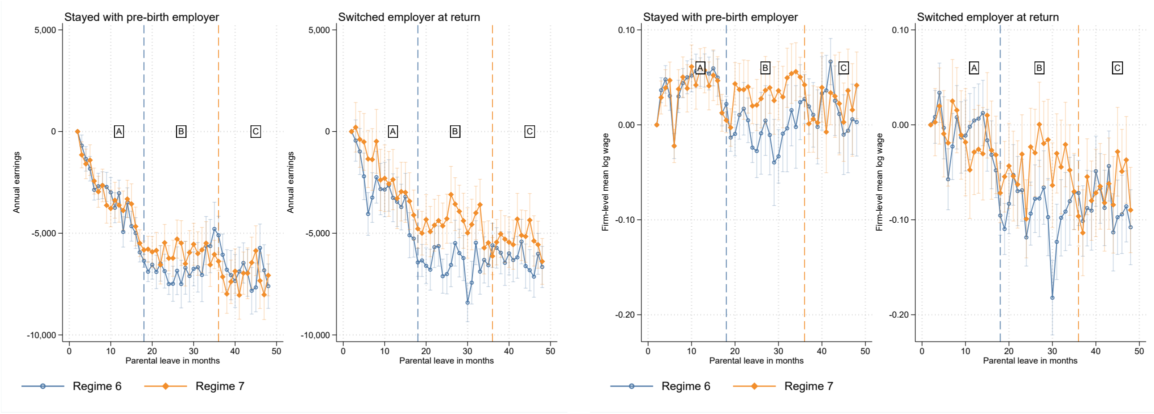
2.5.5 Employer Continuity vs. Switching the Employer

While job protection makes it easier for mothers to continue to work for their pre-birth employer, around one third of mothers chooses to switch to a different employer. As a last step of our analysis, we explore who makes this choice and how the effects of job protection differ between mothers who continue existing employment relationships and mothers who work elsewhere.

On average, not returning to the pre-birth employer is associated with worse outcomes in the long run. Comparing switching and staying mothers who gave birth under parental leave regimes 6 and 7 we find that seven years after childbirth those who switched have lower average earnings (Euro 17 to 17.2 K vs. Euro 20.2 to 20.8 K) and work in establishments that pay lower average log wages (4.5 vs. 4.6) and have lower establishment fixed effects (0.09–0.10 vs. 0.13).[10] As we nevertheless observe mothers making this choice, i.e. accepting the associated losses,

[10] This includes a small fraction (around 3 percent) of mothers who do not return to their pre-birth employer right when re-entering the labor market after childbirth but move back to it later on. We treat them as switching the employer since for job protection to have an effect the initial re-entry to work is the relevant event.

this suggests there are other, not necessarily monetary, incentives. Switching to a different employer can, for instance, be beneficial for mothers if it offers more flexible working hours, is easier to commute to or located closer to a childcare facility. From a theoretical perspective, job protection can be a tool to mitigate the costs of switching the firm as it increases the period during which mothers still have their pre-birth employer as an outside option giving them more time to search for a new employer.



(A) Annual earnings.

(B) Mean log wage in establishment.

Figure 2.7: Effect of extending job protection from 18 to 36 months on annual earnings and mean log establishment wages seven years after childbirth, separately for mothers who stay with their pre-birth employer (left-hand side) and who switch the employer (right-hand side) (regime pairs 6/7).

Notes: The figure plots the results of regressions of annual earnings (left-hand panel) and mean log establishment wage (right-hand panel) on the length of the post-birth labor market break in months, differentiating between mothers who returned to their pre-birth employer and those who switched to a different employer. The regressions are carried out separately for regimes 6 and 7 and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as for indicators for age and year. *Source:* Own estimations using the Ieb data described in Section 2.3.

We provide evidence for this second channel through which job protection improves mothers' labor market outcomes. Figure 2.7 plots for annual earnings and mean log establishment wages the relationship between each outcome and the time spent on parental leave separately for mothers who stay with their pre-birth employer and those who switch (Figure 2.12 in the Appendix adds results for cumulative earnings and establishment fixed effects). For earnings (plotted in Figure 2.7a), we find that the outcomes of staying and switching mothers follow generally similar patterns. Focusing, however, on section B where only mothers from regime 7 return under job protection, we find a markedly greater difference between the regimes. It suggests that mothers who return under job protection to a different employer, while having lower earnings on average, are the group who profits more from job protection in relative terms as their earnings losses for a given time spent on leave are smaller compared to mothers who returned after the end of job protection. The results for mean log wages in the establishment

(Figure 2.7b) indicate that both staying and switching mothers profit from returning under job protection. Consistent with our other results, we do not find substantial differences between mothers from either regime in sections A and C of the figure.

TABLE 2.4: Pooled estimation results of returning under job protection on earnings and establishment-related outcomes in year seven after childbirth for regime pair 6/7 (p-values are given in brackets).

	Full Sample	Stayers	Switchers
Annual earnings	1,115	817	1,915
	[0.000]	[0.002]	[0.000]
Cumulative earnings	9,390	7,188	14,265
	[0.000]	[0.000]	[0.000]
Mean log establishment wage	0.050	0.042	0.056
	[0.000]	[0.000]	[0.000]
Establishment FE	0.018	0.010	0.023
	[0.000]	[0.024]	[0.009]

Notes: The table shows the differences in the losses in outcomes between parental leave regimes 6 and 7 comparing mothers who return to work under job protection with mothers returning after the end of the job protection period (section B as illustrated in Figure 2.2). Positive values indicate mothers in the regime 7 are relatively better off. Differences are obtained from regressions as specified in equation (2.2) controlling for average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) and age. *Source:* Own estimations using the IEB data described in section 2.3.

To quantify the effects of job protection, Table 2.4 lists the differences in relative losses between mothers from parental leave regimes 6 and 7 where only those from the more recent regime returned to work while still being covered by job protection (i.e. we restrict the results to those in section B as illustrated in Figure 2.2). For both annual and cumulative earnings, the offsetting effect of job protection on earnings losses is around twice as large for mothers who switch to a new employer after childbirth. Mothers who return under job protection and stay with their pre-birth employer accumulate around Euro 7.2 K higher earnings in the seven years after childbirth than those who return after the end of the job protection period. For mothers who switch the employer this difference amounts to Euro 14.3 K, even though they have lower earnings in levels. Our findings for mean log establishment wages are similar. Staying with the pre-birth employer and returning under job protection is associated with working in an establishment that pays on average 4.2 percentage points higher wages (around 10 percent of a standard deviation). For switching mothers the differences is 5.6 percentage points, i.e. larger by a factor of 1.3. While without differentiating between staying and switching returning under job protection is associated with working in an establishment that pays a 1.8 percentage points larger wage premium, switching mothers have an advantage of 2.3 percentage points compared to 1 percentage point for those who stay with their previous employer.

Taking our findings together, we conclude that the effect of job protection on employer continuity is more important in absolute terms. However, restricting to the subsamples of mothers who stay with their pre-birth employer and who switch, the relative gains from returning under job protection are greater when mothers switch their employer.

2.5.6 Robustness Checks

For our estimations to be valid we require, first, that mothers do not select between parental leave regimes and, second, that for each duration of parental leave mothers are sufficiently similar across regimes. In this section we assess both assumptions.

Table 2.5: Pooled estimation results of length of parental leave taking on cumulative earnings for the seven years after birth, regimes 2–7. Differences between more recent and previous regime (p-values are given in brackets). Estimation samples are restricted to three months prior and after changes in the parental leave legislation.

	Section			
Regime Pair	A	B	C	Observations
Extension of job protection and benefit payments				
2/3	7,275	21,637	9,375	13,850
	[0.086]	[0.000]	[0.002]	
3/4	1,597	16,251	847	19,103
	[0.632]	[0.000]	[0.771]	
4/5	-655	11,587	4,550	21,794
	[0.845]	[0.002]	[0.128]	
5/6	4,459	18,625	4,330	24,109
	[0.159]	[0.000]	[0.137]	
Extension of job protection alone				
6/7	2,849	10,121	3,119	26,826
	[0.353]	[0.001]	[0.296]	

Notes: The table shows the differences in the losses in cumulative earnings between the more recent and the previous parental leave regime. Positive values indicate mothers in the new regime are relatively better off. The groups are defined as illustrated in Figure 2.2, i.e. in group A mothers in both regimes return while having job protection, in group B only mothers in the more recent regime have job protection and in group C no mother returns under job protection. Differences are obtained from regressions as specified in equation (2.2) controlling for average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) and age. *Source:* Own estimations using the IEB data described in section 2.3.

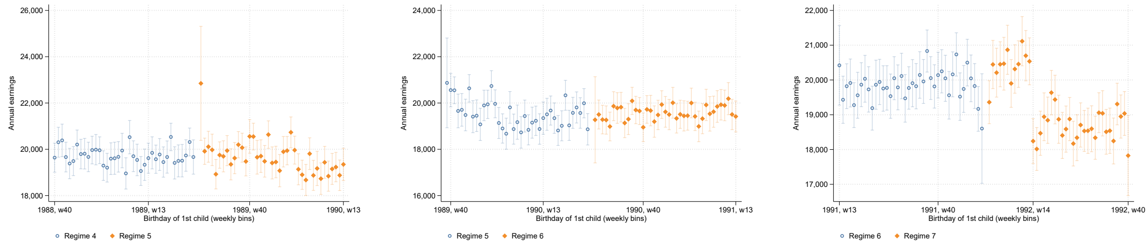
Selection Into Parental Leave Regimes If mothers can strategically postpone childbirth to benefit from increased parental leave and if those who do so are mothers who are on a generally higher earnings path, our estimates for the effects of parental leave policy instruments can be biased upwards. We account for this concern in two ways. First, we restrict our estimation sample to only include mothers who had as little as possible chance to manipulate when they give birth. Precisely, we include only those mothers who gave birth up to three months prior and after the birthday cutoff date for each reform (and leaving out those one week before and after the cutoff).[11] Following Schönberg and Ludsteck (2014) (who refer to Dustmann and Schönberg 2012), these mothers conceived their children already before the reforms were drafted. In Table 2.5, we list estimation results for cumulative earnings in the seven years after childbirth based on the restricted sample. Despite the considerable smaller number of observations, the patterns we observe for the full sample (see Table 2.3) are confirmed. For section A, the differences between the previous and new regime relatively small and at most at the 9 percent level significant whereas for mothers in section B we, again, find large and highly significant differences between the old and the new regime that point out that mothers in the new regime profit from parental leave and job protection by having lower earnings losses.

Figure 2.8 shows results for a second test where we plot the average annual earnings seven years after childbirth by the date of childbirth over a window of nine months before (in blue) and after (in orange) the reforms that moved from the fourth parental leave regime to the seventh.[12] As the figure shows, there are no significant differences in long-run earnings between mothers around the regime cutoffs. For the reform moving from regime 6 to regime 7, we find a drop in average earnings around three months after the reform date. This drop does not show a clear connection to a reform. If further speaks against the concern that higher earning mothers select themselves into regime 7 to profit from an increased duration of job protection. On the contrary, if this drop was related to mothers who take between 19 and 36 months of leave, it would rather decrease our estimate of the effect of job protection. Overall, the results provide evidence against a self-selection of higher earning mothers into more generous parental leave regimes.

Selection Into Parental Leave Durations We now turn to the second potential concern that at a parental leave reform which extends coverage, mothers changes their selection into parental leave durations such the that estimated differences between two subsequent regimes that condition on leave duration are in fact driven by differences in the characteristics of mothers. In Section 2.5.2 we discuss the possible impacts of unobservable factors, here we assess the role of observable ones. All our estimations control for mothers' pre-birth characteristics. Specifically for education, occupation, earnings (average for the three years prior to birth) and federal state as well as for age and year. It could nevertheless be that our results are driven by selection patterns of certain mothers into different lengths of parental leave. To address this concern, Figure 2.9 plots the relationship of observable characteristics of mothers and the

[11] For a list of the birthday cutoff dates, see Table 2.1.

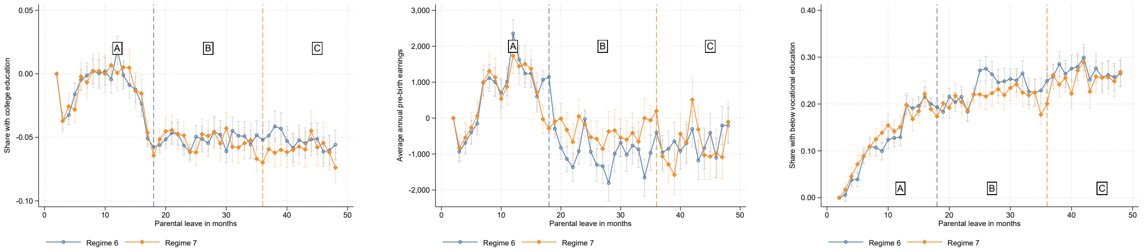
[12] Figure 2.13 in the Appendix shows results for the other reforms from regimes 1 to 4.



(A) Regimes 4/5 (B) Regimes 5/6 (C) Regimes 6/7

FIGURE 2.8: Average annual earnings seven years after birth by birthday (in weekly bins) around parental leave reforms.

Notes: The figure plots the results of regressions of average annual earnings seven years after childbirth on the week of birth of the first child along with 95 percent confidence intervals. The regressions are carried out separately by parental leave regime for windows of nine months around the reform date. *Source:* Own estimations using the IEB data described in section 2.3.



(A) Share with college education. (B) Average earnings over three years prior to birth. (C) Share returning in part-time.

FIGURE 2.9: Pre-birth characteristics of mothers by length of parental leave (regime pair 6/7).

Notes: The figure plots the results of regressions of the respective outcomes on the length of the post-birth labor market break in months. The regressions are carried out separately for regimes 6 and 7 and control for indicators for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

time they spend on parental leave. It focuses on our main identifying variation, the reform in 1992 that extends the job protection period from 18 to 36 months. For the share of college educated mothers we find no difference. Figure 2.14 in the Appendix also plots the remaining categories vocational and below vocational or no education, showing that mothers in regime 7 who return between the two relevant regime cutoffs (i.e. take between 19 and 36 months of leave) are slightly more likely to have below vocational or no education and slightly less likely to own a vocational degree. This indicates that mothers in regime 6 are, if anything, better educated, such that we would expect higher earnings for them. The effect we estimate, would then constitute a lower bound. For average pre-birth earnings, we find that those mothers in regime 7, who return between 19 and 36 months of leave, show slightly smaller decreases in their earnings. These are, however, in almost all cases not significantly different from the estimates for regime 6. It has to be noted that our main specifications control for pre-birth earnings, however, including or omitting it from estimations has little effect on the results. In Panel C of the figure, we further plot results for the share of women who return to the labor market in part-time work. For this variable we, again, find no substantial differences.

Table 2.6: Pooled estimation results of length of parental leave taking on cumulative earnings for the ten years after birth, regimes 2–7. Differences between more recent and previous regime (p-values are given in brackets).

	Section			
Regime Pair	A	B	C	Observations
Extension of job protection and benefit payments				
2/3	1,874	21,430	5,631	265,019
	[0.276]	[0.000]	[0.000]	
3/4	-2,012	8,163	-3,428	145,892
	[0.270]	[0.000]	[0.034]	
4/5	3,485	16,080	13,151	123,584
	[0.105]	[0.000]	[0.000]	
5/6	3,588	13,748	3,752	154,747
	[0.074]	[0.000]	[0.045]	
Extension of job protection alone				
6/7	1,174	10,632	-2,077	169,374
	[0.541]	[0.000]	[0.285]	

Notes: The table shows the differences in the losses in cumulative earnings between the more recent and the previous parental leave regime. Positive values indicate mothers in the new regime are relatively better off. The groups are defined as illustrated in Figure 2.2, i.e. in group A mothers in both regimes return while having job protection, in group B only mothers in the more recent regime have job protection and in group C no mother returns under job protection. Differences are obtained from regressions as specified in equation (2.2) controlling for average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) and age. *Source:* Own estimations using the IEB data described in section 2.3.

Extending the Time Horizon Our results are reported for the seventh year after childbirth. For this year, we expect that mothers already have made the most relevant labor supply decisions after childbirth. They re-entered the labor market after some time of parental leave and the child started kindergarten and, later on, primary school which both potentially change labor supply constraints with respect to childcare. It could still be possible that focusing on the seven years after childbirth misses additional long-run effects, for instance disturbances when a second child is born.

In Table 2.6, we therefore repeat the pooled estimation from Table 2.3 and extend the observation period to the ten years that follow the first childbirth. We find that ten years after birth indeed additional mothers have re-entered the labor market. This, as well as the additional time, lead to results that are quantitatively different but continue to show the pattern that we find substantial gains for mothers from section B returning under the newer parental leave regime while the differences between the new and the old regime for sections A and C are insignificant or in comparison smaller. For our main specification, parental leave regimes 6 and 7, we find that mothers who returned under job protection have accumulated Euro 10.6 K higher earnings than mothers who took similar amounts of leave but returned while not being covered by job protection any more. For sections A and C, where mothers in both regimes face similar policy constraints when returning, the differences are smaller in size by up to a factor of ten and not statistically significant.

2.6 Conclusion

In this paper, we reassess the effect of extending parental leave on the long-run earnings of mothers after their first childbirth. The descriptive relationship between longer parental leave taking and earnings is, in some ways, similar to the effects of extending unemployment benefits[13], i.e. that increased coverage of a wage replacement payment delays return to the labor market and decreases earnings. Although this finding is in line with the theoretical prediction that during longer work interruptions human capital depreciates, most existing studies on parental leave conclude only small effects in the long run. We exploit multiple reforms that extended parental leave coverage in Germany to empirically disentangle the two policy instruments parental leave consists of, job protection and parental benefit payments. This enables us to show that, contrary to unemployment insurance, the job protection instrument of parental leave serves as a mechanism that compensates for parts of the losses mothers incur during longer labor market breaks.

Holding constant the duration of parental leave, we show that around a reform mothers who gave birth under the new, more generous policy regime experience significantly smaller earnings losses. Strikingly, such policy effects are only observable for mothers who return after a parental leave period that is longer than the old regime's coverage period but shorter than the new regime's coverage period. The development of earnings for mothers who return after parental leave shorter than the old regime's coverage is virtually identical across regimes. Focusing on reforms that either extend job protection or benefit payments alone shows that the positive effect on earnings can solely be attributed to extensions of job protection. Increasing the duration of parental benefit payments has no impact on earnings. The magnitude of the effect of job protection is sizeable. In our preferred specification that doubles the duration of job protection from 18 to 36 months the losses in annual earnings seven years after childbirth of mothers affected by the policy change are smaller by Euro 1.1 K or 5.9 percent of mean annual

[13] See, for instance, Schmieder, von Wachter, and Bender (2016) for evidence from Germany and Schmieder and von Wachter (2016) for an overview

earnings. The losses in cumulative earnings over the seven-year period after the first birth are smaller by Euro 9.5 K or 9.7 percent of the mean.

We explain our findings by showing that eligibility to prolonged job protection increases the likelihood that mothers return to their pre-birth employer and, more generally, that they return to employers that pay higher average wages. Nevertheless, employer continuity does not serve as the sole explanation. Differentiating between mothers who stayed with their pre-birth employer and those who switched to a different employer, we find relative gains in annual and cumulative earnings, average firm wage and establishment fixed effect for mothers who return under job protection in both groups. In addition to fostering the continuation of existing employment relationships, this underlines that job protection also improves the outside option of mothers who change their employer. This second channel is larger in relative terms and especially important because the average earnings of mothers who change their employer after childbirth (even though they return to the labor market while still covered by job protection) are lower than for their staying counterparts. This implies that these mothers are willing to bear some costs of switching, for instance because they have to accommodate their employer choice to constraints regarding time with the family, commuting or childcare availability.[14] Our findings underline that, in addition to fostering the continuation of existing employment relationships, job protection plays an important role for mothers who not directly make use of it.

Even though we use a reform in 1992 as our main identifying variation, our findings provide important policy implications. They show that for a policy aim of allowing mothers time with their newborns while supporting them in continuing a career in the labor market, job protection is a sufficient policy instrument. Parental benefit payments are of negligible direct relevance for labor market outcomes (in line with Dahl et al. 2016). Nevertheless, job protection measures have received comparably little attention over the past years. In Germany, there have been no reforms to job protection since the extension in 1992 while other countries provide only relatively short durations. The United States, for instance, offer 12 weeks through the Family and Medical Leave Act (FMLA) which, however, does not cover all employees (Rossin-Slater 2017). Our results point out that, especially in countries that provide relative short amounts of leave, an increase in the job protection period can increase mothers' choice sets while also offsetting some of the earnings losses from prolonged labor market breaks. As interruptions of work are still costly, additional focus has to be put on further mitigating measures such as the widespread provision of care for young children or, to directly counteract human capital depreciation, on training on-the-job that focuses on mothers.

[14] In line with this assumption, Lafférs and Schmidpeter (2021) find for Austria that only a small set of mothers who are mobile and put a greater value on their careers profit from switching the firm after childbirth.

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens (2017). "The Career Costs of Children". In: *Journal of Political Economy* 125 (2), pp. 293–337.

Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016). "Parenthood and the Gender Gap in Pay". In: *Journal of Labor Economics* 34 (3), pp. 545–579.

Angrist, Joshua D. and Guido W. Imbens (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62 (2), pp. 467–475.

Bailey, Martha J., Tanya S. Byker, Elena Patel, and Shanthi Ramnath (2019). "The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from U.S. Tax Data". NBER Working Paper No. 26416.

Baker, Michael and Kevin Milligan (2008). "How Does Job-Protected Maternity Leave Affect Mothers' Employment?" In: *Journal of Labor Economics* 26 (4), pp. 655–691.

Bana, Sara H., Kelly Bedard, and Maya Rossin-Slater (2020). "The Impacts of Paid Family Leave Benefits: Regression Kink Evidence from California Administrative Data". In: *Journal of Policy Analysis and Management* 39 (4), pp. 888–929.

Bellmann, Lisa, Ben Lochner, Stefan Seth, and Stefanie Wolter (2020). "AKM effects for German labor market data". In: *IAB FDZ Methodenreport* 01/2020.

Bergemann, Annette and Regina T. Riphahn (2023). "Maternal employment effects of paid parental leave". In: *Journal of Population Economics* 36, pp. 139–178.

Bertrand, Marianne (2020). "Gender in the Twenty-First Century". In: *American Economic Review: Papers & Proceedings* 110, pp. 1–24.

Blau, Francine D. and Lawrence M. Kahn (2013). "Female Labor Supply: Why Is the United States Falling Behind?" In: *American Economic Review: Papers & Proceedings* 103 (3), pp. 251–256.

– (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature* 55 (3), pp. 789–865.

Brenøe, Anne A., Serena P. Canaan, Nikolaj A. Harmon, and Heather N. Royer (2020). "Is Parental Leave Costly for Firms and Coworkers?" NBER Working Paper No. 26622.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace Heterogeneity and the Rise of West German Wage Inequality". In: *The Quarterly Journal of Economics* 128 (3), pp. 967–1015.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes (2016). "What is the Case for Paid Maternity Leave?" In: *The Review of Economics and Statistics* 98 (4), pp. 655–670.

Dustmann, Christian and Uta Schönberg (2012). "Expansions in Maternity Leave Coverage and Children's Long-Term Outcomes". In: *American Economic Journal: Applied Economics* 4 (3), pp. 190–224.

Frodermann, Corinna, Alexandra Schmucker, Stefan Seth, and Philipp vom Berge (2021). "Sample of Integrated Labour Market Biographies (SIAB) 1975–2019". In: *IAB FDZ Datenreport* (01|2021).

Ginja, Rita, Jenny Jans, and Arizo Karimi (2020). "Parental Leave Benefits, Household Labor Supply, and Children's Long-Run Outcomes". In: *Journal of Labor Economics* 38 (1), pp. 261–320.

Ginja, Rita, Arizo Karimi, and Pengpeng Xiao (2023). "Employer Responses to Family Leave Programs". In: *American Economic Journal: Applied Economics* 15 (1), pp. 107–135.

Huebener, Mathias, Jonas Jessen, Daniel Kuehnle, and Michael Oberfichtner (2021). "A Firm-Side Perspective on Parental Leave". IZA Discussion Paper No. 14478.

Kiessling, Lukas, Pia Pinger, Philipp Seegers, and Jan Bergerhoff (2019). "Gender Differences in Wage Expectations: Sorting, Children, and Negotiation Styles". CESifo Working Paper No. 7827.

Kleven, Henrik, Camille Landais, Johanna Posch, Adreas Steinhauer, and Josef Zweimüller (2019). "Child Penalties across Countries: Evidence and Explanations". In: *American Economic Review: Papers & Proceedings* 109, pp. 122–126.

– (2021). "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation". NBER Working Paper No. 28082. URL: `https://www.henrikkleven.com/research.html` (visited on 04/03/2023).

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). "Children and Gender Inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11 (4), pp. 181–209.

Kluve, Jochen and Sebastian Schmitz (2018). "Back to Work: Parental Benefits and Mother's Labor Market Outcomes in the Medium Run". In: *Industrial and Labor Relations Review* 71 (1), pp. 143–173.

Kowalski, Amanda (2020). "Counting Defiers: Examples from Health Care". URL: `https://sites.lsa.umich.edu/amanda-kowalski/publications/` (visited on 04/03/2023).

Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington (2018). "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?" NBER Working Paper No. 24740.

Lafférs, Lukáš and Bernhard Schmidpeter (2021). "Mothers' job search after childbirth". IZA Discussion Paper No. 14593.

Lalive, Rafael, Analía Schlosser, Adreas Steinhauer, and Josef Zweimüller (2014). "Parental Leave and Mothers' Careers: The Relative Importance of Job Protection and Cash Benefits". In: *The Review of Economic Studies* 81, pp. 219–265.

Lalive, Rafael and Josef Zweimüller (2009). "How Dows Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments". In: *The Quarterly Journal of Economics* 124 (3), pp. 1363–1402.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rassmussen (2017). "Can Women Have Children and a Career? IV Evidence from IVF Treatments". In: *The American Economic Review* 107 (6), pp. 1611–1637.

Müller, Dana, Andreas Filser, and Corinna Frodermann (2022). "Update: Identifying mothers in administrative data". In: *IAB FDZ Methodenreport* 01/2022.

Olivetti, Claudia and Barbara Petrongolo (2016). "The Evolution of Gender Gaps in Industrialized Countries". In: *Annual Review of Economics* 8, pp. 405–434.

– (2017). "The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries". In: *Journal of Economic Perspectives* 31 (1), pp. 205–230.

Raute, Anna (2019). "Can financial incentives reduce the baby gap? Evidence from a reform maternity leave benefits". In: *Journal of Public Economics* 169, pp. 203–222.

Rossin-Slater, Maya (2017). "Maternity and Family Leave Policy". NBER Working Paper No. 23069.

Ruhm, Christopher J. (1998). "The Economic Consequences of Parental Leave Mandates: Lessons from Europe". In: *The Quarterly Journal of Economics* 113 (1), pp. 285–317.

Schmieder, Johannes F. and Till von Wachter (2016). "The Effects of Unemployment Insurance Benefits: New Evidence and Interpretation". In: *Annual Review of Economics* 8, pp. 547–581.

Schmieder, Johannes F., Till von Wachter, and Stefan Bender (2016). "The Effect of Unemployment Benefits and Nonemployment Durations on Wages". In: *The American Economic Review* 106 (3), pp. 739–777.

Schönberg, Uta (2009). "Does the IAB employment sample reliably identify maternity leave taking? A data report". In: *Zeitschrift für Arbeitsmarktforschung* 42 (1), pp. 49–70.

Schönberg, Uta and Johannes Ludsteck (2014). "Expansions in Maternity Leave Coverage and Mother's Labor Market Outcomes after Childbirth". In: *Journal of Labor Economics* 32 (3), pp. 469–505.

Thomas, Mallika (2021). "The Impact of Mandated Maternity Leave Policies on the Gender Gap in Promotions: Examining the Role of Employer-Based Discrimination". URL: `https://sites.google.com/site/mallikathomas/research` (visited on 04/03/2023).

Tô, Linh T. (2021). "The Signaling Role of Parental Leave". URL: `https://linh.to/` (visited on 04/03/2023).

Appendix

2.A Additional Figures and Tables

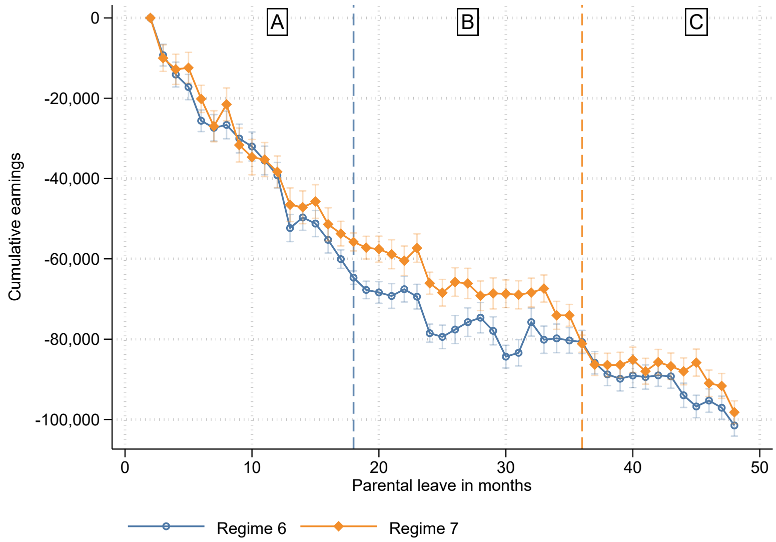


FIGURE 2.10: Effect of extending job protection from 18 to 36 months on cumulative earnings for the seven years after birth, regime pair 6/7.

Notes: The figure plots the results of regressions of annual earnings cumulated over the seven years after childbirth on the length of the post-birth labor market break in months. The regressions are carried out separately for regimes 6 and 7 and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as indicators for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

Table 2.7: Length (in months) and amount (in DM) of additional benefit payments from federal states by regime.

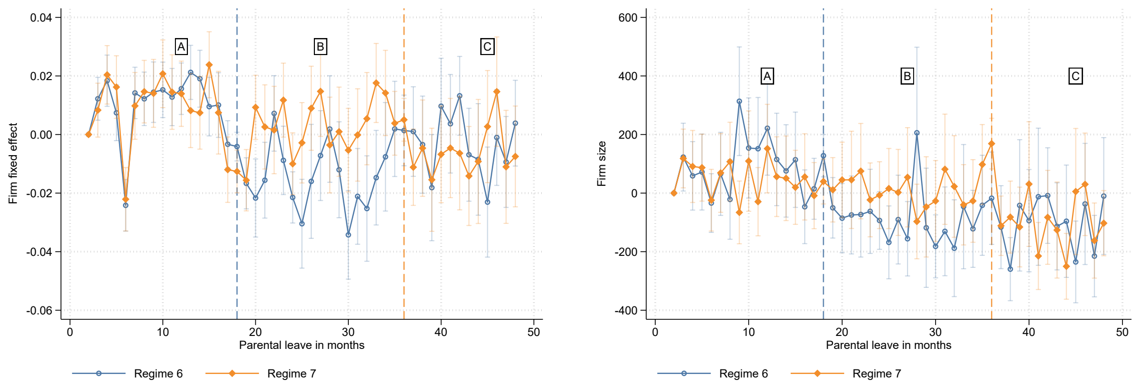
Regime	Births since	Benefit Duration	Benefit Amount
Baden-Württemberg			
I	1986-01-01	12	400
Bavaria			
I	1989-01-01	6	500
II	1993-01-01	12	500

Notes: Representation based on the relevant laws at the state level for Baden-Württemberg and Bavaria (*Landeserziehungsgeldgesetz*). In the early 2000s benefit amount were changed to Euro values. Baden-Württemberg abolished its extra benefits payments in 2012, Bavaria replaced it 2018.

Table 2.8: Pooled estimation results of length of parental leave taking on annual earnings seven years after birth, regimes 2–7. Differences between more recent and previous regime (p-values are given in brackets).

	Section			
Regime Pair	A	B	C	Observations
Extension of job protection and benefit payments				
2/3	222	1,350	537	263,311
	[0.269]	[0.000]	[0.000]	
3/4	15	488	-476	143,727
	[0.945]	[0.073]	[0.014]	
4/5	-162	835	231	111,314
	[0.530]	[0.006]	[0.332]	
5/6	359	1,407	753	127,638
	[0.141]	[0.000]	[0.001]	
Extension of job protection alone				
6/7	205	1,123	-48	147,346
	[0.381]	[0.000]	[0.853]	

Notes: The table shows the differences in the earnings losses between the more recent and the previous parental leave regime. Positive values indicate mothers in the new regime are relatively better off. The groups are defined as illustrated in Figure 2.2, i.e. in group A mothers in both regimes return while having job protection, in group B only mothers in the more recent regime have job protection and in group C no mother returns under job protection. Differences are obtained from regressions as specified in equation (2.2) controlling for average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) and age. *Source:* Own estimations using the IEB data described in section 2.3.

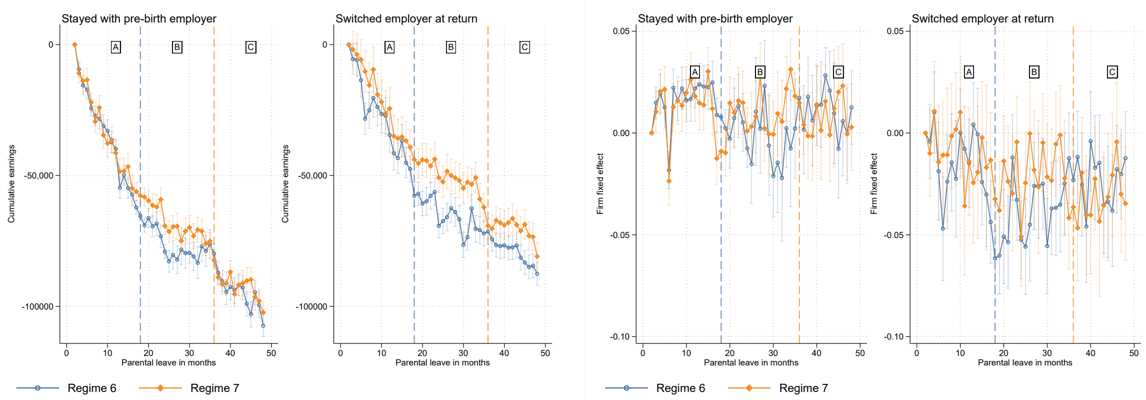


(A) Establishment fixed effect.

(B) Establishment size.

FIGURE 2.11: Effect of extending job protection from 18 to 36 months on establishment fixed-effects and establishment size seven years after birth (regime pair 6/7).

Notes: The figure plots the results of regressions of the establishment fixed-effect, calculated for the period 1985–1992 (Card, Heining, and Kline 2013; Bellmann et al. 2020), as well as establishment size on the length of the post-birth labor market break in months. The regressions are carried out separately for regimes 6 and 7 and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as for indicators for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

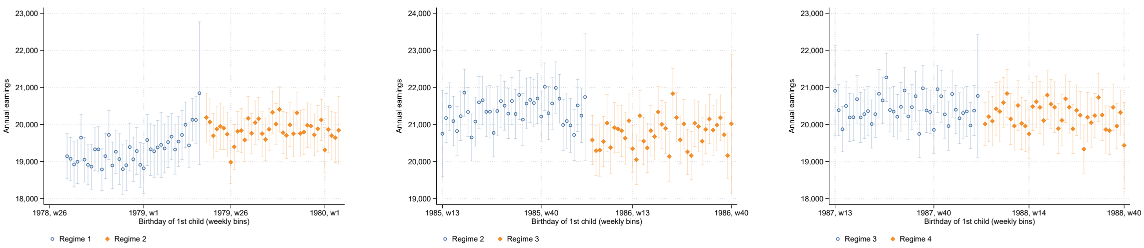


(A) Cumulative earnings.

(B) Establishment fixed effect.

FIGURE 2.12: Effect of extending job protection from 18 to 36 months on cumulative earnings and establishment fixed effects seven years after childbirth separately for mothers who stay with their pre-birth employer (left-hand side) and who switch the employer (right-hand side) (regime pairs 6/7).

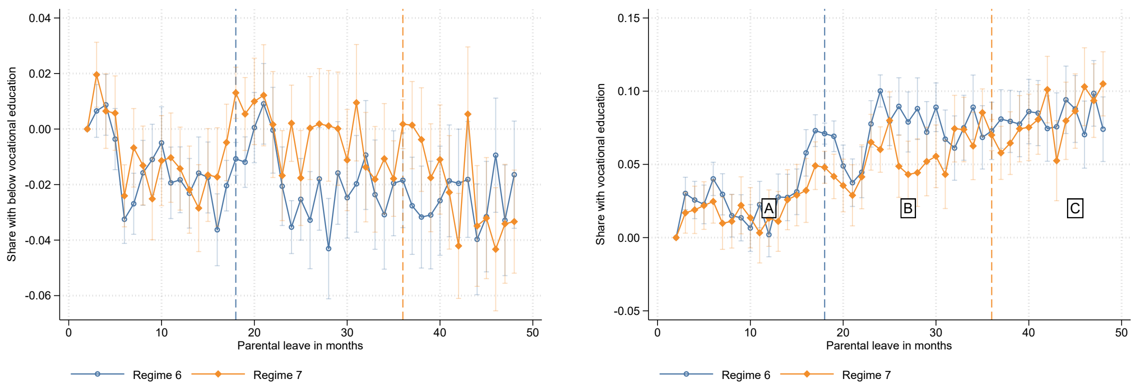
Notes: The figure plots the results of regressions of cumulative earnings (left-hand panel) and establishment fixed effects, calculated for the period 1985–1992 (Card, Heining, and Kline 2013; Bellmann et al. 2020) (right-hand panel), on the length of the post-birth labor market break in months, differentiating between mothers who returned to their pre-birth employer and those who switched to a different employer. The regressions are carried out separately for regimes 6 and 7 and control for the average earnings in the three years prior to childbirth and indicators for education, 2-digit occupation, federal state (all prior to birth) as well as for indicators for age and year. *Source:* Own estimations using the IEB data described in Section 2.3.



(A) Regimes 1/2 (B) Regimes 2/3 (C) Regimes 3/4

FIGURE 2.13: Average annual earnings seven years after birth by birthday (in weekly bins) around parental leave reforms (regimes 1–4).

Notes: The figure plots the results of regressions of average annual earnings seven years after childbirth on the week of birth of the first child along with 95 percent confidence intervals. The regressions are carried out separately by parental leave regime for windows of nine months around the reform date. *Source:* Own estimations using the IEB data described in section 2.3.



(A) Share with no education/below vocational education. (B) Share with vocational education.

FIGURE 2.14: Pre-birth characteristics of mothers by length of parental leave (regime pair 6/7).

Notes: The figure plots the results of regressions of the respective outcomes on the length of the post-birth labor market break in months. The regressions are carried out separately for regimes 6 and 7 and control for indicators for age and year. *Source:* Own estimations using the IEB data described in section 2.3.

3 Child Penalty Estimation and Mothers' Age at First Birth

Joint with Valentina Melentyeva (University of Cologne)

Motherhood remains costly for women's careers and a precise assessment of its effects is crucial to understand the remaining gender inequality, both in the labor market and beyond. Recently, child penalty estimations based on event studies have become a popular tool to assess losses in career outcomes after the first childbirth. We show that mothers' outcomes and characteristics are heterogeneous by the age at which they give their first childbirth. Common event studies pool all mothers together and, therefore, are prone to suffer from biases as discussed in the literature on static and dynamic difference-in-differences models. Since existing heterogeneity-robust estimators do not offer solutions that apply to childbirth as treatment, we propose a novel approach to estimate child penalties that combines a "stacked" difference-in-differences estimation with a restricted set of control groups. Applying it to German administrative data, we show that the conventional approach substantially underestimates the child penalty as it does not take unrealized earnings growth after childbirth into account. We, further, confirm large heterogeneity in child penalties by age at first birth and provide new insights on their extent for cumulative earnings and occupational ranks.

3.1 Introduction

The past century has been marked by major successes of the women's rights movement. In developed countries, women gained—among other achievements—widespread and unrestricted access to education and the labor market. Nevertheless, gender inequality remains prevalent in the labor markets around the world and one particular reason for that persists to these days: motherhood is still costly for women's careers. This has been shown by multiple theoretical and empirical studies, which highlight the crucial role of gender differences in parenthood costs for the remaining gender inequality in labor market outcomes.[1] Correctly tracking the dynamics of the career costs of motherhood and studying the mechanisms behind them is crucial to understanding gender inequality and giving informed policy advice. In this paper, we show that the most popular approach to empirically estimate child penalties is biased if the effects of childbirth are heterogeneous by mothers' age at first birth. We propose a novel solution to address this issue and apply it to study how motherhood affects women's labor market outcomes depending on the timing of their first birth.

The career costs of motherhood have been both subject of research and a recurring topic in public debates for a long time. Just recently, the empirical approach based on event studies—estimating so called child penalties—received widespread attention as it provides a straightforward and intuitive way to visualize the career impact of childbirth. The paper by Kleven, Landais, and Søgaard (2019) popularized the method and gained more than 1,000 citations over the span of three years. Kleven, Landais, and Søgaard crucially contributed to an understanding of the challenges of combining motherhood and career among a wide range of audiences. Researchers have been actively using the method to estimate gender inequality in the labor market, within and across countries, and to evaluate policies. However, the effects of motherhood are heterogeneous across women who give birth at different age. As existing literature shows (see, for instance, Adda, Dustmann, and Stevens 2017; Goldin, Kerr, and Olivetti 2022) and as we confirm, the age at which a woman gives birth to her first child is highly correlated with both pre- and post-birth outcomes as well as human capital levels and other characteristics that are relevant in the labor market. We demonstrate that this heterogeneity introduces biases in event-study-based estimations of child penalties and undermines the validity of the control groups that are commonly used in such estimations.

Building on the emerging literature on static difference-in-differences (DiD) models and their dynamic extensions (see the summaries by J. Roth et al. 2022; de Chaisemartin and D'Haultfoeuille 2022), we show that event studies with childbirth as treatment and heterogeneous treatment effects are prone to yield estimates which do not necessarily reflect the actual impact of childbirth. The underlying problems are typically coined as "forbidden comparisons" and "contamination". "Forbidden comparisons" means that observations from already treated units are employed as control group to proxy for a state without treatment. Applied to the childbirth context, "forbidden comparisons" happen when mothers who have given birth end up as part of the control group. Even though, they are experiencing the costs of motherhood, they are used as counterfactual as if they had no child. "Contamination" only applies to multi-period settings such as event studies. It means that the estimate for each relative time period around an event can not only contain the treatment effect at this specific

[1] See, for example, the review by Andrew et al. (2021), the book by Goldin (2022) and the papers by Adda, Dustmann, and Stevens (2017), Goldin, Kerr, and Olivetti (2022), and Blundell et al. (2021).

period but also the treatment effects from all other periods. In other words, estimates from the post-birth period can be informed by treatment effects from the pre-birth period. The consequence for conventional child penalty event studies is that estimates are likely biased and pre-trends are not informative about parallel trends. These issues can lead both to an over- or underestimation of child penalties, depending on the degree and pattern of heterogeneity and on the composition of a given sample.

The problems of DiD and TWFE models have been addressed by newly developed estimators that are robust to heterogeneity. Their application, however, requires employing very specific control groups; either units that are the last to receive treatment, that have not-yet received treatment or that are never treated. Given the differences among mothers by their age at first birth and the selection into having children, the validity of these control groups hinges on assumptions that are unlikely to hold with childbirth as treatment. We further illustrate that, depending on the control group, child penalty estimates exhibit a great amount of variation across estimators which is driven by the heterogeneity among mothers. Those who have their child early in life are very different from those who have their child later, while childless women are different from mothers.

Therefore, we combine the insights from our discussion of issues in event studies to present a new approach to estimate child penalties. We propose to use a "stacked" DiD design that estimates the effects of childbirth separately for each cohort of mothers (i.e. for each age at first childbirth). In addition, we employ restricted control groups that are specific for each cohort. They consist only of pre-birth observations from not-yet-treated mothers who give birth at slightly older ages. Exploiting the strong correlation between age at first birth and labor market outcomes, this ensures the comparability of treatment and control group for each cohort. It prevents comparisons of mothers with other mothers who are too different and with childless women who are a selected group. Our approach builds on work by Cengiz et al. (2019) who propose "stacking" and Callaway and Sant'Anna (2021) who suggest using not-yet-treated units as control group (though, without making further restrictions that are necessary for childbirth as treatment). It allows us to estimate the unbiased and causal effects of childbirth on post-birth labor market outcomes.

We apply our approach to reassess child penalties in Germany. For cohort-specific effects, we find the absolute earnings losses after childbirth increase strongly in age at first birth which is driven by the higher earnings of older mothers. Child penalties in relative terms show substantially less variation across cohorts. On average, we estimate a child penalty of −85 percent which is 15 percentage points larger than what the conventional approach finds. The underestimation in conventional event studies is largely due to the fact that they cannot capture the unrealized growth of earnings after childbirth since its control group includes already treated mothers. We, additionally, apply our approach to estimate child penalties in cumulative earnings and occupational rank. For cumulative earnings, we find again large differences between cohorts, both in absolute and relative terms. Mothers who give birth at age 26 loose almost 100 percent of their pre-birth cumulative earnings until year four after childbirth, whereas mothers who give birth at 32 loose less than 50 percent. The effects of childbirth on occupational ranks are comparatively small. We interpret them as evidence against widespread occupational downgrading of mothers but in favor of a slowdown in career progress.

Our paper makes multiple contributions. First, to the large literature assessing the effects of motherhood on women's careers using event studies and related methods (Angelov, Johansson, and Lindahl 2016; Kuziemko et al. 2018; Bütikhofer, Jensen, and Salvanes 2018; Fitzenberger, Sommerfeld, and Steffes 2013; Kleven, Landais, Posch, et al. 2019; Bruns 2019; Andresen and Nix 2022). We highlight that age at first childbirth is an important measure for the career orientation of women due to its high correlation with both static measures such as education and the career progress realized until childbirth. Building on this, we show that common event studies are prone to biases and propose a new approach to estimate unbiased, causal effects of childbirth. We use this new approach to show that the post-birth earnings losses mothers incur differ substantially by the age at which they give birth. Our new estimator can be applied to analyze impacts of childbirth, gender earnings inequality (Blau and Kahn 2017) as well the effects of policy reforms. For the latter, it can be particularly useful to assess cohort-specific effects since some policies—for instance expansions in public childcare (Krapf, A. Roth, and Slotwinski 2020; Kleven, Landais, Posch, et al. 2021) that often affect lower-earning women—can have effects that are not uniformly distributed by age at first birth.

Second, we build on the recently emerging literature on DiD and TWFE models. We provide an example of childbirth as a case that is common in the empirical literature and of high policy relevance but where we are confronted with issues stemming from heterogeneous treatment effects as discussed by Goodman-Bacon (2021), de Chaisemartin and D'Haultfoeuille (2020), Sun and Abraham (2021), and Callaway and Sant'Anna (2021) (see also the summaries by de Chaisemartin and D'Haultfoeuille 2022; J. Roth et al. 2022). In addition, childbirth is also an example case where the existing heterogeneity-robust estimators are not fully applicable. Building on the existing findings, we therefore propose a novel approach that addresses the peculiarities of the childbirth setting.

