Women's Career Dynamics and the Effects of Family Policy: Structural and Quasi-Experimental Evidence

Fabian Martin Stürmer-Heiber

Munich 2022

Women's Career Dynamics and the Effects of Family Policy: Structural and Quasi-Experimental Evidence

INAUGURAL-DISSERTATION

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2022

vorgelegt von

Fabian Martin Stürmer-Heiber

Referent:	Prof. Dr. Joachim Winter
Korreferent:	Prof. Dr. Dominik Sachs
Mündliche Prüfung:	24. Januar 2022
Berichterstatter:	Prof. Dr. Joachim Winter
	Prof. Dr. Dominik Sachs
	Prof. Dr. Andreas Peichl
Promotionsabschlussberatung:	02. Februar 2022

To Jana

Acknowledgments

Writing this dissertation has been a fun but challenging endeavor, that benefited greatly from the support and interaction with my outstanding colleagues and friends at LMU.

First and foremost, I want to thank Joachim Winter, my first year mentor, supervisor, and boss for the past four years. I am deeply grateful to him for the opportunity to pursue my PhD at the Chair of Empirical Economic Research. Throughout the years, he granted me the freedom and support to develop and pursue my research agenda, while continuously guiding me with valuable feedback that shaped my path forward.

I am equally grateful to my second supervisor and coauthor, Dominik Sachs. He took me on board at a time when I was struggling with my own ideas a bit and allowed me to directly leverage my skills in our joint project. This gave me the opportunity to benefit from the breadth of know-how in our team and get exposure at renowned conferences. Furthermore, he has been a great counterpart to bounce ideas, pitches, and solve puzzles, while always steering me into productive directions.

A special thanks goes to Andreas Peichl for his excellent feedback and encouragement as part of the public economics group as well as for agreeing to be part of my dissertation committee.

Turning to my coauthors, all of them deserve special credit for this dissertation. The collaboration with David Koll on Chapter 1 has been my most productive teamwork experience yet. Our shared passion for turning rough ideas quickly into code and convincing results has been the fuel to push each other and the project further and further. Hélène Turon kept providing us with the necessary guidance and continued to challenge our ideas to keep us on a productive path. Moreover, I want to thank Ulrich Schneider for the collaboration on Chapter 3. Our lively discussions gave me the clarity and courage to tackle the project's challenges successfully and round it off neatly. It was a tremendous joy to work with all of you and our collaboration enriched my journey substantially.

Furthermore, I want to thank the entire team of the Chair of Empirical Economic Research: Sharing an office with Brendan and Pavel, two coding aficionados like me, has been a true pleasure, as it has been to work and teach alongside with the other PhDs Corinna, Peter, Sebastian, Luisa, Michael, Alexandre, Christoph, and Tobias. I am also thankful to the junior faculty members Greg, Andreas, and Sarah for providing me with a breadth of comments and feedback on my projects. Ines sustained support on the administrative side has been outstanding. It was great to be part of such a fun and encouraging group.

Additionally, I want to thank my MGSE colleagues and the members of the econometrics as well public economics group for all the encouraging conversations. The room 17 group of Vera, Leo, Johannes, Sebastian, Manfei, and Schanzah was the best starting ground I could have asked for. Outside of Munich, Lukas and Ruben have been two extremely valuable voices of encouragement and feedback that have helped me to navigate through academia.

I greatly appreciate generous support from the Elite Network of Bavaria via the Evidence-Based Economics program. Additionally, Chapters 2 and 3 would not have been possible without the excellent support from the team at the research data center of the Federal Employment Agency (FDZ-IAB).

Finally, my greatest gratitude goes to my family. My parents, grandparents, and siblings have been the backbone of support and guidance throughout this journey. However, none of it would have been possible without my partner, Jana, who has always had my back and kept me going with her unwavering support and encouragement. I cannot express how grateful I am for having her by my side over the years.

Fabian Martin Stürmer-Heiber
Munich
September 2021

Contents

List of Figures

List of Tables

Preface

The past century was characterized by the rapid rise of the role of women in society and the economy. Progressive family and labor market policies have been crucial components of this 'grand gender convergence', as Goldin (2014) frames it. As a consequence, the compatibility of having a family and a career has substantially improved. However, the event of childbirth continues to mark a significant point of divergence between women's and men's careers.