The paper is organized as follows. Section 3.2 gives a brief overview of the datasets we use, Section 3.3 provides an overview of the heterogeneity of outcomes and characteristics among different cohorts of mothers, Section 3.4 discusses the issues with the conventional approach to child penalty estimation. In Section 3.5, we suggest a new solution to estimate child penalties and apply it in Section 3.6, Section 3.7 concludes.

3.2 Data

This paper uses survey and administrative data from Germany as both types of data have their strengths that complement each other. The survey data from the German Socio-Economic Panel (Soep) provide the greater level of detail and more characteristics while the Sample of Integrated Labor Market Biographies (Siab) provides the larger sample size along with the precision of administrative data. This section describes both datasets.

The German Socio-Economic Panel (Soep) is a well-established panel study that started in 1984 (Goebel et al. 2019) and surveys around 12,000 households and their members each year. Along with detailed socio-demographic information it provides data on labor force status, labor earnings, working hours, occupations as well as on the household context of mothers. Importantly, it also records full birth histories that allow to identify mothers and when they have given birth. We use the Soep data from the period 1984 until 2020 to provide a systematic overview of the heterogeneity in outcomes and characteristics for mothers by age at childbirth in section 3.3 and for some complementary analyses.

The Sample of Integrated Labor Market Biographies (Siab) is provided by the Institute for Employment Research (IAB). It is a two percent sample drawn from the universe of German workers who are subject to social security contributions (i.e. individuals in self-employment and civil servants are not covered). It includes administrative records of individual labor market biographies of nearly 120,000 mothers for the period 1975 to 2019. It further provides information on employers, occupation and wages. From the latter we construct annual earnings. Since the data are taken from employers' reports to the social security system they have some shortcomings. The main two of them are that, first, wages are only recorded up to the threshold for social security contributions. For wages above that ceiling we apply an imputation method that follows Dauth and Eppelsheimer (2020) who build on work by Dustmann, Ludsteck, and Schönberg (2009) and Card, Heining, and Kline (2013). Second, births cannot be observed directly but have to be imputed following Müller, Filser, and Frodermann (2022). This imputation utilizes the maternity protection period around childbirth that mandates an employment break of at least 14 weeks. Müller, Filser, and Frodermann show that their method identifies around 60 percent of all births in Germany. Since it is applied to smaller subset of births by women in employment who are subject to social security contributions the share of identified births in our sample will be larger. We use this dataset to illustrate "forbidden" comparisons in section 3.4 and in the application of the new approach to estimate child penalties which we propose in section 3.5.

Tables 3.2 and 3.3 in the Appendix provide summary statistics for both datasets. With respect to the key characteristic age at first childbirth the data from the Soep (29 years) and the Siab (28.6 years) give virtually identical values. Similarly, the values for earnings in the pre-birth year are close together (Euro 26.1 K for the SOEP and Euro 23.9 K[2] for the SIAB). The finding of larger earnings in the SOEP data is in line with expectations as the SOEP also includes women in self-employment and civil servants, i.e. two groups who are typically farther up in the earnings distribution.

We use both datasets according to their respective strengths. As the SOEP provides more detailed information on mothers' characteristics, we primarily use it to illustrate how educational levels, occupational rank or working hours of mothers differ by their age at first birth (in most applications grouped into quartiles of age at first birth). Leveraging the larger number of observations in the SIAB allows us to precisely show how the effect of childbirth on maternal earnings changes with increasing age at first birth and to conduct a cohort-specific analysis of child penalties. The findings of this paper are not specific to either dataset. They are rather driven by the fact that mothers exhibit substantial differences in various characteristics depending on their age at the first childbirth (see next section). Our results, including those in section 3.5 where we present an alternative method to assess post-birth earnings losses, are qualitatively and quantitatively similar in both datasets.

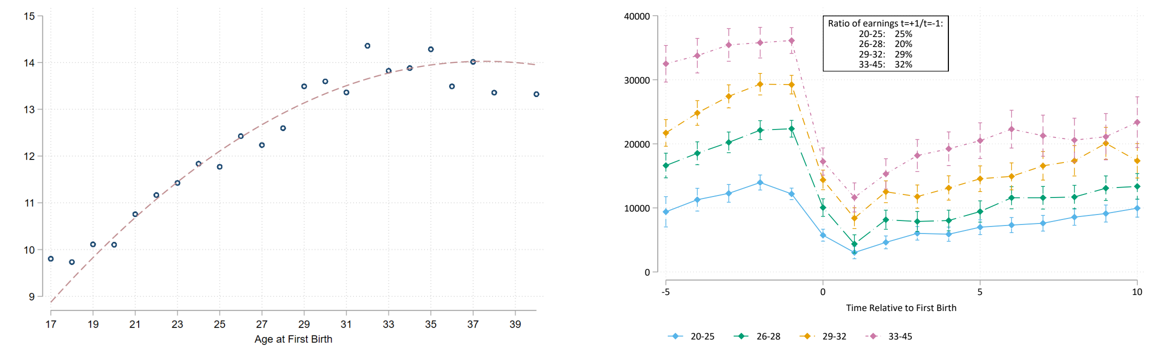
3.3 The Source of Problems: Heterogeneity by Age at First Birth

As this paper largely builds on the observation that women who are older when they become mothers are very different from those who have their first child early in life. This section provides detailed descriptive evidence on the heterogeneous composition of mothers and how this heterogeneity is related to age at first birth. Different types of heterogeneity between younger and older mothers and childless women have been mentioned in many economic papers (Goldin, Kerr, and Olivetti 2022; Wilde, Batchelder, and Ellwood 2010; Adda, Dustmann, and Stevens 2017). Nevertheless, we provide a systematic analysis of the across-cohorts differences in outcomes and relevant covariates to be able to fully support our reasoning on their consequences for the estimation of child penalties.

To start with, we show that late mothers tend to have higher levels of education (Figure 3.1a)—the outcome which is usually decided on during the early stages of career paths and is related to desired fertility (Adda, Dustmann, and Stevens 2017; Doepke et al. 2022). We observe that first-time mothers who are older than 30 on average have completed higher education, while women who become mothers before 25 are more likely to only hold a high school degree.

Furthermore, late mothers have on average a smaller number of children over life than early ones (Figure 3.15 in the Appendix). First-time mothers younger than 25 tend to have more than two children, while those older than 35 are more likely to have one child in total. This

[2] All monetary values are in real terms for the base year 2015.



(A) Average total years of education of mothers by their age at first childbirth.

(B) Average earnings of mothers around birth for four quantiles of age at first birth.

FIGURE 3.1: Heterogeneity in education and earnings among mothers by their age at first birth.

Notes: The left-hand figure shows years of education of mothers by their age at first childbirth as a binned scatter plot with an added quadratic fit. The figure plots average annual labor earnings of mothers (including zero earnings) in time relative to their first birth for four quantiles of the distribution of age at first birth. The first quartile includes mothers aged 20–25 at first birth, the second those age 26–28, the third those aged 29–32 and the fourth those from 33–45. *Source:* Own calculations based on the SOEP.

also means that interpretation of effects estimated around the first childbirth differs for these groups of mothers as the estimates for the younger cohorts also capture the effect of having additional children.

With respect to labor market outcomes, we document in Figure 3.13 in the Appendix that late mothers work more hours and in positions with higher occupational rank, both before and after the childbirth, and are more likely to return to the labor market after becoming mothers. Women who delay the timing of their first birth also tend to have a shorter parental leave break, returning to work faster (see Figure 3.11 in the Appendix).

The heterogeneity described above translates into mothers having different earnings trajectories depending on their age at childbirth. As shown in Figure 3.1b and Figure 3.12 in the Appendix, later mothers tend to have higher pre- and post-birth levels of earnings, larger magnitudes of drops in both absolute and percentage terms, and faster recovery growth rates of their post-birth earnings. These differences also hold if we restrict the sample only to those women who continue to work after the childbirth (Figure 3.13 in the Appendix). These results are in line with Wilde, Batchelder, and Ellwood (2010) who show similar correlations between wages of mothers and childbirth timing.

Overall, we observe that, on average, older fist-time mothers exhibit significantly better labor market outcomes and levels of human capital than younger ones, both before and after the childbirth. These results provide a suggestive evidence that effects of motherhood are potentially heterogeneous depending on age at which women give birth to their first child and that the effects are likely to change over time. As we show in the following section, these differences become the source of biases when estimating child penalties using event study regressions.

3.4 Child Penalty Estimation under Heterogeneity by Age at Birth

In this section, we explain how heterogeneity by age at birth poses a threat to estimating child penalties since it is not just one of many characteristics of mothers but also a timing dimension. As Goodman-Bacon (2021) and Sun and Abraham (2021) have shown, if treatment effects differ across cohorts or if they just change over time, these differences enter the DiD estimates and bias the results. First, we show how these issues materialize in child penalty estimations and lead to significant biases. Second, we discuss how the heterogeneity of treatment effects invalidates the commonly-used control groups and makes new heterogeneity-robust estimators inapplicable to the settings with childbirth as treatment.

3.4.1 The Conventional Approach to Estimate Child Penalties

We start with presenting the common event study setup on which existing child penalty estimations are based. Child penalties aim to quantify and visualize the losses women experience with regard to some outcome—typically earnings—following the birth of their first child. Their estimation usually starts with an event study regressing the outcome on a set of event time dummies and additional control variables. A next step that re-scales the event study coefficients to get percentage changes is not strictly necessary but common.

Most papers that estimate child penalties follow the specification by Kleven, Landais, and Søgaard (2019) (henceforth KLS) who propose to estimate the following regression on a sample of mothers:

$$Y_{it} = \sum_{l,\, l \neq -1} \beta_l \times \mathbb{1}\left[t - t_i^0 = l\right] + \gamma_{it} + \lambda_t + \varepsilon_{it}. \tag{3.1}$$

Here, Y_{it} is the outcome of interest for unit i at calendar time t, t_i^0 is the time when unit i gives the first childbirth. Fixed effects for age (γ_{it}) and calendar time (λ_t) are added to control for life cycle as well as business cycle effects. The coefficients of the relative event time dummies ($\hat{\beta}_l$) then are intended to capture the effect of being l years away from the year of the first childbirth t_i^0 on the outcome of interest. This outcome of interest can be earnings—the one that gains most attention and that most of this paper focuses on—but other continuous (such as wages or working hours) or discrete (for instance, employment status) variables are used as well. In many studies, additional individual fixed effects are added to account for heterogeneity among mothers that is constant over time, including the pre-birth level-differences in outcome.

3.4.2 Child Penalty Estimation as a Case of *Forbidden* Comparisons and *Contamination*

As discussed, among others, by Goodman-Bacon (2021), Sun and Abraham (2021) and Callaway and Sant'Anna (2021), heterogeneity of effects for cohorts treated at different points in time can lead to several issues in the setting of event studies, on which child penalty estimations are based. According to the literature, there are two main sources of heterogeneity-induced biases. The first one are "forbidden" comparisons, in which already-treated units are used as controls, and their post-birth changes in earnings are used as counterfactual trends. Event studies do such comparisons and assign them the weights which depend on sample variance and composition, such that the resulting average can even lie outside of the range of actual effects. Second, in dynamic settings, where effects for multiple periods are estimated, estimates for one relative time period can be "contaminated" by effects from other periods. These problems lead to uninformative pre-trends and biased estimates of treatment effects. We argue that both are likely to arise when using event studies to estimate child penalties.

"Forbidden" comparisons "Forbidden" comparisons are made when a DiD estimator compares units at non-matching points in time. To explain their origin and consequences, we take a step back and consider a static DiD model where only an average treatment effect is of interest. Goodman-Bacon (2021) has shown that a DiD estimator is a variance-weighted average of all possible 2x2 DiD estimators that compare cohorts to each other, i.e. compares each treated group to all other already-treated and non-treated groups. If effects are homogeneous—exactly the same across cohorts and constant over time—the differences between the trends of a treated and the already-treated groups will be zero in each period. Then, the estimate will be unbiased. However, if effects change over time or if they are different across cohorts, the differential trends of the already-treated cohorts will be used as counterfactuals introduce biases.

What does it mean when estimating child penalties? To answer this question, we use a stylized example and illustrate "forbidden" comparisons building on the Goodman-Bacon (2021) decomposition. We take annual earnings as an outcome and the Siab data for this illustration. As simplification, we focus on one treated cohort that gives birth at age 29 and are interested in estimating the effect of childbirth for this cohort for the year 0 (i.e., at age 29) relative to the pre-birth year (at age 28). As further simplification, for illustration purposes, we restrict the sample to one much earlier-treated cohort (that gave birth at age 24), one cohort that gave birth in the previous year at age 28 and one next-treated cohort that will give birth at age 30. The average earnings for these four cohorts over age are plotted in Figure 3.2. It shows that all cohorts exhibit similar earnings trajectories around childbirth: steady growth before birth, large losses in period 0, additional smaller losses in period +1, some recovery in the year

+2 (likely since mothers start re-entering the labor market after maternity leave), and slow growth thereafter.

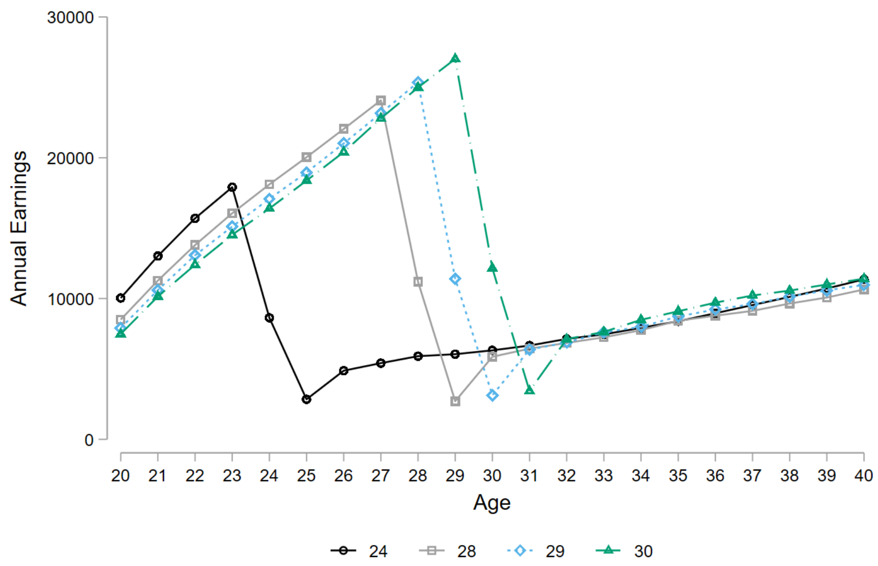


FIGURE 3.2: Average earnings of mothers around birth for four ages at first birth (24, 28, 29, 30)

Notes: The figure plots average annual labor earnings of mothers (including zero earnings for periods of non-participation) by age for four levels of age at first birth (24, 28, 29, 30). *Source:* Own calculations based on the SIAB.

Formally, estimating the treatment effect in year 0 for the chosen treated cohort (giving birth at age 29) means estimating the following regression on a sample that is restricted to the cohorts that give birth at ages 24, 28, 29 and 30 and the age window from 28 to 29:

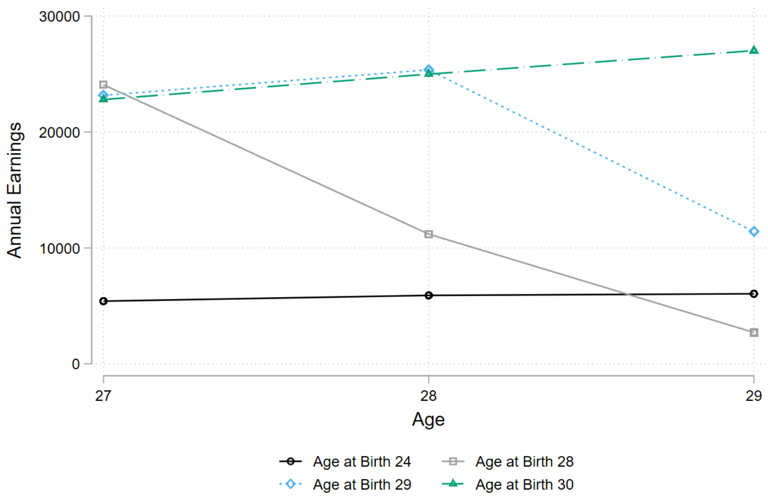
$$Y_{it} = \beta \times \text{treated}_{it} + \gamma_i + \lambda_t + \varepsilon_{it}. \tag{3.2}$$

Here, Y_{it} indicates annual earnings for mother i at age t. Fixed effects for individual (γ_i) and age (λ_t) are included to implement the DiD design, accounting for pre-birth differences between treatment and control groups and the changes in outcomes of the control group over time. treated_{it} is a treatment status indicator, which takes value of 0 if the individual is not-treated yet, switches to 1 when the treatment happens and stays 1 thereafter. The estimate $\hat{\beta}$, then, captures the change in the outcome for the treated group, compared to changes the control group experiences. Estimating this regression using the SIAB sample and the restrictions described above yields the average estimate of Euro −11,562 (see Table 3.1, column "Average").

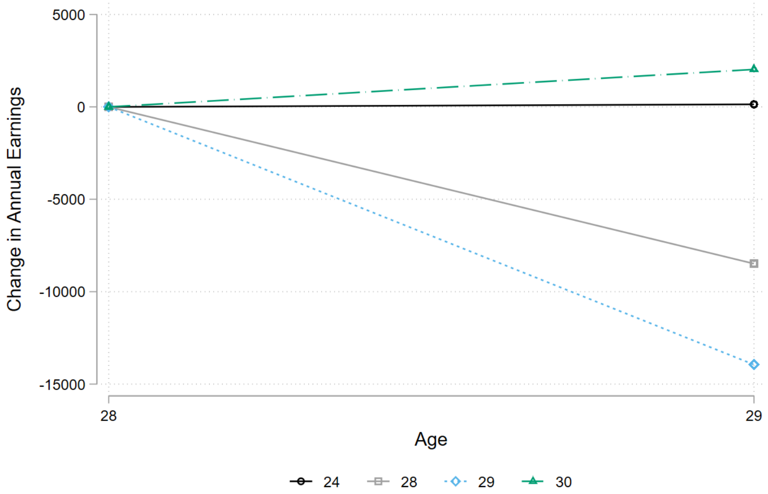
However, decomposing this average DiD estimate shows it is a weighted-average of three comparisons. The cohort giving birth at age 29 is compared to one non-yet-treated cohort that is going to give birth at age 30. This comparison is "clean". It is, further, compared to the already-treated cohorts who gave birth at ages 24 and 28. As a just-treated and two already-treated cohorts are compared, these are "forbidden" comparisons. To see why this happens, we look at the average annual earnings of these four cohorts over estimation window in Figure 3.3a. We observe that the treated cohort (blue line) experiences growth until the age of 28 and then large losses in the year of childbirth (at the age of 29). The not-yet-treated cohort that gives birth at 30 (green line) experiences a steady increase in earnings. The much earlier-treated cohort that gave birth at age 24 (black line) exhibits only very slow earnings growth at age levels 28 and 29. The previously-treated cohort that gave birth at 28 (grey line) continues to face losses at the age of 29, although they become smaller in year after birth compared to the year of childbirth. The comparison that researchers usually intend to make is the one between the treated and the not-treated group—in the figure, between the blue and green line. After the pre-birth differences in levels are taken into account, the trend of the not-yet-treated group is assumed to reflect a counterfactual development of earnings absent children[3]. However, this is not the only comparison that the regression makes.

The main issue is that the regression treats the cohorts that already gave birth at 24 and 28 as control groups, because their treatment dummy (treated_{it}) does not change for them and has the value of 1 in both periods. The coefficient from the treatment dummy ($\hat{\beta}$) captures the response in the outcome to the change in treatment status by 1 unit, which at the age of 29 happens only for the cohort that gives birth at 29. Therefore, all the other cohorts where the treatment status does not change are considered as control groups. In our example, the estimator makes three comparisons that we illustrate in Figure 3.3b (Table 3.1 provides the according estimates). There, we plot the changes in earnings from age 28 to age 29, normalized to the pre-birth period (age 28). The "clean" comparison is the one where the changes in earnings of the treated cohort and the not-yet-treated one are compared (in our example the difference equals Euro −15,976). The forbidden comparisons are the other two that use already-treated mothers as control units. The cohort that gave birth in the previous year (at age 28) is still experiencing losses in its first post-birth year. However, the losses are not constant over time. At age 29, the are smaller than in the year before when the cohort gave birth. If the losses and thus the earnings trajectories were equal over time and equal to the losses of the group treated at age 29, this comparison would yield a difference between the cohorts of zero, hence would not introduce a bias. Instead, since the losses decrease from year 0 to year +1, this comparison yields an estimate of Euro −5,430, i.e. it underestimates the earnings losses for the treated cohort. The second "forbidden" comparison is the one with the cohort that gave birth already at age 24. This cohort exhibits a mostly flat trend in its earnings, highlighting the slow earnings growth of mothers from the second post-birth year onward (see Figure 3.2).

[3] We discuss using not-yet-treated mothers as control group later in this section.



(A) Average earnings in levels for age range 27 to 29.



(B) Changes in average earnings for age range 28 to 29.

Figure 3.3: Average earnings in levels and changes for four ages at first childbirth (24, 28, 29, 30).

Notes: The figure plots average labor earnings for mothers who give birth at age 24, 28, 29 and 30. The upper panel plots earnings in levels, the lower panel plots the changes in earnings between age 28 and 29. Earnings include zero earnings in periods of non-participation. *Source:* Own calculations based on the Siab.

Using this trend as counterfactual leads to an underestimation of the treatment effect as well (we estimate Euro −14,158).

The weights that each of these three comparisons receives in the average estimate depend on the group size and variance of the treatment dummy in each pair of cohorts, i.e. on parameters specific for a given sample (Goodman-Bacon 2021). In our example, the estimate using the two "forbidden" comparisons receives almost two thirds (66.2 percent) of the weight. It introduces a substantial upward bias into the average DiD estimate which equals Euro −11,562, whereas the estimate solely based on the clean comparison yields Euro −15,975. The bias equals Euro 4,413 or 38 percent of the average estimate.

Table 3.1: Decomposition of average estimate: "Clean" and "forbidden" comparisons.

	Average	"Clean" (to 30)	All "Forbidden" (to 24 and 28)	"Forbidden" (to 24)	"Forbidden" (to 28)
Treatment status	−11,562***	−15,976***	−9,307***	−14,158***	−5,430***
	(142)	(173)	(160)	(173)	(202)
Age FEs	Yes	Yes	Yes	Yes	Yes
Person FEs	Yes	Yes	Yes	Yes	Yes
Included cohorts	24, 28, 29, 30	29, 30	24, 28, 29	24, 29	28, 29
Estimation window	28–29	28–29	28–29	28–29	28–29
Weight in average		33.8%	66.2%		

Notes: The table reports the results from estimating the effect of childbirth on average annual labor earnings in the year after birth for the cohort that gives birth at age 29 following Equation (3.2). The first column shows results for the estimated average effect using the sample that includes four cohorts that give birth at ages 24, 28, 29 and 30. The second column shows the results from "clean" comparisons, when only the cohorts 29 and 30—treated and not-yet-treated—are included in the sample. The third column shows the result from "forbidden" comparisons, when only the cohorts 24, 28, and 29—already-treated and treated—are included in the sample. The fourth and fifth column show the coefficients separately for each of the two "forbidden" comparisons. Standard errors are reported in parentheses. *** indicates statistical significance at the 1 percent level. *Source:* Own estimation using the Siab.

In Figure 3.4, we extend this example to include all cohorts of mothers. Generally, the estimate of a given cohort in a given period is composed of comparisons of the treated cohort's changes in earnings to mostly flat post-birth trends of already-treated cohorts, comparably less steep downward trends from the cohort treated just before and to upward sloping trends of not-yet treated mothers. In the figure, the blue line indicates the cohort treated at age 29. The orange line captures the losses of the previously treated cohort (at age 28), the set of almost flat trends comes from cohorts treated earlier (before age 28) and the set of upward trends depicts the earnings development of not-yet-treated mothers who give birth at age 30 and above. All trends that come from the previous and the earlier cohorts lead to an underestimation of the treatment effect. Noteworthy, for different outcomes and samples the bias can go in any direction since it depends on pre- and post-birth trends and characteristics of the sample at hand.

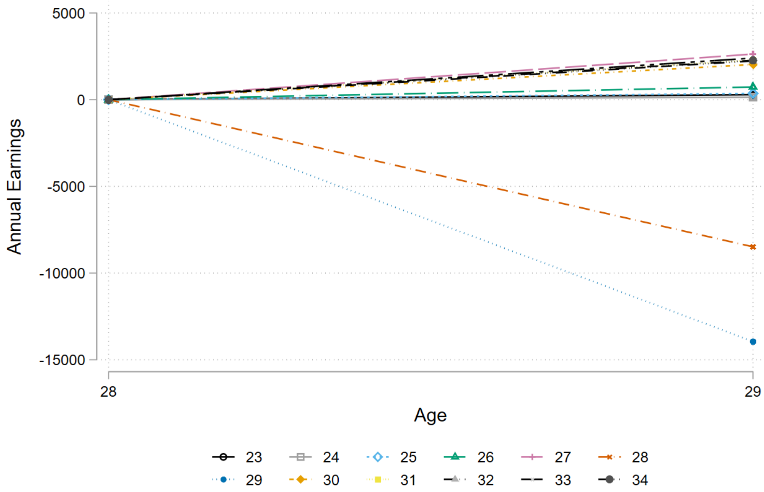


FIGURE 3.4: Changes in earnings of mothers from age 28 to 29 for all ages at birth (23-34)

Notes: The figure plots average annual labor earnings of mothers (including zero earnings) adjusted to pre-birth (at age 28) differences between cohorts, over ages 28-29 for all ages at birth between 23 and 34. *Source:* Own calculations based on the SIAB.

"Contamination" In common applications, dynamic treatment effects rather than an average one are of interest, which, as we discuss next, introduces additional issues. The general finding by Goodman-Bacon (2021) is that in the case of a static DiD with average treatment effect, the estimates can be biased as they are a weighted average of "clean" and "forbidden" comparisons if multiple groups are treated at different points in time and treatment effects are heterogeneous among these groups. As shown by a growing number of recent papers, the same as well as additional problems apply to dynamic DiD models, that move from the static case to having dynamic treatment effects (reviewed by de Chaisemartin and D'Haultfoeuille 2022; J. Roth et al. 2022).

To apply the insights from the TWFE literature to the case of child penalties we, first, have to note which one is the relevant dimension of treatment timing. The literature on TWFE models assumes treatment effect heterogeneity among cohorts defined by *calendar time*. When looking at childbirth as treatment, however, *age at first birth* is the more important dimension. As shown in the previous section, age at first birth is highly correlated with several characteristics relevant for the labor market as well as with the absolute and relative earnings losses after childbirth. In the following, we therefore use age at first birth as the variable to define cohorts of mothers. This means in practice, that we often observe relatively small differences in the effects of childbirth when comparing, for instance, mothers of the same age giving birth in the years 2000 and 2015, but we find much larger differences between mothers who give birth at ages 20 and 35 within the same calendar year.

With this change in the cohort definition, we can apply the analysis by Sun and Abraham (2021) to the case of child penalties. Their setting considers event study regressions of a single, binary and absorbing treatment (i.e. where treatment status can only change from 0 to 1 and stays 1 thereafter) with a staggered roll-out. This treatment description exactly matches the birth of the first child as an event that appears once and has consequences for the entire time thereafter.[4] The set of treated units can be divided into cohorts based on when they receive treatment. This setup can be treated as equivalent to KLS's specification in (3.1) with added individual fixed effects. As pointed out above, age at first birth now serves as the variable grouping individuals into cohorts based on when they receive treatment.[5]

Sun and Abraham (2021) decompose the point estimates $\hat{\beta}_l$ from TWFE regressions (similar to the one in Equation (3.1) with added person fixed effects) and show that for each relative time period l they consist of two parts. The first one is a weighted average of the cohort specific treatment effects in period l, which is similar to the decomposition of the static DiD from the previous illustration and can be biased if there is heterogeneity of effects across cohorts. The second part is again a weighted average, but of the treatment effects from the *other periods* l'. This part is only in place in the dynamic DiD setting and commonly referred to as "contamination". When the single treatment dummy is replace with a set of dummies for each relative time period around treatment interacted with treatment status, the treatment effect for a given period is estimated conditioning on effects from other periods. By including the set of event time dummies, the outcome in each period is adjusted by the estimated average of cohort-specific treatment effects in each period. If these cohort-specific treatment effects are identical, the average reflects an unbiased estimate. If, however, treatment effects vary across cohorts, subtracting an average does not take this heterogeneity into account, hence, it introduces and additional bias.

Sun and Abraham show that if there is heterogeneity in treatment effects across cohorts, then other periods will enter the estimation of a coefficient for a given period with non-zero weights and "contaminate" it. In Figure 3.16 in the Appendix, we provide the evidence that "contamination" is likely to be in place when estimating child penalties. We plot the weights obtained from the decomposition Sun and Abraham propose. Both for a pre-birth (-4) as well as a post-birth ($+5$) time period we find that estimates from other relative time periods (on the

[4] Note, that we only consider the birth of the *first* child which is both for child penalty estimations and in the TWFE literature the common case. Considering higher order births can introduce additional complications. de Chaisemartin and D'Haultfoeuille (2022) provide an econometric assessment of the case of multiple treatments.

[5] For the illustrations in Figures 3.1b, 3.12, 3.13 and 3.14 we restrict to four cohorts to improve readability. In practice, the number of cohorts is determined by the number of different levels of age at first birth in the data at hand.

x-axis) will receive non-zero weights (on the y-axis), if effects are heterogeneous. It is especially of note that the estimate for year +5 after childbirth can include effects from pre-birth time periods. This decomposition confirms that under effect heterogeneity, "contamination" is likely to bias child penalty estimates.

Similar to the static DiD example, there are two main consequences. First, estimated pre-trends are uninformative. If they are flat, this is no clear indication of parallel trends before treatment. Second, especially in the presence of treatment heterogeneity, estimated treatment effects can be biased. It implies that heterogeneity in the effects of childbirth on maternal outcomes can lead to estimates that are neither a numerically correct representation of the actual treatment effect nor an interpretable weighted average of treatment effects of multiple cohorts.

3.4.3 Validity of Control Groups and Applicability of Heterogeneity-Robust Estimators

A growing number of papers propose new estimators that avoid "contamination" and "forbidden" comparisons and—under certain conditions—obtain estimates that are robust to heterogeneous treatment effects. In the following we discuss the most common of these estimators and the assumptions one has to make to apply them.

As the underlying problem of static and dynamic DiD estimations is the inclusion of treated units in a counterfactual, researchers developed various new estimators that clearly specify their control groups. Generally, three types of control groups are suggested by the DiD literature. Units that are the last ones to receive treatment before they become treated (Sun and Abraham 2021), units that never receive treatment (Sun and Abraham 2021; Callaway and Sant'Anna 2021), and units that are not-yet treated (de Chaisemartin and D'Haultfoeuille 2020; Callaway and Sant'Anna 2021). Estimating child penalties using any of these control groups requires to make strong assumptions that they are suitable counterfactuals for each cohort of mothers. Moreover, heterogeneity among these control groups in the setting of childbirth leads to substantial differences in the effects that are estimated and their interpretation.

Last-treated units as control group As we show in Section 3.3, the timing of childbirth is correlated with some basic characteristics of mothers and their labor market outcomes, both before and after birth. It directly follows that those mothers who are the oldest when they have their first child are a selected group. On average, they have a higher socio-economic status, reflected in higher earnings, education levels, occupational ranks and other labor market outcomes. Additionally, they have the shortest post-birth labor market breaks (see Figure 3.11 in the Appendix). Therefore, the outcomes of last-treated mothers cannot serve as a counterfactual for earlier-treated cohorts.

Not-yet-treated units as control group The case of using not-yet-treated mothers as a control group is related to the idea of using outcomes of those mothers who are the oldest when they have children as a counterfactual. Some share of the control group consists of later and the last-treated mothers, i.e. older ones. For them, the above reasoning that later childbirth is associated with fundamental differences in pre- and post-birth characteristics and outcomes continues to hold. In their entirety not-yet-treated units are therefore unsuitable as control group.

However, a control group of not-yet-treated units as well consists of mothers who have their child rather early. For instance, for mothers who have their child at some age a, mothers who give birth at age $a + 1$ are part of the control group. In order to use them as a control group, the necessary assumption is that age at childbirth is quasi-randomly allocated within a relatively small bandwidth of age. Even though we show that age at childbirth and a number of characteristics that are relevant for labor market outcomes are correlated, it seems plausible to assume that even women who put effort in planning when to have children cannot precisely manipulate the exact date. Then, childbirth has a random component such that within the span of a few years mothers are comparable. We build on this feature when constructing the control group in Section 3.5.[6]

Never-treated units as control group Using those who never receive treatment, i.e. never give birth to a child, as the control group implies using childless women or men to obtain the effect of childbirth on mothers.

If one uses men, the estimate will capture the effect of having a child on the outcomes of mothers under the assumptions that men are unaffected by childbirth and that, absent children, women would have the same outcome trajectories as men. In practice, these assumptions are unlikely to hold. First, there are gender differences in career paths because of various kinds of discrimination and different experiences that can have an effect already prior to birth, for instance when choosing a college major. After childbirth, being a father can have both positive and negative effects on outcomes of men. Negative ones can arise in societies that are more gender-equal where the burden of raising children is shared more equally between both parents. Positive effects for fathers are documented as well. Goldin, Kerr, and Olivetti (2022), for instance, show the existence of a "fatherhood premium" that is larger for more time-intensive occupations pointing towards an increase in productivity due to focus on market work (similar to the findings on male marital wage premiums; documented, among others, by Antonovics

[6] This idea is also used by Fitzenberger, Sommerfeld, and Steffes (2013).

and Town 2004).[7] Using the Soep data, we also document that there is a positive correlation between labor market outcomes of fathers and the timing of parenthood, further showing that they are not unaffected by childbirth (see Figure 3.14 in the Appendix).

To use childless women as control units one needs to make the assumption that not having children is a trait that is allocated to women in a quasi-random fashion and is unrelated to their labor market outcomes, such that outcome trajectories for mothers and childless women are similar. This assumption is restrictive and difficult to test.[8] Further, it does not allow for the voluntary decision of being childlessness for career reasons.[9] In practice, childless women represent a very heterogeneous group of those who decided to have no children and those who wanted, but could not have them for different reasons.

Both for men and those childless women who know they will stay childless (for instance due to medical reasons) or explicitly plan to do so, there remains the issue of anticipation. Including them in an event study leads to a situation where the control group knows it will never receive treatment while the treatment group at least anticipates it is very likely to be treated at some point in the future (and possibly plans to influence when). Both groups are thus able to make according decisions such that level-differences and trends differ even more between them.

Comparing estimation approaches Overall, none of the available groups—last-treated, all not-yet-treated or never-treated units—can serve as a suitable control group to estimate child penalties. Heterogeneity-robust estimators that require to use one of them are thus not fully applicable. We now illustrate how different estimators with different choices of control groups lead to a variety of estimates that all diverge from the conventional ones.

In Figure 3.5 we plot the results from estimating earnings losses after childbirth using different estimators. Our starting point is the conventional approach by KLS as in Equation (3.1) (plotted as light blue line) that only considers women who eventually become mothers and includes age and year fixed effects. We then add childless women to the estimation sample (green line) which mostly changes the level of the estimated coefficients and pre-trends, but not their overall slightly downward sloping trend. In a next step, we include individual fixed-effects to account for time-constant heterogeneity (plotted as light orange—for a model that only uses mothers—and pink—a model that includes childless women—lines). We further estimate TWFE models following Callaway and Sant'Anna (2021) (using not-yet-treated mothers as

7 There is the related special case where the control group is constructed from the male partners of mothers (as for instance in Angelov, Johansson, and Lindahl 2016; Andresen and Nix 2022). Here, the within-couple distribution of roles and tasks could introduce an additional source of effect heterogeneity.

8 Lundborg, Plug, and Rassmussen (2017) utilize the special case of IVF treatment success to estimated career effects of childbirth conditional on receiving IVF treatment.

9 Steinhauer (2018) documents interrelations between gender roles, childlessness and choices in the labor market.

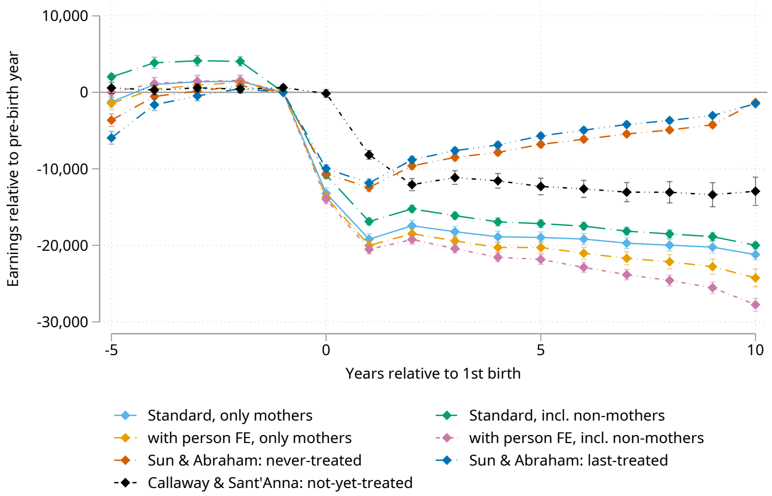


FIGURE 3.5: Comparison of estimates from different estimators and specifications.

Notes: The figure shows event study estimates for the impact of the first childbirth on real annual labor earnings obtained using different specifications and estimators. All models control for age and calendar year. *Source:* Own calculations based on the SOEP.

control units; plotted in black) as well as Sun and Abraham (2021) where we use both never-treated units (plotted in dark orange) and last-treated mothers (plotted in blue) as control groups. We observe a wide range of estimates. Ten years after childbirth, estimated earnings losses range from slightly below zero to almost Euro −30 K (note that the average earnings in the year prior to childbirth are around Euro 26 K). These results point out how substantial the differences between the control groups are such that the resulting estimates show large variation in their levels as well as different trends.

3.4.4 Re-scaling

The second common step when estimating child penalties is to re-scale the event study estimates from levels to changes in percentages. Percentage changes are easier to interpret and allow straightforward comparisons of child penalties across different settings such as countries, points in time or policy regimes. KLS's canonical approach re-scales by calculating

$$P_l = \frac{\hat{\beta}_l}{\mathrm{E}\left[\tilde{Y}_{it}|l\right]}, \tag{3.3}$$

where $\tilde{Y}_{it}$ is the prediction from the previous regression (as in Equation 3.1) when omitting the relative event-time dummies $\hat{\beta}_l$. $\mathrm{E}\left[\tilde{Y}_{it}|l\right]$, the earnings levels that only depend on fixed effects for age and calendar year, are intended to proxy for earnings in a counterfactual state of the world in which a woman does not have children. The child penalty P_l then gives the percentage difference between the earnings of mothers and counterfactual earnings of women without children. Note, that while the event study leads and lags make a comparison relative to the omitted pre-birth time period (typically one year before birth) the denominator changes in relative time around birth, thus introduces a second comparison that is specific for each relative period l.

The re-scaling step introduces an additional source of potential biases as it still contains treatment effects and makes comparisons between units that have been treated at different points in time. When predicting counterfactual earnings based on age and year fixed effects alone, the resulting $\tilde{Y}_{it}$ is not restricted to only use specific observations and therefore consists of observations *both* from the pre- and the post-birth period. Since the share of women who already had their first child increases in age, the composition of the counterfactual in terms of including pre- or post-birth observations changes in age at childbirth as well. Figure 3.6 illustrates this situation by plotting standard child penalty estimates over -5 to $+10$ years around birth based on Equations (3.1) and (3.3) for the four quartiles of the distribution of age at first birth. Along with the child penalty, we plot the share of pre-birth observations that is used at each point in relative time to construct the counterfactual earnings.

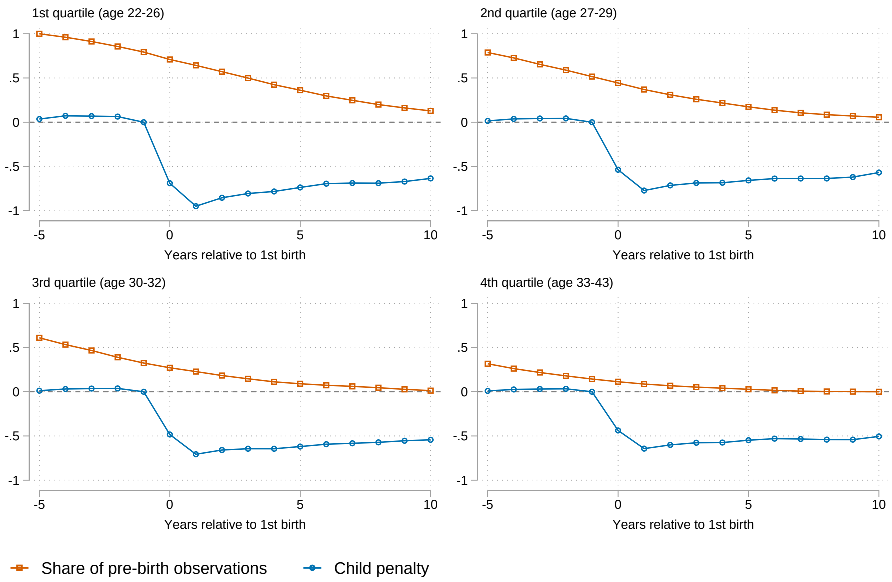


Figure 3.6: Composition of counterfactual earnings E $\left[\tilde{Y}_{ist}|t\right]$.

Notes: Child penalties (blue) and share of pre-birth observations (orange) in the counterfactual by time around first childbirth, plotted for quartiles of age at first birth. *Source:* Own calculations based on the Soep.

For younger mothers, the counterfactual earnings are primarily informed by observations of not-yet-mothers while the counterfactual for older mothers mostly consists of observations from women who are already mothers. For the youngest mothers in the first quartile—for whom the largest penalty is estimated—counterfactual earnings in the post-birth period consist to a large share of observations of other mothers *before* they give birth (for the two years directly after birth above 50 percent). For mothers in the fourth quartile, the counterfactual for their post-birth earnings primarily contains observations of other mothers *after* they have given birth (at the time of birth, the share of pre-birth observations is already below 25 percent and further decreases over time to virtually zero in year seven after birth).[10] In other words, for the majority of the sample the counterfactual for post-birth earnings is constructed from post-birth earnings which do not give an adequate depiction for a situation without treatment. This way of constructing counterfactual earnings leads to smaller counterfactuals for older mothers. Along with their earnings generally being higher, this mechanically generates smaller penalty estimates with increasing age at first birth.

Abstracting from our example in figure 3.6 that is estimated using SOEP data, it is straightforward to formalize the relationship between age at first birth and the composition of counterfactual earnings. Assume, age at first birth A is distributed on some interval $\left[\underline{A}, \bar{A}\right]$. The—dataset specific—distribution function is $F_A(a) = P(A \leq a)$ and gives the share of observations that have given birth at or before age a. Therefore, for any age at first birth a and any post-birth year t, $\sigma_{\text{pre}} = 1 - F_A(a + t)$ gives the share of remaining pre-birth observations in the sample that is not influenced by effects of previous childbirth. By definition of a distribution function, σ_{pre} is decreasing in age, highlighting that fewer and fewer suitable observations to construct counterfactual earnings are available when we look at older mothers.

3.5 The Solution: Stacked DiD with Rolling-Window Control Groups

Having discussed the issues with the common event-study-based approaches to estimate child penalties, we propose a new solution within the DiD framework, which takes into account heterogeneity of treatment effects by age at first childbirth, employs a valid control group and—under certain assumptions—is able to bring the results closer to an estimate of the causal effects of motherhood on labor market outcomes. In this section, we present the new approach along with the required assumptions and apply it to estimate child penalties in the German labor market and study their heterogeneity by age at first birth.

We suggest to combine the concepts of the "stacked" DiD design (Cengiz et al. 2019) and the Callaway and Sant'Anna (2021) estimator with additional restriction imposed on the control group. The main idea behind any DiD design is that a counterfactual trend is borrowed from a control group and that level differences in outcomes between treated and control groups are

[10] Table 3.4 in the Appendix lists the shares of available pre-birth observations for five levels of age at first childbirth.

taken into account. Identification within a DiD design then builds on the general comparability of treatment and control group, which translates into similar outcome trajectories absent treatment. To ensure comparability of both groups in the setting of a child penalty estimation, we propose a "stacked" DiD design combined with a rolling window of cohort-specific control groups over age at birth. Specifically, we construct a control group for each treated cohort from the pre-birth observations of only the closest (in terms of age at birth) not-yet-treated cohorts. Due to the strong correlation between age at birth and labor market outcomes (see Section 3.3), this approach brings the most comparable treated and control mothers together. By creating such valid and "clean" control groups for each treated cohort and "stacking" the cohorts with their control groups by event time, we are able to estimate unbiased cohort-specific treatment effects and avoid making "forbidden" comparisons.

Practically, we apply the following procedure. First, we define an estimation window of relative time around childbirth that is the same for each cohort. The number of post-treatment periods is determined by the number of control cohorts (N_{cc}) that are included in each control group. This number is based on an assumption of how long mothers in the control groups who give birth at an older age remain comparable. We treat childbirth for each cohort as separate sub-events and construct separate sub-samples by taking the observations for each treated cohort by age at birth (a) for the defined event-time window and selecting only pre-birth observations from cohorts that give birth at age $[a + 1, a + N_{cc}]$ that fall in the same estimation window. We then append (i.e. "stack") each sub-sample to the main dataset by event time. This step duplicates some observations such that they can serve as controls for some sub-events of earlier-treated cohorts and as treated in one sub-event when they belong to the treated cohort. In the next step, we generate sub-event-specific event-time dummies that are 0 for control units and 1 for treated units. This means that the same individual in the same event-time might be marked as non-treated if they serve as a control unit and marked as treated in the sub-event when they are the treated cohort. We then use the stacked dataset to estimate the dynamic DiD model as a TWFE regression:

$$Y_{ias} = \sum_{l=-L_{min}, l \neq -1}^{L^{max}} \beta_l^s \times \mathbb{1}[a - a_i^0 = l] \times \mathbb{1}[a_i^0 = s] + \gamma_{as} + \lambda_{is} + \varepsilon_{ias}. \tag{3.4}$$

In the equation, Y_{ias} indicates the outcome of mother i who belongs to cohort s at age a. Contrary to the conventional approach that focuses on calendar time, we treat age as the relevant time dimension.[11] Accordingly, the model includes a set of event-time indicators β_l^s that identify when a mother is l years away from her first childbirth at age a_i^0. The indicator function $\mathbb{1}[a_i^0 = s]$ identifies each cohort of mothers along with the assigned control units,

[11] See also Section 3.4.2.

i.e. it allows the event-time indicators to vary by sub-event. The fixed effects for age and individual, γ_{ts} and λ_{is}, are allowed to vary by sub-event as well (we omit additional indicator functions in the equation for readability). This model then estimates the cohort-specific effects β_l^s on the outcome of interest over the estimation window from L_{min} to L_{max}. Since it allows treatment effects and fixed effects to differ across sub-events (i.e. childbirth for each cohort), it is equivalent to running separate TWFE regressions for each sub-event. We cluster the standard errors at the level of the individual to account for correlation of the error term over time and across duplicated observations of the same mothers.

Most commonly, the outcome Y is used in absolute terms which allows to keep zero earnings during times of non-participation in the sample. To show percentage changes that are often more informative, an additional transformation is required. As we discuss in Section 3.4.4, re-scaling by a counterfactual composed of average age and year fixed effects is problematic. More generally, any transformation involves a choice made by the researcher and different transformations can lead to different results. Depending on the outcome, we suggest one of the following. The first transformations explicitly calculates counterfactual outcomes. Separately for each cohort and event-time, we take the average outcomes of the control group (i.e. women who thus-far have not given birth) and subtract the pre-birth level of the treatment group. This removes potential level-differences between both groups and assumes that mothers, in the counterfactual state without childbirth, follow the trajectory of the control group not-yet but soon-to-be treated women. Dividing the estimates $\hat{\beta}_l^s$ by the counterfactual gives the penalty of having a child relative to not having a child l years after childbirth. This approach stays close to the established notion of the child penalty as it adds an event-time-specific comparison to the comparison relative to a baseline that the event study estimates. A second possibility is to divide the estimates $\hat{\beta}_l^s$ by the cohort-specific pre-birth level of the outcome. Sticking to the pre-birth period as reference point is more suitable to highlight the interruption of a trend (i.e. a mothers' career) by childbirth. We apply and discuss both transformations in the following section.

Even though the cohort-specific estimates discussed so far provide additional information, it is often necessary to calculate an average effect over all cohorts. To this end, we follow Sun and Abraham (2021) and weight the cohort-specific estimates by the sample shares of each cohort:

$$\hat{\beta}_l = \sum_{s=S_{min}}^{S_{max}} \frac{N_s}{N} \times \hat{\beta}_l^s, \tag{3.5}$$

where N_s indicates the number of observations per cohort and N the total number of observations in the dataset. Duplicates due to the data transformation into stacked sub-events are neglected. The result is a weighted average estimate over all cohorts.

The main assumption behind this "stacked" DiD design is that absent treatment, the trends of the treatment and control groups are parallel for each cohort (i.e. sub-event). The similarity of outcome trajectories before treatment is testable by plotting the cohort-specific pre-trends. The comparability of the post-treatment trajectories absent children has to be assumed. It can be assessed by comparing the covariates in the pre-birth periods, which provides evidence for the general comparability of treatment and control groups. Intuitively, this means for our setting to assume that women, who give birth at a certain age, absent children, would have followed the same outcome trajectory as those women, who give birth a few years later, at the same age. The strong correlation of age at birth with both outcomes and relevant covariates provides the foundation for this assumption, because we bring together treated and control units from a narrow window of ages at birth. Age at first birth is arguably a better proxy for career trajectories than, for instance, education as it combines information on the levels of education and the pre-birth realized career.

Since we do not employ all not-yet-treated cohorts, but a smaller subset of them as a unique control group for each cohort, it is now more straightforward to assume that timing of childbirth is quasi-random within the span of the few years covered by the control cohorts. Thus, the restricted control group allows to interpret our estimates as the causal effects of becoming a mother compared to delaying childbirth. The intuitive reasoning behind this assumption is that, even though mothers can select into a time range during which they give birth, it is substantially less likely that they can precisely choose at which age they give birth. Having a general plan to have a child rather early or late in life can be a common trait among women, while ensuring that a child is born in a specific year hinges on a number of factors that are not entirely within a woman's control.

A particular limitation of our approach is that it allows to estimate the costs of having a child only within a medium-run time horizon. The number of post-birth time periods (L_{max}) depends on an assumption on how long the cohorts in the control group remain comparable to the treated cohort. Since labor market outcomes almost monotonically increase in age at birth (see Section 3.3), assuming comparability between mothers who give birth at ages 26 and 28 is easier than for ages 26 and 38. The number of cohorts for which the researcher is willing to make such an assumption determines how many not-yet-treated cohorts can be included in the control group and, consequently, the time horizon the estimation can cover. The width of the window in which cohorts are comparable can be formally tested (currently work in progress) and may differ across cohorts, such that for some cohorts effects for more post-birth periods can be estimated than for others. Noteworthy, the conventional approach implicitly assumes the comparability across all ages at birth by employing all older first-time mothers as control units.

A further advantage of the presented approach is the option to study effect heterogeneity across different cohorts of mothers by age at birth, which is of particular importance in policy evaluation settings. For example, given the positive correlation of earnings and age at first birth, an evaluation of a policy separately for younger and older first-time mothers would be informative about the distributional effects.

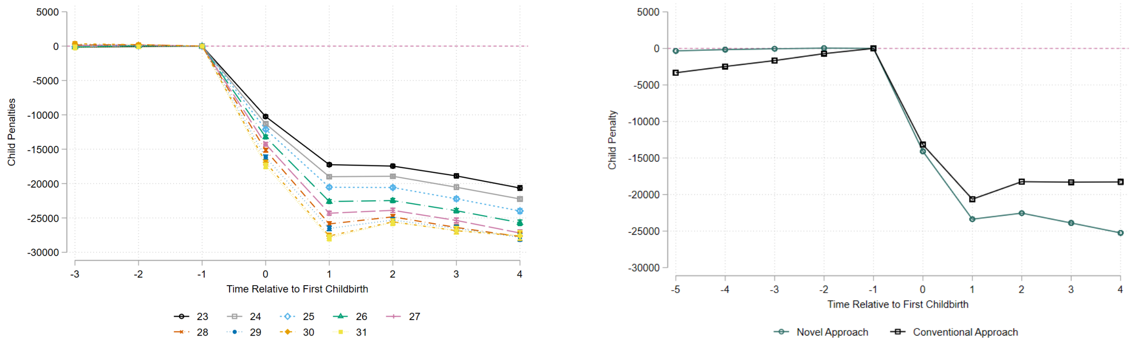
3.6 Application and New Results on Child Penalties

In the following, we provide an application of the stacked DiD approach with restricted control groups as described above to assess child penalties for women in Germany. For this task, we rely on the Siab data (see Section 3.2), which offers a large sample size and, thus, allows to conduct analyses of effect heterogeneity with respect to age at first birth.[12]

Annual Earnings We start by exploring the effects of motherhood on annual earnings, since it is the most commonly used outcome and contains information on both intensive and extensive margin labor supply effects. We estimate Equation (3.4) with annual pre-tax earnings including zeros for years of non-participation as an outcome. Results for cohort-specific and average effects in absolute terms are plotted in Figure 3.7. First, we document that losses in earnings are heterogeneous by age at first birth. As shown in Figure 3.7a, child penalties in absolute terms differ significantly across cohorts and almost monotonically increase in age at first birth. Second, in Figure 3.7b, we compare the average estimates from our approach with a conventional event study. We document that the conventional approach substantially underestimates the child penalty, in year four after birth by around Euro 7 K (or 38 percent). These results are consistent with our discussion in Section 3.4.2 that "forbidden" comparisons of just-treated with already-treated unit miss earnings growth that would have happened without childbirth such that post-birth trends of early-treated mothers bias the counterfactual downwards. They confirm, that "forbidden" comparisons receive a substantial weight in common event studies, which leads to an underestimation of earnings losses. Instead, our approach correctly captures both losses in levels as well as unrealized earnings growth, compared to the control group of the soon-to-be mothers. The downward sloping trend of the estimates that we find for each cohort and for the average, reflects that women do not only lose earnings in levels, but also experience a slow-down of their earnings growth after the first childbirth, which leads to an increase of the child penalty over time. Importantly, the pre-trends are flat and insignificant for all

[12] We prefer the Siab data for this exercise as the large sample size ensures precise and clearly distinguishable estimates for the different cohorts. Using the smaller but easier available Soep data yields estimates that are qualitatively and quantitatively similar.

cohorts. This provides evidence that the parallel trends assumption holds in our application and further that there is no substantial anticipation of treatment.



(A) Post-birth earnings losses by age at first childbirth.

(B) Average post-birth earnings loss.

FIGURE 3.7: Earnings losses after the first childbirth.

Notes: The figure plots the estimates of absolute losses in annual labor earnings after the birth of the first child following the approach described in section 3.5. The left-hand panel reports the estimates for cohort-specific losses as in Equation (3.4). The right-hand panel reports the estimates of the weighted average across-cohorts as in Equation (3.5) and the conventional approach as in Equation (3.1). The estimates are reported for the periods from −3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015. *Source:* Own calculations based on the SIAB.

We transform the estimates to relative terms by dividing them by counterfactual outcomes calculated from the control group's outcomes. With annual earnings as outcome, we prefer this method. It is, first, close to the conventional approach as it makes event-time specific comparisons and, second, it allows for further growth of the counterfactual outcome after treatment such that the result highlights that a part of the child penalty is due to a lower earnings growth of mothers compared to non-mothers. The results are plotted in Figure 3.8 along with the penalties estimated using the conventional approach. We observe that the heterogeneity pattern is now different, with the youngest and oldest first-time mothers having smaller relative penalties than mothers, who give birth at the middle of the distribution of age at first birth. The average child penalty over all cohorts is −85 percent which is substantially larger than the estimate from the conventional approach that is smaller by 15 percentage points (or 20 percent).

Cumulative Earnings Estimating child penalties using annual earnings is the most common approach. Here, we also add an estimation that uses earnings cumulated over the entire working life. This adds a life-cycle perspective by showing the earnings losses after childbirth in relation to what mothers have accomplished before birth and by highlighting how the losses accumulate over time. We plot the results in Figure 3.9. In absolute terms, older first-time mothers who have the higher earnings accordingly experience larger losses than younger ones; penalties

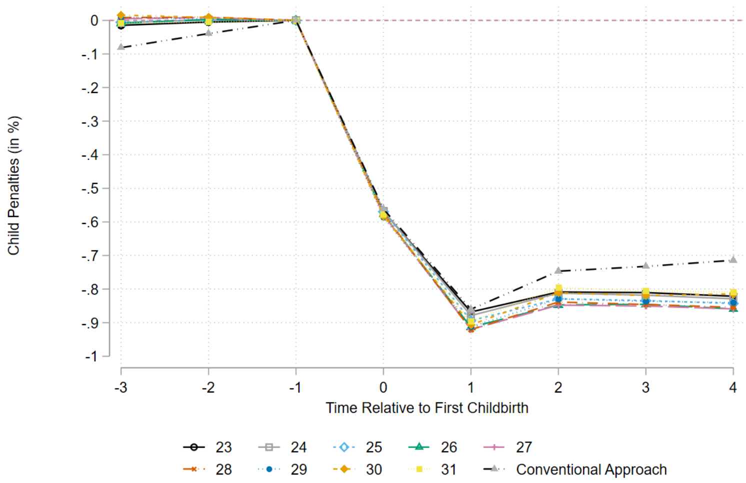
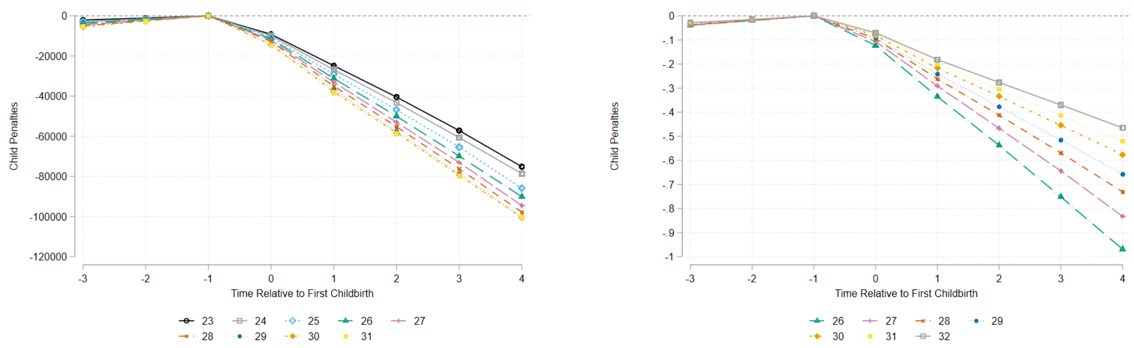


FIGURE 3.8: Earnings losses after the first childbirth (in relative terms).

Notes: The figure plots the estimates of relative losses in annual labor earnings after the birth of the first child following the approach described in section 3.5. The figure reports the estimates for cohort-specific losses as in Equation (3.4) and the conventional approach as in Equation (3.1). The estimates are reported for the periods from −3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015. *Source:* Own calculations based on the SIAB.

increase in age at first birth. We transform the estimate to percentage terms by dividing them by the cohort-specific averages in the pre-birth year. We choose this transformation to provide effects and how they develop relative to the pre-birth reference point. This takes into account that older mothers are at a more advanced career stage where they have accumulated more earnings such that for them the financial burden of having children is less pronounced. Our estimates confirm this. In relative terms, older first-time mothers experience smaller negative effects than younger ones. Until year four after birth, for those who become mothers at 31 the forgone earnings accumulate to 50 percent of what they have accumulated before childbirth. For mothers who give birth at 26, the unrealized earnings accumulate to almost 100 percent of what they have cumulatively earned before childbirth. These findings are in line with previous research showing that more educated and career-oriented women strategically delay childbirth, since it is less costly for careers compared to giving birth early in life (see, for instance, Adda, Dustmann, and Stevens 2017).

Occupational Rank In addition to earnings, we now turn to the question how the occupational rank of those mothers who continue to work after childbirth develops. We measure occupational rank as the median earnings by occupation at the 3-digit level (KldB 2010 Bundesagentur für Arbeit 2021). In contrast to earnings where missing values in periods of non-participation can be replaced with zeros, information on occupational rank while not



(A) Absolute losses in cumulative earnings.

(B) Relative losses in cumulative earnings.

FIGURE 3.9: Cumulative earnings losses after the first childbirth by cohort.

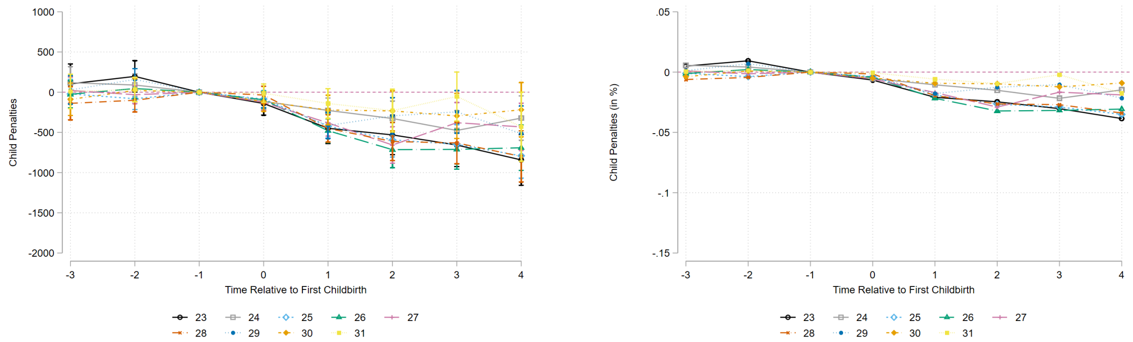
Notes: The figure plots the estimates of losses in cumulative annual labor earnings after the birth of the first child following the approach described in section 3.5. The left-hand panel reports the estimates for cohort-specific losses in absolute terms as in Equation (3.4). The right-hand panel reports the estimates of cohort-specific effects divided by the pre-birth cumulative earnings. The estimates are reported for the periods from −3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015. *Source:* Own calculations based on the SIAB.

working is missing and is not straightforward to replace. Further, since a substantial share of women leaves the labor market after the first childbirth for some time and some even never return, the sample composition changes. We therefore restrict the sample to include only women who eventually return to the labor market. For the time they spend on parental leave, we impute the pre-birth values of occupational rank.[13]

The results for the development of occupation ranks around childbirth are plotted in Figure 3.10. The left-hand panel reports estimates in absolute terms, the right-hand panel the relative changes in occupational rank. As for cumulative earnings, we transform the estimates to percentage values by dividing them by the cohort-specific pre-birth levels to analyze if occupational rank changes compared to what mothers have achieved before childbirth. On average, we observe a small reduction in median earnings by occupation of around Euro 500 or −2.5 percent. The effects are driven by the younger cohorts of first-time mothers, for whom we measure impacts of the first birth of up to Euro 1 K. For older mothers, the estimates are smaller and only marginally significant. Our findings show that women do not generally sort into lower-paying jobs after they become mothers. Instead, mothers experience a slow-down in the growth of their occupational rank compared to the control group, that continues to work and advances in their careers. Motherhood, thus, seems to be more costly for younger

[13] For our data on Germany, this can be motivated by the job protection period after childbirth that allows mothers to return to their pre-birth position at their pre-birth employer. Since job protection is granted for up to three years, this gives rise to the concern that our estimation window might be too short to capture an effect. However, we do not find substantial differences or changing trends between years three and four after birth which makes potential biases unlikely.

first-time mothers in terms of climbing the career ladder. This is consistent with them still being in the early and most active stage—the "job-shopping" phase (Bagger et al. 2014)—of their working lives such that childbirth and the following interruption of work has a greater impact.



(A) Development of occupational ranks in absolute terms.

(B) Development of occupational ranks in relative terms.

FIGURE 3.10: Development of occupational ranks after the first childbirth by age at first birth.

Notes: The figure plots the estimates of absolute (in the left-hand panel) and relative (in the right-hand panel) losses in occupational rank after the birth of the first child following the approach described in section 3.5. Occupational rank is measured as the median earnings by 3-digit occupation categories. The estimates are reported for the periods from −3 to +4, where 0 is the year of the first childbirth. All monetary values are deflated to the base year 2015. *Source:* Own calculations based on the SIAB.

3.7 Conclusion

Estimating child penalties based on event studies has quickly become a widely used and valuable tool to assess the labor market effects of childbirth on mothers. It comes, however, at the cost of making strong assumptions. In this paper, we show that the considerable amount of heterogeneity in both maternal outcomes and characteristics by age at first childbirth leads to a violation of these assumptions and introduces issues similar to those documented in the literature on staggered difference-in-differences models. By making "forbidden" comparisons to already-treated units as well as by introducing "contamination", i.e. the possibility that the treatment effects for one period are confounded by effects from other periods, child penalty event studies can produce biased estimates. In some settings such issues can be addressed by using heterogeneity-robust estimators. We discuss that they are not fully applicable with childbirth as treatment as they employ control groups that are not comparable.

Instead of relying on conventional event studies, we propose a novel approach to estimate child penalties. We use a "stacked" DiD design with an additional restriction of the control group. We construct cohort-specific control groups from observations of only the closest not-yet-treated mothers. This rolling window of included control cohorts ensures comparability of treatment and control group which is crucial to justify the parallel trends assumption. This approach allows us to estimate the unbiased and causal effects of childbirth, both specific for each cohort and on average.

Our application that revisits the estimation of child penalties in the German labor market highlights the common event studies substantially underestimate the earnings losses following the first childbirth. We, further, show heterogeneity by cohort and new results on the child penalty in occupational rank. We conclude that earnings losses after birth are just to a small extend driven by occupational downgrading, while unrealized growth in earnings is an important driver of underestimated penalties when using the common approach.

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens (2017). "The career costs of children". In: *Journal of Political Economy* 125 (2), pp. 293–337.

Andresen, Martin Eckhoff and Emily Nix (2022). "What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples". In: *Journal of Labor Economics*. forthcomming.

Andrew, Alison, Oriana Bandiera, Monica Costa-Dias, and Camille Landais (2021). "Women and men at work". In: *IFS Deaton Review of Inequalities*.

Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016). "Parenthood and the Gender Gap in Pay". In: *Journal of Labor Economics* 34 (3), pp. 545–579.

Antonovics, Kate and Rober Town (2004). "Are All the Good Men Married? Uncovering the Sources of the Marital Wage Premium". In: *American Economic Review* 94 (2), pp. 317–321.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin (2014). "Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics". In: *American Economic Review* 104 (6), pp. 1551–1596.

Blau, Francine D. and Lawrence M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature* 55 (3), pp. 789–865.

Blundell, Richard, Monica Costa-Dias, David Goll, and Costas Meghir (2021). "Wages, Experience, and Training of Women Over the Life Cycle". In: *Journal of Labor Economics* 39 (S1), S275–S315.

Bruns, Benjamin (2019). "Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap". In: *American Economic Journal: Applied Economics* 11 (2), pp. 74–113.

Bundesagentur für Arbeit, ed. (2021). *Klassifikation der Berufe 2010 – überarbeitete Fassung 2020*. Nürnberg.

Bütikhofer, Aline, Sissel Jensen, and Kjell Salvanes (2018). "The role of parenthood on the gender gap among top earners". In: *European Economic Review* 109, pp. 103–123.

Callaway, Brantly and Pedro H. C. Sant'Anna (2021). "Difference-in-Differences with multiple time periods". In: *Journal of Econometrics* 225, pp. 200–230.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace Heterogeneity and the Rise of West German Wage Inequality". In: *The Quarterly Journal of Economics* 128 (3), pp. 967–1015.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs". In: *The Quarterly Journal of Economics* 134 (3), pp. 1405–1454.

Dauth, Wolfgang and Johann Eppelsheimer (2020). "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide". In: *Journal for Labour Market Research* 54.

de Chaisemartin, Clément and Xavier D'Haultfoeuille (2022). "Two-way Fixed Effects and Differences-in-Differences Estimators with Several Treatments". URL: `https://sites.google.com/site/clementdechaisemartin/` (visited on 04/03/2023).

de Chaisemartin, Clément and Xavier D'Haultfoeuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". In: *American Economic Review* 110 (9), pp. 2964–2996.

de Chaisemartin, Clément and Xavier D'Haultfoeuille (2022). "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey". NBER Working Paper No. 29691.

Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt (2022). "The economics of fertility: A new era". NBER Working Paper No. 29948.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). "Revisiting the German Wage Structure". In: *The Quarterly Journal of Economics* 124 (2), pp. 843–881.

Fitzenberger, Bernd, Katrin Sommerfeld, and Susanne Steffes (2013). "Causal effects on employment after first birth – A dynamic treatment approach". In: *Labor Economics* 25, pp. 49–62.

Goebel, Jan, Markus M. Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp (2019). "The German Socio-Economic Panel (SOEP)". In: *Jahrbücher für Nationalökonomie und Statistik* 239 (2), pp. 345–360.

Goldin, Claudia (2022). "Career & Family: Women's Century-Long Journey Toward Equity. By Claudia Goldin. Princeton: Princeton University Press, 2021. Pp v, 325., hardcover." In: *The Journal of Economic History* 82 (3), pp. 915–916.

Goldin, Claudia, Sari Pekkala Kerr, and Claudia Olivetti (2022). "When the Kids Grow Up: Women's Employment and Earnings across the Family Cycle". NBER Working Paper No. 30323.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225 (2), pp. 254–277.

Kleven, Henrik, Camille Landais, Johanna Posch, Adreas Steinhauer, and Josef Zweimüller (2019). "Child Penalties across Countries: Evidence and Explanations". In: *American Economic Review: Papers & Proceedings* 109, pp. 122–126.

– (2021). "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation". NBER Working Paper No. 28082. URL: `https://www.henrikkleven.com/research.html` (visited on 04/03/2023).

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). "Children and Gender Inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11 (4), pp. 181–209.

Krapf, Matthias, Anja Roth, and Michaela Slotwinski (2020). "The Effect of Childcare on Parental Earnings Trajectories". CESifo Working Paper No. 8764.

Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington (2018). "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?" NBER Working Paper No. 24740.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rassmussen (2017). "Can Women Have Children and a Career? IV Evidence from IVF Treatments". In: *The American Economic Review* 107 (6), pp. 1611–1637.

Müller, Dana, Andreas Filser, and Corinna Frodermann (2022). "Update: Identifying mothers in administrative data". In: *IAB FDZ Methodenreport* 01/2022.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe (2022). "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature". URL: `https://psantanna.com/research/` (visited on 04/03/2023).

Steinhauer, Adreas (2018). “Working Moms, Childlessness, and Female Identity”. SciencesPo, LIEPP Working Paper No. 79.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. In: *Journal of Econometrics* 225 (2), pp. 175–199.

Wilde, Elizabeth Ty, Lily Batchelder, and David T. Ellwood (2010). “The mommy track divides: The impact of childbearing on wages of women of differing skill levels”. NBER Working Paper No. 16582.

Appendix

3.A Additional Tables and Figures

Table 3.2: Summary statistics of mothers one year prior to their first childbirth (Soep).

	Mean	SD	Min.	p5	p50	p95	Max.
Labor earnings	26,079	17,730	0	0	26,072	53,412	182,917
Share fulltime	0.71	0.45	0	0	1	1	1
Share not working	0.10	0.30	0	0	0	1	1
Experience (fulltime)	5.41	4.33	0	0	5	13	24
Experience (parttime)	0.89	1.95	0	0	0	5	17
Total years of education	12.83	2.81	7	9	12	18	18
Age at first birth	29.02	4.59	21	22	29	37	45
N	1,992						

This table provides the summary statistics for the Soep dataset.

Table 3.3: Summary statistics of mothers one year prior to their first childbirth (Siab).

	Mean	SD	Min.	p5	p50	p95	Max.
Labor earnings	23,908	17,702	0	0	24,367	51,349	343,804
Share fulltime	0.81	0.40	0	0	1	1	1
Share not working	0.15	0.35	0	0	0	1	1
Experience (fulltime)	5.01	4.08	0	0	4	13	20
Age at first birth	28.57	4.18	22	22	28	36	40
N	151,979						

This table provides the summary statistics for the Siab dataset.

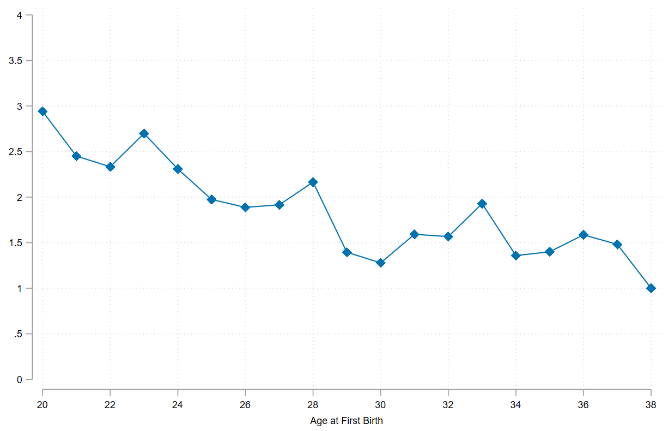


Figure 3.11: Average length of employment break by age at childbirth (conditional on returning to work).

Notes: The figure plots the average length of the post-birth employment break for mothers by age at childbirth, conditional on returning to work. *Source:* Own calculations based on the SOEP.

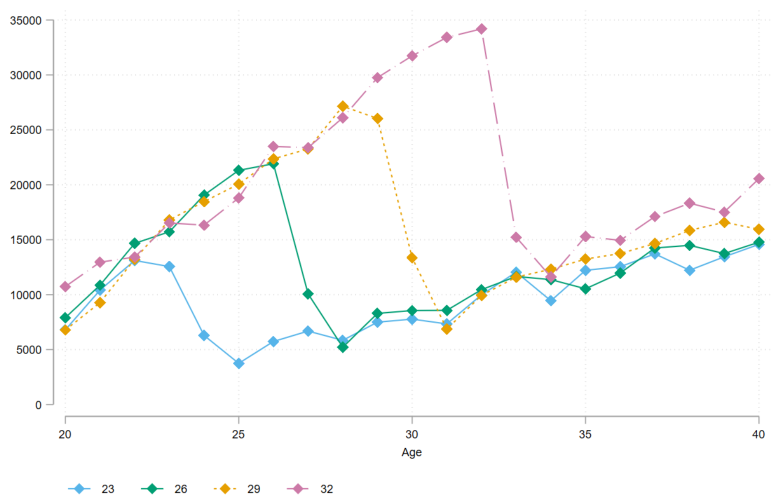
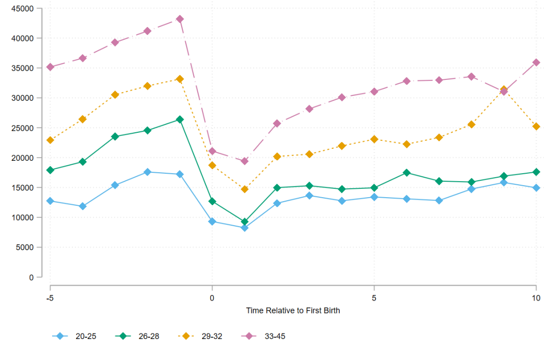
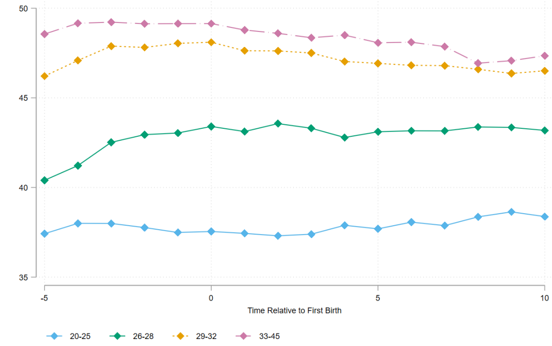


Figure 3.12: Average earnings of mothers around birth for four ages at first birth, incl. zeros

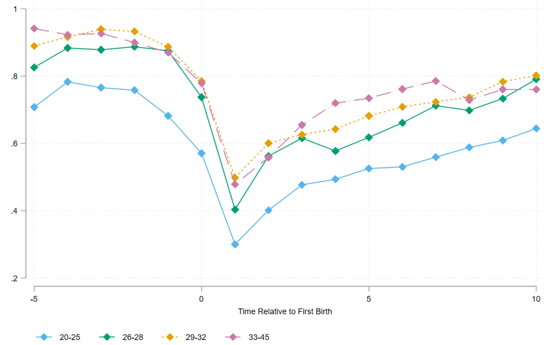
Notes: The figure plots average annual labor earnings of mothers (including zero earnings) by age for four ages at first birth (23, 26, 29, 32). *Source:* Own calculations based on the SOEP.



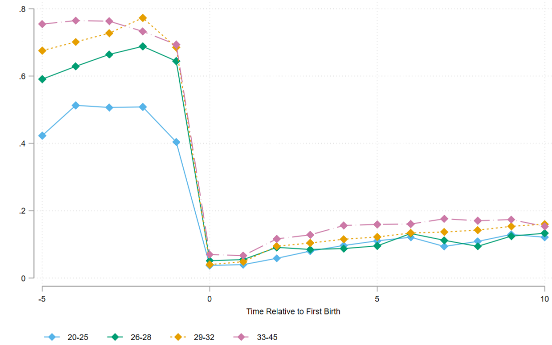
(A) Average annual earnings of *working* mothers.



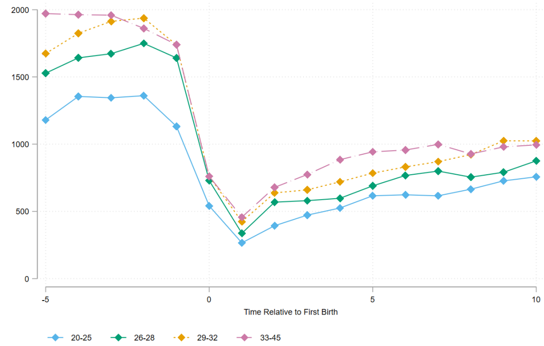
(B) Standard international occupational prestige scale (SIOPS).



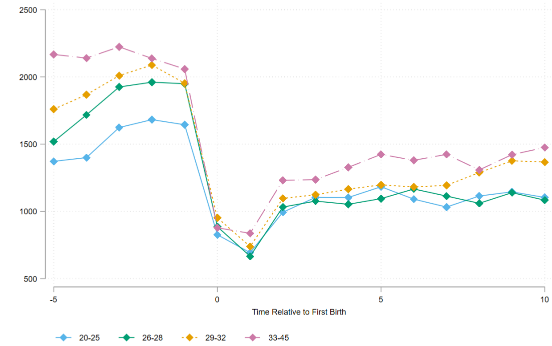
(C) Share working mothers.



(D) Share of mothers working full-time.



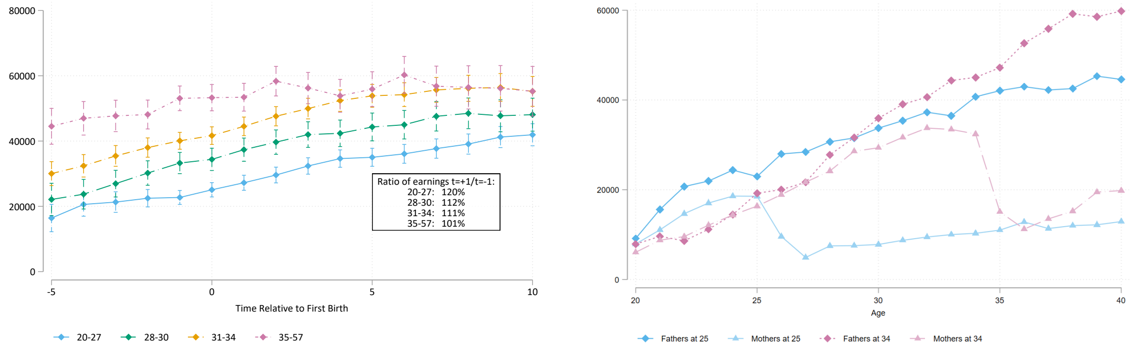
(E) Annual working hours for mothers (including zeros).



(F) Annual working hours for mothers conditional on working.

FIGURE 3.13: Annual earnings and working hours, occupational rank, labor force and full-time status of mothers by their age at first childbirth.

Notes: The figure shows a measure of annual earnings and working hours, occupational rank, labor force and full-time statuses of mothers for four quantiles of the distribution of age at first birth. The first quantile includes mothers aged 20–27 at first birth, the second those age 28–30, the third those aged 31–34 and the fourth those from 35–57. *Source:* Own calculations based on the SOEP.



(A) Annual earnings of fathers by event time.

(B) Annual earnings of fathers and mothers by age (childbirth at the age of 24 and 35).

Figure 3.14: Average earnings of fathers and mothers around birth.

Notes: The left figure plots average annual labor earnings of fathers in time relative to their first birth for four quantiles of the distribution of age at first birth. The first quantile includes fathers aged 20–27 at first birth, the second those age 28–30, the third those aged 31–34 and the fourth those from 35–57. The right figure plots annual labor earnings of mothers and fathers who became parents at the age of 24 and 35. *Source:* Own calculations based on the Soep.

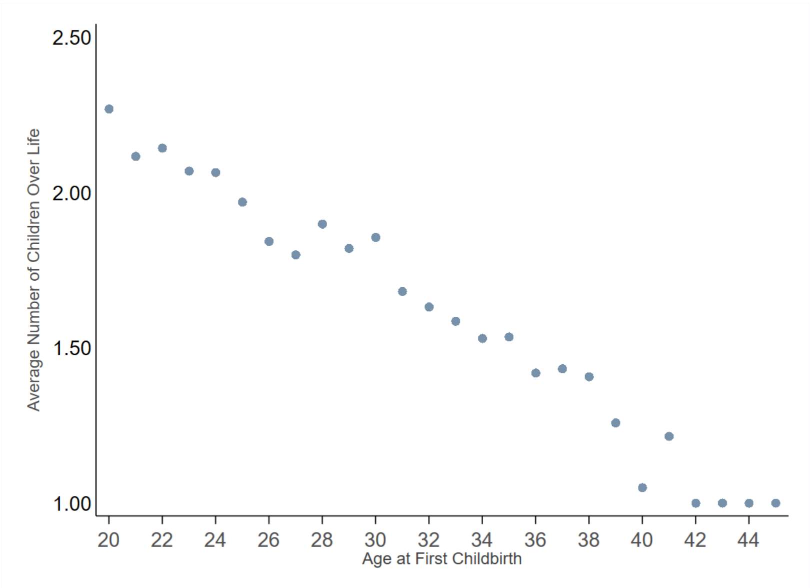
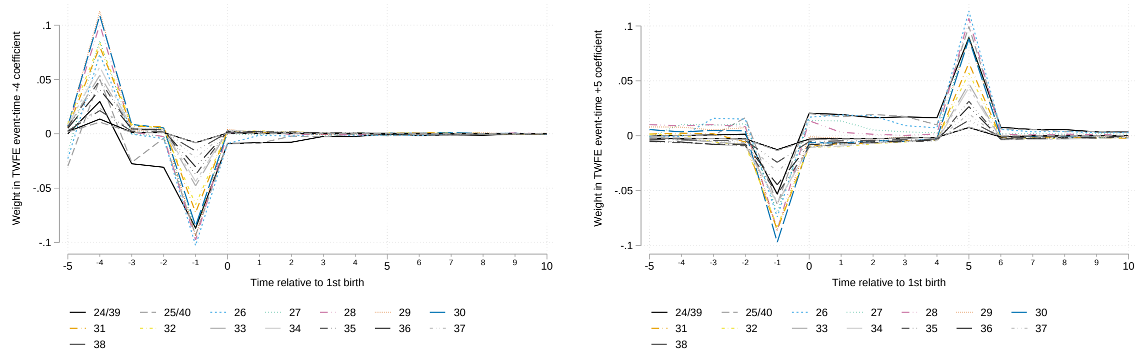


Figure 3.15: Average number of children over life by age at childbirth.

Notes: The figure plots average number of children in total over life for mothers by age at first birth. *Source:* Own calculations based on the Soep.



(A) 4 years before childbirth.

(B) 5 years after childbirth.

FIGURE 3.16: Sun and Abraham (2021) decomposition of weights: "contamination" from other periods.

Notes: The figure plots the weights obtained from the Sun and Abraham (2021) decomposition for the relative time periods −4 (left-hand panel) and +5 (right-hand panel). The weights are calculated with the `eventstudyweights` Stata module provided by Sun and Abraham (2021). *Source:* Own calculations based on the SOEP.

TABLE 3.4: Share of pre-birth observations to construct counterfactual earnings by relative time and age at first childbirth.

Time relative to 1st birth	Age at 1st birth				
	24	26	29	32	35
−5	1.000	1.000	0.733	0.537	0.310
−4	1.000	0.866	0.679	0.455	0.268
−3	1.000	0.794	0.582	0.373	0.216
−2	0.866	0.733	0.537	0.310	0.181
−1	0.794	0.679	0.455	0.268	0.136
0	0.733	0.582	0.373	0.216	0.106
1	0.679	0.537	0.310	0.181	0.083
2	0.582	0.455	0.268	0.136	0.071
3	0.537	0.373	0.216	0.106	0.060
4	0.455	0.310	0.181	0.083	0.042
5	0.373	0.268	0.136	0.071	0.020
6	0.310	0.216	0.106	0.060	0.006
7	0.268	0.181	0.083	0.042	0.008
8	0.216	0.136	0.071	0.020	0.000
9	0.181	0.106	0.060	0.006	0.000
10	0.136	0.083	0.042	0.008	0.000

The table reports the share of pre-birth observations that are available to construct counterfactual earnings as in Equation (3.3) by relative time around childbirth and for 5 levels of age at first childbirth. The age levels correspond to the 10th, 25th, 50th, 75th and 90th percentiles of age at first childbirth. *Source:* Own calculations based on the SOEP.

4 Wage Inequality Consequences of Expanding Public Childcare

Family policies are important in shaping the labor supply of women. Hence, for the female part of the population, they play a role as labor market institutions. While the impact of such institutions as unions or minimum wages on wage inequality has extensively been studied, family policies and their effect on inequality have received much less attention. This paper, therefore, assesses the impact of a large expansion of public childcare in Germany on wage inequality. Exploiting regional variation in childcare supply, I show that in regions with stronger increases in childcare wage inequality among women increased less strongly compared to regions with smaller increases. This is primarily driven by the lower half of the wage distribution and qualitatively similar for both full- and part-time workers. Larger expansions in childcare, however, do not contribute to a further closing of the gender wage gap, suggesting they are associated with a more negative selection of women into the workforce.

4.1 Introduction

A large body of literature documents the impact of changes in labor market institutions on inequality of wages (Fortin and Lemieux 1997). Common examples for such institutions are labor unions or minimum wage regulations. Most of the existing studies focuses on the male part of the workforce since women face a different set of constraints when making decisions on human capital acquisition and labor supply. Almost all constraints women face additionally to men are—at least indirectly—related to the costs of having children. The impact children have on female careers is large, long-lasting and well-documented, ranging from direct earnings losses due to maternity leave breaks or part-time work (Angelov, Johansson, and Lindahl 2016; Adda, Dustmann, and Stevens 2017; Kleven, Landais, and Søgaard 2019) to reductions in fertility (Doepke and Kindermann 2019). Consequently, policies that change women's calculus with respect to work and fertility choices play a role as labor market institutions as well. In this paper, I therefore evaluate which impact a large expansion of publicly provided childcare in

Germany had on wage inequality. I look, both, at the dispersion of wages among women but also at the gender wage gap as a measure for between-gender inequality.

Over the course of the past decades most developed countries saw striking increases in female labor force participation (Olivetti and Petrongolo 2016). One enabling factor in increasing female labor supply was the expansion of publicly provided childcare which made reconciling work and family easier. While a direct effect of childcare on inequality (such as observed for union membership of minimum wages) is unlikely, this paper's hypothesis is that easier availability of childcare promoted changes in female labor supply such that the composition of the female workforce and, in turn, inequality in wages changed.

Germany provides a very well suited setting to assess the impact of childcare on wage inequality. Before the expansion of childcare, female labor force participation was relatively low and likely selective. One of the major family policies was parental leave. By granting job protection and benefit payments during leave—both was extended multiple times until 1993—it set incentives for prolonged labor market absence of women after childbirth. When in 1996 Germany introduced the legal right to a slot in public childcare for children of kindergarten age (usually starting at age three until the start of primary school at age five or six), this marked a shift in policy towards promoting labor supply of women, especially mothers, by lowering the opportunity costs of work. This shift happened at a time when the length of parental leave taking had started to decrease, indicating a general shift to more positive attitudes towards working mothers and thus making it more likely that women utilized the additional childcare. Further, the 1990s are a period during which the participation of women in the labor market substantially increased—to a large degree due to work in part time—and wage inequality (among both genders) increased as well. To connect these observations, I assess how the changes in the composition of the female workforce and the development of wages differ between regions with differential increases in public childcare supply. I, further, decompose the changes in wage inequality within female workers and between genders to show the impact of workforce composition. For most analyses, I make a distinction between full- and part-time working women to highlight similarities and differences between both groups.

I use administrative data on wages and characteristics of workers along with data on county-level childcare supply to provide the following main findings on the development of wage inequality between 1986 and 2010. First, in regions with larger increases in childcare supply more women select into part-time work, female workers are more often from the middle of the education distribution and work in more stable jobs. Second, while wage inequality overall increases, stronger regional increases in childcare are associated with smaller increases in inequality. In 1986, counties that increase their childcare supply stronger show the larger levels of wage inequality, until 2010 this relationship has reversed. This is primarily driven by workers from the lower end of the wage distribution. For 1986, an (at this time future) increase in childcare at the county-level by one standard deviation (or 13.36 additional slots per 100 children) is associated with a p50–p15 log wage gap that is larger by 21 percent of a standard deviation. In 2010, however, the same increase in childcare is associated with a p50–p15 wage gap that is smaller by 8.5 percent of a standard deviation. Third, stronger regional increases in childcare are further associated with larger fractions of the increase in inequality being explained by compositional changes in the workforce; this is as well driven by the lower end of the wage distribution. Among these compositional changes, those that are related to changes in the participation decision appear to be more relevant than changes in the choices of women who already work. Lastly, I move from a within-gender perspective to comparing women and men. The gender wage gap decreased over the observation period. However, in regions with stronger increases in childcare supply smaller fractions of this decrease can be explained by changes in the workforce composition, i.e. an inverse relationship between inequality and workforce composition compared to within-gender inequality. Together with the findings for within-female inequality, this points out that an increasing supply with public childcare affected primarily the participation decisions of women with lower earnings potential. Compared with the initial lower end of the female wage distribution, those who provided additional labor were a more positive selection such that inequality among women increased less. At the same time, when comparing with male workers, they were negatively selected such that they had a smaller contribution to the fall of the gender pay gap. For part-time workers, the findings are broadly similar with the relationship between larger increases in childcare and smaller increases in inequality being stronger.

This paper builds on and contributes to several strands of the literature. There is a number of studies analyzing the development of wage inequality in Germany that focus on male workers. Dustmann, Ludsteck, and Schönberg (2009) focus on the roles of composition, declining unionization and polarization between occupations, whereas Card, Heining, and Kline (2013) highlight sorting between workers and firms. Dauth, Findeisen, Moretti, et al. (2022) extend studying worker-firm sorting to assess spatial wage inequality. Dustmann, Lindner, et al. (2022) and Bossler and Schank (2023) analyze the effects of the German minimum wage. While West Germany often is often the sole focus of studies Brüll and Gathmann (2020) analyze wage inequality in East Germany. With respect to inequality between women and men Antonczyk, Fitzenberger, and Sommerfeld (2010) assess gender wage inequality in Germany focusing on workforce composition and unionization. They and especially the results of Biewen, Fitzenberger, and Lazzer (2018) put emphasis on personal characteristics such as education or experience in explaining rising inequality. Bruns (2019), building on Card, Cardoso, and Kline (2016), analyzes the growing role of firms for wage setting (among other factors due to the decline in collective bargaining coverage) and the gender inequality contribution of sorting of workers to firms. Drechsel-Grau et al. (2022) combine social security and tax data to calculate inequality series on income of both genders. In comparison to most existing work, this paper focuses on female workers. It is the first to relate wage inequality to a family policy, namely an expansion in public childcare. In addition, it provides the first results on wage inequality among part-time working women who—both during the 1990s and 2000s as well as today—account for a large part of the female workforce. My findings on the gender wage gap highlight the role of selection into work for women (Olivetti and Petrongolo 2008; Mulligan and Rubinstein 2008) and connects that to a policy change.

I, further, contribute to the literature evaluating the consequences of expansions in the supply of public care for children of kindergarten age. Existing international evidence paints a positive to mixed picture of the employment effects of childcare (see, among others, Baker, Gruber, and Milligan 2008, Cascio 2009 and Havnes and Mogstad 2011; Olivetti and Petrongolo 2017 and Albanesi, Olivetti, and Petrongolo 2022 give overviews of the literature) where substitution from informal to public childcare appears to weaken the link. Bauernschuster and Schlotter (2015) provide an important finding on which this paper builds as they analyze the expansion of childcare slots in Germany in the 1990s at the micro-level. They conclude that it led to marked increases in maternal labor supply, both on the intensive and the extensive margin. I extend these findings by looking at the entire wage distribution and address the question how additional women joining the workforce as well as changing behavior of those who were already part of it led to compositional changes and, in turn, affected wage inequality.