Alongside of potential shifts in preferences, the necessity of some form of childcare implies a large change in the time or budget constraints of households. The consequence is in most cases that mothers reduce their labor supply to some degree when young children are around. This reduction in working hours is one of the main drivers of the divergence in career development and the sizeable earnings losses that can be observed for women after childbirth (Kleven, Landais, Posch, et al. 2019). Furthermore, the reduction in labor supply has long-lasting effects on future earnings, as human capital accumulation is slowed or halted.

To be able to design effective family policies that can close the gender gap further, it is crucial to thoroughly understand the just described constraints and long-run effects. In this dissertation, I therefore pursue three different angles of women's career dynamics before and after childbirth, that highlight the effects of current family policies and point to alternative directions for them.

The first chapter focuses on the phase when young children are present and evaluates the subsidization of public childcare. The second chapter shifts the focus onto young childless women and maps out how their career development is affected by family policies directed at mothers. The third chapter provides a more general take on a labor supply choice that is very common among women with children of all ages, namely part-time work, and the division of labor among spouses.

In terms of methodology, the family and labor economics literature draws upon two main strands: quasi-experimental methods and structural methods. Quasi-experimental methods leverage as-good-as-random variation in the implementation of policies to identify causal effects on the outcomes of interest. Additionally, these methods also allow to exploit that policies may have altered the incentive structures of endogenous choices in a plausibly exogenous fashion. This allows researchers to break the endogeneity and estimate the causal effect of a particular choice on a set of outcomes. Overall, these methods enable the identification of the effects of interest without making explicit assumptions on the structure of the decision problem, as identification stems from variation that is orthogonal to the decision factors. However, quasi-experimental methods are naturally limited in their scope by the scope of the observed policies and cannot yield insights on unobserved alternative policies.

Structural methods, on the other hand, allow to construct such counterfactuals and conduct simulations with a broad array of policy variants. They do require to fully specify the decision problem with all constraints and how its specifics are governed by 'structural (preference) parameters'. These parameters are then estimated based on the observed choices in the data. Under the assumption that the structural parameters are invariant to changes in the policy environment, the estimates then allow to simulate individuals' behavior under alternative policies.

Throughout this dissertation, I combine both methods to establish credible evidence on the causal effects of different choices and policies. Specifically, the first chapter estimates a structural life cycle model, whereas the second and third chapter exploit historical policy variation to identify the effects of interest.

In terms of data, past research has relied mostly on survey data. Long running panel studies such as the German Socio-Economic Panel continue to be invaluable sources of life cycle trajectories combined with rich background information. Therefore, they are especially useful to estimate structural models such as in Chapter 1, as they allow researchers to incorporate, for example, the effects of heterogeneous family backgrounds on choices over the life cycle.

In recent years, a second source of data has gained increasing popularity for labor market research: administrative data from government sources. These data sets are constructed, for example, from tax and pension registries or reports to supervisory government bodies. They are usually characterized by a large number of observations and high information quality, as the data collection process typically penalizes misreporting. This makes them especially attractive for studies that focus on sensitive information such as wages, which are likely to be reported with measurement error in survey studies.[1] Chapters 2 and 3 fall into this category and are conducted with data from the German Federal Employment Agency. Nevertheless, these data sets often contain only limited background information as such are not part of the data collection processes. In cases where rich background data is necessary for the analysis, the combination of survey and administrative data through record linkage can be a fruitful strategy. Such record linkage, which is showcased in the third chapter, is however rarely possible due to data protection regulation and a lack of common identifiers across data sets.

The remainder of this section provides an overview of the three chapters in this dissertation. Each chapter is a self-contained paper with supplementary material provided in a corresponding appendix. A consolidated bibliography for all three chapters is provided at the end.

Chapter 1, which is joint work with David Koll, Dominik Sachs, and Hélène Turon, focuses on the subsidization of public childcare. As subsidized public childcare improves the compatibility of having a family and a career, it increases maternal life cycle earnings and tax payments. This generates a return for the government that (partially) offsets the cost of subsidization. Driven by this perspective, we address the following questions: How much should childcare be subsidized given this fiscal externality? How should subsidies vary with income? We estimate a dynamic discrete choice model of female labor supply and childcare decisions on German panel data. The model incorporates a large amount of heterogeneity: beyond heterogeneous preferences for domestic childcare and leisure, and in terms of education levels, wages, as well as informal childcare availability, we also account for heterogeneity in the timing of birth(s) and number of children.