The remainder of this paper is organized as follows. Section 4.2 describes the data I use, Section 4.3 gives an introduction to the German labor market during the 1990s and 2000s and describes the changes in family policy, especially the expansion of childcare supply. Section 4.4 explains which effects of additional childcare options on wage inequality can be expected under which circumstances. Section 4.5 shows changes in workforce composition and Section 4.6 presents the main results for the development of wage inequality over time, decomposes these changes, separately for full- and part-time working women and by regions with lower or higher increases in their childcare supply. It, further, decomposes the decrease in the gender wage gap by regional increases in childcare supply. Section 4.7 concludes.

4.2 Data

This paper uses individual-level, administrative data on workers and combines them with information on the availability and utilization of childcare facilities at the regional level. This section describes the datasets.

4.2.1 Individual-level Labor Market Data

As most work on wage inequality in Germany this paper relies on the *Sample of Integrated Employment Biographies* (Siab, Frodermann et al. 2021) provided by the *Institute for Employment Research* (Iab) as its main source of data on worker and their characteristics. The Siab is a two-percent-sample from the German social security records containing information on workers who are subject to social security contributions for the period 1975–2019.

Data Preparation As the data is sourced from notifications by employers to the federal employment agency which do not explicitly consider the data needs for research purposes, some limitations apply. They are discussed in the following.

Wages above the social security threshold are censored. Earnings are only subject to social security contributions up to an upper limit. If earnings exceed this limit, the associated daily wages are right-censored such that they need to be imputed based on individual-level characteristics.[1] All results reported in the following are based on these imputed daily wages.

Compared to work that focuses on men, censored wages are a much less severe concern for the results in this paper which puts its emphasis on women. Card, Heining, and Kline (2013) report a share of around 2 to 3 percent of censored wages for women, whereas around 10 percent of male wages are censored. In this paper's sample up to 3 percent of female wages are censored on average (see Table 4.1), in some regions up to 4 percent (see Table 4.2).

Workers in part-time work. Most existing studies on wage inequality restrict their samples to those working full time. As their primary focus are male workers this restriction applies only to very small fractions of the data. For this paper, I choose to keep all workers, both in full- and in part-time work, in the sample. Dropping women in part-time work would reduce the sample substantially as during the 1990s and early 2000s between around 25 and 40 percent of women worked part time. As part-time work among women is a choice that is likely related to circumstances such as the presence of children and the availability of childcare, focusing on those in full-time work would reduce the overall representativeness of the analysis. To allow for different effects for full- and part-time workers, the analysis treats them separately.

It has to be noted that the indicator for part-time work in the Siab data has limitations. In 2011 the notification procedure to record workers in part-time has changed. Fitzenberger and Seidlitz (2020) argue that for the period prior to 2011 the share of actual part-time workers was larger than the one identified in the Siab. They also propose a correction which, however, can only be used for workers who are part of the sample in 2012 (i.e. in a year where their

[1] Section 4.A in the Appendix provides are description.

part-time status is recorded correctly). As the probability to observe a person in 2012 decreases with greater distance in time and this paper focuses on a period starting in the mid 1980s, this correction cannot be applied. In practice, this means that while the full-time sample can include women in part-time work and thus some bias cannot be fully ruled out, the sample of part-time working women only includes those who actually work in part-time. There is no indication that the amount of false part-time reports changed substantially over time prior to 2011. Fitzenberger and Seidlitz (2020) compare observed and corrected wage percentiles between 2000 and 2010 finding differences in levels, though no different trends. As this paper primarily focuses on changes in inequality a bias that some full-time workers in fact work in part-time is unlikely to have a large impact on its results.

Another shortcoming of the Siab data is that there is no information on working hours beyond the part-time indicator. When measuring how inequality changes over time this poses the threat that changes in inequality among part-time workers are driven by changes in intensive margin labor supply. Since this concern cannot be addressed directly with the Siab, I use survey data from the German Socio-Economic Panel (Soep) to assess working hours of full- and part-time workers. Figure 4.12 in the Appendix plots contracted working hours of women in full- and part-time for the sample period between 1986 and 2010. Working hours of full-time workers vary little, both over time and within each year. Part-time workers show a larger spread in hours which is likely to explain some fraction of cross-sectional wage disparity. Over time, however, the variation in working hours does not show substantial changes. Wanger (2020) also documents only small changes in mean working hours of women in part-time and additionally points out that hours changes for part-time workers are often driven by workers in marginal employment, i.e. a group that this study drops from the sample (see below). It is thus unlikely as well that changes in working hours drive changes in inequality.

Break in the notification procedure for workers in marginal employment. As of April 1999, employers are required to submit employment notifications for workers in marginal employment which leads to an influx of additional workers into the sample (Frodermann et al. 2021). Even though workers in marginal employment can be identified based on their wages before the reporting change, they have to be dropped to ensure a sample that is consistent over time.

Workers in East Germany. Family policy in the German Democratic Republic (Gdr) put an emphasis on the wide-spread availability of childcare to promote maternal labor supply. The expansion of childcare the western part of Germany experienced thus does not have an analog in the East. Family policies in the Gdr are further associated with generally different attitudes towards female employment such that East and West Germany are hardly comparable.[2] Therefore, I only keep workers in West Germany (excluding Berlin) in the sample.

TABLE 4.1: Summary statistics for women in 1986 and 2010.

	1986		2010	
	Mean	SD	Mean	SD
Log daily wage	4.13	0.49	4.25	0.57
Share censored wage	0.02	0.12	0.03	0.16
Age	37.84	11.21	42.26	10.23
Share part-time	0.26	0.44	0.40	0.49
Education				
Share no vocational degree	0.24	0.43	0.08	0.28
Share vocational degree	0.71	0.46	0.77	0.42
Share university degree	0.04	0.19	0.14	0.34
Years in employment	7.79	3.58	15.89	8.79
Years in current job	5.70	4.00	7.31	7.03
Observations	132,550		164,493	

Notes: Summary statistics for women in regular employment, age 21–60 in the years 1986 and 2010. The log of daily wages is given in Euro, inflation adjusted to 2015 as base year. *Source:* Own calculations using the SIAB data described in Section 4.2.1.

Sample Characteristics The final dataset is a panel of West German workers in regular employment who are between 21 and 60 years old. Workers in vocational training, marginal employment, interns and others in less common employment relationships are dropped. The data covers the period 1986 until 2010. I chose 1986 as starting point as this is the first year for which data on childcare is available (see below). I extend the observation period beyond the last data point for childcare (2002) until 2010 to be able to capture long-run effects. These are a likely possibility because during most of the increases in childcare supply in the mid 1990s slots had to be rationed (see Bauernschuster and Schlotter 2015, and Section 4.3). Widespread use of public childcare thus cannot be expected to immediately have in impact. Both for mothers who intended to utilize it as well as for all women, it takes time to adapt decision making on having children and labor supply. The duration of parental leave, for instance, is typically decided on before the birth of the child such that the decision has to take into account the childcare supply around three years before the child can enter kindergarten. Extending the analysis to years beyond 2010 is not possible because of the break in the reporting of part-time work in 2011.

Table 4.1 provides summary statistics of the SIAB data in 1986 and 2010. Throughout the paper, monetary values are expressed in Euro, inflation adjusted to the base year 2015; wages always

[2] Becker, Mergele, and Woessmann (2020) argue that marked differences in gender norms between the eastern and western parts of Germany even predate the creation of the GDR. Boelmann, Raute, and Schönberg (2022) exploit these differences to provide evidence on the relationship between gender norms and the labor supply of mothers.

refer to log imputed wages. The table points out some of the changes the female workforce underwent over time. The share of women who work in part-time strongly increased (from 26 to 40 percent). Education levels increased as well. In 2010, only on third of the initial 24 percent of women without a vocational degree remains, the share of women with vocational education increases from 71 to 77 percent and the share of women who own a high school degree more than triples (from 4 to 14 percent).

4.2.2 County-level Data on Childcare

Since the SIAB data do not allow to observe individual take-up of childcare, I use regional differences in the supply of care for children of kindergarten age as a proxy for individual exposure, comparable with an intention-to-treat approach in a policy analysis setting. The main explanatory variable is the county-level change in the ratio of kindergarten slots to the number of children between age three and five. It is constructed by combining data from the regional database of the German Youth Institute (Bertram, Bayer, and Bauereiss 1993) and the German Federal Statistical Office (Statistisches Bundesamt 2023c; Statistisches Bundesamt 2023b). I describe the change in childcare supply in more detail in the next section.

4.3 Institutional Background

This sections starts with giving a broad overview of the main changes in the German labor market between the 1980s and the 2010s to, then, narrow its focus to policy changes affecting especially female workers. Its last part provides a description of this paper's object of investigation, the expansion of public childcare in the mid 1990s.

Trends in the German Labor Market Over the 1990s and early 2000s the German economy and the labor market faced several difficulties until a recovery in the later 2000s. The oil price shocks at the beginning of and economic turmoil during the 1980s led to high unemployment rates that persisted over the following decade. The German reunification and the associated costs put additional pressure on the economy. Starting in the mid of the 1990s, we observe marked increases in wage inequality, both for men and for women (Dustmann, Ludsteck, and Schönberg 2009). An explanation that is often put forward argues that trade exposure—especially with Eastern Europe following the fall of the Iron Curtain—and increased competition had negative economic impacts (following Autor, Dorn, and Hanson 2013). Dauth, Findeisen, and Suedekum (2014) instead argue, that in Germany job losses due to trade exposure were caused in some industries while new jobs in other industries offset these negative impacts. Goldschmidt and Schmieder (2017) highlight the contribution of domestically outsourcing certain occupations to wage inequality. Starting in 2003 Germany implemented a set of labor market reforms intended to foster flexibility of employment (coined "Hartz IV"). In the following, the economy recovered and unemployment decreased which, however, was not necessarily an effect of these reforms. Dustmann, Fitzenberger, et al. (2014) point instead to the decline in unionization and the share of workers covered by collective bargaining agreements starting in the mid 1990s. By allowing for more decentralized wage setting, this contributed to lower wages, mostly at the lower end of the distribution, which in turn increased competitiveness and contributed to the overall recovery of the German economy.

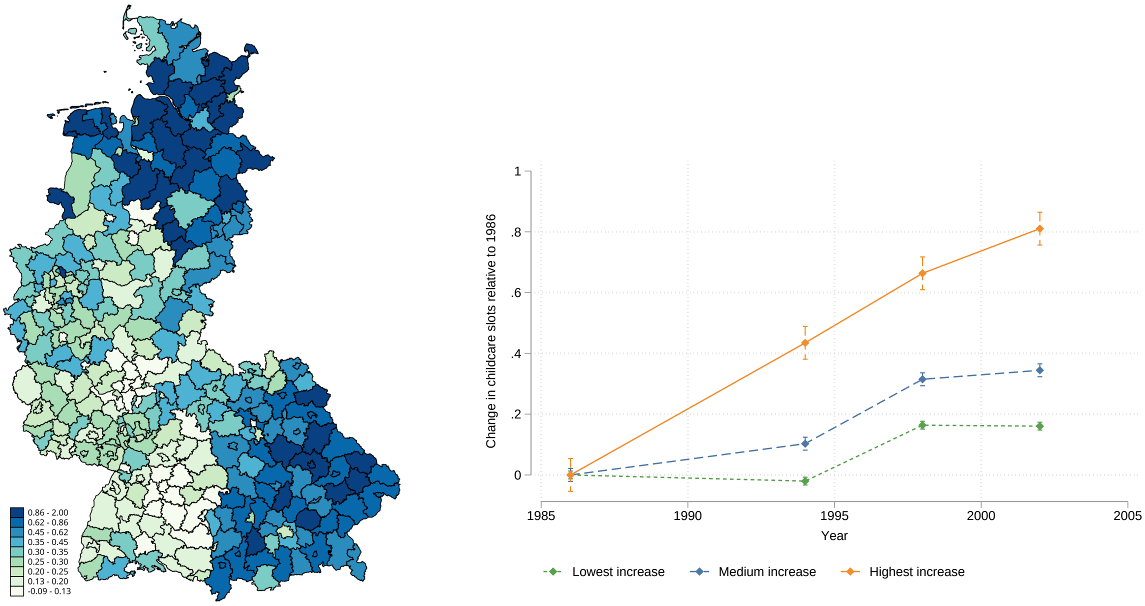
Female Workers and Changes in Family Policy Against the backdrop of these broad trends, the female part of the workforce experienced additional changes. Most notable is the increasing number of women who entered the workforce, though a substantial fraction of them chose to work in part-time (see also the descriptive statistics in Table 4.1 as well as Section 4.5).

As women take on the main responsibility for raising children and household work (still today but even more so in the 1990s), family policies can play an especially important role for female labor supply. Before the expansion of childcare in the mid of the 1990s, parental leave was the arguably most important family policy. Starting in 1979, Germany first introduced six months of job protected maternity leave during which maternity benefits were paid. Multiple reforms extended the duration of both job protection and benefits as well as they introduced eligibility for men, among whom the take-up is, however, small. Between 1993 and 2006 parental leave was in terms of duration most generous offering 36 months of job protection and up to 24 months of benefit payments. The incentives this policy set were ambiguous as, on the one hand, it promotes employer continuity which can be career improving, but, on the other hand, it encourages longer labor market breaks after childbirth that can be harmful for maternal careers through losses of human capital. As benefits were paid as lump sums until 2006 their theoretical effect of delaying the re-entry to the labor market is of relatively greater importance for lower earning mothers. Schönberg and Ludsteck (2014) find overall small effects of this policy on maternal labor market outcomes which Findeisen et al. (2023) attribute to the effect of job protection. An additional reform in 2007 tied the level of benefit payments to mothers' pre-birth earnings from which the higher earning mothers gained relatively more.[3]

Contrary to the incentive to take longer post-childbirth labor market breaks the average duration of maternity leave starts to decline for children born since 1994 (see Figure 4.14 in the Appendix) suggesting that the preferences of mothers started shifting towards an increased willingness to participate in the labor market, even in the presence of small children. The

[3] Bergemann and Riphahn (2023) and Kluve and Schmitz (2018) assess the labor market consequences of these reforms, Raute (2019) shows that fertility increased for higher earning women.

increase in childcare supply, thus, happened in a environment where mothers generally appear to be willing to make use of it.



(A) Geographical distribution of increase in childcare supply between 1986 and 2002, normalized by 1986 values.

(B) Increase in childcare supply between 1986 and 2002, normalized by values in 1986 and grouped by counties with low, medium and high increases.

FIGURE 4.1: Increase in childcare supply for children of kindergarten age between 1986 and 2002.

Notes: The right-hand figure documents the increase in childcare supply relative to levels in 1986 along with 95-percent confidence bands. Absolute levels are plotted in Figure 4.13 in the Appendix. Childcare supply is measured at the county-level as the number of slots in public childcare for children of kindergarten age per 100 children aged three to five. The left-hand figure shows the geographical distribution at the county-level of the increases in childcare supply. *Source:* Own representation using the county-level data described in Section 4.2.2.

Expansion of Childcare A further work incentive for mothers was set by the expansion of childcare that this paper uses as its main variation of interest. Starting as of 1996[4], children from the age of three to school entry (which can be between age five and seven, depending on the federal state and the parents' decisions) had the legal right to a place in kindergarten. This right entitled children and their parents only to half-day care; full-day childcare options were sparse at that time. As Bauernschuster and Schlotter (2015) point out, the supply with childcare could not meet the demand such that initially rationing measures had to be implemented. Importantly for this paper, childcare supply was expanded by different extents and at different speed at the regional level. The map in Figure 4.1a plots the county-level relative increase in the number of available slots in public childcare for children of kindergarten age per 100 children

[4] The relevant reform changed the social security legislation in the *SGB VIII, Achtes Buch Sozialgesetzbuch*.

of age three to five between 1986 and 2002, normalized to zero in 1986. It shows that the largest increases were concentrated in Bavaria in the south and Lower Saxony and Schleswig-Holstein in the north. In these counties the number of available childcare slots on average doubled, in some counties even tripled. The smallest increases, in some counties even slight decreases, can be found in Baden-Württemberg and Hesse. Especially the difference between Bavaria and Baden-Württemberg stands out as in both regions conservative gender norms are more prevalent. This suggests that regional differences in gender norms are unlikely to be a main driver of this paper's findings.

This expansion of childcare had significant effects on maternal labor supply. Bauernschuster and Schlotter (2015) estimate that childcare attendance of the youngest child increases mothers' labor supply both on the extensive (by around 37 percentage points) and on the intensive margin (by around 14 hours per week).

Identifying Variation Most analyses in this paper use aggregated county-level changes in childcare supply and relate them to changes in workforce composition and wage inequality. I group the 324 West German counties[5] into terciles with 108 counties each. The variable to build the terciles is the increase at the county-level in the number of childcare slots available to children of kindergarten age between 1986 and 2002. It differentiates between three groups of counties, henceforth referred to as *regions*, with low, medium and high increases in their childcare supply. Figure 4.1b illustrates the relative increase in childcare supply by region. In the low- and medium-increase regions the increase over time amounts to 16, respectively 34 percentage points. The largest increases happen between 1994 and 1998. Between 1986 and 1994, childcare supply in the low-increase regions even decreases slightly while it increases by 10 percentage points in the medium-increase regions. The high-increase regions show a steady increase over the entire period of observation, in 2002 it amounts to around 81 percentage points relative to 1986 levels. It has to be noted that the regions did not start at equal levels of childcare. Figure 4.13 in the Appendix plots the absolute increases in the number of childcare slots by region. It indicates that larger increases are associated with lower initial levels, while in 2002 all regions reached similar levels. The stronger increase in childcare in some regions is thus rather a catch-up process than an overtaking of other regions.

Apart from these observable initial differences between the regions, there can be unobservable ones. For instance, women's preferences regarding labor supply can differ as well as gender norms of women, their families or of employers. To take them into account, this paper primarily aims to explain changes in workforce composition and inequality by changes in childcare supply. As long as regional differences are constant over time, this removes them similar to fixed effects. Partitioning the sample into three regions allows for a sufficiently large

[5] In German *Landkreise*, referring to the NUTS-3-level.

number of observations in each region to provide robust graphical results and to calculate DFL weights while ensuring common support between baseline and target year. To provide a further validation of the results that are obtained across regions, I provide additional results at the county-level in Section 4.6.2.

4.4 Potential Effects of Childcare on Wage Inequality

Easier and more widespread access to childcare options can generally be expected to lower women's opportunity costs of working, especially for mothers. How this in turn impacts wage inequality is, ex ante, ambiguous. It rather depends on which group of women reacts in which ways. To give a background for the later analysis, this section provides an overview how women's labor supply can be affected by increased childcare options and the related effects on wage inequality.

The following considerations make some assumptions. First, women can be ranked by their earnings potential. This applies to those who work and those who do not work. Second, larger supply with public childcare decreases the price of childcare. Here, price refers to the monetary costs of childcare as well as the non-monetary opportunity costs of organizing care for a child. The latter can include that public and informal care by relatives can easier be combined or that childcare is more conveniently to access, for instance due to being located closer to the home or the workplace. Third, I assume sufficient labor demand, i.e. that increasing female labor supply translates into employment, and rule out general equilibrium effects such that more labor supply by women due to childcare does not lead to changes in wages.

Mothers with young children are the likeliest beneficiaries of an increased availability of childcare. For them, two channels can be distinguished. The first one emphasizes labor supply on the extensive margin. In a simplified model case, mothers work if their net benefits of working are positive. Holding other factors constant, the net benefits of working refer to the (potential) earnings net of childcare costs. With lower childcare prices, more mothers have positive net benefits of working such that labor force participation increases. Mothers either revert a non-participation decision or they decrease the duration of a post-childbirth labor market interruption. If childcare costs are similar for all mothers, this will not affect mothers with sufficiently high earnings potential as they had positive net benefits of working already before the childcare expansion. Instead, such a mechanism will rather expand the workforce towards those mothers with lower earnings potential. Their increased participation, then, will increase inequality.

The second channel focuses on increased flexibility. This is especially relevant if some level of informal care, for instance by relatives, is already present, because public childcare only covered half of the day.[6] As before, an expansion of public care lowers the price of childcare in general but, in addition, also increases the flexibility of mothers as, in total, more care is available and easier to plan. Additional flexibility can impact mothers by allowing them easier commuting, such that they have a broader choice of employers, by being less depended on the family-friendliness of an employer or by giving them additional time. These aspects ease the constraints they face when making labor supply decisions. This can lead to an increase in participation in the labor market, but also to additional working hours of mothers who already participate. It, further, can lead to fewer downgrades in terms of hours, occupation or employer when making a participation decision after childbirth. Increasing flexibility decreases the opportunity costs of working such that it more likely affects women with lower earnings potential. As, however, the availability of informal care is unlikely to strongly depend on earnings potential, the effect of additional public care is not necessarily restricted to them. It can also extend upwards in the distribution as well as to those women who are already part of the labor market. Nevertheless, the effect is likely to decrease in strength in earnings potential since better earning women have more opportunities to organize and pay for childcare, even in absence of public care provision.

These channels are not mutually exclusive. Given the results by Bauernschuster and Schlotter (2015) who find labor supply effects of the German childcare expansion on the extensive and intensive margin, it is plausible to expect a combination of both.

Turning away from mothers, an expansion of childcare can, more indirectly, also benefit women before they have children. For them, not the direct but the expected opportunity costs of working decrease.[7] With increasing prospects of a career that can be sustained in the long run, investments into education are more likely to pay off such that the probability of young women pursuing higher degrees and selecting into higher-paying occupations increases as well. After women have entered the labor market, better long-run career prospects can affect employer choice, effort put into work or on-the-job training. Under the assumption that concerns that center around the costs of having children are more relevant for women with lower intrinsic career preference, i.e. have a lower earnings potential, they can be expected to have the larger gains from childcare. Therefore, this indirect channel rather decreases wage inequality among women.

6 Informal care likely plays a substantial role during the observation period since a market for private childcare was almost nonexistent in the early and mid 1990s (see Bauernschuster and Schlotter 2015).

7 This rests on the assumption that these women expect to have children at some point in the future. Given that for the cohorts who are of childbearing age during the 1990s the share of those who eventually become mothers is close to or above 80 percent (Statistisches Bundesamt 2019), this is likely to be the case.

4.5 Increasing Childcare Supply and Changes in Workforce Composition

Over the 1990s we see a number of changes in the German labor market, both regarding the development of wages and the composition of the female workforce. This section describes the compositional changes and relates them to changes in the supply of childcare at the regional level. Women in regions with larger childcare expansions show stronger increases in part-time work, they are more often in the middle of the education distribution and have accumulated longer tenure in their current job. As most results in this section focus on changes over time relative to the baseline year 1986, Table 4.2 lists summary statistics in levels by region for 1986 and 2010. The regions do not enter the observation period with similar characteristics. Regions where childcare supply increased stronger have the lower wage levels, both in 1986 and 2010. They also have slightly higher shares of part-time working women; these differences become more pronounced until 2010. Education levels are mostly similar across regions. Experience measures are smaller in regions with higher childcare increases.

Table 4.2: Summary statistics for women by regional increase in childcare supply, years 1986 and 2010.

	1986			2010		
	Low	Medium	High	Low	Medium	High
Log daily wage	4.17	4.12	4.03	4.30	4.24	4.13
Share censored wage	0.02	0.01	0.01	0.04	0.02	0.01
Age	37.93	37.91	37.46	42.14	42.31	42.44
Share part-time	0.25	0.26	0.27	0.38	0.40	0.44
Education						
Share no vocational degree	0.24	0.23	0.26	0.09	0.08	0.08
Share vocational degree	0.71	0.71	0.70	0.75	0.77	0.82
Share university degree	0.04	0.04	0.03	0.16	0.13	0.10
Years in employment	7.84	7.76	7.70	15.84	15.8	16.14
Years in current job	5.72	5.71	5.64	7.25	7.26	7.55
Observations	61,654	46,904	23,299	74,126	58,639	31,692

Notes: Summary statistics (means) for women in regular employment, age 21–60 in the years 1986 and 2010. The table differentiates between regions with low, medium and high increases in childcare supply as described in Section 4.3. The log of daily wages is given in Euro, inflation adjusted to 2015 as base year. *Source:* Own calculations using the Siab data described in Section 4.2.1.

Part-time Work The likely most prominent change in the composition of the female workforce is the increase in part-time work. Between 1986 and 2010 the share of women who are recorded as working in part-time in the Siab data rose from 26 to 40 percent.[8] This trend is

[8] The German Federal Statistical Office (Statistisches Bundesamt 2023a) records an increase from 29 to 46 percent during this time period which gives an impression of the amount of underreporting in the Siab data. Contrary to the dataset

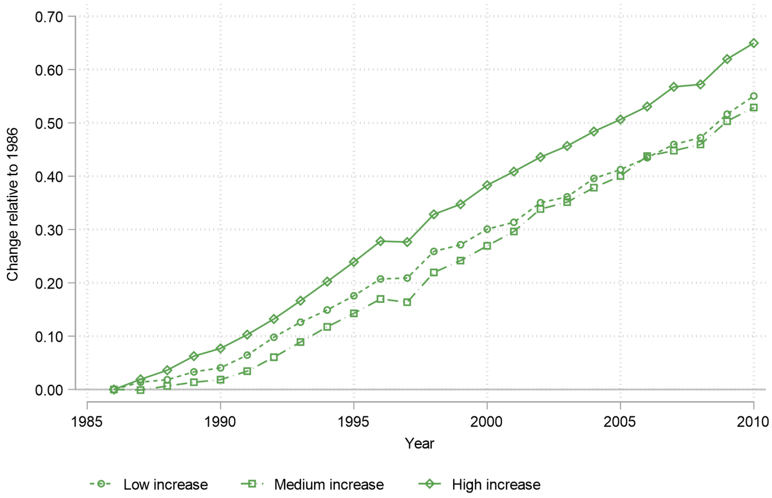


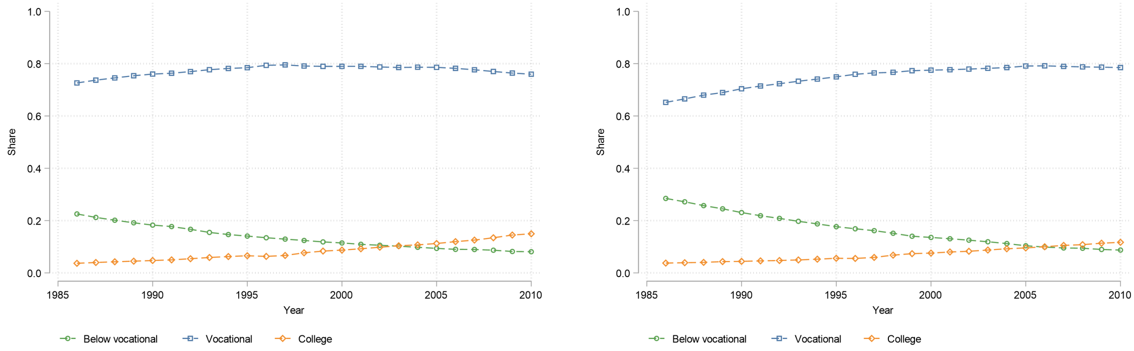
FIGURE 4.2: Change in part-time work relative to 1986 by year and development of childcare supply.

Notes: Changes in part-time work of the female workforce over time by development of childcare supply. Plots indicate changes relative to 1986. Observations are grouped by the position of a region in the distribution of the change in childcare supply between 1986 and 2002. The tercile of regions with the largest increases is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

observable in all regions, although Figure 4.2 shows that there are clear differences in how strong it is. Regions with large increases in childcare supply have the initial slightly higher share of part-time workers (27 vs. 25 percent on average) but also show the strongest increase by 63 percent to 44 percent on average. The change in the other regions with low or medium sized increases in childcare supply is around ten percentage points smaller and less distinct from another at between 52 to 54 percent.

Education As shown in Figure 4.3, the workforce composition with regard to education improved. This is mostly driven by the a substantial decrease in the share of workers with low education levels (from 23 to 8 percent) and an increase in the share of workers with university-level education (from 4 to 14 percent). Nevertheless, most workers—between around 71 and 77 percent—own a vocational degree. Their share increased as well until 1998, but is decreasing since then. In general, the development of the educational composition is similar for full- and part-time workers. For the latter, the decrease in less than vocational respectively the increase in vocational education is stronger while university-level education increases to a smaller degree.

used in this paper the data from the Statistical Office additionally include workers above the age of 60 as well as those in marginal employment. These are both groups for whom the part-time share is relatively large. Taking this into account implies that underreporting by the SIAB is even smaller than suggested by the numbers given here.

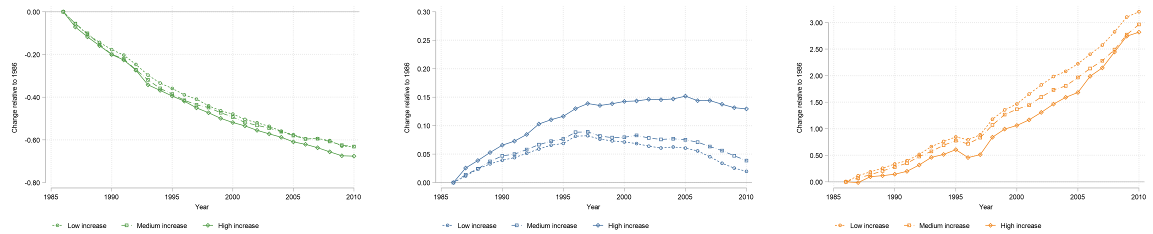


(A) Full-time workers   (B) Part-time workers

FIGURE 4.3: Shares of education levels of the female workforce between 1986 and 2010.

Notes: Education of women by year. Education below the vocational level is plotted in green, vocational education in blue, and university education in orange. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

Differentiating by regional increases in childcare supply yields qualitatively similar trends but with noticeable differences. Figure 4.4 plots the change in the education levels of full-time workers relative to 1986 for regions with low, medium and high increases in childcare supply. The extent of the regional changes are ordered such that those regions with the smallest childcare increases show the smallest changes in below vocational and vocational education, whereas where childcare increased the strongest the reduction in below vocational respectively the increase in vocational education is the strongest. As the initial education levels across the regions show only little differences, this led to substantial changes in the education composition. The difference between low and high-increase regions almost continuously widens over time, in 2010 it amounts to −4 percentage points for below vocational, 11 percentage points for vocational and −39 percentage points for university education. For vocational education, it stands out that the slight decrease since the late 1990s observed in the full sample is almost entirely driven by regions with only low or medium increases in childcare supply. Such an ordering can as well be found for university education. Here, however, greater expansions of childcare coincide with the lowest increases in the share of highly educated women who participate in the labor market. Results for women who work in part-time show similar trends (see Figure 4.15 in the Appendix). For them, the overall increase in vocational education is stronger, though the differences between the regions are smaller. In sum, this findings indicate a greater degree of upgrading from the bottom to the middle of the education distribution in regions with larger childcare expansions. In comparison, in the other regions a relatively larger fraction of the female workforce can be found in the tails of the education distributions, especially the upper one.



(A) Below vocational education (B) Vocational education (C) University or equivalent education

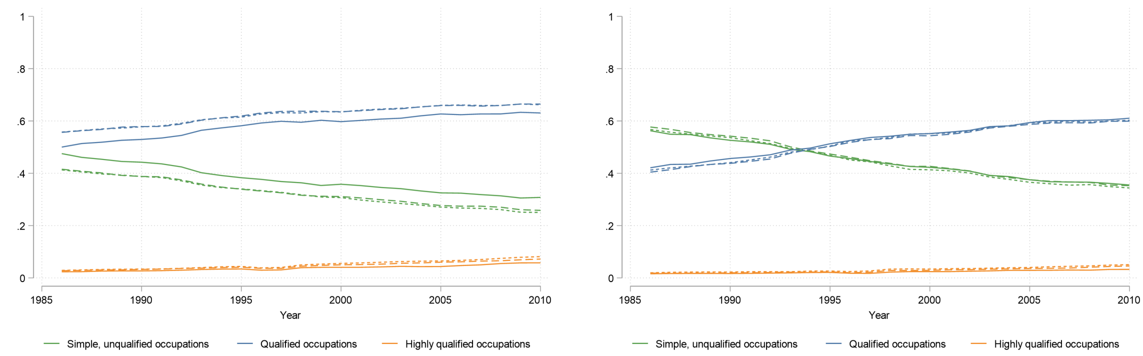
FIGURE 4.4: Changes in education of female full-time workers relative to 1986 by year and development of childcare supply.

Notes: Changes in the education levels of the female full-time workforce over time by development of childcare supply. Plots indicate changes relative to 1986. Observations are grouped by the position of a region in the distribution of the change in childcare supply between 1986 and 2002. The tercile of regions with the largest increases is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. Results for part-time workers are plotted in Figure 4.15 in the Appendix. *Source:* Own estimations based on the SIAB data described in Section 4.2.1.

Selection into Occupations and Associated Skills Figure 4.5 plots the shares of women in three occupation categories by increase in regional childcare supply. The categories are derived from the occupation groups constructed by Blossfeld (1985) and indicate occupations that require on average low, medium and high levels of qualification.[9] For both full- and part-time working women, there is clear evidence for upgrading in terms of occupations. Over time, the share of women working in low-qualified occupations decreases, while there are more women working in qualified and highly qualified occupations. This is consistent with the findings for education. Differentiating by the increase in childcare supply reveals substantially different levels between regions among full-time workers (plotted in the left-hand panel), though no different trends. In regions with high childcare increases, more women are in low qualified occupations and fewer in qualified and highly qualified ones. Part-time workers (depicted in the right-hand panel) show initially higher shares of women in low qualified occupations. Their trends for all categories are similar to the ones of full-time workers. There are only negligible regional differences.

To provide an additional perspective on occupations, I also plot how the workforce composition in terms of the required skills within occupations develops. This measure is derived from the fifth digit in the occupation classification KldB 2010 (Bundesagentur für Arbeit 2021). It differentiates between unskilled and semi-skilled, skilled and two levels of complex tasks workers have to carry out. I aggregate complex tasks into one category since they appear only in small proportions among women. In comparison to the measure used in Figure 4.5 which compares *between* occupations, this measure of tasks also captures differences *within* an occupation. A

[9] Table 4.7 in the Appendix provides the mapping between the Blossfeld groups and the categories by qualification, Table 4.8 (in the Appendix as well) reports average wages and employment shares by occupation categories.



(A) Full-time workers

(B) Part-time workers

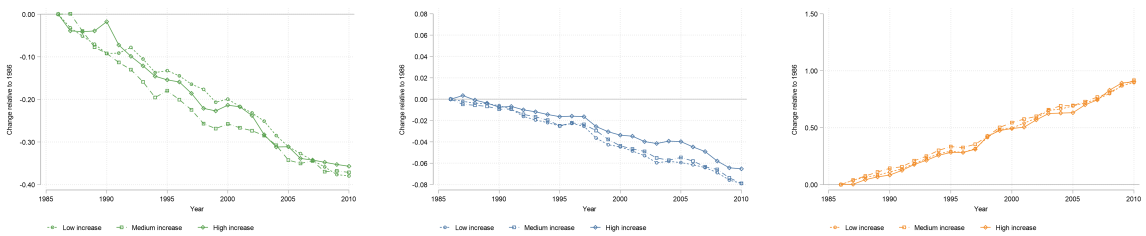
FIGURE 4.5: Shares of women working in occupations of different qualification levels by development of childcare supply between 1986 and 2010.

Notes: Occupations of women by year change in the supply of childcare. Occupations requiring simple, unqualified work plotted in green, those requiring qualified work in blue, and highly qualified occupations in orange. The tercile of regions with the largest increases in childcare is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

nurse, for instance, can fall into all skill categories depending on their specialization and exact place of work. This measure is more suitable to detect changes over the career trajectory, for instance due to receiving further training or being promoted.

Figure 4.6 plots how the task-related skill levels of female full-time workers change over time by regional increases in childcare supply relative to 1986. It confirms the substantial amount of upgrading that education and occupations show as well. Complex and highly complex tasks almost double in their frequencies (from 10 to 19 percent), unskilled tasks become less frequent by around 37 percent (from 7 to 4.5 percent). For both of these skill levels, regional differences are negligible. The share of women in skilled tasks decreases, though less strongly compared to the other two skill groups (from 83 to 77 percent). Regions with low or medium increases in childcare show the largest decreases by up to 8 percent. In regions with higher increases in childcare, the decline in skilled tasks is less pronounced. This divergence starts in the early 1990s; up until 2010 the decrease is only 6.5 percent.

Part-time workers (plotted in Figure 4.17 in the Appendix) show similar decreases in unskilled tasks. Their trend for skilled tasks is not a continuous decline but rather follows a reverse u-shape, that reaches its peak in the mid of the 1990s. The overall changes until 2010 are small, though. Similar to full-time workers, in regions with high childcare increases, skilled tasks are slightly more frequent but the regional differences are less clear. For complex and highly complex tasks, the high childcare increase regions stand out, where they increase by 122 percent. In medium- and low-increase regions the change in more complex tasks is just 101, respectively 98 percent.



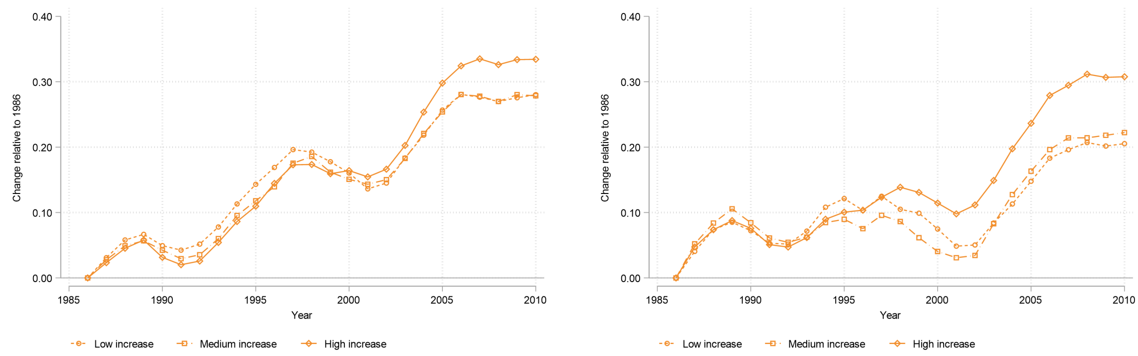
(A) Un-/semiskilled tasks (B) Skilled tasks (C) Complex and highly complex tasks

FIGURE 4.6: Changes in the skill levels of female full-time workers relative to 1986 by year and development of childcare supply.

Notes: Changes in the skill level derived from occupations of the female full-time workforce over time by development of childcare supply. Plots indicate changes relative to 1986. Observations are grouped by the position of a region in the distribution of the change in childcare supply between 1986 and 2002. The tercile of regions with the largest increases is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. Results for part-time workers are plotted in Figure 4.17 in the Appendix. *Source:* Own estimation based on the SIAB data described in Section 4.2.1.

Overall, there are almost no regionally different trends in between-occupation selection. Within occupations, however, full-time workers in regions with high increases in childcare tend to remain in the middle of the skill distribution. This can be found for part-time workers as well, though less pronounced. Instead, for them larger childcare increases are associated with larger shares of women working in jobs that demand more complex tasks.

Tenure and Work Experience Work experience of women increases between 1986 and 2010. Figure 4.16 in the Appendix plots the development in absolute terms for full- and part-time workers. This is, in parts, mechanically related to age. As Table 4.1 shows, average age increases over time. Table 4.2 points out that in 1986 age decreases slightly in the regional increase in childcare supply whereas in 2010 it increases. This matches expectations of an effect of childcare, as it provides women with the opportunity to easier continue work after childbirth and parental leave. Instead of total experience, this section focuses on tenure in the current job which is the more informative measure for current wages as it accounts for the time to accumulate firm and job-specific human capital. The descriptives in Table 4.2 show a relationship that larger increases in childcare are associated with shorter tenure in the current job in 1986 but with longer tenure in 2010. The graphical representation of the changes in tenure over time in Figure 4.7 confirm this. In regions with high increases in childcare supply, both full- and part-time workers show larger increases in their tenure in the current job. For full-time workers, the high-increase regions start to diverge from the others in the early 2000s, in 2010 the increase in tenure is larger by 5.4 percentage points. The trend of part-time workers is similar but for them the high increase regions diverge already since the late 1990s. Their increase until 2010 is by 10.2 percentage points larger compared to regions with low-level increases in childcare.



(A) Full-time workers (B) Part-time workers

FIGURE 4.7: Change in tenure in the current job relative to 1986 by year and development of childcare supply.

Notes: Changes in years spent in the current job of the female workforce over time by development of childcare supply. Plots indicate changes relative to 1986. Observations are grouped by the position of a region in the distribution of the change in childcare supply between 1986 and 2002. The tercile of regions with the largest increases is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. Results for total work experience, tenure in the current firm and tenure in the current job that do not differentiate by childcare are plotted in Figure 4.16 in the Appendix. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

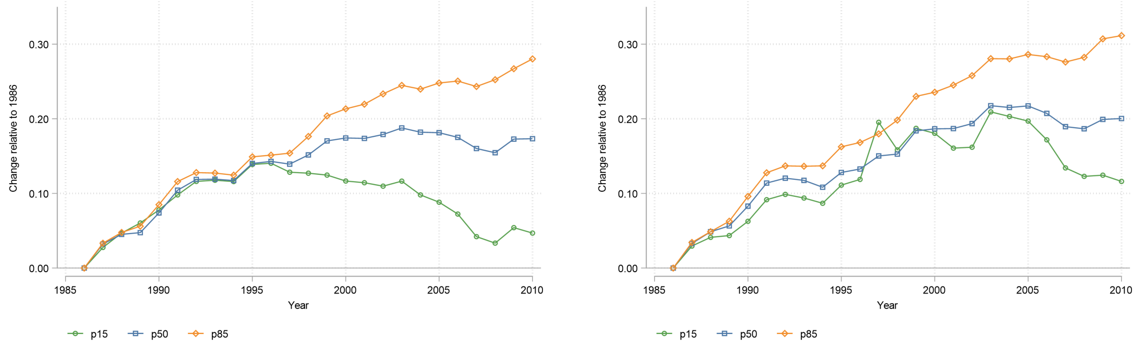
Together with the findings on skills, these results suggest that larger childcare supply is associated with more stable employment relationships among women, which, in turn, induces upgrades in the skill levels of tasks that women perform in their jobs. This finding is more pronounced for women who work in part-time which is consistent with the larger increases of part-time work where childcare increases are larger, as well as with the overall higher prevalence of part-time work among mothers.

4.6 Consequences for Wage Inequality

Having shown the changes in workforce composition, in this section I collect results on the evolution of wage inequality. First, I focus on women and show the development of wages and wage inequality and how this is related to childcare and workforce composition. Then, I extend the perspective to the gender wage gap.

4.6.1 Changes in Wages Over Time

This section, first, shows how wages develop over time, both for all female workers in West Germany as well as separately by regions with low, medium and high increases in their childcare supply.



(A) Full-time workers

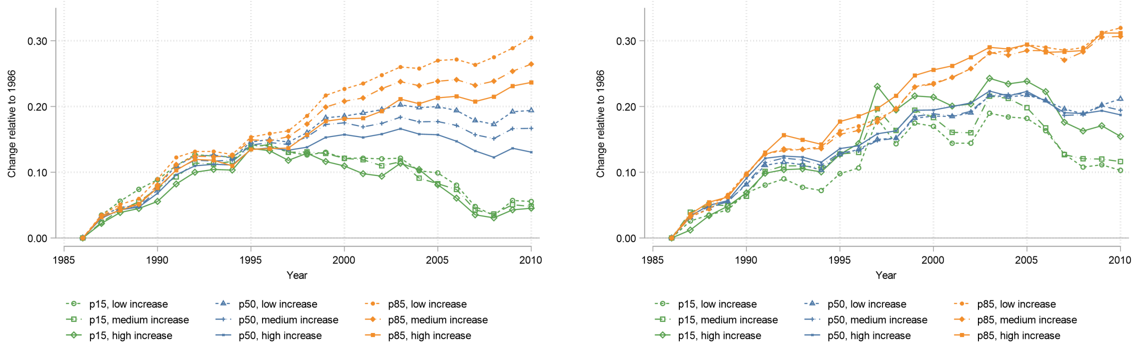
(B) Part-time workers

FIGURE 4.8: Development of the 15th, 50th and 85th log wage percentiles of women, 1986–2010.

Notes: The figure plots changes in percentiles of log daily wages between 1986 and 2010. All values are normalized to 0 in 1986. Panel A focuses on female full-time workers, Panel B on female part-time workers. The 15th percentile is plotted in green, the 50th in blue and the 85th in orange. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

Development of the 15th, 50th and 85th Log Wage Percentiles Figure 4.8 starts by plotting the development of the 15th, 50th and 85th log wage percentiles. The changes for full-time working women in the left-hand panel are similar to the findings in others studies, qualitatively they also follow the trends for male workers (see Dustmann, Ludsteck, and Schönberg 2009; Biewen, Fitzenberger, and Lazzer 2018). Until 1997, the three wage percentiles grew in an almost parallel fashion by around 14 percent. Since 1998 the percentiles diverge. Wages in the 85th percentile continue to grow strongly (by another 12.7 percent to 28 percent in 2010), while the growth of median wages is only modest (additional 3.4 to overall 17.4 percent) and wages in the 15th percentile even have decreased since by 8.2 percent to a total growth of 4.6 percent.

Wages of part-time working women (plotted in the right-hand panel) show a different trajectory. Between 1986 and 2010, they increased stronger than those of full-time workers by 11.6 (p15), 20.1 (p50) and 31.1 percent (p85). In contrast to full-time wages, the percentiles start to diverge earlier, around the end of the 1980s. After some flattening between 1991 and 1994, wages continue to grow until 2003. Between 1993 and 1999, especially the 15th wage percentile stands out with high growth rates. After 2003, only the high wages continue to increase. Median part-time wages, however, remain mostly constant after 2003 and wages in the 15th percentile—which in some years before 2003 increased even stronger than those in the 50th percentile—start to decrease. The longer growth of part-time wages, which is for all percentiles larger than for full-time workers, together with the increased number of women working in part-time, points out that the additional part-time labor supply was met by a sufficient demand for work, providing an additional incentive for women to join the labor market. This finding is, further, consistent with the hypothesis that more mothers, who otherwise would have stayed out of the labor market and who are better selected compared to the preexisting part-time workforce, take up part-time work.



(A) Full-time workers (B) Part-time workers

FIGURE 4.9: Development of the 15th, 50th and 85th log wage percentiles of women by change in childcare supply, 1986–2010.