We use the model to evaluate a recent expansion of public childcare for 0 – 2 year olds that effectively ended the previous rationing of childcare slots. We find that this policy fully paid for itself even though the childcare slots are subsidized at 80% on average, making a strong case for the policy as it was essentially implemented at no cost. But does this also justify that childcare slots are subsidized at such a high rate? We find that the rate of subsidization is likely too high because a small untargeted increase barely pays for itself: only 5.9% of the additional governmental expenses are recovered through

[1] This measurement error may stem for example from individuals unwillingness to disclose their actual wages, from rounding, or from individuals not recalling their precise wages.

increased tax revenues. Hence, the very high subsidies cannot be solely grounded on the fiscal externality argument.

We then turn to the redistribution that is entailed in the progressive childcare fee schedule and show that high subsidies for high-income families can barely be grounded on efficiency rationales. Specifically, we compare the childcare fee schedule to the income tax schedule in terms of each schedules' efficiency cost of redistribution. We find that making the childcare fee schedule more progressive would imply efficiency costs that are one-third lower than what society currently pays for redistribution through the tax system. In other words, if society would like to solve the trade-off between equity and efficiency for the childcare-using population according to the same standards as it does for the general population, childcare fees should be more progressive.

Chapter 2 focuses on the early career development of women. I first use German administrative social security data to map out patterns in the promotion rates of women prior to first birth, i.e., *potential* mothers, and compare them to young men. I document large differences in promotion rates between occupations and find that, after controlling for selection into occupations and firms, young men are promoted at significantly higher rates than young women. The estimates of the gender gap in promotions range from 1.6 to 2.8 percentage points, which translates into young women being promoted 10% to 20% less often.

I then use reform-driven childcare policy variation to study how expected future labor supply patterns affect the career progress of young women. Specifically, I exploit quasi-random variation induced by a staggered expansion of public childcare, a policy aimed at mothers, to identify the causal effect of public childcare availability on women's pre-birth promotion probabilities. I find that a 10 percentage point increase in public childcare availability improves the promotion probability of young women by 1.7 percentage points, which translates into a 10% increase.

With the observed response to the policy variation as a lens into the employer-employee interaction, the results show that expected future labor supply plays an important role for early career progress. This underlines that statistical discrimination is a plausible driver of the early career gender gap in promotion rates. Additionally, the results show that childcare policy has an important impact that goes beyond the well-documented effect of increasing mothers' labor market attachment: My findings are consistent with public childcare effectively reducing leave-related costs for employers and thereby boosting women's pre-birth career prospects through a reduction of statistical discrimination.

Chapter 3, which is joint work with Ulrich Schneider, focuses on part-time wage penalties. The very unequal distribution of working hours among spouses is a major contributor to large earnings losses for women after childbirth. Therefore, we investigate how various working hours choices affect wage levels and growth rates. This perspective sheds light on the earnings implications of a more equal distribution of working hours among spouses. To conduct this investigation, we leverage a combination of rich administrative data on earnings with survey data on hours to construct a high-quality panel of hourly wages for females in Germany. We control for selection into different hour categories by exploiting numerous reforms of the German tax and transfer system that varied in their impact on different earnings.

Our selection-corrected estimates of the part-time penalty in hourly wages show substantial heterogeneity, ranging from -2% to -18% compared to full-time. The heterogeneity is especially driven by large penalties for working low hours in part-time (≤ 16h), for which penalties are more than twice as large as for higher hours. Furthermore, we show that working high part-time hours (> 24 to ≤ 34h) does not only imply sizeable penalties as well, but can even carry higher ones than medium part-time hours (> 16 to ≤ 24h). In terms of yearly wage growth rates, we find selection-corrected penalties ranging between -0.7 and -1.9 percentage points compared to an average wage growth rate in full-time of 3.30%. These penalties are again the largest for working only a few hours, and decrease with increasing working hours, but not in a linear fashion. Wage growth penalties remain significant even for working many hours within part-time, which highlights that there are hours-independent elements in the firm's cost function of part-time.

In conclusion, our findings of sizeable penalties for high part-time hours indicate that moving to a more equal division of labor among spouses is quite costly: It would imply a net earnings loss and potentially moving to a lower wage growth trajectory. Part-time penalties are therefore creating sizeable disincentives to split working hours equally among spouses.