Notes: The figure plots changes in percentiles of log daily wages between 1986 and 2010. All values are normalized to 0 in 1986. Panel A focuses on full-time workers, Panel B on part-time workers. Both panels differentiate between regions with low (plotted as short dashed lines), medium (longer dashed lines) and high (solid lines), increases in childcare supply. The 15th percentile is plotted in green, the 50th in blue and the 85th in orange. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

Development of Wage Percentiles and Increases in Childcare Supply Turning to an analysis that relates the development of wages and childcare, Figure 4.9 plots the development of the 15th, 50th and 85th wage percentile separately for regions with high, medium and low increases in their childcare supply. The development of full-time wages (plotted in the left-hand panel) is qualitatively similar across all regions and follows the trajectory observed in the full sample (see Figure 4.8), i.e. wage growth across percentiles starts to develop in different ways around 1998. There are almost no regional differences for low wages in the 15th percentile. For the 50th and 85th percentile, regional differences are small until 1997 when regional wage growth starts to diverge. The regions exhibit a clear pattern that increases in wages and childcare are inversely related. Both median and high wages increase stronger where childcare supply increases were the smallest, while those regions with the largest changes in childcare supply show the lowest wage gains. In 2010, the wage growth of the 85th percentile ranges between 30.5 and 23.6 percent; for median wages between 19.4 and 13.1 percent.

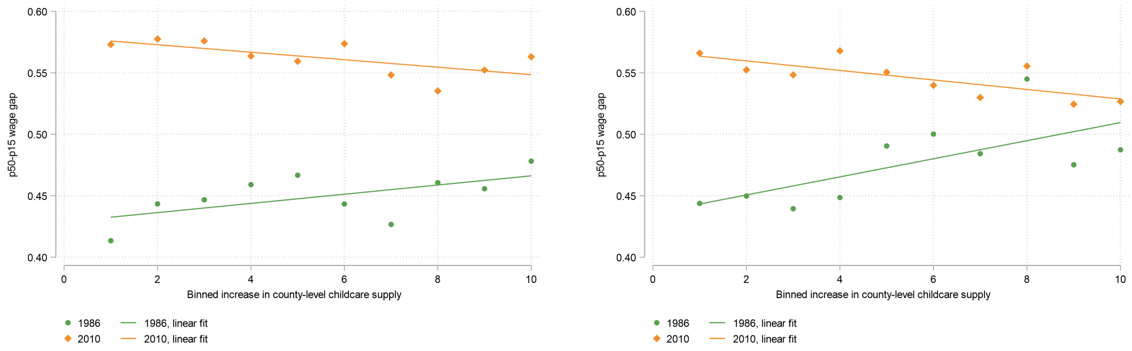
Wages of part-time working women (plotted in the right-hand panel) follow a different profile with respect to their association with childcare. In the upper and middle part of the wage distribution, the 85th and the 50th percentile show slightly larger increases between the early 1990s and 2005 in those regions with greater changes in childcare. After 2005 all regions develop almost similarly. Wage growth until 2010 is similar across regions, for the median between around 19 and 21 percent, for the 85th percentile between 31 and 32 percent. What stands out more, is the lower part of the wage distribution. The trajectory of the 15th wage percentile is, in general, less smooth compared to the median and the 85th percentile which is suggestive of a greater degree of heterogeneity in the underlying part of the female workforce. Despite the more uneven trajectory, a pattern that in regions with larger childcare increases wage gains are larger between 1990 and 2005 is clearly observable for the 15th percentile. Between 1996 and 2006, the increase in low wages in regions with higher childcare increases even exceeds that of median wages. Until 2010, low wages in high-increase regions grew by 15.4 percent; in low-increase regions by 10.3 percent. The finding of relatively large growth for low-wage part-time workers from the full sample is amplified in regions with stronger increases in childcare supply.[10]

4.6.2 Increasing Childcare Supply and Wage Inequality

Having shown that wages of both full- and part-time working women develop differentially with respect to the regional change in childcare supply, this section turns to inequality in wages. Here, I depart from sorting counties into three regions. Instead, I show graphical results over ten bins of counties and regression results directly at the county-level. Figure 4.10 shows binned scatter plots for the relationship between the regional increase in childcare supply and the size of the p50–p15 gap in log wages. On the x-axis there are ten bins of counties by their increase in childcare supply as illustrated in Figure 4.1. The outcome on the y-axis is the size of the p50–p15 log wage gap. The relationship between both variables is plotted for the years 1986 (in green) and 2010 (in orange). Since the increase in childcare supply is a time-constant measure, each bin contains the same counties in each year. Those in the first bin increase their childcare supply between 1986 and 2002 by on average 6.5 slots per 100 children, while for counties in the tenth bin the increase is on average 54 additional slots.

The left-hand panel of Figure 4.10 plots results for full-time working women. Over time, the relationship between the increase in childcare supply and lower-tail wage inequality clearly changed. In 1986, those counties where the (at this time future) increases in childcare are larger, show slightly larger wage gaps. This relationship reverses over time. In all bins, the wage gaps increase and those bins with the smallest increases in childcare exhibit the largest increases of the p50–p15 wage gap such that the relationship between care supply and inequality is negative in 2010. For part-time workers (plotted in the right-hand panel of Figure 4.10) the relationship between childcare supply increases and inequality is qualitatively similar, though

[10] Both for women in full- and part-time work, the ranking of regions with respect to their change in childcare supply is mostly constant over time during the 2000s. This allows to conclude that the choice of 2010 as last year to analyze is unlikely to impact the results.



(A) Full-time workers

(B) Part-time workers

FIGURE 4.10: p50–p15 wage gap of women in 1986 and 2010 by binned regional increase in childcare supply.

Notes: The figure plots the p50–p15 gap in log daily wages of women by binned regional increase in childcare supply. The relationship in 1986 is plotted in green, results for the year 2010 are plotted in orange. Each bin contains 32–33 counties where those in the first bin increase their childcare supply by on average 6.5 additional slots per 100 children and those in the tenth bin by on average 54 additional slots. Figure 4.18 in the Appendix plots the analog for the p85–p15 wage gap. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

markedly more pronounced. This holds both for the positive relationship in 1986 and for the negative one in 2010. While inequality on the lower end of the wage distributions reacts to changes in childcare, no such relationship can be found for the upper end, i.e. the p85–p50 wage gap. Results for overall inequality, measured by the p85–p15 wage gap, show a similar but less strong pattern as the p50–p15 gap (plotted in Figure 4.18 in the Appendix). Figure 4.19 in the Appendix checks the relationship between childcare supply increases and wage inequality for male workers in full-time. It shows that for all three percentile gaps, wage inequality is smaller in counties that experience larger increases in childcare supply. Apart from an overall increase in levels, there is no change over time, i.e. while childcare supply is expanded. This indicates there is no relationship between both measures for men.

As a next step, I quantify the graphical findings. Table 4.3 reports the county-level relationships between different log wage gaps and childcare increases in terms of standard deviations. The results are obtained from regressing each respective log wage gap on the absolute increase in childcare supply in the years 1986 (columns 2 and 5) and 2010 (columns 3 and 6). The regression results are converted to effects for an increase in childcare by one standard deviation, i.e. by 13.36 additional slots per 100 children (the mean increase is 28.3 slots per 100 children). Columns 4 and 7 report the difference between the years. As before, increase in childcare supply is a time-constant measure. The aim of this task is thus to assess whether and how the relationship between inequality within counties and childcare supply changed over time.

For full-time working women in 1986, a larger increase in childcare slots is associated with a significantly larger p85–p15 log wage gap. Precisely, for an increase in childcare by one standard

TABLE 4.3: County-level relationship between increase in childcare supply and wage inequality (in standard deviations), 1986 and 2010.

	Full-time workers			Part-time workers		
	1986	2010	$\Delta_{2010\text{-}1986}$	1986	2010	$\Delta_{2010\text{-}1986}$
p85–p15	0.202***	−0.032	−0.234	−0.070	−0.126***	−0.056
p85–p50	0.057	0.060	0.003	−0.089**	−0.046	0.043
p50–p15	0.210***	−0.085*	−0.295	0.131**	−0.119***	−0.250

The table reports the relationship between the increase in childcare supply and percentile wage gaps for female workers in 1986 and 2010. The values for each year indicate by which fraction of a standard deviation the percentile gap would change if the increase in childcare supply was stronger by one standard deviation (additional 13.36 slots per 100 children). The columns marked with Δ indicate the change from 1986 to 2010. All results are obtained from linear regressions on the county level, weighted with each county's observation share. Analog results for male workers in full-time are given in Table 4.9 in the Appendix. */**/*** indicate significance at the 10/5/1 percent levels. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

deviation, the p85–p15 wage gap increases by around 20 percent of a standard deviation. In 2010, a similar increase in childcare is associated with a decrease in the p85–p15 wage gap by around 3 percent of a standard deviation. For the p50–p15 wage gap, the effect is more pronounced; in 1986 I observe an increase by 21 percent of a standard deviation, but for 2010 a decrease by 8.5 percent of a standard deviation. Both values are statistically significant. There is virtually no change in inequality in the upper end of the wage distribution (measured by the p85–p50 wage gap).

For part-time workers, the finding that larger increases in inequality are associated with a decrease in wage inequality is confirmed. As for full-time workers, this is driven by inequality in the lower end of the wage distribution. Consistent with Figure 4.9 which documents that the differences in wage percentiles between regions are smaller for part-time workers than for full-time workers, effects of childcare documented in Table 4.3 are smaller as well. Larger increases in childcare by one standard deviation are related to a slightly smaller p85–p15 wage gap in 1986 (by 7 percent of a standard deviation); until 2010 this has numerically almost doubled to a significantly negative effect of 13 percent. Results for the p85–p50 wage gap are again comparatively small. For lower wages, a positive association of a 13 percent larger p50–p15 wage gap per standard deviation increase in childcare in 1986 decreases to −12 percent of a standard deviation in 2010.

Table 4.9 in the Appendix reports analog results for men working in full-time that, consistent with the graphical evidence in Figure 4.19, show that there is no relationship between the increase in childcare supply and changes in inequality over time for male workers.

In summary, this section shows that larger increases in childcare supply at the county-level are related to smaller increases in wage inequality among women between 1986 and 2010. This result is primarily driven by the lower part of the wage distribution. It is further consistent across different levels of aggregation; either bins of around 30 counties or in regressions where the single county is the smallest entity. This suggests that other results that are aggregated over three regions, each consisting of around 100 counties, do not suffer from biases by the broader level of aggregation. As the results are based on regional differences in changes (in inequality), they are, further, not driven by time-constant differences between regions, for instance by wage- or education-levels (see also Table 4.2).

4.6.3 Impact of Workforce Composition

So far, I have shown that regions with larger increases in their childcare supply exhibit changes in the composition of their female workforce that are for some characteristics (such as part-time work, medium-level education or experience) more pronounced than in regions with lower childcare supply increases. At the same time, these regions show smaller increases in wage inequality among female workers that are mostly driven by the lower part of the wage distribution. Having established these two aspects, I now turn to analyzing how they are related. This section assesses the contribution of workforce composition to changes in wage inequality. To this end, I decompose inequality growth between 1986 and 2010 in a composition and a price component. I provide multiple decompositions for full-time and part-time workers as well as on the regional level differentiating between the entirety of West Germany and three regions with low, medium and high increases in their supply with childcare slots (as defined in Section 4.3).

To quantify the effects of observable worker characteristics, I rely on the reweighting approach introduced by DiNardo, Fortin, and Lemieux (1996) (DFL). This approach allows to construct counterfactual wage distributions that hold the characteristics of the workforce constant at the levels of a given baseline. It follows the idea of the Oaxaca-Blinder decomposition (Oaxaca 1973; Blinder 1973) as it decomposes differences between two groups (in this paper's case between two years) into one part, that is due to differences in the characteristics of the groups and into a second—residual—part, but extends it from differences in means to the entire distribution such that other statistics as variance or quantiles can be calculated. The latter property is particularly important to assess changes in specific parts of the wage distribution.[11] Specifically, I use the composition of the workforce in the baseline year $t' = 1986$ to estimate weights ψ_z to reweight the wage distribution in year $t = 2010$ to obtain the distribution of wages that would have prevailed if the workforce composition in terms of observable characteristics z had remained at its baseline level in year t'. Comparing how actually observed and reweighted distributional statistics change between years t' and t allows to decompose the total change into two effects. First, the *composition effect* is given as the change in a statistic

[11] Section 4.B in the Appendix gives a formal description of the decomposition.

that can be explained by changes in observable characteristics. For instance, when a workforce becomes better educated and thus receives higher wages. Second, the unexplained part of the total effect is commonly referred to as wage structure or *price effect*. This is due to changes in how the labor market values both observed and unobserved characteristics of workers. The underlying assumption to identify the contribution of workforce composition changes on wages is that the relationship between characteristics z and wages does not change because of the change in z, i.e. that there are no general equilibrium effects.[12]

In the following, I use several decompositions. First, I differentiate between full-time and part-time workers. Second, I differentiate within full- and part-time workers, between a full sample and region-specific subsamples. The full sample includes all working women in West Germany. The region-specific samples split them with respect to regions with low, medium and high increases in childcare supply (as described in Section 4.3). The DFL weights are always calculated and applied specific for each subsample. Splitting the data into subsamples has the advantage of being able to account for two factors. First, it allows for different initial conditions in each subsample. For instance, for lower wages or different education levels of the workforce (as shown in Table 4.2). Second, it is able to capture differential trends across subsamples, both in the development of wages as well as workforce composition (as shown in Section 4.5). It has to be noted, that this approach shows indirect effects of changes in the childcare supply on inequality, i.e. that childcare, first, affects workforce composition in terms of changing participation decisions and sorting into, for instance, occupations which, in turn, has an impact on inequality. An approach that follows the suggestion by DiNardo, Fortin, and Lemieux (1996) to assess inequality consequences of unionization (or other binary-coded changes) directly is not applicable. It would, instead of an indirect one, require that there is a direct effect of childcare supply on wages, similar to union membership usually implying to be covered by collective bargaining which typically leads to higher wages.

In addition to differentiating between subsamples of the data, I use two sets of explanatory variables to calculate the DFL weights. The first set includes three education and five age categories as well as all interactions between them. The second set of explanatory variables consists of those in the first one and adds interactions between age group and experience in the current job (where the interaction accounts for the potential mechanical relation between age and experience) as well as indicators for occupation (at the 3-digit-level) and industry (1-digit-level). Compared to the first one, the second set of variables includes characteristics that are more likely subject to recent choices made by workers. Given participation in the labor market, a worker's choices rather affect their job, occupation or industry while changes in terms of education are much less common. This holds especially since individuals in vocational

[12] This assumption is common for most decomposition methods (Fortin, Lemieux, and Firpo 2011).

training and marginal employment (that could cover a number of students who work parallel to their studies) are excluded. A comparison of these two sets of explanatory variables thus allows to make a distinction between mechanical compositional changes that happen when workers change participation status and changes that happen when workers, in addition to their participation decision, sort to different employers or into different jobs. It has to be noted though, that both sets of explanatory variables identify the joint effects of all included variables. The numerical difference between results from the second and first set cannot necessarily be attributed exactly to the added variables because for neither of them a bias due to omitted variables can be fully ruled out.

Table 4.4: Observed and reweighted changes in wage inequality among full-time working women, 1986–2010.

		I: education, age		II: education, age, experience, occupation, industry	
	Observed change	Composition	Price	Composition	Price
Full sample					
p85–p15	0.233	0.108 (46%)	0.125	0.124 (53%)	0.109
p85–p50	0.106	0.046 (43%)	0.061	0.059 (56%)	0.047
p50–p15	0.127	0.063 (50%)	0.064	0.065 (51%)	0.062
By regional increase in childcare supply					
p85–p15					
Low	0.250	0.094 (38%)	0.156	0.108 (43%)	0.142
Medium	0.216	0.112 (52%)	0.104	0.129 (60%)	0.087
High	0.195	0.104 (54%)	0.090	0.115 (59%)	0.078
p85–p50					
Low	0.110	0.041 (38%)	0.068	0.052 (47%)	0.058
Medium	0.096	0.049 (51%)	0.047	0.070 (72%)	0.027
High	0.106	0.037 (35%)	0.069	0.051 (48%)	0.054
p50–p15					
Low	0.140	0.053 (38%)	0.087	0.056 (40%)	0.084
Medium	0.119	0.062 (52%)	0.057	0.059 (50%)	0.060
High	0.089	0.068 (76%)	0.021	0.064 (74%)	0.023

Notes: Observed and reweighted changes in inequality measures for log imputed daily wages between 1986 and 2010. The observed change is decomposed into a composition effect (columns 3 and 5) and a price (wage structure) effect (columns 4 and 6). Percentage values indicate the contribution of the composition effect to the observed change. Price effects correspond to the change from observed values in 1986 to reweighted values in 2010. Estimates in Panel I use three education and five age categories as well as all possible interactions to estimate Dfl weights, in Panel II experience in the current job along with its interaction with the age categories, occupation (3-digit) and industry (1-digit) identifiers are added. The lower part of the table reports results separately for regions with low, medium and high increases in their childcare supply between 1986 and 2002. *Source:* Own estimations using the Siab data described in Section 4.2.1.

Full-time Working Women Table 4.4 lists actual and reweighted changes in the log wage gaps between the 85th and 15th, the 85th and 50th as well as the 50th and 15th percentile for women in full-time work between 1986 and 2010.[13] The upper part of the table shows results for the full sample while the lower part differentiates between regions with respect to their childcare supply increases. In the full sample, the p50–p15 wage gap, representing the lower part of the wage distribution, increases stronger (by 12.7 log points) than the p85–p50 wage gap (10.6 log points). This adds up to an overall increase in the p85–p15 wage gap by 23.3 log points. Decomposing with respect to age and education (see Panel I of the table) explains between 43 (p85–p50) and 50 percent (p50–p15) of inequality increases. When adding experience, occupation and industry to the decomposition (see Panel II), these shares increase to 56, respectively 51 percent. Thus, for the upper end of the wage distribution the additional explanatory variables explain a substantial additional part of wage inequality increases. In the lower part of the wage distribution, the mechanical factors age and education, i.e. those that pick up changes in participation, are more important.

In the lower part of Table 4.4, I collect region-specific results where the weights are calculated separately for regions with low, medium and high increases in their childcare supply. Overall, the observed changes in percentile wage gaps are within a similar ranger compared to the full sample analysis. While the regional differences for the p85–p50 wage gap are relatively small, they are more pronounced for the p50–p15 gap. Here, larger increases in childcare supply are associated with smaller increases in wage inequality (14 log points in regions with low increases compared to 8.9 log points in regions with a high increase). This finding is consistent with the graphical results in Figure 4.9 that show smaller increases in the 85th and 50th percentiles of wages in regions with larger increases in childcare while the 15th percentile shows little regional differences. The reweighting analysis for the lower part of the wage distribution shows that workforce composition accounts for a fraction of the inequality change that increases in childcare supply changes. Depending on the choice of the explanatory variables around 40, 50 and more than 70 percent of the increase in the p50–p15 wage gap are explained. The overall pattern for low wages qualitatively mostly translates to the p85–p15 wage gap. It is less pronounced, as there is no relationship between childcare increases and inequality for higher wages. Both in the full sample as well as in the region-specific decomposition adding experience, occupation and industry adds markedly greater additional explanatory power to the composition effect in the upper end of the wage distribution, whereas the additional variables yield only minor changes for the lower end. This confirms the finding from the full sample that for the group of low earning women the participation decision and the associated mechanical factors age and education are the primary drivers of increasing inequality.

[13] The corresponding levels for full- and part-time workers are reported in Table 4.10 in the Appendix. It is in line with the findings from Section 4.6.2, especially that the p50–p15 wage gap in 1986 increases in childcare supply changes while it decreases in 2010.

Table 4.11 in the Appendix tests the robustness of the reweighting procedure by repeating the decomposition in Panel II of Table 4.4, but with finer, 4-digit occupation codes and adds indicator variables for four groups of required skills in the occupation. The results for full-time working women are recorded in Panel I. As the more detailed variables lead to some loss of observations, the results differ, but the patterns described above remain generally unchanged.

Table 4.12 in the Appendix repeats the decomposition for full-time working men. The overall increase in wage inequality between 1986 and 2010 as measured by the p85–p15 wage gap is larger for male workers compared to female workers (27.7 vs. 23.3 log points). Further similarities are that inequality in the lower end of the wage distribution shows the stronger increase (15.1 vs. 12.5 log points) and that the p50–p15 wage gap tends to increase less in regions with larger additional childcare supply, though the relationship is markedly weaker than for women. Importantly, the share of the difference between 1986 and 2010 in all wage gaps that can be explained by workforce composition does not show an association with regional childcare supply increases. Thus, there is no sign of a common trend affecting both genders that drives the decomposition results.

Part-time Working Women Table 4.5 reports observed and reweighted inequality changes for women working in part-time. The results for the full sample in the upper part of the table indicate an increase in overall wage inequality by 19.8 log points. Compared to full-time workers (23.3 log points) this is a smaller increase but broadly within a similar range. The larger part of the increase in the p85–p15 wage gap that is due to workers in the upper end of the wage distribution is a clear difference to full-time workers. The p85–p50 gap increases by 10.9 log points whereas the p50–p15 wage gap increases by only 8.8 log points. This is in line with the graphical evidence in Figure 4.8 which shows relatively large increases in the 15th wage percentile for a major part of the observation period. Among the results on reweighted wages two noteworthy aspects stand out. First, for all wage gaps, the share of the inequality increase that workforce composition can explain is substantially smaller than in the full-time sample (see Table 4.4). Depending on the measure and the explanatory variables, between 5 and 18 percent of the increase in inequality can be explained by compositional changes, i.e. around three to ten times less compared to full-time workers. Given the sizeable compositional changes that can be observed for part-time workers, this results is counterintuitive at first. One potential explanation is that an increase of the variation in working hours is responsible for growing inequality such that the missing information on hours would constitute an omitted variable bias. However, as shown in Section 4.2.1, working hours of part-time working women show larger cross-sectional variation than those of full-time workers, though there are no notable increases over time that could explain this raise in inequality. An other explanation is the large influx of women into part-time work itself. As the share of women working in

TABLE 4.5: Observed and reweighted changes in wage inequality among part-time working women, 1986–2010.

		I: education, age		II: education, age, experience, occupation, industry	
	Observed change	Composition	Price	Composition	Price
Full sample					
p85–p15	0.198	0.016 (8%)	0.181	0.024 (12%)	0.174
p85–p50	0.109	0.012 (11%)	0.097	0.007 (6%)	0.102
p50–p15	0.088	0.004 (5%)	0.084	0.016 (18%)	0.072
By regional increase in childcare supply					
p85–p15					
Low	0.218	0.016 (7%)	0.202	0.033 (15%)	0.184
Medium	0.196	0.017 (9%)	0.179	0.023 (12%)	0.174
High	0.154	0.013 (8%)	0.141	0.016 (10%)	0.137
p85–p50					
Low	0.105	0.014 (14%)	0.091	0.010 (9%)	0.095
Medium	0.111	0.013 (12%)	0.098	0.013 (12%)	0.097
High	0.122	0.010 (8%)	0.112	0.009 (7%)	0.112
p50–p15					
Low	0.113	0.001 (1%)	0.112	0.024 (21%)	0.089
Medium	0.086	0.004 (5%)	0.082	0.010 (12%)	0.076
High	0.032	0.004 (11%)	0.029	0.007 (22%)	0.025

Notes: Observed and reweighted changes in inequality measures for log imputed daily wages between 1986 and 2010. The observed change is decomposed into a composition effect (columns 3 and 5) and a price (wage structure) effect (columns 4 and 6). Percentage values indicate the contribution of the composition effect to the observed change. Price effects correspond to the change from observed values in 1986 to reweighted values in 2010. Estimates in Panel I use three education and five age categories as well as all possible interactions to estimate DFL weights, in Panel II experience in the current job along with its interaction with the age categories, occupation (3-digit) and industry (1-digit) identifiers are added. The lower part of the table reports results separately for regions with low, medium and high increases in their childcare supply between 1986 and 2002. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

part-time increased by on average 50 percent, this likely introduced additional changes in the composition of workers beyond those that are observable and can be used in the decomposition. In addition, this finding also underlines that working in full-time is a particular choice, that only a selected part of the female workforce makes—for instance, women before they have their first child or who do not have children as well as those who are distinctly career-oriented. For such a group, unobservable factors are less likely to play a role. Part-time working women, on the other hand, are likely to have more diverse backgrounds as not working full-time after childbirth was a common choice during the observation period that was made by mothers from a variety of backgrounds. Since during this time mothers decreased their time on parental leave and with childcare additionally promoting maternal labor supply, increasing heterogeneity in unobservable variables becomes more relevant.

Second, the impact of adding more explanatory variables (results are listed in Panel II) is clearly different between the lower and upper part of the wage distribution. For the p85–p15 wage gap, adding experience, occupation and industry decreases the fraction of its change explained by composition from 11 to 6 percent. For the p50-p15 wage gap, on the other hand, it more than triples the contribution of composition from 5 to 18 percent. The relative size of these changes is large compared to full-time workers. It provides further indication on the greater heterogeneity of part-time workers such that a reweighting procedure that only relies on age and education is more likely to miss some of the changes in the workforce. The direction of the bias differs. For higher earners, age and education alone explain larger increases than with additional variables, suggesting counteracting effects of added explanatory variables. In the lower part of the wage distribution, I observe the reverse, such that experience, occupation and industry rather contribute to a raise in inequality.

The lower part of Table 4.5 lists separate results by regions with low, medium and high increases in childcare supply. The p85–p15 wage gap of part-time working women is smaller in regions where childcare supply increased stronger. For low-increase regions, I measure 21.8 log points, for medium-increase regions 19.6 and for high-increase regions 15.5 log points. These overall increases are the result of differential patterns in the upper and lower part of the wage distribution. The p85–p50 wage gap increases slightly in childcare supply (from 10.5 to 12.2 log points), the p50–p15 wage gap decreases substantially (from 11.3 to just 3.2 log points). The findings for the impact of workforce composition indicate, again, differential effects of age and education. These more mechanical factors (see Panel I) explain fractions of the p85–p50 gap that decrease in childcare supply changes (from 14 to 8 percent). For the p50–p15 gap, the composition with respect to age and education explains increasing shares, just 1 percent in low-increase regions and 11 percent in high-increase regions. Noting that age and education are mostly deterministic once a person entered the labor market, these findings show that participation decisions of higher earning women in part-time have a relatively smaller effect on inequality where childcare supply increased stronger whereas I find the reverse for the lower part of the wage distribution. Adding experience, occupation and industry as additional explanatory variables to the decomposition (see Panel II) leads to no changes or even a decrease in the share of the inequality increase explained by composition for the p85–p50 wage gap. For the p50–p15 wage gap, more variables add substantial amounts of explanatory power. Moving from low to high childcare supply increases, the decrease the inequality share explained by composition by 21, 2.4 and 2 times to levels between 12 and 22 percent. The large increase in regions with small changes in childcare supply stands out and thus has to be treated with some caution. Nevertheless, these results provide evidence that the impact of factors that are more strongly related to workers' choices beyond participation have effects that decrease in their relative relevance in regions with larger changes in childcare supply. This further underlines the importance of the mainly participation-related factors age and education, which increases with larger childcare supply changes.

Similar to full-time workers, repeating the reweighting with additional explanatory variables (see Panel II. of Table 4.11 in the Appendix) does not lead to noteworthy changes.

The decomposition of the growth in wage inequality allows to draw several conclusions. Observable worker characteristics are distinctly more relevant to explain the increase in wage inequality among full-time working women. For the more heterogeneous group of part-time workers, there is a greater degree of residual changes. Further, there is no or only little association between childcare and inequality in the upper part of the wage distribution while for women with lower wages, inequality decreases with additional childcare supply, both for full- and part-time workers. Larger shares of these smaller increases can be explained by mechanical factors that are mostly related to changes in participation. This suggests that for women with lower earnings potential, the incentives to take up work provided by additional childcare options are more relevant. Changes in decision making by women who are already part of the workforce but opt, for instance, into different occupations show no clear relationship with childcare. The larger relevance of participation decisions, in particular in the lower part of the wage distribution, is consistent with the institutional details of the childcare expansion. It did not start at zero levels, but rather provided additional supply that was relatively affordable, especially in comparison to, at this time, scarce alternatives on the private market. Therefore, it likely addressed a demand among those women with lower earnings potential and thus the lowest willingness to pay for childcare.

4.6.4 Inequality Between Women and Men

Even though larger expansions in childcare supply are associated with less strongly increasing levels of wage inequality, it is, ex ante, ambiguous if and how it changed the position of women relative to men. To assess this question and to provide further context to the previous findings, this section studies the development of the gender wage gap. I restrict attention to workers in full-time since part-time working men likely follow different selection patterns compared to their female counterparts and their number is too small to serve as a comparison group.

Figure 4.11 plots the evolution of the raw (plotted in green) and adjusted (plotted in orange) gender gaps in log wages for regions with low, medium and high increases in their childcare supply. It shows that those regions with the largest increase of their childcare supply initially have the largest raw gender differences in pay (around 44 vs. 42 percent in other regions). Regions with medium or low increases differ only slightly, a pattern that persists over the entire observation period, during which pay differences decrease substantially. The most pronounced decrease is in the early 1990s when the wages of men stagnate or even decline (see

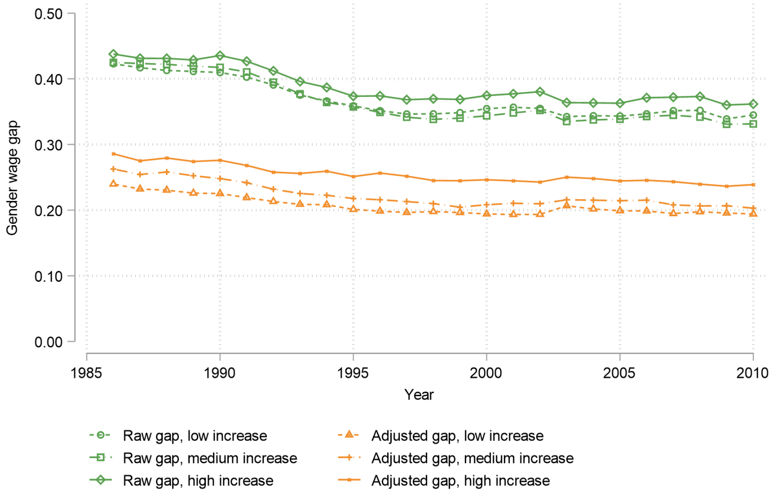


FIGURE 4.11: Raw and adjusted gender wage gaps of full-time workers by development of childcare supply between 1986 and 2010.

Notes: The figure plots the raw (in green) and adjusted (in orange) gender wage gaps, i.e. the average difference between the log wages of men and women. The plot differentiates between regions with low (plotted as short-dashed lines), medium (plotted as long-dashed lines) and high increases (plotted as solid lines) in childcare supply. Adjusted gender gaps control for interactions of age group with education, experience in employment and experience in the current job as well as for occupation (3-digit-level), industry (1-digit-level) and if an individual's wage is censored. To improve readability, the gender gap is shown as a positive number. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

Dustmann, Ludsteck, and Schönberg 2009) while the wages of women, on average, continue to increase (see Figure 4.8). This leads to a marked decrease of the gender gap by around 6 to 7 percentage points. Until 2010, the raw gaps fall to levels between 33 and 36 percent.

The adjusted gender wage gap indicates the remaining, unexplained pay differences after controlling for age, experience (total and in the current job), occupation, industry and a dummy for censored wages. Over time, it decreases in all regions. In 1986, the adjusted gender gap ranges between 24 and 28.6 percent, in 2010 between 19.4 and 23.9 percent. For each year though, its levels increase clearly in the regional change in childcare. This correlation suggests that the finding that the overall position of women in the labor market in regions where childcare increased stronger is worse in terms of observable characteristics (see Table 4.2) extends to residual gender inequality as well.

To further assess the role of workforce composition in comparison to other factors, I use the DFL weights to decompose the change in the gender wage gap. Precisely, I recalculate the raw gender wage in 2010 with the characteristics of the female part of the sample reweighted to match its characteristics in 1986. Observations from men are not reweighted. This approach assess the contribution of the compositional changes in the female workforce to the decrease in the gender wage gap. Table 4.6 reports the results for the full sample and by regional childcare increase.

Table 4.6: Observed and reweighted changes in the gender gap in log wages among full-time workers, 1986–2010.

	Observed			I: age, education		II: age, education, experience, occupation, industry	
	1986	2020	$\Delta_{2010\text{-}1986}$	Composition	Price	Composition	Price
Full sample	0.426	0.341	−0.085	−0.089 (105%)	0.004	−0.073 (86%)	−0.012
By regional increase in childcare supply							
Low increase	0.423	0.345	−0.078	−0.100 (128%)	0.022	−0.085 (109%)	0.007
Medium increase	0.425	0.331	−0.094	−0.083 (88%)	−0.011	−0.068 (72%)	−0.026
High increase	0.438	0.362	−0.076	−0.064 (84%)	−0.012	−0.046 (61%)	−0.030

Notes: Columns 2–4 of the table reports estimates of the raw gender gap in log wages in 1986 and 2010 along with the difference between the two years. Columns 5–8 decompose the change in the gender gap into a composition and a price effect using the Dfl weights described in Section 4.6.3. Percentage values indicate the contribution of the composition effect to the observed change. Estimates in Panel I use three education and five age categories as well as all possible interactions to estimate Dfl weights, in Panel II experience in the current job along with its interaction with the age categories, occupation (3-digit) and industry (1-digit) identifiers are added. The lower part of the table reports results separately for regions with low, medium and high increases in their childcare supply between 1986 and 2002. *Source:* Own estimations using the Siab data described in Section 4.2.1.

In the full sample for all workers in West Germany, the gender wage gap decreases by 8.5 percentage points. The composition effect measured when reweighting with age and education exceeds the actual decrease slightly, indicating that composition alone would have led to an even stronger reduction but that the price effect dampens it. Adding experience, occupation and industry to the weights reduces the impact of composition to 86 percent of the total change. This implies that the female workforce became more similar to the male one in terms of age and education[14] but this is counteracted by differences with respect to experience or selection into occupations which have an increasing effect on the gender wage gap.

The decomposition of the gender gap by regional childcare increases is reported in the lower part of Table 4.6. The absolute decreases of the region-level gender gaps show no clear relationship with childcare. A distinct pattern, however, emerges for the size and contribution of the composition effect. For both sets of explanatory variables, workforce composition explains substantially larger shares of the reduction of the gender wage gap in regions with lower increases in childcare supply (128 vs. 84 percent and 109 vs. 61 percent respectively). This result is a somewhat reverse picture compared to the reweighting analysis of wage inequality within female workers (see Table 4.4) where composition explains larger shares of the changes in the lower end of the wage distribution when childcare supply increased stronger. It underlines

[14] Note, that this does not imply the same levels of pay given the same observable characteristics of men and women. Though, it contributes to a decrease in average pay differences.

that those compositional changes that reduced the increase in wage dispersion among women, at the same time, did not contribute to further reductions of wage inequality between genders. This is consistent with the previous findings that larger increases in childcare supply rather affect low-earning women. It, further, provides an example for the relationship between rather negative selection into employment and the gender gap. As shown, for instance, by Mulligan and Rubinstein (2008) and Olivetti and Petrongolo (2008), those women who work are on average a positive selection from all women such that measures of the gender wage gap include a selection bias. When a policy—such as an expansion of childcare provision—draws more women into employment, the selection bias decreases such that observed wage differentials between women and men increase.

4.7 Conclusion

In this paper, I assess the relationship between a large expansion of public childcare and wage inequality among female workers as well as between women and men. I utilize a major policy change in Germany over the 1990s to provide the following findings on wage inequality for the period from 1986 to 2010.

First, in regions where childcare supply increases stronger, more women select into part-time work and medium levels of education are more common, whereas in other regions there are more female workers from the lower and upper end of the education distribution. Women also exhibit higher levels of experience and work in more stable jobs where childcare increased stronger.

Second, increases in wage inequality among women are lower where childcare supply was expanded more. This is primarily driven by the lower end of the wage distribution.

Third, in regions with larger increases in childcare, larger fractions of the change in wage inequality can be explained by changes in workforce composition while in other regions residual effects have larger impacts. This is, again, driven by the lower part of the wage distribution. The mechanical factors age and education which are mostly affected by changes in participation decisions play a larger role in explaining changes in wage inequality than factors as experience, occupation or industry choices.

Fourth, even though inequality within women decreases with stronger increases in childcare, the opposite is true for inequality between women and men, i.e. the gender wage gap. Both the raw and adjusted gender gap in wages is the largest in those regions with the strongest increases in childcare supply. Consistently, in these regions workforce composition explains the smallest fraction of the overall decrease of the gender gap over time.

My findings are overall similar for women in full- and part-time. For the latter, the relationship between larger increases in childcare and smaller increases in inequality is stronger. Taken together, they suggest that childcare had an impact on wage inequality mostly by changing the participation decisions of women with lower earnings potential.

While I conclude a contribution of additional childcare to less unequal wage among women, I find no indication that it improved the overall composition of the female labor force relative to men beyond existing trends such that the gender wage gap remained at higher levels in regions with larger increases in childcare. Therefore, this childcare expansion provides an example for a family policy that improved female labor supply such that women are better off, both due to current earnings but also in the long-run due to increased pension contributions. At the same time, however, there is no direct contribution to a lower gender wage gap as more participation of women rather decreased the selection bias in measuring female wages.

It has to be noted that some limitations apply to this paper's results. I focus on a period during which more women become part of the workforce, they upgrade in terms of education and occupations and mothers reduce the time they spend on parental leave. The expansion of public childcare slots, thus, was one of multiple changes related to female labor supply. Hence, it rather contributed to general trends such that they were stronger in regions with larger increases in childcare supply but did not lead to shifting trends. Further, the policy change addressed primarily a demand for childcare of those with lower willingness to pay for it, i.e. women with lower potential wages. My results, therefore, do not necessarily generalize to other expansions in care for children of different age level as their mothers likely have different preferences and face different constraints. For this particular childcare expansion, my findings are nevertheless consistent with expectations and with the institutional details, as the reform increased relatively affordable care for children of age three to school entry. It, thus, did not address the needs of highly career-oriented women who tend to re-enter the labor market sooner after childbirth. Instead, it provided incentives to take up or expand work for women with lower potential earnings.

The specific nature of the policy change this paper studies opens opportunities for further research. Since 2013 for instance, children of age one and above have a legal claim to a slot in public childcare in Germany. Presumably, this reform addressed the needs of more high-earning women such that consequences for wage inequality are likely different. Hence, it is worthwhile to extend the analysis of wage inequality to other reforms in childcare as well as to other family policies.

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens (2017). "The Career Costs of Children". In: *Journal of Political Economy* 125 (2), pp. 293–337.

Albanesi, Stefania, Claudia Olivetti, and Barbara Petrongolo (2022). "Families, Labor Markets, and Policy". NBER Working Paper No. 30685.

Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016). "Parenthood and the Gender Gap in Pay". In: *Journal of Labor Economics* 34 (3), pp. 545–579.

Antonczyk, Dirk, Bernd Fitzenberger, and Katrin Sommerfeld (2010). "Rising wage inequality, the decline of collective bargaining, and the gender wage gap". In: *Labor Economics* 17, pp. 835–847.

Autor, David, David Dorn, and Gordon H. Hanson (2013). "The China Syndrome: Local Labor Market Effects of Import Competition in the United States". In: *American Economic Review* 103 (6), pp. 2121–2168.

Baker, Michael, Jonathan Gruber, and Kevin Milligan (2008). "Universal Child Care, Maternal Labor Supply, and Family Well-Being". In: *Journal of Political Economy* 116 (4), pp. 709–745.

Bauernschuster, Stefan and Martin Schlotter (2015). "Public child care and mothers' labor supply – Evidence from two quasi-experiments". In: *Journal of Public Economics* 123, pp. 1–16.

Becker, Sascha O., Lukas Mergele, and Ludger Woessmann (2020). "The Separation and Reunification of Germany: Rethinking a Natural Experiment Interpretation of the Enduring Effects of Communism". In: *Journal of Economic Perspectives* 34 (2), pp. 143–171.

Bergemann, Annette and Regina T. Riphahn (2023). "Maternal employment effects of paid parental leave". In: *Journal of Population Economics* 36, pp. 139–178.

Bertram, Hans, Hiltrud Bayer, and Renate Bauereiss (1993). *Familien-Atlas: Lebenslagen und Regionen in Deutschland*. Opladen: Leske + Budrich.

Biewen, Martin, Bernd Fitzenberger, and Jakob de Lazzer (2018). "The role of employment interruptions and part-time work for the rise in wage inequality". In: *IZA Journal of Labor Economics* 7 (10).

Blinder, Alan S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates". In: *The Journal of Human Resources* 8 (4), pp. 436–455.

Blossfeld, Hans Peter (1985). *Bildungsexpansion und Berufschancen*. Frankfurt: Campus.

Boelmann, Barbara, Anna Raute, and Uta Schönberg (2022). "Wind of Change? Cultural Determinants of Maternal Labor Supply". URL: `https://annaraute.wordpress.com/research/` (visited on 01/27/2023).

Bossler, Mario and Thorsten Schank (2023). "Wage Inequality in Germany after Minimum Wage Introduction". In: *Journal of Labor Economics* 41 (3). Forthcoming.

Brüll, Eduard and Christina Gathmann (2020). "Evolution of the East German Wage Structure". ZEW Discussion Paper No. 20-081.

Bruns, Benjamin (2019). "Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap". In: *American Economic Journal: Applied Economics* 11 (2), pp. 74–113.

Bundesagentur für Arbeit, ed. (2021). *Klassifikation der Berufe 2010 – überarbeitete Fassung 2020*. Nürnberg.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016). "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women". In: *The Quarterly Journal of Economics* 131 (2), pp. 633–686.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace Heterogeneity and the Rise of West German Wage Inequality". In: *The Quarterly Journal of Economics* 128 (3), pp. 967–1015.

Cascio, Elizabeth U. (2009). "Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools". In: *The Journal of Human Resources* 44 (1), pp. 140–170.

Dauth, Wolfgang and Johann Eppelsheimer (2020). "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide". In: *Journal for Labour Market Research* 54.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum (2022). "Matching in Cities". In: *Journal of the European Economic Association* 20 (4), pp. 1478–1521.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum (2014). "The Rise of the East and the Far East: German Labor Markets and Trade Integration". In: *Journal of the European Economic Association* 12 (6), pp. 1643–1675.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach". In: *Econometrica* 64 (5), pp. 1001–1044.

Doepke, Matthias and Fabian Kindermann (2019). "Bargaining over Babies: Theory, Evidence, and Policy Implications". In: *American Economic Review* 109 (9), pp. 3264–3306.

Drechsel-Grau, Moritz, Andreas Peichl, Johannes Schmieder, Kai D. Schmid, Hannes Walz, and Stefanie Wolter (2022). "Inequality and Income Dynamics in Germany". In: *Quantitative Economics* 13, pp. 1593–1635.

Dustmann, Christian, Bernd Fitzenberger, Uta Schönberg, and Alexandra Spitz-Oener (2014). "From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy". In: *Journal of Economic Perspectives* 28 (1), pp. 167–188.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp vom Berge (2022). "Reallocation Effects of the Minimum Wage". In: *The Quarterly Journal of Economics* 137 (1), pp. 267–328.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). "Revisiting the German Wage Structure". In: *The Quarterly Journal of Economics* 124 (2), pp. 843–881.

Findeisen, Sebastian, Jörg Heining, Lukas Riedel, and Sebastian Siegloch (2023). "Unpacking Parental Leave: The Role of Job Protection". Unpublished.

Fitzenberger, Bernd and Arnim Seidlitz (2020). "The 2011 break in the part-time indicator and the evolution of wage inequality in Germany". In: *Journal for Labour Market Research* 54 (1).

Fortin, Nicole M. and Thomas Lemieux (1997). "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" In: *Journal of Economic Perspectives* 11 (2), pp. 75–96.

Fortin, Nicole M., Thomas Lemieux, and Sergio Firpo (2011). "Decomposition Methods in Economics". In: *Handbook of Labor Economics*. Ed. by David Card and Orley Ashenfelter. Vol. 4a. Elsevier.

Frodermann, Corinna, Alexandra Schmucker, Stefan Seth, and Philipp vom Berge (2021). “Sample of Integrated Labour Market Biographies (SIAB) 1975–2019”. In: *IAB FDZ Datenreport* (01|2021).

Gartner, Hermann (2005). “The imputation of wages above the contribution limit with the German IAB employment sample”. In: *IAB FDZ Methodenreport* 2/2005. Ed. by Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit.

Goldschmidt, Deborah and Johannes F. Schmieder (2017). “The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure”. In: *Quarterly Journal of Economics* 132 (3), pp. 1165–1217.

Havnes, Tarjei and Magne Mogstad (2011). “Money for nothing? Universal child care and maternal employment”. In: *Journal of Public Economics* 95, pp. 1455–1465.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). “Children and Gender Inequality: Evidence from Denmark”. In: *American Economic Journal: Applied Economics* 11 (4), pp. 181–209.

Kluve, Jochen and Sebastian Schmitz (2018). “Back to Work: Parental Benefits and Mother’s Labor Market Outcomes in the Medium Run”. In: *Industrial and Labor Relations Review* 71 (1), pp. 143–173.

Müller, Dana, Andreas Filser, and Corinna Frodermann (2022). “Update: Identifying mothers in administrative data”. In: *IAB FDZ Methodenreport* 01/2022.

Mulligan, Casey B. and Yona Rubinstein (2008). “Selection, Investment, and Women’s Relative Wages over Time”. In: *The Quarterly Journal of Economics* 123 (3), pp. 1061–1110.

Oaxaca, Ronald (1973). “Male-Female Wage Differentials in Urban Labor Markets”. In: *International Economic Review* 14 (3), pp. 693–709.

Olivetti, Claudia and Barbara Petrongolo (2008). “Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps”. In: *Journal of Labor Economics* 26 (4), pp. 621–654.

– (2016). “The Evolution of Gender Gaps in Industrialized Countries”. In: *Annual Review of Economics* 8, pp. 405–434.

– (2017). “The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries”. In: *Journal of Economic Perspectives* 31 (1), pp. 205–230.

Raute, Anna (2019). “Can financial incentives reduce the baby gap? Evidence from a reform maternity leave benefits”. In: *Journal of Public Economics* 169, pp. 203–222.

Schönberg, Uta and Johannes Ludsteck (2014). “Expansions in Maternity Leave Coverage and Mother’s Labor Market Outcomes after Childbirth”. In: *Journal of Labor Economics* 32 (3), pp. 469–505.

Statistisches Bundesamt, ed. (2019). *Kinderlosigkeit, Geburten und Familien – Ergebnnisse des Mikrozensus 2018*.

– ed. (2023a). *Tabelle 12211-9010: Abhängig Erwerbstätige: Deutschland, Jahre (bis 2019), Beschäftigungsumfang, Geschlecht*. URL: `https://www-genesis.destatis.de/genesis/online` (visited on 04/03/2023).

– ed. (2023b). *Tabelle 12411-02-03-4: Bevölkerung nach Geschlecht und Altersgruppen*. URL: `https://www.regionalstatistik.de/genesis/online` (visited on 01/26/2023).

– ed. (2023c). *Tabelle 22541-01-01-4: Einrichtungen, genehmigte Plätze, tätige Personen*. URL: `https://www.regionalstatistik.de/genesis/online` (visited on 01/26/2023).

Wagner, Gerd G., Joachim R. Frick, and Jürgen Schupp (2007). "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements". In: *Schmollers Jahrbuch* 127 (1), pp. 139–169.

Wanger, Susanne (2020). *Entwicklung von Erwerbstätigkeit, Arbeitszeit und Arbeitsvolumen nach Geschlecht*. IAB-Forschungsbericht 16|2020. Institut für Arbeitsmarkt und Berufsforschung.

Appendix

4.A Imputation of Wages Above the Social Security Threshold

To impute right-censored wages, this paper builds on the wage imputation by Dauth and Eppelsheimer (2020) who follow the two-step approaches by Dustmann, Ludsteck, and Schönberg (2009) and Card, Heining, and Kline (2013). The imputation is done separately by year, education group (no vocational training, vocational training and university or similar) and gender. The first step fits Tobit regressions of observed log wages on variables for experience (linear and squared terms) and different age profiles for older and younger workers. Predictions based on the regression coefficients $\hat{\beta}$ and observable characteristics $\mathbf{X}$ as the expected value $\mathrm{E}\left[ln(\text{wage})\right] = \mathbf{X}\hat{\beta}$ are likely to exhibit a too strong correlation with $\mathbf{X}$ since they neglect the contribution of unobservable factors. Therefore, a normally distributed random term is added to the expected value. Assuming that wages follow a log-normal distribution, the additional random term is chosen such that for each individual i the following equation holds.

$$ln\left(\text{wage}_i^{\text{imputed}}\right) = \epsilon_i\hat{\sigma} + \mathbf{X}_i\hat{\beta} \tag{4.1}$$

ϵ_i is drawn from the distribution of wages above the censoring limit and $\hat{\sigma}$ refers to the standard deviation of the residuals in the Tobit regression. Gartner (2005) and Dauth and Eppelsheimer (2020) describe the procedure in more detail.

The imputed wages are then used to calculate mean wages at the worker- and establishment-level where always the contribution of the current observation is omitted. In the second step, the Tobit regressions from the first step are repeated, but with these mean wages as additional control variables. They serve as proxies for time-constant effects at the worker- and establishment level, i.e. follow the idea of controlling for worker- and establishment fixed effects. After the second regression, again a random term as described above is added to the prediction from the regression and wages are adjusted to be not larger than ten times the 99th percentile of the predicted wage distribution.

4.B DiNardo, Fortin, and Lemieux Reweighting

DiNardo, Fortin, and Lemieux (1996) start by assuming that each individual observation (of wages w and individual characteristics z at a time t) is an element of a joint distribution at a given point in time, $F(w, z|t)$. The density of wages at t, $f_t(w)$, is then equal to the integral of the density of wages that is conditional on z at time t_w, $f(w|z, t_w)$, over the distribution of characteristics $F(z|t_z)$ at time t_z (where t_w and t_z refer to the points in time when w and z are measured). Formally, this can be written as

$$\begin{aligned} f_t(w) &= \int_z dF(w, z|t_{w,z} = t) \\ &= \int_z f(w|z, t_w)\, dF(z|t_z = t) \\ &\equiv f(w; t_w = t, t_z = t) \end{aligned} \tag{4.2}$$

where the last line refers to the density of wages that is actually observed, i.e. when both w and z are measured at time t. For this paper, the aim of the reweighting procedure is to obtain a counterfactual wage distribution for time t that would have prevailed if individual characteristics z remained unchanged at the levels at time t'. Formally, the counterfactual density of wages at time t given the characteristics z being measured at time t' is given by $f(w; t_w = t, t_z = t')$. By definition, this is unobservable. Using the second line of equation (4.2) to rewrite the counterfactual density gives

$$\begin{aligned} f(w; t_w = t, t_z = t') &= \int_z f(w|z, t_w = t)\, dF(z|t_z = t') \\ &\equiv \int_z f(w|z, t_w = t)\, \psi_z(z)\, dF(z|t_z = t). \end{aligned} \tag{4.3}$$

This expresses the counterfactual wage density as the integral of the density of wage at time t over the distribution of individual characteristics at time t'. Multiplying with $\frac{dF(z|t_z=t)}{dF(z|t_z=t)}$ allows to rewrite the expression as the integral of the wage density at time t over the distribution of z at time t, weighted by ψ_z. The weights are given as

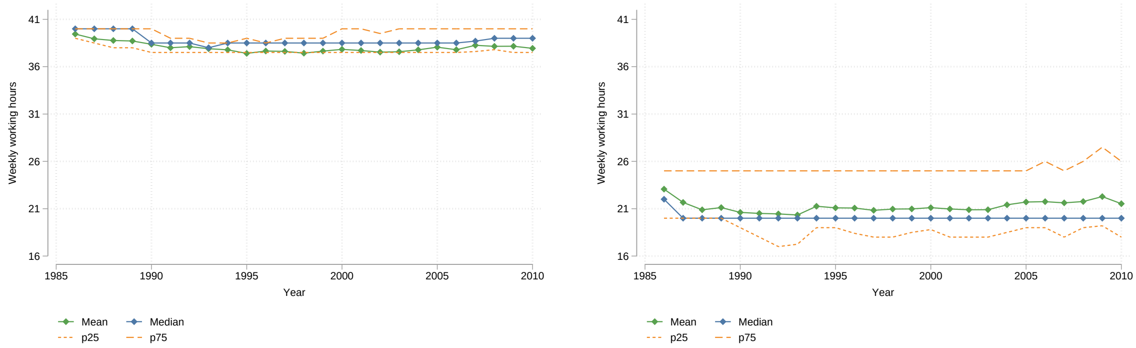
$$\begin{aligned} \psi_z \equiv \frac{dF(z|t_z = t')}{dF(z|t_z = t)} &= \frac{\Pr(z|t_z = t')}{\Pr(z|t_z = t)} \\ &= \frac{\Pr(t_z = t'|z)/\Pr(t_z = t')}{\Pr(t_z = t|z)/\Pr(t_z = t)}. \end{aligned} \tag{4.4}$$

To go from the first to the second line of equation 4.4, Bayes' rule is applied to express $\Pr(z|t_z = t')$ as $\Pr(t_z = t'|z)$. In the last line of equation (4.4), the conditional probabilities can be estimated using a logit or probit model that pools the observations from t and t' and estimates the probability of t' conditional on z. The unconditional probabilities are obtained as the sample shares of observations from time t and t'.

The estimated weights $\hat{\psi}_z$ can be used to calculated weighted statistics for wages in year t. The observed change between years t' and t can be decomposed into a *composition effect* and a *wage structure effect* (commonly also referred to as *price effect*) (Fortin, Lemieux, and Firpo 2011).

For instance, for the percentile difference between the 85th and the 15th percentile PD_{85-15} the composition effect is given as $\mathrm{PD}^{t,c}_{85-15} - \mathrm{PD}^{t}_{85-15}$, i.e. the reweighted counterfactual percentile difference in year t net of its unweighted observed analog. The wage structure effect is given as $\mathrm{PD}^{t'}_{85-15} - \mathrm{PD}^{t,c}_{85-15}$, i.e. as the difference between the observed value in the base year and the counterfactual value in year t. Note, that for the common case of increases in both the observed statistic and the reweighted statistic where the reweighted one is smaller than the observed one both effects are defined as negative numbers. This highlights their counterfactual nature as they report by how much a statistic would have been smaller in year t if workforce composition or wage structure had remained at their levels in year t'. To improve readability and to highlight that both effects typically contribute to increases, this paper displays them as positive numbers.

4.C Additional Figures and Tables



(A) Full-time workers (B) Part-time workers

FIGURE 4.12: Weekly working hours of women between 1986 and 2010 (SOEP data).

Notes: The figure plots usual weekly working hours of women aged 21 to 60 in regular employment in West Germany between 1986 and 2010. *Source:* Own estimations using the SOEP v37 data (Wagner, Frick, and Schupp 2007). Sample restrictions similar to those for the SIAB data described in Section 4.2.1 are applied.

TABLE 4.7: Occupation categories derived from the Blossfeld occupation groups.

Simple, unqualified	Qualified	Highly qualified
Agricultural	Qualified manual	Engineers
Simple manual	Technicians	Professionals (service sector)
Simple service sector	Qualified service sector	Managers
Simple clerks, office jobs	Semi-professionals	
	Qualified office and administrative	

Notes: The table shows the mapping of occupation groups defined by Blossfeld (1985) to the broader occupation classifications by qualification level that is used in Figure 4.5.

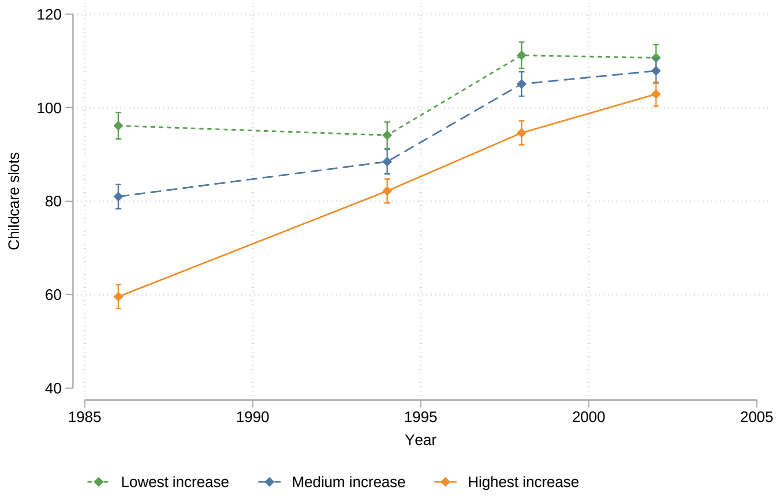


Figure 4.13: Number of childcare slots for children of kindergarten age between 1986 and 2002, grouped by counties with low, medium and high increases.

Notes: The figure documents the increase in childcare supply between 1986 and 2010 along with 95-percent confidence bands. Childcare supply is measured at the county-level as the number of slots in public childcare for children of kindergarten age per 100 children aged three to five. Values above 100 do not indicate excess supply with childcare but rater that kindergarten age can include age levels up to age seven. As population data by age is not available for this group, the number of children of age three to five is used for scaling. The analyses in this paper use the relative supply measure plotted in Figure 4.1. *Source:* Own representation using the county-level data described in Section 4.2.2.

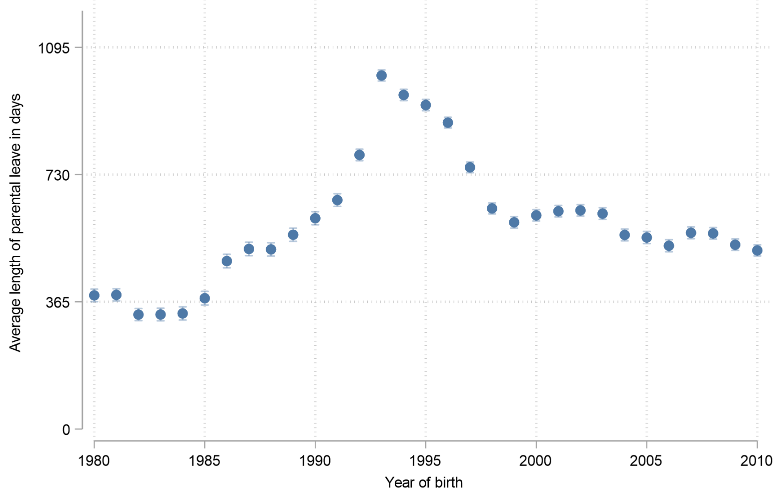
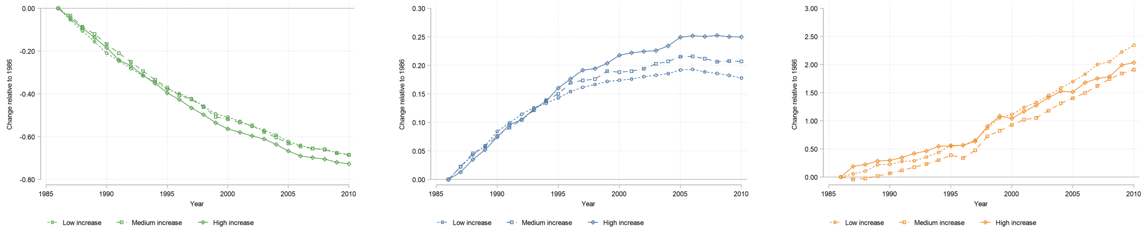


Figure 4.14: Average length of parental leave taking by birthday of the first child (1980–2010).

Notes: Average length of parental leave in days after the birth of the first child. The plot restricts observations to mothers who return to the labor market after at most six years after childbirth. Without this restriction the levels for the years before 1993 are similar to those in 1993, the trend that the length of parental decreases starting around 1994 remains unaffected. Mothers are identified based on the length of their absence from work following Müller, Filser, and Frodermann (2022). *Source:* Own estimation using the Siab data described in Section 4.2.1.



(A) Below vocational education (B) Vocational education (C) University or equivalent education

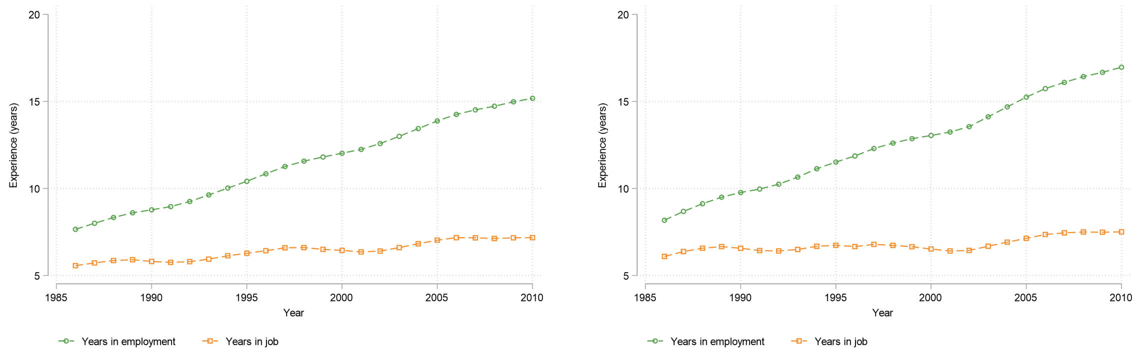
Figure 4.15: Changes in education of female part-time workers relative to 1986 by year and development of childcare supply.

Notes: Changes in the education levels of the female part-time workforce over time by development of childcare supply. Plots indicate changes relative to 1986. Observations are grouped by the position of a region in the distribution of the change in childcare supply between 1986 and 2002. The tercile of regions with the largest increases is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. *Source:* Own estimations based on the SIAB data described in Section 4.2.1.

Table 4.8: Average wages and employment shares of occupation categories.

	Simple, unqualified	Qualified	Highly qualified
Wage			
1986	4.21	4.41	4.79
2002	4.30	4.60	5.01
Employment share			
1986	46.2%	51.3%	2.5%
2010	29.9%	63.8%	6.3%

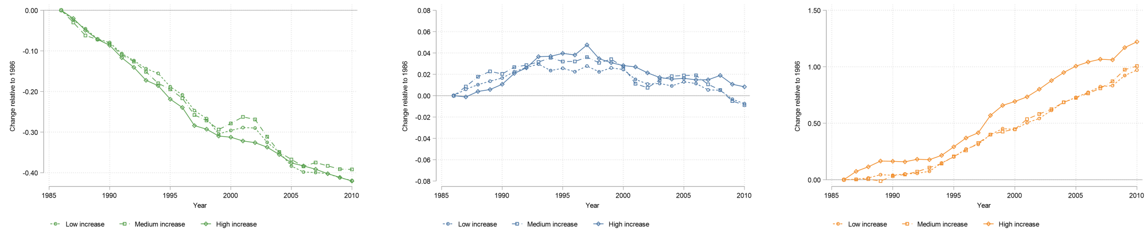
Notes: The table reports the average log daily wages of women per occupation group in 1986 and 2010. Occupation groups are defined as described in Section 4.5 and Table 4.7. *Source:* Own estimations based on the SIAB data described in Section 4.2.1.



(A) Full-time workers (B) Part-time workers

Figure 4.16: Work experience of the female workforce between 1986 and 2010.

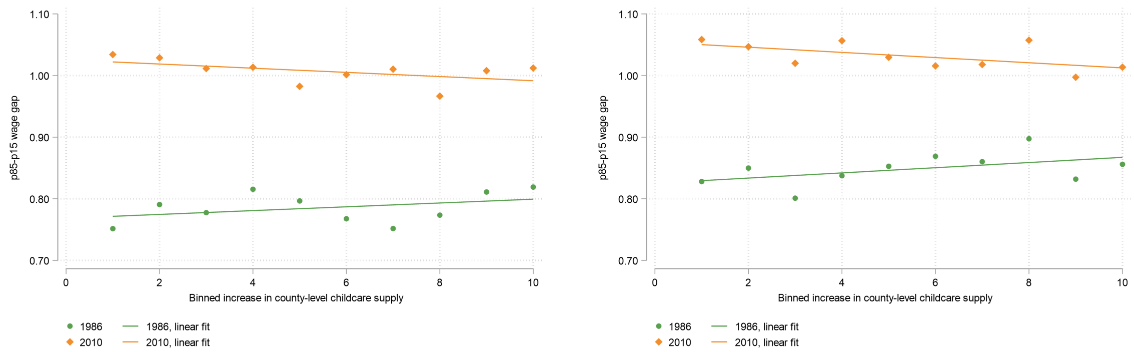
Notes: Development of work experience measures over time. Total work experience is plotted in green, tenure in the current job in orange. *Source:* Own estimations using the SIAB data described in Section 4.2.1.



(A) Un-/semiskilled tasks (B) Skilled tasks (C) Complex and highly complex tasks

FIGURE 4.17: Changes in the skill levels of female part-time workers relative to 1986 by year and development of childcare supply.

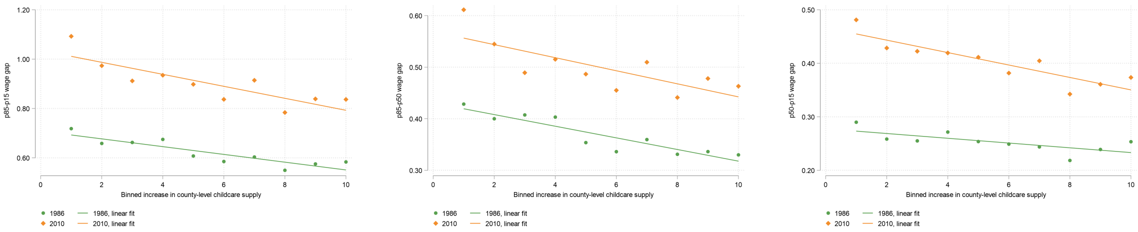
Notes: Changes in the skill level derived from occupations of the female part-time workforce over time by development of childcare supply. Plots indicate changes relative to 1986. Observations are grouped by the position of a region in the distribution of the change in childcare supply between 1986 and 2002. The tercile of regions with the largest increases is plotted as a solid line, the second tercile as a long-dashed line, and the tercile of regions with the smallest increases as a short-dashed line. *Source:* Own estimation based on the SIAB data described in Section 4.2.1.



(A) Full-time workers (B) Part-time workers

FIGURE 4.18: p85–p15 wage gap of women in 1986 and 2010 by binned regional increase in childcare supply.

Notes: The figure plots the p85–p15 gap in log daily wages of women by binned regional increase in childcare supply. The relationship in 1986 is plotted in green, results for the year 2010 are plotted in orange. Each bin contains 32–33 counties where those in the first bin increase their childcare supply by on average 6.5 additional slots per 100 children and those in the tenth bin by on average 54 additional slots. *Source:* Own estimations using the SIAB data described in Section 4.2.1.



(A) p85–p15 wage gap (B) p85–p50 wage gap (C) p50–p15 wage gap

FIGURE 4.19: p85–p15, p85–p50 and p50–p15 wage gaps of full-time working men in 1986 and 2010 by binned regional increase in childcare supply.

Notes: The figure plots the p85–p15 gap in log daily wages of full-time working men by binned regional increase in childcare supply. The relationship in 1986 is plotted in green, results for the year 2010 are plotted in orange. Each bin contains 32–33 counties where those in the first bin increase their childcare supply by on average 6.5 additional slots per 100 children and those in the tenth bin by on average 54 additional slots. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

TABLE 4.9: County-level relationship between increase in childcare supply and wage inequality (in standard deviations), 1986 and 2010, male workers.

	Male full-time workers		
	1986	2010	$\Delta_{2010\text{-}1986}$
p85–p15	−0.539***	−0.532***	0.007
p85–p50	−0.493***	−0.413***	0.080
p50–p15	−0.444***	−0.522***	−0.078

The tables reports the relationship between the increase in childcare supply and percentile wage gaps for male workers in 1986 and 2010. Each value indicates by which share of a standard deviation the percentile gap would change if the increase in childcare supply was stronger by one standard deviation (additional 13.36 slots per 100 children). The row marked with Δ indicates the change from 1986 to 2010. All results are obtained from linear regressions on the county level, weighted with each county's observations share. */**/*** indicate significance at the 10/5/1 percent levels. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

TABLE 4.10: Percentile gaps in log wages of women, 1986 and 2010.

	I: full-time workers		II: part-time workers	
	1986	2010	1986	2010
Full sample				
p85–p15	0.791	1.023	0.834	1.032
p85–p50	0.343	0.449	0.374	0.484
p50–p15	0.448	0.575	0.460	0.548
By regional increase in childcare supply				
p85–p15				
Low	0.775	1.025	0.812	1.030
Medium	0.792	1.008	0.837	1.033
High	0.793	1.008	0.863	1.017
p85–p50				
Low	0.339	0.449	0.376	0.481
Medium	0.345	0.441	0.374	0.485
High	0.339	0.445	0.362	0.484
p50–p15				
Low	0.436	0.576	0.436	0.549
Medium	0.447	0.567	0.463	0.549
High	0.454	0.543	0.501	0.533

Notes: The table reports percentile gaps in log imputed daily wages in 1986 and 2010. Panel I reports values for full-time workers, Panel II for part-time workers. The lower part of the table reports wage gaps separately for regions with low, medium and high increases in their childcare supply between 1986 and 2002. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

Table 4.11: Observed and reweighted changes in wage inequality among women with additional explanatory variables, 1986–2010.

	I. Full-time workers			II. Part-time workers		
	Observed change	Composition	Price	Observed Change	Composition	Price
Full sample						
p85–p15	0.221	0.110 (50%)	0.111	0.196	0.023 (12%)	0.173
p85–p50	0.102	0.055 (54%)	0.047	0.108	0.009 (8%)	0.100
p50–p15	0.118	0.055 (46%)	0.064	0.087	0.014 (16%)	0.073
By regional increase in childcare supply						
p85–p15						
Low	0.241	0.099 (41%)	0.142	0.217	0.031 (14%)	0.186
Medium	0.204	0.116 (57%)	0.088	0.194	0.025 (13%)	0.169
High	0.197	0.110 (56%)	0.086	0.154	0.019 (12%)	0.135
p85–p50						
Low	0.107	0.047 (44%)	0.059	0.104	0.008 (8%)	0.095
Medium	0.093	0.065 (70%)	0.028	0.109	0.015 (14%)	0.094
High	0.108	0.051 (47%)	0.057	0.123	0.011 (9%)	0.112
p50–p15						
Low	0.134	0.052 (39%)	0.083	0.113	0.022 (20%)	0.091
Medium	0.110	0.051 (46%)	0.060	0.085	0.010 (11%)	0.075
High	0.089	0.059 (67%)	0.030	0.031	0.008 (27%)	0.023

Notes: Observed and reweighted changes in inequality measures for log imputed daily wages between 1986 and 2010. The observed change is decomposed into a composition effect (columns 3 and 6) and a price (wage structure) effect (columns 4 and 7). Percentage values indicate the contribution of the composition effect to the observed change. Price effects correspond to the change from observed values in 1986 to reweighted values in 2010. Estimates in Panel I are for full-time workers, Panel II reports them for part-time workers. All specifications use three education and five age categories as well as all possible interactions, experience in the current job along with its interaction with the age categories, occupation (4-digit), four groups of skill requirements in the occupation and industry (1-digit) to estimate Dfl weights. The lower part of the table reports results separately for regions with low, medium and high increases in their childcare supply between 1986 and 2002. *Source:* Own estimations using the Siab data described in Section 4.2.1.

TABLE 4.12: Observed and reweighted changes in wage inequality among full-time working men, 1986–2010.

		I: education, age		II: education, age, experience, occupation, industry	
	Observed change	Composition	Price	Composition	Price
Full sample					
p85–p15	0.277	0.075 (27%)	0.202	0.105 (38%)	0.171
p85–p50	0.125	0.060 (48%)	0.066	0.054 (43%)	0.071
p50–p15	0.151	0.016 (10%)	0.136	0.051 (34%)	0.100
By regional increase in childcare supply					
p85–p15					
Low	0.309	0.101 (33%)	0.208	0.136 (44%)	0.172
Medium	0.274	0.078 (29%)	0.196	0.107 (39%)	0.166
High	0.257	0.075 (29%)	0.182	0.109 (42%)	0.150
p85–p50					
Low	0.139	0.084 (61%)	0.054	0.080 (58%)	0.058
Medium	0.130	0.064 (49%)	0.066	0.060 (46%)	0.070
High	0.130	0.065 (50%)	0.066	0.067 (51%)	0.064
p50–p15					
Low	0.170	0.016 (10%)	0.154	0.056 (33%)	0.114
Medium	0.144	0.014 (10%)	0.130	0.047 (33%)	0.096
High	0.126	0.010 (8%)	0.116	0.042 (33%)	0.086

Notes: Observed and reweighted changes in inequality measures for log imputed daily wages between 1986 and 2010. The observed change is decomposed into a composition effect (columns 3 and 5) and a price (wage structure) effect (columns 4 and 6). Percentage values indicate the contribution of the composition effect to the observed change. Price effects correspond to the change from observed values in 1986 to reweighted values in 2010. Estimates in Panel I use three education and five age categories as well as all possible interactions to estimate DFL weights, in Panel II experience in the current job along with its interaction with the age categories, occupation (3-digit) and industry (1-digit) identifiers are added. The lower part of the table reports results separately for regions with low, medium and high increases in their childcare supply between 1986 and 2002. *Source:* Own estimations using the SIAB data described in Section 4.2.1.

5 Government Expenditure in the Dina Framework: Allocation Methods and Consequences for Post-Tax Income Inequality

Joint with Holger Stichnoth (ZEW Mannheim and University of Strasbourg)

About half of government expenditure in the United States takes the form of in-kind collective expenditure (e.g., education, defense, infrastructure). In studies of post-tax inequality based on the Dina framework, this expenditure is allocated either proportionally to post-tax cash income or as a lump-sum allocation, and the level of inequality is fairly sensitive to this choice. This paper provides direct evidence on how public education spending (a substantial part of collective expenditure) is actually distributed. An allocation proportional to post-tax cash income is clearly rejected, while a lump-sum allocation is found to provide a good approximation.

5.1 Introduction

The United States and many other countries have seen an increase in income inequality in recent decades that has received attention from academic researchers and the general public alike. The measurement of income inequality has traditionally relied on micro data from surveys or administrative tax records. These data, however, capture only about 60 percent of macro totals from national accounts, so a substantial share of national income has been missing from the debate about inequality. In an important contribution, Piketty, Saez, and Zucman (2018) propose a method for constructing distributional national accounts (Dina) that measure how the entire national income is distributed among individuals.

When computing post-tax income, this approach requires the allocation of the entirety of government expenditure to individuals. In recent years, about half of government expenditure in the United States has taken the form of in-kind consumption expenditure (e.g., education, defense, infrastructure); depending on the year, this represents between 16 percent and 20 percent of national income. Piketty, Saez, and Zucman assume that this collective in-kind expenditure is distributed proportionally to post-tax disposable income. This means that, by construction, an important part of national income is assumed to be distributionally neutral. The Dina Guidelines (Alvaredo et al. 2020) explicitly recognize the difficulty surrounding the allocation of government consumption expenditure, calling it "approximate and exploratory." As shown by Blanchet, Chancel, and Gethin (2022), Bozio et al. (2022), and Bruil et al. (2022), the level of post-tax inequality is fairly sensitive to this assumption. We confirm this for the US study by Piketty, Saez, and Zucman. When we replace their proportionality assumption with a lump-sum allocation, the Top 10 percent share of national income decreases by about 5 percentage points, while the share of the Bottom 50 percent increases by roughly the same amount. As a result, the gap between the income shares of the Top 10 percent and the Bottom 50 percent is reduced by half, from about 20 to 10 percentage points in the most recent years.[1]

In light of this sensitivity, the contribution of the present paper is to provide direct evidence on how an important fraction of government consumption expenditure is actually distributed in the United States. We focus on public spending on education, which makes up about 30 percent of collective expenditure and 5 percent of national income in most OECD countries, and is much easier to assign individually than defense or infrastructure expenditure. Our data for the United States are from the 2017 wave of the American Community Survey (ACS). In addition to the large sample size (about 3.2 M individuals in 1.4 M households), the ACS has the advantage that participants are legally obligated to answer the survey questions. The ACS has information on whether household members are currently in education, and, importantly for our purpose, distinguishes between public and private institutions. Finally, the ACS includes individuals in group quarters, which is key for measuring public expenditure that goes to college students who no longer live with their parents. Annual public expenditure per student (net of tuition fees) at different levels of education is taken from the OECD.

We find that, for education at least, public expenditure is *not* proportional to income. On the contrary, average public education spending is highest in the poorest income decile and lowest in the richest decile. The differences are not great, however, so a lump-sum allocation provides a good approximation, at least when income is measured using the equal-split assumption of the Dina framework (i.e., household income is divided by the number of adults aged 20 and above).[2] When equivalized household income is used instead, the negative income gradient is steeper and the approximation is less accurate.

[1] Our calculations are documented in Section 5.A of the Appendix.

[2] Two caveats apply. First, the American Community Survey does not provide the comprehensive income measure that is the *raison d'être* of the Dina approach. However, additional income components such as imputed rents and especially undistributed profits are concentrated among the higher deciles, thus leading to an even greater departure from proportionality. The second caveat is that the ACS provides pre-tax income, while Piketty, Saez, and Zucman (2018) assume that government consumption expenditure is proportional to post-tax income. However, when we simulate post-tax income based on the ACS pre-tax measure and the NBER's TAXSIM model (Feenberg and Coutts 1993), we still clearly reject the proportionality assumption.

These results are strongly driven by age effects. The most striking case are college students who no longer live with their parents. They receive substantial public expenditure while having low current income. But public spending at other levels (pre-primary, primary, secondary) also has an age component, as parents with kindergarten- or school-age children are typically still below the peak of their age-income profiles.

In our second contribution, we examine two justifications for an allocation of collective expenditure proportionally to income that have been proposed in the Dina literature. Piketty, Saez, and Zucman (2017) argue for a proportional allocation by pointing to the positive correlation between public education spending and lifetime earnings. Using the American Community Survey and proxying for lifetime earnings using earnings at age 40–45 (where the rank correlation with lifetime earnings is maximal), we quantify this argument by showing that the 10 percent of individuals with the highest earnings have received average public education spending of \$335 K, about 1.4 times the amount of the bottom 50 percent (\$234 K). The allocation is still not proportional to earnings, however; proportionality would require a factor of about 14. More importantly, adjusting for age effects in public education spending, but not in earnings, capital income, or certain cash transfers, would be inconsistent with the Dina framework, which so far has adopted a strictly cross-sectional perspective.

The Dina Guidelines (Alvaredo et al. 2020) argue that a lump-sum allocation would overestimate the extent of redistribution because of the unequal access to education observed in most countries. While the American Community Survey does not allow us to address this point, we use the Panel Study of Income Dynamics (PSID) to show that more public education spending indeed goes to children of more educated parents. The difference between individuals from the most and the least privileged background with respect to parents' education is \$66 K on the father's side and \$68 K on the mother's side. However, while these inter-generational patterns are arguably more important than the cross-sectional results for the distributional debate, they again do not provide the right empirical basis for an allocation of government expenditure in the cross-section.

Related literature Following the paper by Piketty, Saez, and Zucman (2018) for the United States, the Dina approach has been applied to other countries. Garbinti, Goupille-Lebret, and Piketty (2018) study pre-tax income inequality in France using a Dina approach, and Bozio et al. (2022) extend this to post-tax income and compare France with the United States. Using a simplified approach, Blanchet, Chancel, and Gethin (2022) create distributional national accounts for the member countries of the European Union. Other applications of the Dina framework are for Austria (Jestl and List 2020), China (Piketty, Yang, and Zucman 2019), Germany (Bach, Bartels, and Neef 2021), the Netherlands (Bruil et al. 2022), and Sweden (Hammar et al. 2020). In a related effort, the OECD and Eurostat set up an expert group to disaggregate the household sector in the system of national accounts; see Zwijnenburg (2019) for a comparison with the Dina approach.

Our paper contributes to the discussion about methodological issues in the measurement of income inequality in the Dina framework and beyond. Note that we focus exclusively on the effect of government (in-kind) consumption expenditure and remain silent on the debate about issues in the measurement of pre-tax income, such as the allocation of business profits or untaxed pension income (Auten and Splinter 2022; Saez and Zucman 2020).[3]

There is a literature on the distribution of public (in-kind) expenditure (Gillespie 1965; Musgrave, Case, and Leonard 1974; Smeeding 1977; O'Higgins and P. Ruggles 1981; P. Ruggles and O'Higgins 1981; Smeeding et al. 1993; Wilson, Lambright, and Smeeding 2006; Garfinkel, Rainwater, and Smeeding 2006; Marical et al. 2006; Callan, Smeeding, and Tsakloglou 2008; Horton and Reed 2010; O'Dea and Preston 2012; Verbist, Förster, and Vaalavou 2012; Zwijnenburg, Bournot, and Giovannelli 2017) which pre-dates the Dina approach. Several of these studies also study public education spending. The patterns found in these studies are consistent with our results. In particular, none of the studies find that the allocation of public education spending is proportional to cash income. We contribute to this literature by using a much larger dataset that distinguishes between public and private education as well as different levels of education (pre-primary, primary, secondary, tertiary) and that includes students in group quarters, which is important for the allocation of public spending on tertiary education. We also contribute by linking our findings to the Dina literature. In particular, we break down education spending by individualized income for adults age 20 and above, using the "equal-split" approach of Piketty, Saez, and Zucman (2018). The older studies used equivalized household income instead, which we include as a robustness check. In independent work, Bruil et al. (2022) also study the distribution of education and other in-kind transfers using both the equal-split approach and the approach based on equivalized household incomes. Finally, while existing studies examine public spending in the cross-section, we additionally distinguish by lifetime earnings and by the socio-economic status of the parents.

This earlier literature has raised the important question of whether government in-kind expenditure should be measured at cost or should rather measure the increase in individual welfare that results from the expenditure (see O'Dea and Preston 2012, on this and other methodological issues). With an assignment based on cost, inefficiencies in the provision of public services show up as income, and there is no accounting for different needs of individuals. However, attempts to measure welfare instead of income or to account for different needs by adjusting equivalence scales (Paulus, Sutherland, and Tsakloglou 2010; Aaberge, Bhuller, et al. 2010; Aaberge, Langørgen, and Lindgren 2013; Aaberge, Eika, et al. 2019) depart from the

[3] There is a also a debate about measurement issues regarding wealth inequality, see Saez and Zucman (2016), Smith et al. (2019) and Saez and Zucman (2020).

Dina framework, which – following the practice in national accounts – measures government expenditure on a cost basis. Moreover, we see the issue of valuation as orthogonal to the question of correctly determining who receives the public expenditure in the first place.

The remainder of this paper is organized as follows. Section 5.2 describes the data and methods we use in our empirical study of how public education spending in the United States is actually allocated across the income distribution. Section 5.3 presents our results. We focus on the distribution in the cross-section, which is the perspective that has been adopted in the Dina literature, but also report the distribution by life-time earnings (proxied for by earnings at age 40–45). Finally, in a supplementary analysis based on PSID data, we study how public education expenditure varies by parents' educational attainment. Section 5.4 concludes.

5.2 Methods and Data

5.2.1 Overview

Given that the level of post-tax income inequality is sensitive to the assumption about how government consumption expenditure is allocated, we provide direct evidence on how an important part of this expenditure is actually distributed. We focus on public spending on education, which makes up about 5 percent of national income in the US and in most OECD countries and is much easier to assign individually than defense or infrastructure expenditure.

Our method for allocating public education expenditure is straightforward. We use a micro dataset—the American Community Survey 2017—that allows us to observe the income of the household and that has information on who in the household currently attends a *public* educational institution, distinguishing pre-primary, primary, secondary, and tertiary education. We then multiply the number of students per household with the average public expenditure for students of the respective education level, which we take from the OECD's "Education at a Glance" database.

Following the Dina framework, our main analysis is cross-sectional, i.e. we study the distribution of public education expenditure by current income. In addition, we analyze public education expenditure by lifetime earnings, proxied for by earnings at age 40–45. However, based on another dataset—the Panel Study of Income Dynamics, PSID (Survey Research Center 2022)—we also adopt an intergenerational perspective and document how the expenditure differs by parents' education and occupational prestige.

5.2.2 American Community Survey

Our main source of individual-level microdata is the American Community Survey (ACS). The ACS is conducted by the United States Census Bureau to collect information similar to the decennial census. Our data for the year 2017 is from the public use file of the ACS provided by IPUMS USA (S. Ruggles et al. 2020). It provides information on around 3.2 M individuals in 1.4 M households. In addition to the large sample size, the ACS has the advantage that—unlike in other datasets such as the Current Population Survey—respondents are legally obligated to answer the survey questions.

Enrollment The ACS has information on whether household members are currently enrolled in an educational institution, and, importantly for our purpose, distinguishes between public and private institutions. Moreover, the ACS includes individuals in group quarters including college dormitories, which is key for measuring public expenditure that goes to college students who no longer live with their parents.

The ACS provides a very accurate picture of the number of individuals enrolled in the education system (Figure 5.11 in the Appendix). For public institutions at the pre-primary, primary, and secondary levels in 2017, our own calculations based on the ACS result in 51.4 M students. The OECD (OECD Statistics 2020) and the National Center for Education Statistics (De Brey et al. 2021) report values of 50.6 M and 50.7 M, respectively. At the tertiary level, our ACS number is 16.8 M, which is a little higher than the value of 14.6 M reported by the OECD and the NCES.[4] For completeness, Figure 5.11 also shows the number of students in private education, although we do not include these students when allocating public education expenditure. Private education is empirically relevant only at the pre-primary level (kindergarten) and then again at the tertiary level. Our ACS numbers are again close to the OECD values, while the numbers reported by the NCES are slightly lower.

Income concept Income is measured in the ACS as an aggregate of personal income from different sources—including wages (including bonuses), income from self-employment and businesses, pensions, capital income, and social security payments—of all household members above the age of 15.[5] For individuals in group quarters, such as students in college dormitories, the concept of household income does not apply and only personal income is reported. The

[4] In a robustness check, we scale down the ACS numbers accordingly.

[5] "Personal income, or 'money income,' as per the Census Bureau, is the income received on a regular basis (exclusive of certain money receipts such as capital gains and lump-sum payments) before payments for personal income taxes, Social Security and Medicare taxes, union dues, etc. It includes income received from wages, salary, commissions, bonuses, and tips; self-employment income from own nonfarm or farm businesses, including proprietorships and partnerships; interest, dividends, net rental income, royalty income, or income from estates and trusts; Social Security or Railroad Retirement income; Supplemental Security Income (SSI); any cash public assistance or welfare payments from the state or local welfare office; retirement, survivor, or disability benefits; and any other sources of income received regularly such as Veterans' (VA) payments, unemployment and/or worker's compensation, child support, and alimony." (`https://www.pewresearch.org/social-trends/2018/07/12/methodology-15/`). The income components such as wage or business income are top-coded at the 99.5^{th} percentile of the respective federal state. Higher values are coded as the state-specific average of all values above the threshold.

period of reference for the income measurement are the previous twelve months. Note that, as the ACS is administered throughout the year, this means that the income in most cases does not correspond to a calendar year. Also, despite the legal obligation to answer the survey, some of the individual income components are actually imputed by the data provider. In a robustness check, we drop all households in which more than half of household income is based on an imputation.

Regarding the comparison with the Dina approach, two additional caveats are in order. First, while the American Community Survey provides a fairly comprehensive measure of income, it falls short of the Dina approach, in which pre-tax income sums up to the whole of national income. However, additional income components such as imputed rents and especially undistributed profits are concentrated among the higher deciles, and including them would likely lead to an even greater departure from proportionality than what we find below for our income measure.

The second caveat is that the ACS provides pre-tax income, while Piketty, Saez, and Zucman (2018) assume proportionality of government consumption expenditure to post-tax disposable income. However, when we simulate post-tax income using NBER TAXSIM, we still clearly reject the proportionality assumption.[6]

Unit of measurement In our main specification, we follow Piketty, Saez, and Zucman (2018) and the other Dina studies and measure income at the level of adult individuals aged 20 and above. For couples, we apply an equal-split rule, i.e. each adult gets assigned the same share of household income, while children are disregarded. We apply this rule also in cases in which there are more than two adults in the household (e.g., children over 20 or other relatives).

Another approach in the measurement of inequality takes the household as the unit of measurement and accounts for differences in household size and age composition through equivalence scales. This is the approach that is used in the older literature on the measurement of in-kind transfers (see Section 5.1), and we report results in this tradition as well. When doing so, we apply the commonly used modified OECD equivalence scale, which assigns a value of 1 to the first adult in the household, of 0.5 to each additional household member aged 14 and above, and of 0.3 to each child below the age of 14.

Summary statistics Table 5.1, panel A, shows summary statistics for our main sample of adults age 20 and above. These represent about 2.4 M, or 75 percent of the 3.2 M individuals—adults and children—in the ACS. The average age in our sample is 48.4 years. Age is highest in the

[6] TAXSIM (Feenberg and Coutts 1993) simulates the tax liability for federal, state, and payroll taxes. We use TAXSIM version 32 (`https://users.nber.org/~taxsim/taxsim32/`). The simulations are for tax units, which we identify in our ACS household sample following the procedure outlined by Samwick (2013). We assume that all married couples file jointly.

second and tenth deciles and first falls and then rises in the deciles in between. The first decile is not part of this U shape and stands out for having the lowest average age.

The first and the tenth deciles have the smallest household size on average (2.70 and 2.71). In between, the pattern is an inverted U, with a maximum of around 3.20 in deciles three and four. The differences in household size are mostly driven by the intensive margin. With the exception of the first decile, where the share of adults with children is not only much lower than the average but also noticeably less than in the decile just above, there is little variation in this share across the other deciles, with a slight increase towards the upper range of the income distribution. The age of the youngest child likewise increases with income. The same patterns with respect to age and household composition also hold when grouping individuals based on their post-tax disposable income (panel B).

Turning to the income measures themselves, the mean value of pre-tax income in our ACS sample for the year 2017 is \$43.9 K per adult, and the median is \$32.2 K. Our median is reasonably close to the value of \$36.0 K reported by Piketty, Saez, and Zucman (2018) for 2014, while our mean is much lower than the \$64.6 K that they find using their more comprehensive measure of income (see Figure 5.6 above). For the mean, we can also compare the values by decile. The difference is mostly driven by the richest decile, where the average reported by Piketty, Saez, and Zucman is almost twice as high as the one we compute based on the income concept from the American Community Survey, which does not include imputed rents and undistributed profits (and additionally is right-censored). If these were to be included, we would find an even bigger departure from a proportional allocation than we actually do.

Note that the income values in the first decile are very low, with a median of \$6.3 K and a mean of \$5.6 K per year. In the study by Piketty, Saez, and Zucman, the mean is even lower at \$1.3 K. Like them, we find a number of zero or even negative values in the ACS data. The zeros are often for young adults who report receiving private transfers, which are not part of the standard ACS income measure that we use. When dropping the negative values or all values below the 1st percentile, the results regarding the income-gradient of public education spending are essentially unchanged.

For post-tax disposable income, we find a mean of \$32.7 K and a median of \$26.3 K based on the ACS data and our simulation using TAXSIM. Piketty, Saez, and Zucman have a mean of \$46.5 K. The difference again arises in the upper half of the income distribution, especially in the top decile. For the bottom 50 percent, where imputed rents and undistributed profits play not much of a role, our ACS + TAXSIM measure is fairly close to what Piketty, Saez, and Zucman find. With the exception of the bottom decile, our numbers are a bit lower than theirs even for this group, however, despite our use of a more recent year (2017 vs. 2014).

Table 5.1: Summary Statistics

	Decile										
	1	2	3	4	5	6	7	8	9	10	Total
A. Pretax income											
Ann. Income (Median)	6.3	13.1	18.5	23.8	29.3	35.5	43.2	53.5	70.0	118.0	32.2
Ann. Income (Mean)	5.6	13.1	18.5	23.7	29.3	35.6	43.4	53.7	71.0	147.2	43.9
Ann. Income (Mean, PSZ)	1.3	9.6	16.0	23.0	31.1	41.3	53.6	69.6	96.8	303.9	64.6
Household Size	2.70	3.10	3.19	3.20	3.12	3.08	2.97	2.86	2.79	2.71	2.97
Children in HH (0/1)	0.30	0.35	0.37	0.37	0.36	0.37	0.37	0.38	0.39	0.40	0.37
Age	46.8	50.3	49.3	48.7	48.1	47.7	47.3	47.4	48.2	50.1	48.4
Age Youngest Child	7.4	7.4	7.6	7.9	8.2	8.3	8.3	8.5	8.7	8.8	8.1
B. Posttax cash income											
Ann. Income (Median)	6.3	12.5	16.8	20.6	24.6	28.9	34.0	40.3	50.2	78.6	26.7
Ann. Income (Mean)	5.6	12.5	16.8	20.7	24.6	28.9	34.0	40.4	50.7	95.8	33.0
Ann. Income (Mean, PSZ)	3.5	11.9	17.4	22.4	27.5	33.5	40.8	50.5	66.8	190.3	46.5
Household Size	2.63	3.08	3.21	3.24	3.14	3.07	2.96	2.89	2.79	2.70	2.97
Children in HH (0/1)	0.25	0.33	0.38	0.39	0.38	0.38	0.37	0.39	0.39	0.39	0.37
Age	46.7	49.1	47.7	47.4	47.8	47.9	48.1	48.3	49.5	51.4	48.4
Age Youngest Child	7.8	7.6	7.5	7.6	8.0	8.2	8.3	8.4	8.6	8.9	8.1

Notes: The table shows summary statistics for our estimation sample with respect to income and household composition. With the exception of the final column (which is for the sample as a whole), the columns report values within deciles. The cells report mean values; for annual income (abbreviated as ann. income), the median is shown as well. The income measures are compared with the values reported by Piketty, Saez, and Zucman (2018). The upper panel divides adult (aged 20 and above) into deciles based on their pre-tax income. In the lower panel, the deciles are based on post-tax cash income instead. Pre-tax income is taken directly from the ACS, while post-tax cash income is simulated using TAXSIM. *Source:* Own calculations based on the American Community Survey 2017. For comparison with the Dina approach, in households with more than one adult, household income is divided by the number of adults (equal-split). The income reported in the table is annual income in thousand US Dollars. N=2,375,184 adult individuals (aged 20 and above). For the sake of presentation and given the large sample size, standard errors are omitted. HH: Household. PSZ (Piketty, Saez, and Zucman 2018): Pre-tax income from Appendix Tables II-B4 (deciles 1-9, computed as the average over percentiles 0-9, 10-19 etc.) and II-B3 (overall mean and decile 10). Post-tax cash income from Appendix Tables II-C4b (deciles 1-9, computed as the average over percentiles 0-9, 10-19 etc.) and II-C3e (overall mean and decile 10). Piketty, Saez, and Zucman report averages for post-tax cash income by percentile of post-tax income (including non-cash transfers), while the averages for post-tax cash income in our ACS data are computed for deciles of post-tax cash income only, because the focus of our paper is to probe the assumption that the non-cash componenents (in our case, education) are distributed proportionally to post-tax cash income. However, as Piketty, Saez, and Zucman allocate non-cash transfers proportionally to post-tax cash income, the deciles for post-tax cash income and post-tax income should coincide in their case.