Chapter 1

Equity and Efficiency of Childcare Subsidies: A Structural Approach*

Abstract: Childcare policies improve the compatibility of family & career and can increase maternal life cycle earnings & tax payments. How much should childcare be subsidized given this fiscal externality? How should subsidies vary with income? We estimate a dynamic discrete choice model of labor supply and childcare decisions of heterogeneous families with German panel data. We evaluate a recent expansion of public childcare slots and find that this program paid for itself through the fiscal externality. Increasing subsidies further by marginally lowering fees per slot would only be 6% self-financing. Increasing redistribution with more progressive subsidies comes at low efficiency costs.

*This chapter is based on joint work with David Koll, Dominik Sachs, and Hélène Turon.

1.1 Introduction

Numerous studies have evaluated early childcare programs with respect to their net fiscal cost. García et al. (2020), for example, evaluate programs targeted at low-income families taking into account, e.g., children's and parents' earnings.[1] They find that the long-run effects on earnings imply increases in net tax revenue that exceed the direct costs of the programs. Consistent with that, Hendren and Sprung-Keyser (2020) document that most targeted childcare programs in the US were also fully self-financing. For universal childcare programs, which also entail substantial subsidies to high-income families, much lower self-financing rates have been documented.[2]

A key limitation of such reduced form studies is, however, that they can only evaluate the impact of actual policies. A structural approach, on the other hand, allows to experiment with different features of counterfactual childcare policies and address questions like: At what rate should public childcare be subsidized? How should the rate of subsidization vary with household income?

In this paper, we use a structurally estimated dynamic household model to provide answers to such policy questions. We consider a nationwide public childcare program in Germany and focus on the effect on maternal earnings over the life cycle. First, we evaluate a recent expansion of public childcare for 0 – 2 year olds that effectively ended the previous rationing of childcare slots. We find that this creation of new slots in a situation of excess demand fully paid for itself despite the slots being subsidized at a rate of 80% on average. This makes a strong case for the policy as it was effectively implemented at no cost. But does this also justify that public childcare slots are subsidized at such a high rate in the non-rationed status quo? We find that the rate of subsidization is likely too high, because a small untargeted subsidy increase from the current level barely pays for itself: only 5.9% of the additional government expenses are recovered through increased tax revenues. Hence, the very high subsidies cannot be solely grounded on the fiscal externality argument through maternal labor supply.

We then turn to the redistribution that is entailed in the progressive childcare fee schedule: subsidies decrease with increasing household income to a minimum of around 50%. We show that the high subsidies for high-income families can barely be grounded

[1] Specifically, they focus on the Carolina Abecedarian Project and the Carolina Approach to Responsive Education.

[2] Baker, Gruber, and Milligan (2008, 2019) analyze such a universal childcare program in Quebec. They find positive effects on parental earnings, but the implied tax revenue increase amounted to (only) 40% of the program costs. However, they do not take potential long-run effects on parental earnings into account.

on efficiency rationales. Specifically, we compare the childcare fee schedule to the income tax schedule in terms of each schedules' efficiency cost of redistribution. We find that making the childcare fee schedule more progressive would imply efficiency costs that are one-third lower than what society currently pays for redistribution through the tax schedule. In other words, if society would like to solve the trade-off between equity and efficiency for the childcare-using population according to the same standards as it does for the general population, childcare fees should be more progressive. The current system would, however, be optimal if societal preferences for redistribution from high to low-income households were significantly lower for the childcare-using population.

Model. To arrive at these results, we set up a dynamic discrete choice model with unitary households that is tailored to the German context: In particular, we consider a model with rich heterogeneity in the timing and spacing of births of up to three children across 3-year model periods. This allows us to explicitly distinguish between public childcare for 0 – 2 and 3 – 5 year olds as well as after school care for 6 – 8 year olds. Households decide how to provide care for their children and whether the mother works full-time, part-time, or does not work at all, while fathers always work full-time. Labor supply decisions affect future wages, which allows us to capture the career penalties for working less than full-time. Children can be cared for by the mother at home, through the use of informal childcare by, e.g., grandparents, or through the use of public childcare services.

A distinct feature of our model is the large amount of heterogeneity. Households differ with respect to their education, wages, and family composition. They also differ in three unobserved dimensions: their preference for domestic childcare, their taste for leisure of the female spouse, and their access to free informal childcare. Accounting for this heterogeneity is crucial because it allows us to capture the heterogeneity in households' responses to changes in childcare policies. For example, mothers with three children respond differently from mothers with just one child. Furthermore, for mothers with high wages, the incentives to return to the labor market soon after childbirth are higher than for mothers with low wages. On the other hand, mothers with low wages may be more responsive to changes in the childcare fee schedule. Conditional on the observable differences, the way mothers respond to policies also depends on unobserved characteristics, which our structural approach allows us to capture.