5.2.3 Public Expenditure on Education

Per-student values Annual public expenditure on education in the United States in 2017 is taken from the OECD's "Education at a Glance" database (OECD Statistics 2020), subsection "Educational finance indicators". The information is available for different levels of education, based on the ISCED 2011 classification (Table 5.2). Total public expenditure in 2017 is \$56 B at the pre-primary level (ISCED 0), \$296 B at the primary level (ISCED 1), \$328 B at the secondary level (ISCED 2–3), and \$308 B at the tertiary level.[7] In line with the practice of national accounting and the Dina approach, expenditure is valued at cost, as opposed to the valuation that students or their parents put on this expenditure, which is much more difficult to measure.

The numbers are for "all expenditure types" in the OECD nomenclature. This includes both current expenditure (a large share of which are salaries and wages) and capital outlays, but excludes R&D as well expenditure for ancillary services. R&D expenditure is relevant only at the tertiary level, where it amounts to \$37 B in 2017. As part of our robustness checks, we use both a narrower (only current expenditure) and a broader (all expenditure types plus R&D and ancillary services) definition of public education spending. This has little effect on the results, which are mostly driven by differences in enrollment across the income distribution.

The OECD calculates expenditure per student on the basis of full-time equivalents. In these calculations, students in part-time education—relevant only at the pre-primary and the tertiary level—are assumed to represent one-third of a full-time equivalent. Since we do not observe part-time student status in the ACS, we assign the expenditure per full-time equivalent to all students.

Public per-student expenditure in 2017 is around \$13 K at both the pre-primary and the primary level and slightly higher (\$14.5 K) at the secondary level.[8]

At the tertiary level, expenditure per student is \$29.1 K. This is an average over 2-Year and 4-Year colleges. As part of our robustness checks, we try to distinguish between the two categories. While the annual per-student expenditure can be calculated by going back to the NCES data on enrollment and expenditure, which is more detailed than what the OECD provides, there is no information in the ACS on the type of college. However, the ACS distinguishes between undergraduate studies on the one hand and graduate and professional schools on the other. In a robustness check, we assign all graduate students to 4-Year colleges, and randomly assign undergraduates to either 2-Year or 4-Year colleges, based on the relative importance of the two types as reported by the NCES.

7 We abstract from post-secondary non-tertiary education, where annual public expenditure in 2017 is a mere \$1.2 B.

8 The distribution of funds among the ISCED levels 0, 1, 2, and 3 are estimated by the OECD. The National Center for Education Statistics (De Brey et al. 2021) reports only a single value for these levels. We run a robustness check in which we discard the small differences and use the NCES number.

Table 5.2: Public Expenditure on Education

	Per student ($ K)	Total ($ B)			
	OECD	ACS	OECD	NCES	NIPA
Pre-primary	13.0	38	56	n.a.	n.a.
Primary	13.0	283	296	n.a.	n.a.
Secondary	14.5	384	328	n.a.	n.a.
Sum	.	705	681	681	666
Tertiary	29.1	488	308	335	288
Total	.	1.192	990	1.016	954

Notes: The table reports the per-student values for annual public expenditure on education that we use in our calculations (column 1) and the aggregates that we find when combining these values with our ACS enrollment data (column 2). These aggregates are compared with statistics published by the OECD, the National Center for Education Statistics (NCES), and National Accounts (NIPA). Sources: OECD: OECD Statistics (2020). Post-secondary non-tertiary education (ISCED level 4) is negligible and omitted for simplicity. OECD per-student expenditure is for full-time equivalents. The OECD assumes that part-time students receive one-third of a full-time equivalent. The distinction between full-time and part-time is only relevant at the pre-primary and the tertiary levels. ACS: Own calculations based on enrollment as observed in the American Community Survey 2017, combined with the per-student expenditure numbers of the OECD. NCES: National Center for Education Statistics, Digest of Education Statistics 2019 (De Brey et al. 2021), Table 236.10: Summary of expenditures for public elementary and secondary education and other related programs, by purpose: Selected years, 1919-20 through 2016-17. NIPA: Table 3.16. Government Current Expenditures by Function (Data published on March-26-2021 in connection with the Third estimates for 2020 Q4). NIPA Codes: Total: G16029; Elementary and secondary: G16030; Tertiary = Higher (G16031; $195 B) + Libraries (G16032; $13 B) and Other (G16033; $81 B).
Remarks: (1) The US national sources only report aggregates for elementary and secondary education. The breakdown into pre-primary, primary, and secondary is estimated by the OECD. (2) Piketty, Saez, and Zucman (2018) report a lower value for education spending, which corresponds to current expenditure only (NIPA code G17019). The value reported in their paper is $762 B in 2014 (see Section 5.A above). The NIPA value for 2014 has since been updated to $789 B. In the 2020 update of their analysis, Piketty, Saez, and Zucman use the same measure that we employ in this paper (NIPA Code G16029) and report values of $884 B for 2014 and $956 B for 2017 (see PSZ2020AppendixTablesI(Aggreg).xlsx, Sheet DataIncome, Column PF, available at `http://gabriel-zucman.eu/usdina/`, which is almost identical to the $954 B reported in the table, the small difference being likely due to an update of the NIPA data.

The education expenditure includes all levels of government—this is important as most public education spending in the US occurs at the state and local levels. The OECD only provides the national average of education spending. As part of our robustness checks, we use averages by state provided by the National Center for Education Statistics.[9]

Unfortunately, we do not have data on per-capita public expenditure at the sub-state level that would allow us to capture differences between richer and poorer school districts or neighborhoods. This means that the differences in public spending by income that we document are driven by different enrollment rates, different propensities to choose public vs. private institutions, and, in the robustness check, by differences across states. We do *not* capture any remaining income-related variation in per-capita spending. As this remaining variation is likely positively related to income, this means that we do not capture one component that would work towards the proportionality assumption used as the benchmark in the DINA approach. However, we find such a strong departure from proportionality that the within-state differences in per-capita spending would have to be implausibly high in order to justify the assumption. Moreover, at least at the level of school districts, the difference by income is less pronounced than one might think, and is characterized by a U-shape instead of a monotonous increase with income. Average per-pupil expenditure in public elementary and secondary schools is \$12.9 K in low-poverty districts, \$11.2 K in middle-low poverty districts, \$10.8 K in middle-high poverty districts, and \$13.0 K in high-poverty districts (De Brey et al. 2021, Table 236.85). There is a rural-urban divide: while in cities high-poverty districts have substantially higher public per-pupil spending than low-poverty districts, the difference is smaller in suburban districts and turns in favor of low-poverty districts in towns and rural areas.

Aggregates Table 5.2 also shows aggregate annual expenditure. Our own numbers—obtained from combining the enrollment observation in the ACS with the OECD values for per-student expenditure—are compared with the OECD aggregates, information from the National Center for Education Statistics (NCES), and with the national accounts (NIPA) data published by the Bureau of Economic Analysis, which is the source that Piketty, Saez, and Zucman (2018) use. (They only report the total, without the breakdown by education level.)

At ISCED levels 0-3 (pre-primary, primary, secondary), we obtain an aggregate expenditure of \$705 B, close to the \$681 B reported by both the OECD and the NCES, and only about

[9] State-level information is taken from the National Center for Education Statistics, Digest of Education Statistics 2019 (De Brey et al. 2021). For the pre-primary, primary, and secondary levels, we use the values from Table 236.75: Total and current expenditures per pupil in fall enrollment in public elementary and secondary schools, by function and state or jurisdiction: 2016-17. To compute public per-student expenditure at the tertiary level, we divide total expenditure (Table 334.20: Total expenditures of public degree-granting postsecondary institutions, by level of institution, purpose of expenditure, and state or jurisdiction: 2014-15 through 2017-18) by the number of students (Table 304.15: Total fall enrollment in public degree-granting postsecondary institutions, by state or jurisdiction: Selected years, 1970 through 2018).

5 percent higher than the NIPA figure of \$666 B. That our value is slightly higher than the OECD figure is due to two factors. First, as shown in Figure 5.11, the ACS enrollment numbers are slightly higher than what is reported by the OECD (51.4 M vs. 50.6 M). Second, some of the children in pre-primary education attend kindergarten only part of the day. When computing full-time equivalents, the OECD assigns individuals in part-time education a weight of 0.3. In the ACS, we do not observe part-time status, and assign all individuals the full-time equivalent expenditure reported by the OECD. This amounts to the assumption that all individuals are in fact in full-time education, which leads us to overestimate the annual expenditure.

Both factors are aggravated at the tertiary level. In the ACS, there are 16.8 M students enrolled in public tertiary institutions, while the OECD and the NCES report 14.6 M students, a difference of 2.2 M or about 15 percent (Figure 5.11).[10] Moreover, the part-time share is even higher than for pre-primary education.[11] As a result of both factors, our estimate of annual public expenditure at the tertiary level of \$488 B is substantially higher than the numbers reported by the OECD, the NCES, and the Bureau of Economic Analysis, which range between \$288 B and \$335 B.

As part of our robustness checks, we address these issues by rescaling the enrollment numbers in the ACS so that we have the same number of full-time equivalent students as the OECD. We do this both in a neutral way—by assuming that the excess number of full-time equivalents is independent of income—and as a bounds analysis in which we assume that the excess mass is concentrated in either the bottom or the top half of the income distribution.

5.3 Results

5.3.1 Distribution of Public Education Spending

Allocation Based on Actual Enrollment Figure 5.1 shows how public education spending in the United States in 2017 is distributed among the deciles of the income distribution.[12] Following Piketty, Saez, and Zucman (2018), the distribution is for adults age 20 and above; in households with more than one adult, income is split equally. Income is pre-tax income as reported in the American Community Survey; below, we report results when we use simulated

[10] Part of the difference is probably due to students at private non-profit institutions. According to the NCES, 1.1 M students attended such an institution in 2017. If some of these declared to be in a public institution in the ACS because they equated not-for-profit with public, this could explain part of the higher number of students at public institutions that we find. Note, however, that we also overestimate the total number of students at the tertiary level, so the measurement issue does not only concern the classification of institutions into public or private.

[11] According to the OECD, 1.6 M out of 5.1 M children (68 percent) in pre-primary education attend kindergarten only part-time. At the tertiary level, there are 6.4 M part-time students (43 percent of the total 14.6 M). At the primary and secondary levels, all pupils attend school full-time.

[12] The numerical values are reported in Table 5.4 in the Appendix.

post-tax income instead, as a first step towards the more comprehensive measure of post-tax cash income used by Piketty, Saez, and Zucman.

Public education spending is highest in the first decile—with an average of $6.0 K per adult—and lowest in the tenth decile of the pre-tax income distribution, where the average is $4.3 K. In deciles 2 to 9, the means of per-capita spending are fairly close together, at between $4.5 K and $4.8 K. The overall average is $4.8 K. The Bottom 50 percent of the pre-tax income distribution receive an average of $4.9 K per year in terms of public education spending, followed by the Middle 40 percent with $4.7 K, and, as noted, the Top 10 percent with $4.3 K.

Turning to the different levels of education, we see little differences by income for pre-primary and primary education. Per-capita expenditure on secondary education tends to grow with income, with an average of $1.4 K allocated to each adult in decile 1 and about $1.9 K in deciles 9 and 10. Public spending on tertiary education shows the opposite pattern. It is the driver behind the regressivity of public education spending, being concentrated in the bottom decile of the income distribution, where average annual spending is $3.3 K, more than three times the average in the top decile ($1.0 K). The high average in the poorest decile is mostly explained by college students who no longer live with their parents. By contrast, the public expenditure on students who are still in the parental household is spread out much more evenly across the income distribution.

Comparison with Proportional and Lump-Sum Allocations Figure 5.2 contrasts the actual distribution based on the American Community Survey with the proportional allocation used by Piketty, Saez, and Zucman (2018). For public spending on education (about 30 percent of collective expenditure in the United States), a proportional allocation is clearly not a good assumption. It implies annual per-capita spending of $0.6 K in the poorest decile, only a tenth of the value that we find based on actual enrollment data from the American Community Survey. At the top of the income distribution, the proportionality assumption allocates $18.4 K to each adult in the richest decile, more than four times the value based on the ACS. Furthermore, as pointed out in Section 5.A in the Appendix, given the unequal distribution of pre-tax income even within the top decile, a proportional allocation implies implausibly high per-capita values among individuals in, say, the Top 1 percent or Top 0.1 percent of the distribution.

As microdata on enrollment in education is easily available for the United States and other countries, we believe that the precision of the Dina approach can be improved at little cost by replacing the proportionality assumption with an allocation based on actual enrollment. An even easier fix consists in replacing the allocation proportional to post-tax cash income—which the Dina Guidelines recommends as the benchmark—by a lump-sum allocation. As Figure 5.2 shows, assigning the mean of $4,783 to each adult is a good approximation to the distribution based on actual enrollment.

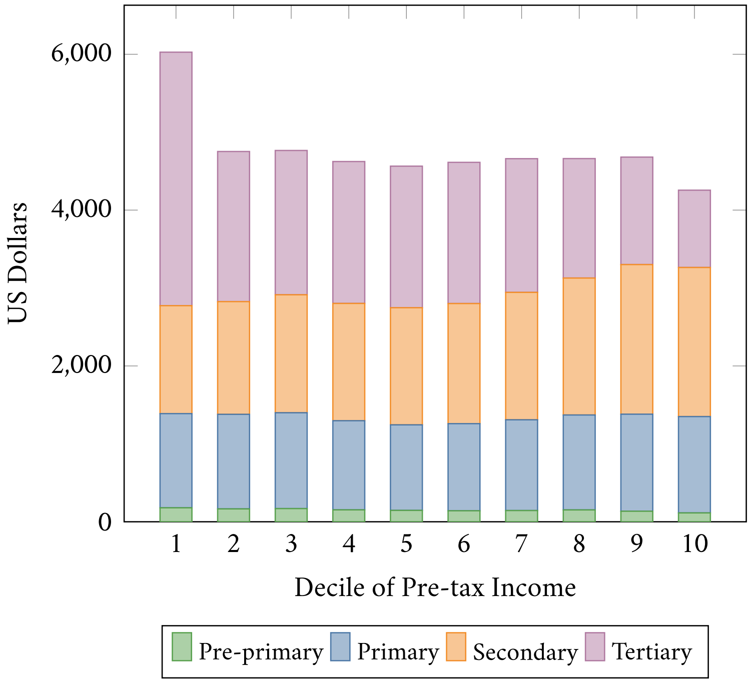


FIGURE 5.1: Public Education Spending by Pre-tax Income, Allocated Based on Actual Enrollment

Notes: The figure shows how public education spending in the United States in 2017 is distributed among the deciles of the pre-tax income distribution. For each decile, the bars show the average values of annual public education spending (in 2017 US Dollars) at the pre-primary, primary, secondary, and tertiary levels of education. *Source:* Enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is summed up at the household level, and the resulting sum is split equally among adults aged 20 and above in the household. Household income is likewise split equally among all adults.

Figure 5.2 also includes the distribution that arises from allocating public education spending as a lump-sum transfer per child below the age of 20, as in the robustness check in the paper by Piketty, Saez, and Zucman (2018). This assumption performs much better than the proportional allocation. The differences with respect to our baseline results arise from the fact that this shortcut method does not take into account the differences in per-capita expenditure by level of education (tertiary education is much more expensive than the rest, at least in the United States), and especially that it does not capture public spending that goes to college students age 20 and above.

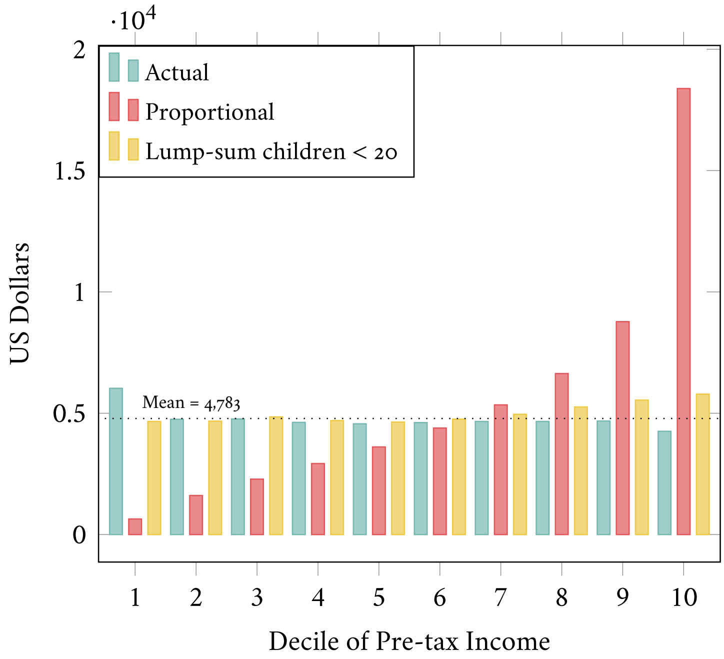


Figure 5.2: Public Education Spending by Pre-tax Income: Comparison of Allocation Methods

Notes: The figure compares the actual distribution of public education spending (in green, this is the same distribution as in Figure 5.1) with the distributions that result from an allocation that is proportional to pre-tax income (yellow) and from a lump-sum transfer (gray) to all children below age 20, irrespective of actual enrollment and disregarding the differences in per-capita spending between pre-primary, primary, secondary, and tertiary education. The figure also shows the value of $4,783 that would result from a lump-sum allocation to all adults. *Source:* Enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is summed up at the household level, and the resulting sum is split equally among adults aged 20 and above in the household. Pre-tax household income is likewise split equally among all adults.

Regressivity Driven by Age Effects The regressivity of public education spending in the cross-section is strongly driven by age effects (Figure 5.3). Individuals aged 20 to 24 receive a lot of education spending on average, mostly for their own (tertiary) education. At the same time, they have by far the lowest current income of all age groups. Pre-primary and primary education does not play a large role in this age group, as the share of parents is still low. Spending on secondary education is a bit higher because some individuals are still in secondary education themselves.

In the age group 25–29, average public education spending is much lower. (Own) tertiary education is still significant, but less so than for individuals in their early 20s. Secondary education also drops in importance, while public expenditure on pre-primary and primary education starts building up as individuals in this group have more (and older) children than in the age group just below.

In the older age groups, the share of parents and the age of their children continue to rise, as reflected in the increasing public expenditure at the pre-primary, primary, and secondary levels. While the first two peak in the age group 35–39, spending on secondary and tertiary education continues into age groups 40–44 and 45–49, respectively. At later ages, expenditure falls for them as well as children leave the parental household. The maximum of total public education spending is reached in the age group 40–44. Pre-tax income, by contrast, peaks at age 45–49, and is still fairly high thereafter while public education spending declines steeply for individuals in their late 40s and in their 50s. Together with the high level of public education spending for the poorest age group 20–24, this phase shift drives the regressivity of public education spending in the cross-section.

5.3.2 Robustness Checks

Post-tax Cash Income So far, our results have been for pre-tax income, which is directly observable in the American Community Survey. However, Piketty, Saez, and Zucman (2018) assume that education and other collective expenditure items are allocated proportionally to post-tax cash income. We therefore run a robustness check in which we use our measure of post-tax cash income—simulated using TAXSIM—to divide adults into deciles (Figure 5.12 and Table 5.4 in the Appendix). As for pre-tax income, the proportionality assumption is rejected, while a lump-sum allocation is a reasonable approximation except for the bottom and the top of the income distribution.

Household Equivalence Income As noted in Section 5.1, there is an older literature that augments the standard survey measures of disposable (money) income by different components of public in-kind spending, often with a cross-country focus. These studies measure income at the household level and attempt to make households of different size and age composition comparable through equivalence scales. As Figure 5.13 in the Appendix shows, public education spending remains regressive when adopting such a household perspective.[13] The average

[13] The numerical values are reported in Table 5.4 in the Appendix.

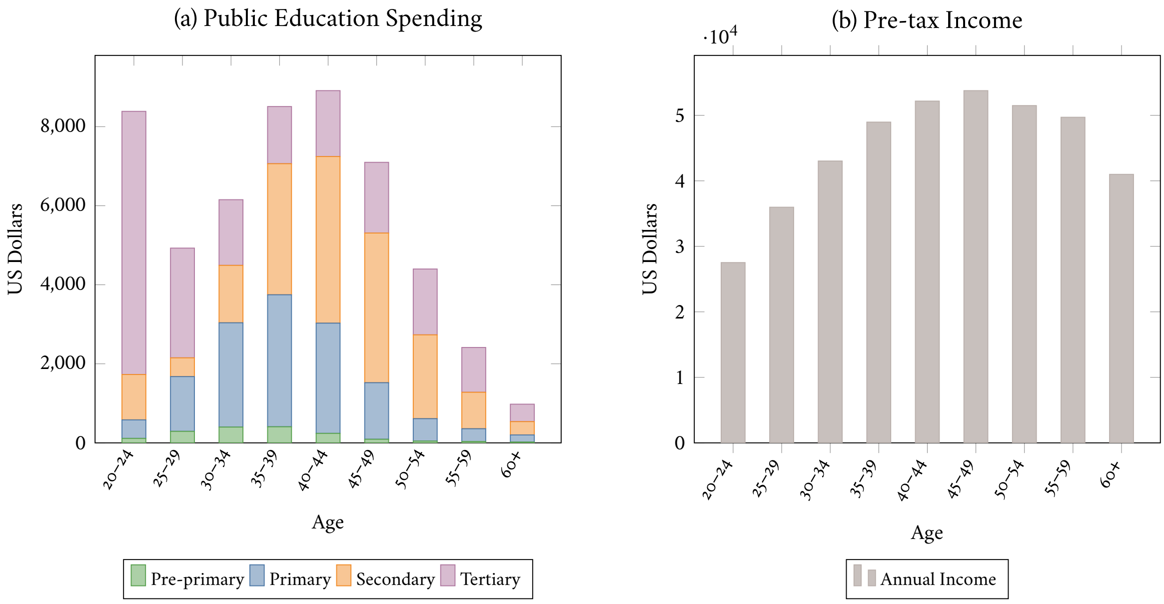


FIGURE 5.3: Public Education Spending and Pre-tax Income by Age

Notes: The left panel of the figure shows how the average value of public education spending (allocated based on actual enrollment) differs by age. The right panel depicts average pre-tax income for the same age categories. *Source:* Enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is summed up at the household level, and the resulting sum is split equally among adults aged 20 and above in the household. Pre-tax household income is likewise split equally among all adults.

amount of public education spending received is now higher as the transfers are measured at the household level and not divided equally among adults. When the deciles are defined based on pre-tax income, average spending declines throughout the distribution. For a distribution based on post-tax cash income, a lump-sum allocation is a decent approximation for the bottom three or four deciles, but the upper half of the distribution is again characterized by a negative relationship between public education spending and household income.

The finding that public education spending declines with household income is in line with the study by Zwijnenburg, Bournot, and Giovannelli (2017) who report the percentage of total education spending by quintiles of household disposable income for the United States and several other countries. In the United States in 2012, 25.4 percent of public education spending goes to households in the bottom quintile, compared with 15.3 percent in the top quintile. In our data for 2017, the shares are similar, but the regressivity is even more pronounced: 26.3 percent of public spending goes to the 20 percent of households with the lowest post-tax cash income, while the richest 20 percent receive 11.1 percent of the total.

Other Checks We also ran a number of other, more technical robustness checks. As noted above, despite the legal obligation to answer the survey, some of the individual income components are actually imputed by the data provider. When we drop all households in which more than half of household income is based on an imputation (slightly less than 20 percent of our sample), the results are virtually unchanged (Table 5.5). The same holds when we drop all households with negative income or all households with income below the 1st percentile. When the threshold is increased to the 2.5th percentile, average public education spending in the first decile is reduced from \$6.0 K to \$5.4 K, but is still higher than in all other deciles. Dropping all households whose income is above the 99.5th percentile likewise has no effect on the results.

As seen in Figure 5.11, the ACS slightly overestimates the enrollment in educational institutions by comparison with the numbers reported by the OECD and the NCES. When we scale down our ACS enrollment numbers to meet the NCES numbers, average public education spending goes down in all deciles, but the negative relationship with income is preserved.

In our main specification, we use a single value for per student expenditure at the different levels of education, as the OECD does not provide information on within-country variation. When we use state-specific values from the NCES instead, the difference between the first and the tenth deciles is slightly reduced, but the poorest decile still receives substantially more public education spending. A lump-sum allocation is again a good approximation for the deciles in between.

The OECD calculates expenditure per student on the basis of full-time equivalents; students in part-time education—relevant only at the pre-primary and the tertiary level—are assumed to represent one-third of a full-time equivalent. In the ACS, there is no information on whether individuals are enrolled only part-time, and we assign the expenditure per full-time equivalent to all students in our main specification. As a robustness check, we randomly assign part-time status based on the share of part-time students reported by the OECD. This brings down the average expenditure by decile, but leaves the negative income gradient intact.

The OECD only reports a single number for annual per-student expenditure at the tertiary level, which is an average over 2-Year and 4-Year colleges. In a robustness check, we assign all graduate students to 4-Year colleges, and randomly assign undergraduates to either 2-Year or 4-Year colleges, based on the relative importance of the two types as reported by the NCES. Average expenditure is higher than in the main specification, but the relationship between income and expenditure remains the same.

In our main specification, public expenditure per student includes both current expenditure (a large share of which are salaries and wages) and capital outlays, but excludes R&D—which is relevant only at the tertiary level—as well expenditure for ancillary services. Alternatively, we have used a narrower (only current expenditure) and a broader (all expenditure types plus R&D and ancillary services) definition of public education spending. This changes the level of expenditure, but has little impact on the income gradient.

5.3.3 Beyond the Cross-Section

The Dina literature invokes two arguments for assigning public education spending proportionally to post-tax cash income: the unequal access to education by parental income—e.g., Alvaredo et al. (2020, p. 65) or Saez and Zucman (2020, p. 33)—and "a lifetime perspective where everybody benefits from education, and where higher earners attend better schools and for longer" (Piketty, Saez, and Zucman 2017, p. 27/28).

In the following, we show that individuals with higher lifetime earnings (proxied for by earnings at age 40–45) have indeed received substantially more public education spending in the past. We also show—based on PSID data—that more public education spending goes to children whose parents have a higher socio-economic status (proxied for by educational attainment). In both cases, we depart from the cross-sectional perspective we have adopted so far. In particular, we do not consider the public education spending received in a single year, but the sum of spending received in the education system. We classify individuals by their highest degree and assume that a given degree implies that the individual has passed through all the stages below, and that everyone needed the same number of years to complete each stage.[14] This is admittedly a simplification. For instance, not every child attends kindergarten, and some students repeat a year in school or take longer to finish a bachelor's or master's degree,

[14] The details of our mapping between the highest degree observed in the ACS and the number of years spent at the different ISCED levels are presented in Table 5.6 in the Appendix.

and this variation is likely correlated with both lifetime earnings and parents' socio-economic status. However, with our data there is little we can do about this, and the differences that we find are so large that they are robust to different assumptions. A potentially more important qualification is that we do not observe whether individuals completed their education abroad. We have no information about this in our data, and assume that the entire schooling was obtained in the United States. Another shortcut that we take is to use the 2017 per-student values for public education expenditure (Table 5.2) although the cohort of individuals that we consider—40-45-year-olds in 2017, i.e. people born in the early and mid-1970s—obtained their education in the past. Given that we consider a cohort of only six years and that our interest is in the gradient and not the level of spending, this assumption should be fairly innocuous as well. Finally, moving beyond the cross-section—i.e., current educational enrollment—means that we cannot distinguish between public and private institutions anymore. We assume that all individuals obtained their degrees in the public education system. This means that we overestimate the level of expenditure and, more importantly, the income gradient, as graduating from a private college is positively correlated with both own lifetime earnings and parents' socio-economic status.

Differences by Lifetime Earnings We proxy for lifetime earnings using current earnings of individuals aged 40–45. At this age, the rank correlation between current earnings and lifetime earnings reaches its maximum (e.g. Haider and Solon 2006; Bönke, Corneo, and Lüthen 2015). As we now consider earnings and not income, we do not use the equal-split assumption that we adopt in the cross section, but directly use the personal earnings information available in the ACS.

Figure 5.4 shows how the highest degree and public education spending received vary with earnings. As expected, the highest degree is positively correlated with earnings (Panel a). While in the bottom half of the earnings distribution most individuals have at most a high school diploma or attended college without obtaining a degree, the share of people with a bachelor's, master's, or professional and doctor's degree increases in the upper half of the earnings distribution.

When translating these differences in degrees into differences in public education spending received, there is—unlike in the cross-section—a positive income (or, more precisely, earnings) gradient. The 10 percent of individuals with the highest earnings have received average public education spending of $335 K, about 1.4 times the amount of the bottom 50 percent ($234 K). The allocation is still not proportional to earnings, however; proportionality would require a factor of about 14 ($196 K vs. $14 K).

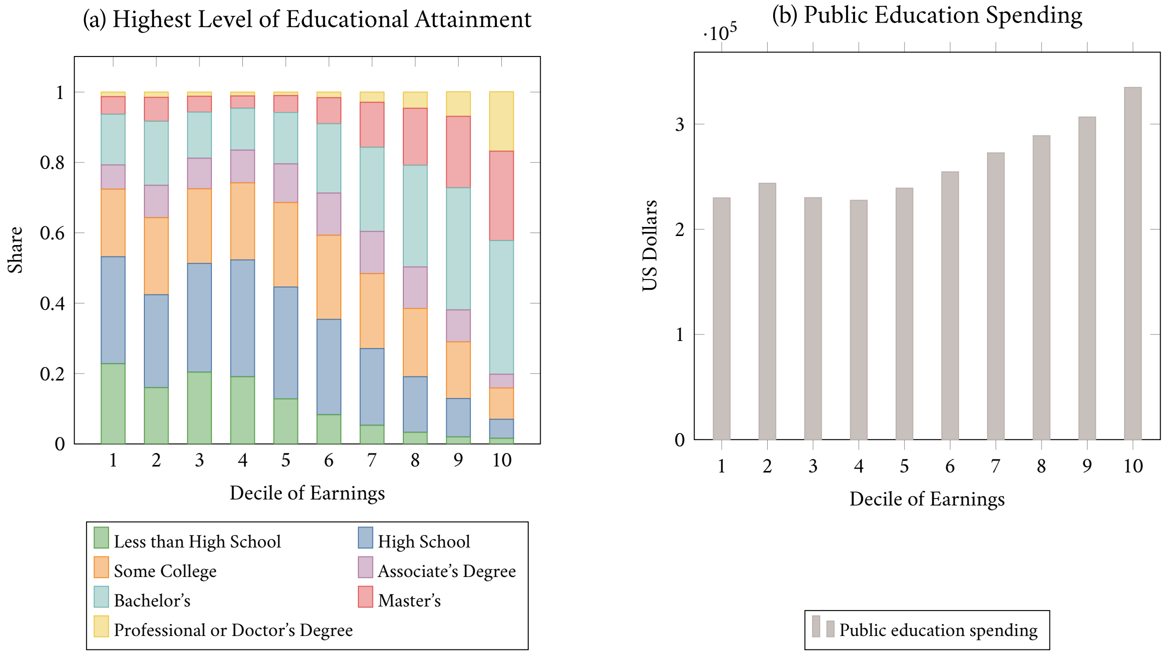


Figure 5.4: Highest Degree and Public Education Spending by Current Earnings, Individuals Aged 40–45

Notes: The figure shows the highest degree (left panel) and public education spending by current earnings (right panel) for individuals aged 40–45. *Source:* Own calculations based on the American Community Survey 2017. When calculating public education spending, we assume that a given degree implies that the individual has passed through all the stages below, and that everyone needed the same number of years to complete each stage (see Table 5.6 in the Appendix for details). We also assume that all individuals have attended only public educational institutions. Each year in the education system is multiplied with the per-capita value of public education spending taken from the OECD (see Table 5.2). We use the 2017 values of per-capita spending although the individuals who were 40–45 years old in 2017 obtained their education in earlier years.

Intergenerational Perspective The second argument invoked in the DINA literature for assigning public education spending proportionally to post-tax cash income is the unequal access to education by parents' income or, more generally, socio-economic status (SES). Studies documenting this inequality are legion. Children from a more advantaged socio-economic background tend to go to better schools and are more likely to attend college. We show that these differences indeed produce a positive relationship between parents' SES and the public education expenditure that their children receive. Unfortunately, we do not observe parents' SES in the American Community Survey. We therefore make use of the Panel Study of Income Dynamics (PSID) instead, which we again combine with information from the OECD on current public expenditure per student.[15]

Like in the analysis based on lifetime earnings, we restrict the sample to individuals aged 40–45 in 2017. As we do not observe individual trajectories, we again assume that individuals followed a stylized path to their highest degree (no grade retentions etc.).

Figure 5.5 shows that individuals whose parents attended college received substantially more public education spending than children of parents with a high school degree or no degree at all. As almost all individuals attended school at least until grade 8, differences start to arise for upper secondary education (ISCED level 3). Individuals whose mother (father) did not complete high school received around \$50 K (\$46 K) in public spending for upper secondary education. If a parent attended college, public spending at the upper secondary level was higher by around \$7.5 K (mothers) and \$11.5 K (fathers). The differences at the tertiary level (ISCED levels 5–8) are more pronounced. Among individuals whose parents have no high school degree, only around 13 percent completed college (the share is about the same both for maternal and paternal education). This group therefore has a low (unconditional) average of public education spending at the tertiary level of \$32 K (mothers) and \$37 K (fathers). By contrast, individuals where one or both parents attended college received an average of around \$90 K in terms of public spending on tertiary education.

Total public spending on education was on average \$267 K for the cohort considered here. The difference between individuals from the most and the least privileged background with respect to parents' education is \$66 K on the father's side and \$68 K on the mother's side.

DINA as a Cross-Sectional Approach That children from already more privileged backgrounds receive almost \$70 K more in public education expenditure is more important for the distributional debate than the regressive pattern of public expenditure found in any given year, which, as seen above, is strongly driven by age effects. However, the positive association between public expenditure and parental SES or own lifetime earnings does not provide a justification for allocating public education expenditure proportionally to income in the DINA

[15] The data are described in Section 5.B in the Appendix.

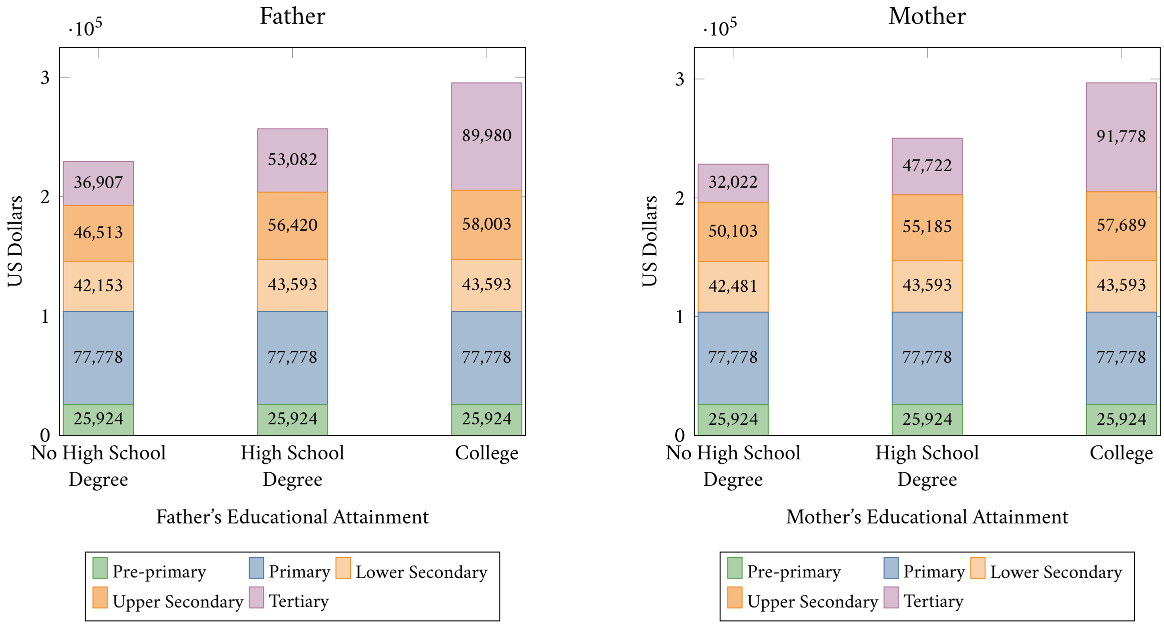


FIGURE 5.5: Public Education Spending by Parents' Education. Individuals Aged 40–45

Notes: The figure depicts average public education spending by parents' education for individuals aged 40–45. The left panel distinguishes by the education of the father, the right panel by the education of the mother. *Source:* Own calculations using the Panel Study of Income Dynamics (PSID) 2017. When calculating public education spending, we assume that a given degree implies that the individual has passed through all the stages below, and that everyone needed the same number of years to complete each stage. We also assume that all individuals have attended only public educational institutions. Each year in the education system is multiplied with the per-capita value of public education spending taken from the OECD (see Table 5.2). We use the 2017 values of per-capita spending although the individuals who were 40–45 years old in 2017 obtained their education in earlier years.

approach. So far, the approach has been exclusively cross-sectional, and departing from this cross-sectional perspective only for public education spending seems ad hoc. After all, age effects are also present in earnings or capital income, but are not adjusted for when measuring pre-tax income. Likewise, many cash transfers such as family benefits or in-kind transfers such as Medicare are also age-dependent, but are assigned to current recipients in the Dina approach.

5.4 Conclusion

In the distributional national accounts (Dina) created by Piketty, Saez, and Zucman (2018) and others, government collective expenditure (e.g., education, defense, infrastructure) is typically allocated proportionally to post-tax cash income, which renders half of government expenditure distributionally neutral and implies large differences in the per-capita value of collective expenditure. The level of post-tax inequality is fairly sensitive to this assumption. When the expenditure is allocated on a lump-sum basis instead—an assumption that the recent version of the Dina Guidelines (Alvaredo et al. 2020) suggests as an alternative to the proportional allocation—the gap in post-tax income shares between the Top 10 percent and Bottom 50 percent is reduced by half. The trend in US post-tax income shares is hardly affected by the assumptions, however.

The main contribution of our paper is to provide evidence on how an important part of collective expenditure is actually distributed. We find that, when adopting the cross-sectional perspective of the Dina approach, public education spending goes disproportionately to the bottom half of the income distribution. This pattern is strongly driven by age effects. There is indeed a positive relationship between public education spending and lifetime earnings or parents' socio-economic status, but even the relationship with earnings is far from being proportional. More importantly, the last two patterns do not provide an empirical basis for the cross-sectional Dina approach. Adjusting for age effects only for public education, but not for other items such as earnings, capital income, family cash transfers, or Medicare, would introduce an inconsistency into the framework.

Based on our findings, we conclude that public education expenditure should not be allocated proportionally to post-tax cash income as recommended in the Dina Guidelines. As microdata on education is widely available, an allocation based on actual enrollment can improve the distributional analysis of post-tax income at little extra cost. This recommendation is in line with the OECD–Eurostat Expert Group on Disparities in a National Accounts framework (EG DNA), which also argues for an allocation based on actual use (Zwijnenburg 2019). An even easier improvement is to allocate public education spending as a lump-sum transfer, which—at least in the US context of 2017—provides a good approximation of the actual distribution.

Given that a proportional allocation implies very high per-capita values for individuals with high incomes, we believe that a lump-sum allocation is the preferable benchmark for the remaining parts of government consumption expenditure (defense, infrastructure) as well. However, given the difficulty of assigning these other items to households and individuals, reporting results for both a lump-sum and a proportional allocation is probably a reasonable compromise. Another option is to resort to an income concept such as disposable personal income that takes only money transfers and certain in-kind transfers such as Medicare and Medicaid into account while avoiding the assignment of government consumption expenditure altogether (Gindelsky 2022). Whether this or the more comprehensive Dina income concept is more useful depends on the question at hand.

In our analysis, differences in public education spending result from differences in enrollment and in the choice of public or private institutions. In a robustness check, we also exploit differences in average spending across states. We do *not* capture any remaining income-related variation in per-capita spending. As this remaining variation is likely positively related to income, this means that we do not capture one component that would work toward the proportionality assumption used as the benchmark in the Dina approach. However, we find such a strong departure from proportionality that the within-state differences in per-capita spending would have to be implausibly high in order to justify the assumption. Moreover, at least at the level of school districts, the difference by income is less pronounced than one might think, and is characterized by a U-shape instead of a monotonous increase with income. Still, incorporating more fine-grained information on per-capita spending would further increase the precision of the Dina approach and is a useful direction for future research.

References

Aaberge, Rolf, Manudeep Bhuller, Audun Langørgen, and Magne Mogstad (2010). "The Distributional Impact of Public Services When Needs Differ". In: *Journal of Public Economics* 94, pp. 549–562.

Aaberge, Rolf, Lasse Eika, Audun Langørgen, and Magne Mogstad (2019). "Local Governments, In-kind Transfers, and Economic Inequality". In: *Journal of Public Economics* 180, p. 103996.

Aaberge, Rolf, Audun Langørgen, and Petter Lindgren (2013). "The Distributional Impact of Public Services in European Countries". Statistics Norway Research Department Discussion Paper No. 746.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Blanchet, Lucas Chancel, Luis Bauluz, Matthew Fisher-Post, Ignacio Flores, Bertrand Garbinti, Jonathan Goupille-Lebret, Clara Martínez-Toledano, Marc Morgan, Theresa Neef, Thomas Piketty, Anne-Sophie Robilliard, Emmanuel Saez, Li Yang, and Gabriel Zucman (2020). "Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database". Paris. World Inequality Lab.

Auten, Gerald and David Splinter (2022). "Income Inequality in the United States: Using Tax Data to Measure Long-term Trends". Working Paper.

Bach, Stefan, Charlotte Bartels, and Theresa Neef (2021). "Distributional National Accounts: A Macro-Micro Approach to Inequality in Germany". DIW and Freie Universität Berlin.

Blanchet, Thomas, Lucas Chancel, and Amory Gethin (2022). "Why is Europe More Equal than the United States?" In: *American Economic Journal: Applied Economics* 14 (4), pp. 450–518.

Bönke, Timm, Giacomo Corneo, and Holger Lüthen (2015). "Lifetime Earnings Inequality in Germany". In: *Journal of Labor Economics* 33 (1), pp. 171–208.

Bozio, Antoine, Bertrand Garbinti, Jonathan Goupille-Lebret, Malka Guillot, and Thomas Piketty (2022). "Predistribution vs. Redistribution: Evidence from France and the U.S." World Inequality Lab, Paris.

Bruil, Arjan, Céline van Essen, Wouter Leenders, Arjan Lejour, Jan Möhlmann, and Simon Rabaté (2022). "Inequality and Redistribution in the Netherlands". Working Paper.

Callan, Tim, Timothy Smeeding, and Panos Tsakloglou (2008). "Short-run Distributional Effects of Public Education Tranfers to Tertiary Education Students in Seven European Countries". In: *Education Economics* 16 (3), pp. 275–288.

De Brey, Cristobal, Thomas D. Snyder, Anlan Zhang, and Sally A. Dillow (2021). "Digest of Education Statistics 2019 (NCES 2021-009)". Washington, DC. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Feenberg, Daniel and Elisabeth Coutts (1993). "An Introduction to the TAXSIM Model". In: *Journal of Policy Analysis and Management* 12 (1), pp. 189–194.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2018). "Income Inequality in France, 1900–2014: Evidence from Distributional National Accounts". In: *Journal of Public Economics* 162, pp. 63–77.

Garfinkel, Irwin, Lee Rainwater, and Timothy Smeeding (2006). “A Re-examination of Welfare States and Inequality in Rich Nations: How In-kind Transfers and Indirect Taxes Change the Story”. In: *Journal of Policy Analysis and Management* 25 (4), pp. 897–919.

Gillespie, W. Irwin (1965). “Effect of Public Expenditures on the Distribution of Income”. In: *Essays in Fiscal Federalism*. Ed. by Richard A. Musgrave. Washington, DC: Brookings Institution, pp. 122–186.

Gindelsky, Marina (2022). “Do Transfers Lower Inequality Between Households? Demographic Evidence from Distributional National Accounts”. In: *Economic Inquiry*, pp. 1–25.

Haider, Steven and Gary Solon (2006). “Life-Cycle Variation in the Association between Current and Lifetime Earnings”. In: *American Economic Review* 96 (4), pp. 1308–1320.

Hammar, Olle, Paula Roth, Daniel Waldenström, and Gabriel Zucman (2020). “Distributional National Accounts in the Welfare State: Sweden, 1930–2016”. In progress.

Horton, Tim and Howard Reed (2010). *Where the Money Goes: How We Benefit from Public Services*. Trade Unions Congress.

Jestl, Stefan and Emanuel List (2020). “Distributional National Accounts (DINA) for Austria, 2004–2016”. World Inequality Lab Working Paper, 2020/23.

Marical, Francois, Marco Mira d’Ercole, Maria Vaalavou, and Gerlinde Verbist (2006). “Publicly-provided Services and the Distribution of Resources”. In: *OECD Social, Employment and Migration Working Papers* 45.

McGonagle, Katherine A., Robert F. Schoeni, Narayan Sastry, and Vicki A. Freedman (2012). “The Panel Study of Income Dynamics: overview, recent innovations, and potential for life course research”. In: *Longitudinal and Life Course Studies* 3 (2), pp. 268–284.

Musgrave, Richard A., Karl E. Case, and Herman Leonard (1974). “The Distribution of Fiscal Burdens and Benefits”. In: *Public Finance Quarterly* 2 (3), pp. 259–311.

O’Dea, Cormac and Ian Preston (2012). “The Distributional Impact of Public Spending in the UK”. London. IFS Working Paper W12/06.

O’Higgins, Michael and Patricia Ruggles (1981). “The Distribution of Public Expenditures and Taxes among Households in the United Kingdom”. In: *Review of Income and Wealth* 27 (3), pp. 298–326.

OECD Statistics (2020). *Education at a Glance – Dataset “Educational finance indicators”*. URL: `https://stats.oecd.org` (visited on 09/23/2021).

Paulus, Alari, Holly Sutherland, and Panos Tsakloglou (2010). “The Distributional Impact of In-Kind Public Benefits in European Countries”. In: *Journal of Policy Analysis and Management* 29 (2), pp. 243–266.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2017). *Distributional National Accounts: Methods and Estimates for the United States. Data Appendix*. URL: `http://gabriel-zucman.eu/usdina/` (visited on 09/23/2021).

– (2018). “Distributional National Accounts: Methods and Estimates for the United States”. In: *Quarterly Journal of Economics* 133 (2), pp. 553–609.

Piketty, Thomas, Li Yang, and Gabriel Zucman (2019). "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015". In: *American Economic Review* 109 (7), pp. 2469–2496.

Ruggles, Patricia and Michael O'Higgins (1981). "The Distribution of Public Expenditure among Households in the United States". In: *Review of Income and Wealth* 27 (2), pp. 137–164.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2020). *IPUMS USA: Version 10.0 [dataset]*. Minneapolis, MN: IPUMS. URL: `https://doi.org/10.18128/D010.V10.0`.

Saez, Emmanuel and Gabriel Zucman (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data". In: *Quarterly Journal of Economics* 131 (2), pp. 519–578.

– (2020). "Trends in US Income and Wealth Inequality: Revising After the Revisionists". NBER Working Paper No. 27921.

Samwick, Andrew A. (2013). "Donating the Voucher: An Alternative Tax Treatment of Private School Enrollment". In: *Tax Policy and the Economy* 27 (1), pp. 125–160.

Smeeding, Timothy (1977). "The Antipoverty Effectiveness of In-Kind Tranfers". In: *The Journal of Human Resources* 12 (3), pp. 360–378.

Smeeding, Timothy, Peter Saunders, John Coder, Stephen Jenkins, Johan Fritzell, Aldi J. M. Hagenaars, Richard Hauser, and Michael Wolfson (1993). "Poverty, Inequality, and Family Living Standards Impacts across Seven Nations: The Effect of Noncash Subsidies for Health, Education and Housing". In: *Review of Income and Wealth* 39 (3), pp. 229–256.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick (2019). "Capitalists in the Twenty-first Century". In: *Quarterly Journal of Economics* 134 (4), pp. 1675–1745.

Survey Research Center, ed. (2022). *Panel Study of Income Dynamics, Public Use Dataset*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

Verbist, Gerlinde, Michael Förster, and Maria Vaalavou (2012). "The Impact of Publicly Provided Services on the Distribution of Resources: Review of New Results and Methods." In: *OECD Social, Employment and Migration Working Papers* 130.

Wilson, Kathryn, Kristina Lambright, and Timothy Smeeding (2006). "School Finance, Equivalent Educational Expenditure, and Income Distribution: Equal Dollars or Equal Chances for Success?" In: *Education Finance and Policy* 1 (4), pp. 396–424.

Zwijnenburg, Jorrit (2019). "Unequal Distributions: EG DNA versus DINA Approach". In: *AEA Papers and Proceedings* 109, pp. 296–301.

Zwijnenburg, Jorrit, Sophie Bournot, and Federico Giovannelli (2017). "OECD Expert Group on Disparities in a National Accounts Framework – Results from the 2015 Exercise". Paris. OECD Statistics Directorate Working Paper No. 76.

Appendix

5.A Revisiting Piketty, Saez, and Zucman (2018)

5.A.1 Overview

In an important methodological contribution, Piketty, Saez, and Zucman (2018) create distributional national accounts that make income measures from tax and survey data consistent with the macro totals published in national accounts. They study pre-tax and post-tax income inequality in the United States for the years 1913–2014 and document a massive increase for both types of inequality since 1980.

In this section of the Appendix, we show that their findings regarding the *level* of post-tax inequality are sensitive to their assumption regarding the allocation of government consumption expenditure. Based on their publicly available data, we show that with a different assumption—a lump-sum allocation of collective expenditure instead of an allocation proportional to post-tax cash income—, the gap in the shares of post-tax national income accruing to the Bottom 50 percent and the Top 10 percent is reduced by half in recent years, from 20 to 10 percentage points.[16] The effect of the allocation rule on post-tax income shares is of the same order of magnitude as in the study by Blanchet, Chancel, and Gethin (2022) for a number of European countries.

5.A.2 Post-tax Income Inequality and Government Expenditure

Figure 5.6 summarizes the distribution of U.S. national income in 2014, the most recent year in their study. Piketty, Saez, and Zucman allocate all items of national income to adults age 20 and above. In couples, the income is assumed to be split equally. The mean value of national income by adult in 2014 is $65 K. By construction, the mean is the same for pre-tax and post-tax income, which are alternative ways of allocating the same total national income.[17] Pre-tax income is distributed very unequally: the 10 percent of adults with the highest pre-tax income receive 47 percent of the total, while the Bottom 50 percent receive only 13 percent. This translates into an average pre-tax income of about $300 K among the Top 10 percent (47/10 times the mean income of $65 K), compared with $16 K for the bottom half of the pre-tax

[16] The results, code, and most of the micro data are available at `http://gabriel-zucman.eu/usdina/`. We use the November 2017 vintage, which corresponds to the published version (Piketty, Saez, and Zucman 2018). The series have since been updated to more recent years, improved, and revised (to incorporate changes in the underlying National Accounts data). These changes are documented in `https://gabriel-zucman.eu/files/PSZUpdates.pdf`. The part of the analysis that we focus on in this article – the allocation of government consumption expenditure – has not been affected by the updates.

[17] The small difference in Figure 5.6—63,632 vs. 64.633—is due to rounding.

income distribution. The Middle 40 percent receive almost exactly their population share of 40 percent, and accordingly have an average pre-tax income close to the overall mean.

Income after taxes and transfers is only slightly less unequally distributed. The share of the Top 10 percent decreases from 47 percent to 39 percent, while the shares of the Middle 40 percent and the Bottom 50 percent increase by 2 and 6 percentage points, respectively.

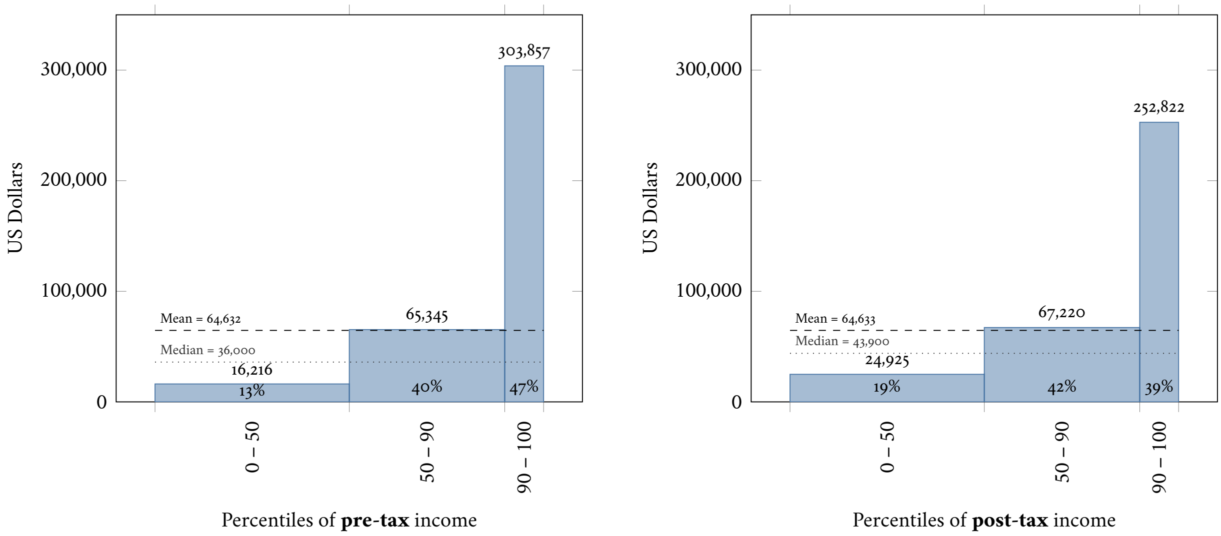


FIGURE 5.6: Distribution of Pre-tax and Post-tax National Income

Notes: The figure shows how national income is distributed among adults aged 20 and above in the United States in 2014. The figure depicts the income shares of the Bottom 50 percent, Middle 40 percent, and Top 10 percent, both for total pre-tax income (left panel) and for post-tax income (right panel), as well as the overall mean and median and the mean within each group. *Source:* Own calculations based on Piketty, Saez, and Zucman (2018). Pre-tax income: Appendix Tables II-B1, II-B3, II-B13. Post-tax income: Appendix Tables II-C1, II-C3, II-C13.

Figure 5.7 shows how post-tax income is divided between two broad categories – income net of taxes on the one hand and transfers on the other – and how each is divided among the Top 10 percent, Middle 40 percent, and Bottom 50 percent. Overall, 66.5 percent of U.S. national income in 2014 corresponds to income net of taxes, while the remaining 33.5 percent are transfers. The share of national income that goes to the Bottom 50 percent is made up of 5.1 percent of income net of taxes and 14.1 percent of transfers, yielding a total of 19 percent. For the two other groups, post-tax income is mostly income net of taxes, but transfers play a role as well. In fact, the Top 10 percent receive more than twice their population share in terms of transfers (22.7 percent), while the Bottom 50 percent receive less than half of all transfers (42.1 percent).

This surprising result is explained by the way in which government transfers are allocated to individuals in Piketty, Saez, and Zucman (2018)'s analysis. Figure 5.8 breaks down government transfers into several underlying categories. Overall, these transfers amount to $5.072 B or

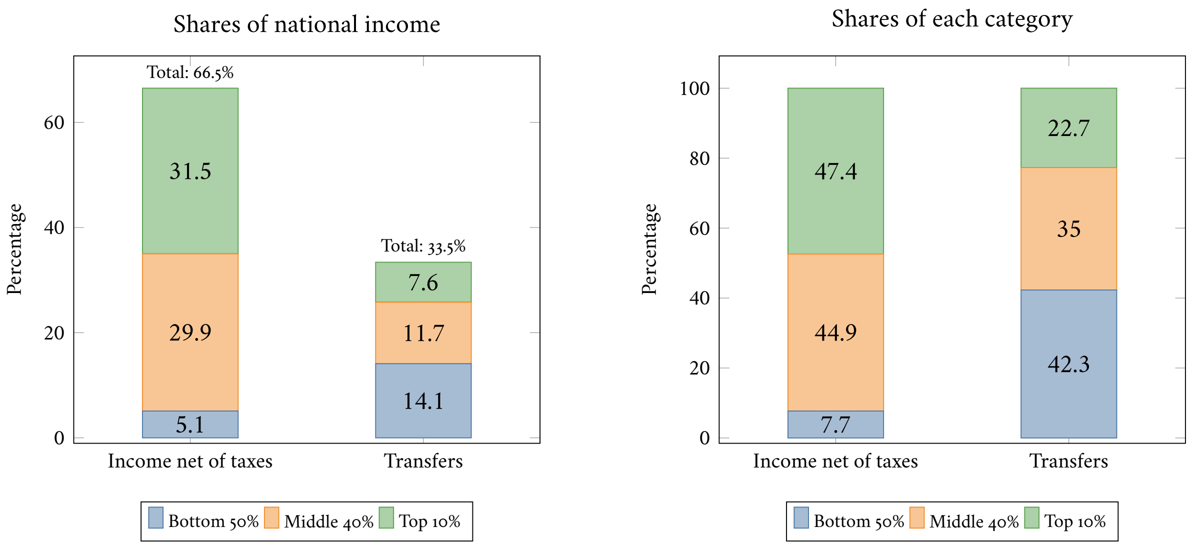


Figure 5.7: Decomposition of Post-tax Income

Notes: The figure shows a decomposition of post-tax income into income net of taxes and transfers. The left panel shows that income net of taxes makes up 66.5 percent of national income, while transfers make up the remaining 33.5 percent. The right panel decomposes both categories of national income. The figure shows, for example, that the transfers that accrue to the Bottom 50 percent represent 14.1 percent of national income and 42.3 percent of all transfers. *Source:* Own calculations for the United States in 2014 based on Piketty, Saez, and Zucman (2018), Appendix Table II-C2.

33.5 percent of total national income ($15.154 B) in 2014. Piketty, Saez, and Zucman treat about half of these transfers ($2.515 B) as individualized. This category in turn can be divided into cash transfers and in-kind transfers. The cash transfers are Social Security pension and non-pension (disability insurance, unemployment insurance), social assistance benefits in cash (refundable tax credits, veterans' benefits, workers' compensation, food stamps, supplemental security income, TANF/AFDC, and some smaller programs). These are assigned based on rules and on the recipient status observed in the Current Population Survey (CPS). Individualized in-kind transfers are mostly Medicare (assigned based on rules: age or receipt of disability insurance) and Medicaid (assigned based on the CPS). Note that some of these transfers (pension benefits, disability, and unemployment insurance) are already included in pre-tax income and are thus not counted towards as government redistribution in the definition of Piketty, Saez, and Zucman, which is limited to the difference between pre-tax and post-tax income.

The other half of government transfers ($2.558 B) represents in-kind transfers in three domains: education, defense, and a catch-all other category, which includes roads, public transportation and more generally the physical as well as legal and administrative infrastructure. These are items of government consumption expenditure. They represent goods and services and not a cash flow from the government to individuals. In accordance with the practice of national accounting, they are valued at the monetary cost of providing them (net of fees for

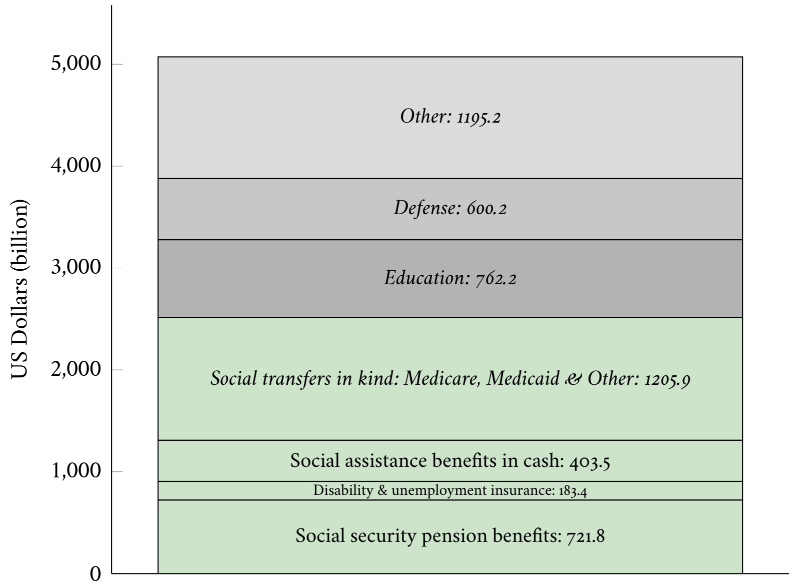


FIGURE 5.8: Categories of Government Expenditure

Notes: The figures shows the different categories that make up total government expenditure. Individualized transfers are shown in green, collective expenditure in gray. Transfers in kind are italicized, the remaining items are cash transfers. Social assistance in cash comprises refundable tax credits, SNAP, SSI, TANF/AFDC, and various smaller programs. *Source:* Own calculations for the United States in 2014 based on Piketty, Saez, and Zucman (2018), Appendix Table I-SA11.

their use), as opposed to the monetary equivalent of the benefit that individuals attach to them, which is much more difficult to measure. Citing the difficulty of observing who receives these goods and services (except for education), Piketty, Saez, and Zucman opt to allocate all of them proportionally to post-tax disposable income, which is pre-tax income minus taxes plus individualized monetary transfers.

This choice makes half of government spending distributionally neutral by assumption, and implies extremely unequal amounts of government consumption expenditure per capita (Figure 5.9a). As the 50 percent of adults with the lowest post-tax disposable income receive 18.0 percent of the total, they get assigned the same share of government consumption expenditure, which corresponds to less than $4 K per person and year. By contrast, each adult in the Top 10 percent is assumed to receive $45 K per year in terms of public spending on education, defense, public transportation, roads, and other infrastructure, despite more frequently using private-sector alternatives, at least for education and transportation. At the very top per capita values are even higher. The 0.01 percent with the highest incomes each receive more than $4 M per year.

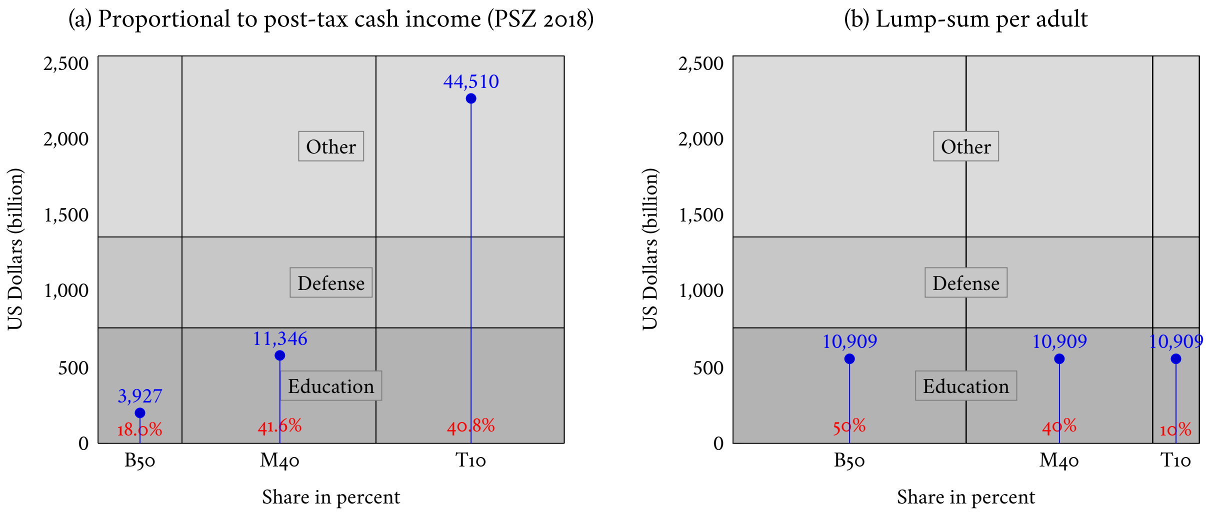


FIGURE 5.9: Comparison of Assumptions About Collective Expenditure

Notes: The figure contrasts Piketty, Saez, and Zucman (2018)'s assumption regarding the allocation of collective expenditure with the alternative of a lump-sum allocation. Piketty, Saez, and Zucman allocate collective expenditure proportionally to post-tax cash income (left panel). With this assumption, the Bottom 50 percent of the post-tax income distribution receive 18.0 percent of collective expenditure, while the Middle 40 percent receive 41.6 percent, and the Top 10 percent 40.8 percent. This implies a per-capita value of $3,927 in the bottom half of the distribution, compared with $44,510 among the Top 10 percent. The right panel shows an alternative assumption in which each adult receives the same share of collective expenditure, which corresponds to a per-capita value of $10,909. With this assumption, the share of collective expenditure that goes to the three groups is equal to their population share. *Source:* Own calculations for the United States in 2014 based on Piketty, Saez, and Zucman (2018), Appendix Tables I-SA11, II-C1b.

5.A.3 Consequences for Post-tax Income Shares

Levels With a lump-sum allocation, each adult gets assigned the same value of annual government consumption expenditure of $11 K (see Figure 5.9b). This assumption leads to a substantial change in the level of post-tax inequality. With a lump-sum allocation, the gap in the post-tax income shares between the Bottom 50 percent and the Top 10 percent is reduced by half in 2014. When each adult is allocated the same amount of government consumption expenditure, the share of the Bottom 50 percent is higher by about 5 percentage points, and the share of the Top 10 percent is reduced by about the same magnitude compared with an allocation that is proportional to post-tax disposable income (see Figure 5.14). As a result, the gap in the income shares of the two groups is reduced from about 20 to 10 percentage points. The share of the Middle 40 percent is almost unaffected. The effect of the allocation rule on the income shares is of the same order of magnitude as in the study by Blanchet, Chancel, and Gethin (2022) for a number of European countries.

Trends The sensitivity of the level of post-tax income inequality to the assumptions regarding the allocation of government consumption has not always been highlighted enough in the Dina literature[18] and motivates our analysis of how this expenditure (or parts thereof) is actually distributed. However, the key finding of Piketty, Saez, and Zucman, namely the sharp increase not only in pre-tax, but also post-tax inequality over the past four decades or so, also holds with a lump-sum allocation of government consumption expenditure.

As Figure 5.10 shows, replacing the proportional allocation with a lump-sum allocaton leads to a parallel shift in the series for the national income shares of the Bottom 50 percent and the Top 10 percent. With the lump-sum allocation, the series intersect both in the mid-1960s and the mid-1980s. However, given that the population shares of the two groups differ, an identical share of national income means that the average post-tax income of the Top 10 percent is five times larger than for the Bottom 50 percent. In 2014, the ratio of average incomes is 10.1 with a proportional allocation and 6.9 with a lump-sum allocation (Figure 5.15).

There are two reasons for the parallel shift. First, the share of collective expenditure in national income has been fairly stable between 15 and 20 percent over the period considered here. Second, while the income shares based on a proportional allocation merely reflect the trends observed for post-tax disposable income, the series for the lump-sum allocation is based on population shares that are time-constant by construction (Top 10 percent, Middle

[18] Piketty, Saez, and Zucman (2018) do run a robustness check in which they assign public education spending not proportionally to post-tax income, but as a function of the number of children. This check does not take into account the differences in per-capita expenditure by level of education (tertiary education is much more expensive per capita than primary and secondary education, at least in the United States) and, importantly, it does not capture public spending that goes to college students age 20 and above. They only report the consequences for the average income of the Bottom 50 percent and not the change in the income shares of all three groups.

40 percent, Bottom 50 percent) and thus cannot capture any real movements in the allocation of government consumption expenditure either. We do not believe that the allocation of these expenditure items has changed so much since 1980 that the divergence in post-tax income shares documented by Piketty, Saez, and Zucman would be altered substantially.

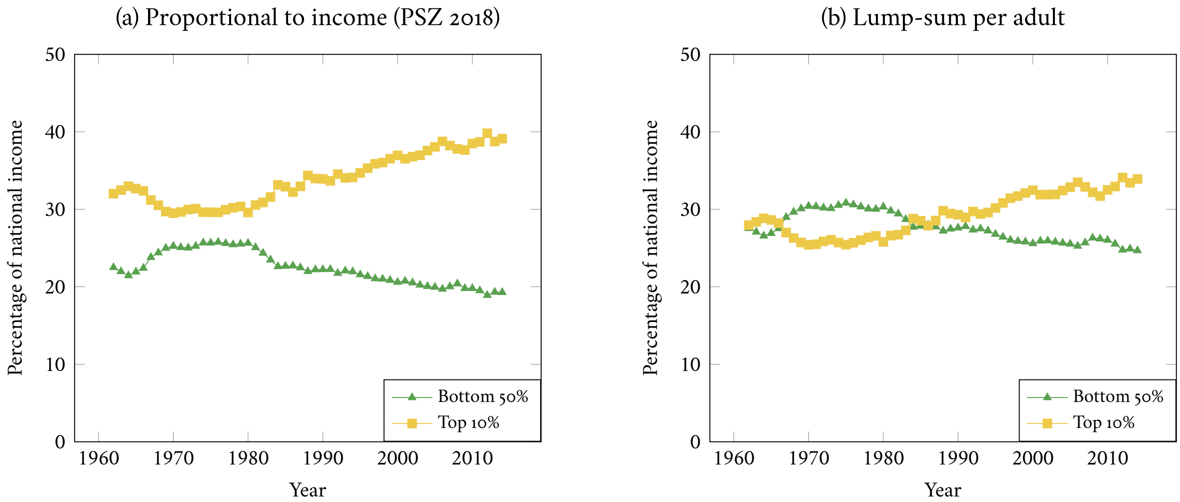


Figure 5.10: Effect of the Assumptions on Post-tax Income Shares, 1962–2014

Notes: The figure shows how the assumption regarding the allocation of collective expenditure affects the distribution of post-tax income in the United States over the years 1962–2014. Each panel shows the share of the Bottom 50 percent and the Top 10 percent. The left panel is for the assumption adopted by Piketty, Saez, and Zucman (2018), i.e. an allocation of collective expenditure that is proportional to post-tax cash income. The right panel shows the income shares that result from assuming a lump-sum allocation. *Source:* Own calculations based on Piketty, Saez, and Zucman (2018), Appendix Tables I-SA11, II-C1b, II-C2, II-C3b.

5.B Supplementary Analyses: PSID Data Linking Parents and Children

For some of the supplementary analyses, we draw on additional data from the Panel Study of Income Dynamics (PSID), a well-established panel study that began to survey 5,000 families in 1968 (McGonagle et al. 2012). As with the ACS, we use the 2017 wave. The PSID is much smaller than the ACS, but tracks individuals after they leave their original household, which allows us to link parents' and children's educational attainment in many cases.

The PSID provides information on the highest grade or year of school someone has completed and, if applicable, on the type of college degree (associate's, bachelor's, master's, PhD). Like the ACS, the PSID does not record complete educational histories. We therefore assume that a given degree implies that the individual has passed through all the stages below, that everyone needed the same number of years to complete each stage (see Table 5.6 in the Appendix for details), and that all education was received in the United States.

We use the PSID only for the intergenerational analysis in Section 5.3.2, where we focus on individuals aged 40–45 in 2017. As a check on the data, we compare summary statistics between the PSID and individuals from the same age group in the ACS (Table 5.3). The check is important because we can link information on education between parents and children for only about half of individuals in our age group. Reassuringly, the table shows that summary statistics for both samples are very close, which suggests that selection is not a major issue.

Table 5.3: Summary Statistics: Comparison of ACS and PSID, Individuals Aged 40–45

	ACS		PSID	
Age	42.5	(0.004)	42.5	(0.059)
Share Female (%)	50.9	(0.001)	48.1	(0.016)
Annual Labor Income ($)	51,143	(143)	51,007	(1,889)
High School Education (%)	23.8	(0.001)	26.5	(0.015)
Associate's Degree (%)	9.3	(0.001)	9.7	(0.010)
Bsc. Degree (%)	21.4	(0.001)	20.5	(0.013)
Msc. Degree (%)	10.5	(0.001)	10.3	(0.010)
Total Education Transfers ($)	261,440	(172)	269,224	(2,055)
N	216,278		925	
N, weighted	23.787 M		11.962 M	

Notes: The table compares means (and standard errors in parentheses) of some key variables for individuals aged 40–45 across the ACS and the PSID data. The ACS data are used in Figure 5.4, the PSID data are used in Figure 5.5. The difference in the number of weighted observations is due to missing values for parental education in the PSID. Without conditioning on education information for at least one parent being present, the PSID has 1,770 observations and 23.858 M weighted observations, very close to the ACS number.

5.C Additional Tables and Figures

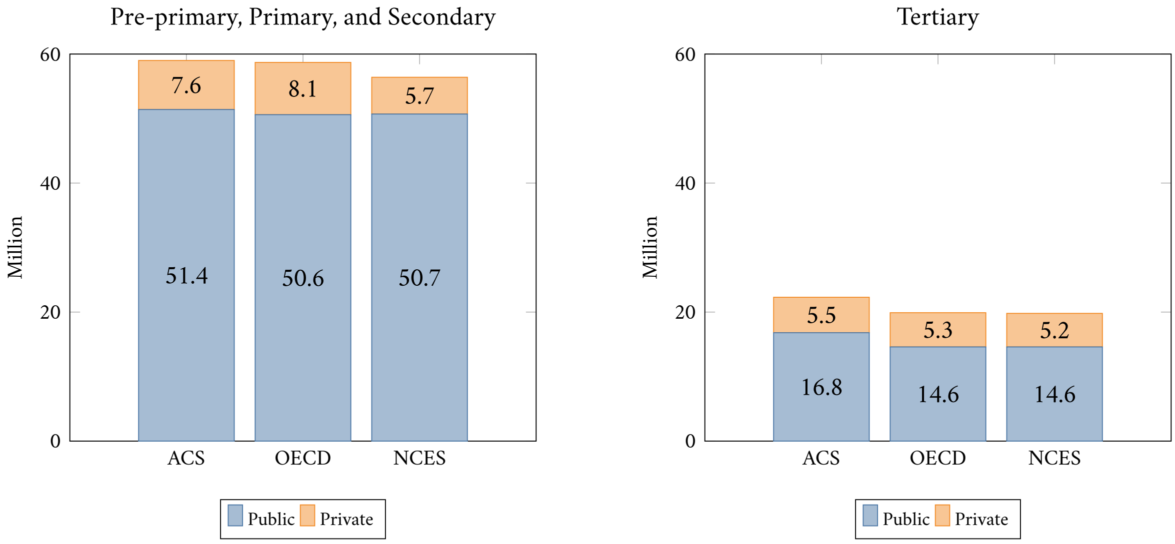


FIGURE 5.11: Enrollment in Educational Institutions, United States 2017

Notes: The figure compares our ACS-based numbers for the enrollment in educational institutions in the United States in 2017 with statistics published by the OECD and the National Center for Education Statistics (NCES). The left panel shows the number of students enrolled in pre-primary, primary, or secondary education, the right panel is for tertiary education. A distinction is made between public and private institutions. *Source:* Own calculations based on the American Community Survey 2017. OECD: Education at a Glance 2020 (OECD Statistics 2020), Table: Enrollment data adjusted to the financial year. Sum of students in full-time and part-time education. Part-time is only non-zero at the pre-primary and the tertiary levels. Students in post-secondary non-tertiary education not included (110 K are enrolled in public institutions, 273 K in private institutions). NCES: National Center for Education Statistics, Digest of Education Statistics 2019 (De Brey et al. 2021), Table 105.30: Enrollment in elementary, secondary, and degree-granting postsecondary institutions, by level and control of institution: Selected years, 1869-70 through fall 2029.

TABLE 5.4: Public Education Spending by Income: Detailed Results

	Income Decile									
	1	2	3	4	5	6	7	8	9	10
A. Pre-tax Income (Adults)										
Pre-primary	179	165	169	153	146	141	144	152	136	114
Primary	1,209	1,214	1,230	1,144	1,098	1,118	1,165	1,218	1,245	1,237
Secondary	1,386	1,447	1,514	1,505	1,504	1,542	1,635	1,758	1,920	1,913
Tertiary	3,253	1,927	1,852	1,821	1,817	1,812	1,716	1,534	1,379	992
Total	6,027	4,752	4,765	4,623	4,564	4,613	4,660	4,661	4,680	4,256
B. Post-tax Cash Income (Adults)										
Pre-primary	138	141	162	174	164	157	153	163	138	110
Primary	935	978	1,189	1,277	1,217	1,231	1,254	1,294	1,293	1,209
Secondary	1,102	1,248	1,430	1,604	1,591	1,678	1,709	1,877	1,953	1,926
Tertiary	3,402	2,046	1,889	1,834	1,798	1,742	1,652	1,468	1,312	976
Total	5,577	4,413	4,670	4,888	4,769	4,808	4,768	4,802	4,696	4,221
C. Equivalized Pre-tax Income (Households)										
Pre-primary	379	409	378	333	306	271	241	189	154	124
Primary	2,701	3,074	2,874	2,627	2,222	2,176	1,923	1,716	1,481	1,290
Secondary	3,254	3,890	3,890	3,495	3,275	3,152	2,691	2,428	2,049	1,723
Tertiary	6,771	3,447	3,476	3,524	3,455	3,462	3,147	2,953	2,566	1,686
Total	13,106	10,820	10,618	9,978	9,258	9,061	8,002	7,286	6,250	4,822

Notes: The table shows how public education spending in the United States in 2017 is distributed among the deciles of the income distribution. All values in 2017 US Dollars. For the sake of presentation and given the large sample size, standard errors are omitted. The deciles are based on pre-tax income (panel A), post-tax cash income (panel B), and equivalized pre-tax household income (panel C). The same information is presented in graphical form in Figure 5.1 in the main text and Figures 5.12 and 5.13 in the Appendix. *Source:* Enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is summed up at the household level. In panels A and B, the resulting sum is split equally among adults aged 20 and above in the household, and household income is likewise split equally among all adults. Panel C reports public education spending at the household level instead, and deciles are based on equivalized pre-tax household income, using the modified OECD equivalence scale, which assigns a value of 1 to the first adult in the household, of 0.5 to each additional household member aged 14 and above, and of 0.3 to each child below the age of 14. Pre-tax income is directly taken from the American Community Survey, while post-tax cash income is simulated using TAXSIM v32.

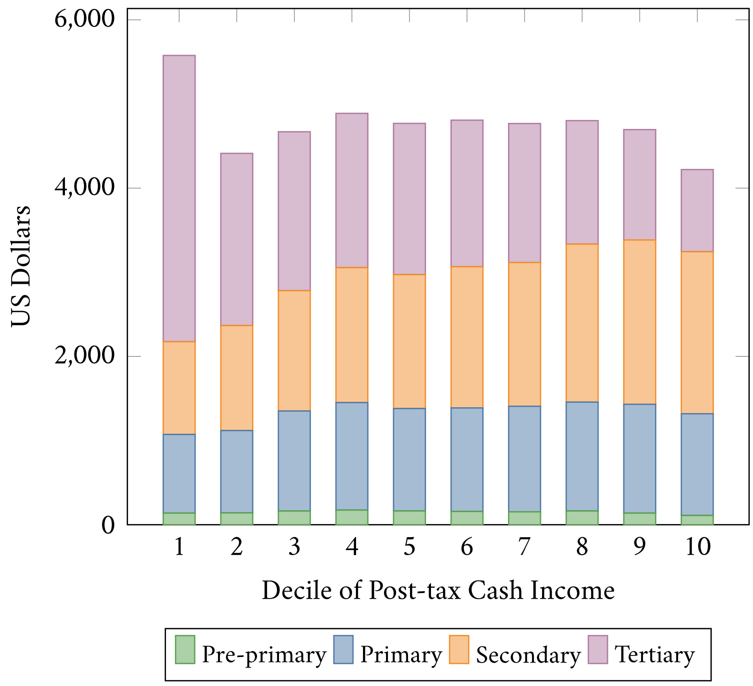


FIGURE 5.12: Public Education Spending by Post-tax Cash Income, Allocated Based on Actual Enrollment

Notes: The figure shows how public education spending in the United States in 2017 is distributed among the deciles of the post-tax cash income distribution. For each decile, the bars show the average values of annual public education spending (in 2017 US Dollars) at the pre-primary, primary, secondary, and tertiary levels of education. *Source:* Enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is summed up at the household level, and the resulting sum is split equally among adults aged 20 and above in the household. Household post-tax cash income is simulated using TAXSIM v32, and is likewise split equally among all adults.

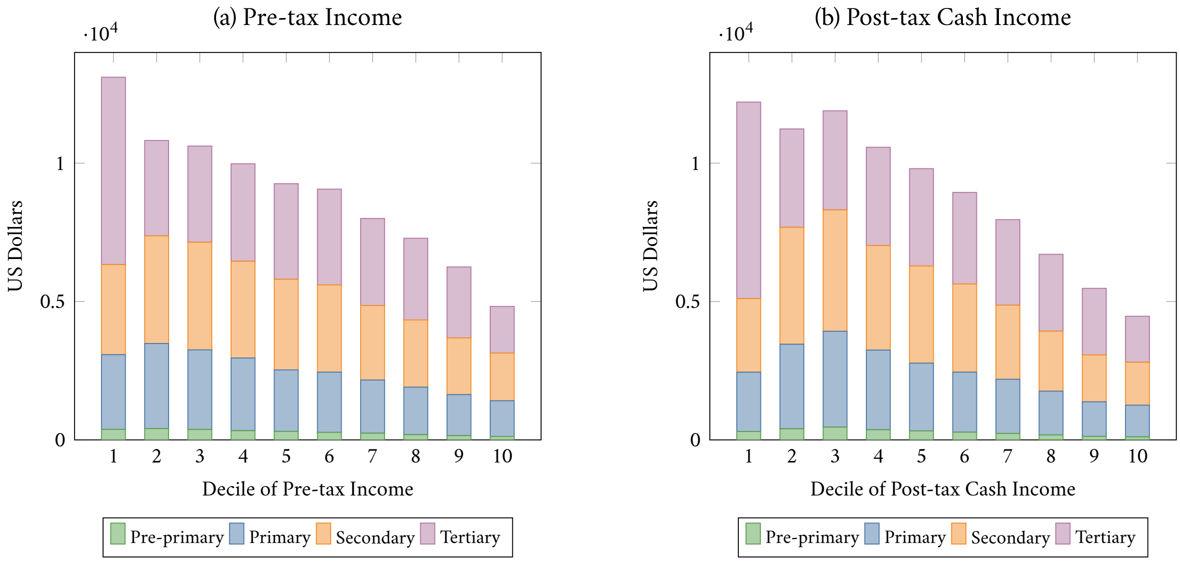


FIGURE 5.13: Public Education Spending by Equivalized Household Income, Allocated Based on Actual Enrollment

Notes: The figure shows how public education spending in the United States in 2017 is distributed among households over deciles of the equivalized household income distribution. Left panel: deciles based on pre-tax income. Right panel: deciles based on post-tax cash income simulated using TAXSIM v32. For each decile, the bars show the average values of annual public education spending (in 2017 US Dollars) at the pre-primary, primary, secondary, and tertiary levels of education. *Source:* Enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is then summed up at the household level. Household income is equivalized using the modified OECD equivalence scale, which assigns a value of 1 to the first adult in the household, of 0.5 to each additional household member aged 14 and above, and of 0.3 to each child below the age of 14.

TABLE 5.5: Robustness Checks

	Decile of Pre-tax Income									
	1	2	3	4	5	6	7	8	9	10
Main specification	6.0	4.8	4.8	4.6	4.6	4.6	4.7	4.7	4.7	4.3
Drop if > 50% of income imputed	5.9	4.7	4.8	4.6	4.5	4.6	4.7	4.7	4.6	4.2
Drop if income negative	6.0	4.7	4.8	4.6	4.6	4.6	4.7	4.7	4.7	4.3
Drop if income < 1%	6.0	4.7	4.8	4.6	4.6	4.6	4.7	4.7	4.7	4.3
Drop if income < 2.5%	5.4	4.7	4.8	4.6	4.6	4.6	4.7	4.7	4.7	4.3
Drop if income > 99.5%	6.0	4.8	4.8	4.6	4.6	4.6	4.6	4.7	4.7	4.3
Enrollment as in NCES	5.5	4.4	4.4	4.3	4.2	4.3	4.3	4.4	4.4	4.0
Variation across states	5.6	4.4	4.5	4.4	4.3	4.3	4.4	4.5	4.6	4.4
Full-time equivalents	4.6	3.8	3.8	3.7	3.7	3.7	3.8	3.8	3.9	3.6
2/4-year college	6.8	5.2	5.2	5.1	5.0	5.1	5.1	5.1	5.1	4.6
Current expenditure	5.5	4.3	4.3	4.2	4.1	4.2	4.2	4.2	4.2	3.9
All expenditure	7.2	5.5	5.5	5.4	5.3	5.3	5.4	5.3	5.3	4.8

Notes: The table summarizes the results of our robustness checks. For comparison, results for the main specification are shown in the first row as well. The amounts reported in the table are annual public education transfers in thousand US Dollars. *Source:* In the main specification, enrollment in public educational institutions is taken from the American Community Survey 2017. Each pupil or student is assigned the per-capita value of public education spending taken from the OECD (see Table 5.2). Public education expenditure is summed up at the household level, and the resulting sum is split equally among adults aged 20 and above in the household. Household income is likewise split equally among all adults. The robustness checks modify the measurement of enrollment, of per-capita expenditure, or of the household income that enters the computation of the deciles. For details, see Section 5.3.2.

Table 5.6: Construction of Educational Trajectories

	Years Spent at ISCED Level									
Highest Degree (ACS)	0	1	2	3	4	5	6	7	8	Total
No schooling completed	0	0	0	0	0	0	0	0	0	0
ISCED 0										
Nursery school, preschool	2	0	0	0	0	0	0	0	0	2
ISCED 1										
Kindergarten	2	1	0	0	0	0	0	0	0	3
Grade 1	2	2	0	0	0	0	0	0	0	4
Grade 2	2	3	0	0	0	0	0	0	0	5
Grade 3	2	4	0	0	0	0	0	0	0	6
Grade 4	2	5	0	0	0	0	0	0	0	7
Grade 5	2	6	0	0	0	0	0	0	0	8
ISCED 2										
Grade 6	2	6	1	0	0	0	0	0	0	9
Grade 7	2	6	2	0	0	0	0	0	0	10
Grade 8	2	6	3	0	0	0	0	0	0	11
ISCED 3										
Grade 9	2	6	3	1	0	0	0	0	0	12
Grade 10	2	6	3	2	0	0	0	0	0	13
Grade 11	2	6	3	3	0	0	0	0	0	14
12th grade, no diploma	2	6	3	3	0	0	0	0	0	14
Regular high school diploma	2	6	3	4	0	0	0	0	0	15
GED or alternative credential	2	6	3	4	0	0	0	0	0	15
Some college, but less than 1 year	2	6	3	5	0	0	0	0	0	16
ISCED 4										
Associate's degree, type not specified	2	6	3	4	2	0	0	0	0	17
ISCED 5										
1 or more years of college credit, no degree	2	6	3	4	0	2	0	0	0	17
ISCED 6										
Bachelor's degree	2	6	3	4	0	0	4	0	0	19
ISCED 7										
Master's degree	2	6	3	4	0	0	4	2	0	21
Professional degree beyond a bachelor's degree	2	6	3	4	0	0	4	2	0	21
ISCED 8										
Doctoral degree	2	6	3	4	0	0	4	2	4	25

Notes: The table documents how we map the information on the highest degree in the American Community Survey (ACS) 2017 into educational trajectories. The rows correspond to the values of the variable "Highest degree" (educd) in the ACS. The question reads: "What is the highest degree or level of school this person has completed?". As our method is retrospective and we do not have information on grade repetition or, more generally, the individual pathways to a given degree, we assign the same number of years to all individuals with the same degree. For instance, individuals with a regular high school diploma are assumed to have spent two years at ISCED level 0, six years at ISCED level 1, three years at ISCED level 2, and four years at ISCED level 2. Individuals with a bachelor's degree are assigned the same trajectory plus four years at ISCED level 6, and a master's degree would add two years at ISCED level 7. The last column of the table gives the total number of years thus obtained. The number is meant as a summary measure only. When computing the public expenditure for each degree, we multiply the number of years *at each ISCED level* with the corresponding OECD per-student expenditure from Table 5.2.

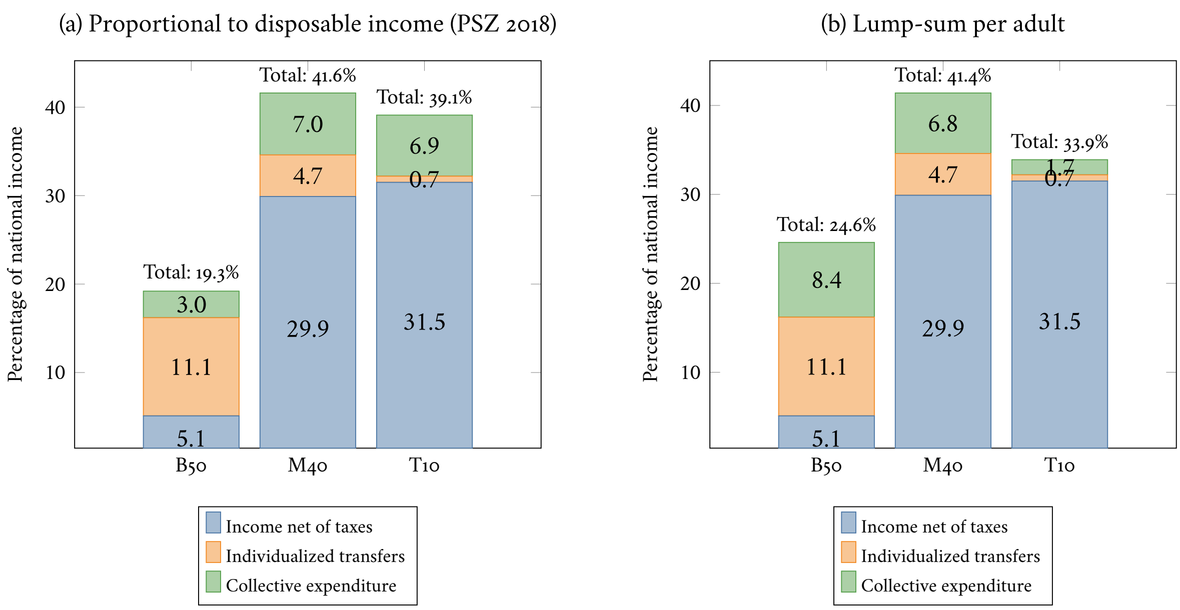


Figure 5.14: Effects of the Assumptions on the Distribution of Post-tax Income

Notes: The figure shows how the assumption regarding the allocation of collective expenditure affects the distribution of post-tax income. When collective expenditure is allocated based on post-tax cash income as in Piketty, Saez, and Zucman (2018), the Bottom 50 percent receive 19.3 percent of national post-tax income, while the Middle 40 percent receive 41.6 percent, and the Top 10 percent receive 39.1 percent (left panel). Under the alternative assumption in which each adult receives the same amount of collective expenditure, the shares are 24.6 percent, 41.4 percent, and 33.9 percent instead (right panel). *Source:* Own calculations for the United States in 2014 based on Piketty, Saez, and Zucman (2018), Appendix Tables I-SA11, II-C1b, II-C2, II-C3b.

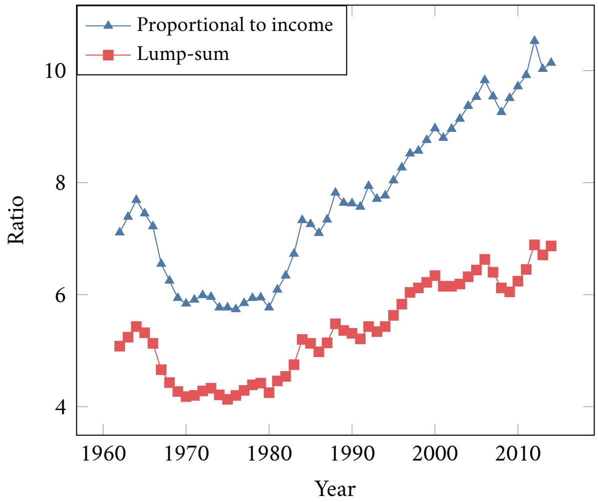


Figure 5.15: Effect of the Allocation Rules on the Ratio of Average Incomes, 1962–2014

Notes: The figure shows how the assumption regarding the allocation of collective expenditure affects the ratio of average post-tax incomes of the Bottom Top 10 percent and the Bottom 50 percent in the United States over the years 1962–2014. The blue graph is for the assumption adopted by Piketty, Saez, and Zucman (2018), i.e. an allocation of collective expenditure that is proportional to post-tax cash income. The red graph shows the ratio that results from assuming a lump-sum allocation. *Source:* Own calculations based on Piketty, Saez, and Zucman (2018): Appendix Tables I-SA11, II-C1b, II-C2, II-C3b.