Estimation. The estimation of the model is carried out predominantly with panel data from the German Socio-Economic Panel. In a first step, we estimate reduced form relationships: i) how childcare fees vary with income and family structure, ii) a stochastic fertility process conditional on age and education, and iii) Mincerian wage equations that account for dynamic wage penalties for working part-time or staying out of the labor market.

For the second part of the estimation, we use the explicit structure of the dynamic model. We apply a maximum likelihood approach and account for measurement errors in public childcare hours and wages. We estimate the joint distribution of the unobserved preferences for domestic childcare and for female leisure, as well as the access to informal childcare. We show that the estimated model fits a number of empirical moments well and yields reasonable participation and hours elasticities.

Based on the estimated joint distribution of unobserved heterogeneity, we simulate long-term behavioral responses to changes in childcare policies and evaluate the implied equity-efficiency trade-off as well as the implications for public finances.

Quantitative policy analysis. First, we evaluate a recent public childcare expansion for 0 – 2 year olds in Germany.[3] Our results show that the increase in publicly provided childcare was 103% self-financing. One-third of the fiscal externality can be attributed to dynamic career effects. For 45% of the children who got enrolled, mothers increased their labor supply – a number that is comparable to quasi-experimental evidence for a comparable reform in the 1990s (Bauernschuster and Schlotter 2015). This implies that the average mother who started to work due to the policy paid taxes and social security contributions that are around twice the average subsidy paid per childcare slot. The fiscal effects are large because of the high marginal tax rates that these women face. This is due to the joint taxation of married couples in Germany, see, e.g., Bick and Fuchs-Schündeln (2018), which results in marginal tax rates around 50% for the majority of our sample.

Second, we quantify the fiscal externality of a marginal increase in subsidies to evaluate to what extent the argument that childcare subsidies pay for themselves still holds. Specifically, we consider an untargeted increase in childcare subsidies by lowering the fee that households have to pay for a full-time slot uniformly by 50 Euro per month. We find that this increase in subsidies is only 6% self-financing. For mothers who are marginal in their labor supply decision w.r.t. to such an increase, we find that they pay an additional 1.68 Euro in taxes for every Euro of subsidies that they receive. However, the share of inframarginal mothers, i.e., those who receive a windfall gain through higher

[3] The expansion efforts of the federal government were initiated in 2005 and more supporting legislation followed in 2007 and 2008.

subsidies but whose labor supply is unaffected, is very high. We show that for each Euro spent on a marginal mother, the government has to pay 41 Euro to inframarginal mothers. In an alternative counterfactual, we find that targeting a similar increase in subsidies to full-time working mothers is more than 50% self-financing. The higher self-financing degree is driven by substantially lower spending on inframarginals.

Third, we turn to the question of how subsidies should vary with income. This is generally a thorny normative question that requires adding up monetary gains and losses of different households. We circumvent this issue in the following way: We compare the childcare fee schedule to the income tax schedule in terms of their efficiency costs of redistribution. For this purpose, we consider a small budget-neutral increase in progressivity which increases taxes (respectively childcare fees) for households with above-median income and decreases taxes (respectively childcare fees) for households with below-median income. We then quantify the marginal excess burden of these reforms, which captures the following intuitive concept: For each Euro taken from above-median income households, how much reaches those with below-median income and how much is lost due to lower labor supply incentives? For the tax schedule, we find that that 69 cents reach households with below-median income. If the reform is conducted for the childcare fee schedule, we find that this number is 79 cents. In other words, the marginal excess burden is 31 cents per Euro for the tax schedule and only 21 cents per Euro for the childcare fee schedule. Hence, the efficiency costs of redistribution are one-third lower for the childcare fee schedule than for the tax schedule.

To understand the intuition behind this result, note that both reforms disincentivize labor supply through an increase in the effective marginal tax rate on labor income: The tax reform increases the marginal tax rate and the childcare fee reform increases the marginal price for one hour of childcare. The childcare fee reform, however, also changes the absolute price for one hour of childcare. We show that households with below-median income are more likely to be marginal in their decision to work with respect to the childcare fee reduction than above-median households facing a childcare fee increase. This partly compensates the labor supply disincentive for below-median income households and mitigates the efficiency cost of redistribution in the childcare fee schedule.

Finally, we suggest two possible policy implications: The first simple conclusion is that society has a weaker desire to redistribute from above-median to below-median income in the childcare-using population relative to the general population. In that case, it would be optimal to have a lower marginal excess burden in the childcare fee schedule. Alternatively, if the government has the same desire for redistribution in the childcare-using population as it has in the general population, then the childcare fee schedule should be made more progressive.

The remainder of this paper is organized as follows: Next, we provide a short review of the literature. In Section 1.3 we present the institutional background and a number of stylized facts, which motivate our analysis. Section 1.4 contains the model setup and Section 1.5 presents the estimation of its components. In Section 1.6 we illustrate the fit of the estimated model. Section 1.7 presents our policy analysis and fiscal calculations and Section 1.8 concludes.

1.2 Related Literature

This paper relates to three strands of literature: reduced form studies on the effects of public childcare on maternal labor supply, structural work modelling household decision making, and the public finance literature.

Reduced form studies that show short to medium-term positive effects of (subsidized) public childcare on female labor supply in the German context include Bauernschuster and Schlotter (2015), Gathmann and Sass (2018), Busse and Gathmann (2020), and K.-U. Müller and Wrohlich (2020).[4] Bauernschuster and Schlotter (2015) leverage the 1996 introduction of a legal entitlement to a place in kindergarten for children above 3 for identification and find a positive impact of public childcare on maternal employment. K.-U. Müller and Wrohlich (2020) find a similar effect by exploiting spatial and temporal variation in a recent public childcare expansion for below 3-year-olds. Focusing on a policy reform which increased the implicit price of childcare in the East German state Thuringia, Gathmann and Sass (2018) show that in response to the reform, public childcare attendance dropped and maternal labor supply declined, especially for vulnerable subgroups such as single, low-income, and low-skilled parents. Busse and Gathmann (2020) use regional variation

[4]Positive effects of public childcare on women's labor market outcomes have been documented for a number of countries. See, e.g., Baker, Gruber, and Milligan (2008), Bettendorf, Jongen, and Muller (2015), Givord and Marbot (2015), and Nollenberger and Rodríguez-Planas (2015). One of the few studies that evaluates long-term effects on maternal labor supply is Haeck, Lefebvre, and Merrigan (2015) who focus on the universal childcare reform in Québec. They find positive effects on maternal labor force participation over a ten year horizon. For other countries, however, researchers have also found contrasting results. Havnes and Mogstad (2011) and Kleven, Landais, Posch, et al. (2020) found at most modest effects of public childcare provision on female earnings in Norway and Austria respectively.

in childcare fees in Germany combined with birthday cut-offs to evaluate the introduction of free universal public childcare, which took place in some selected West German states at different points in time. They find an increase in public childcare attendance of 2 – 3 year old children, which led to higher labor market attachment of mothers. The responses are stronger for vulnerable subgroups such as poorer families and low-skilled parents.

In contrast to the just described studies, our paper focuses on capturing the long-run fiscal effects of subsidized public childcare.[5] These long-run effects are crucial to evaluate the net fiscal cost of childcare subsidies as well as the equity-efficiency trade-off embedded in the childcare fee schedule.

Structural work that focuses on the particular German setup includes Bick (2016), Geyer, Haan, and Wrohlich (2015), and H. Wang (2019). The former paper studies the effect of childcare policies, while the latter two jointly study the impact of public childcare and parental leave policies. All three find that an increase in public childcare availability raises maternal labor force participation. In addition, Haan and Wrohlich (2011) use variation in taxes, transfers, and childcare fees to pin down the labor supply incentives in a structural model. The authors find that higher childcare subsidies increase the labor supply of highly educated women. Also related to our paper are the studies by Guner, Kaygusuz, and Ventura (2020) and Laun and Wallenius (2021) that focus on other countries. Guner, Kaygusuz, and Ventura (2020) study the welfare effects of different child-related transfers in the US context. Laun and Wallenius (2021) examine the effects of family policies in a structural household model for Sweden and find heterogeneous but generally positive effects of childcare subsidies on maternal employment. Our model, in particular the way we model the childcare need and the taste for domestic childcare, builds on Turon (2019). Lastly, the paper is related to Adda, Dustmann, and Stevens (2017), who set up a structural model to determine the career costs of children, and Blundell, Costa Dias, et al. (2016), who structurally estimate the returns to experience for women in the UK (and thereby also the wage penalties from not working full-time).

Our approach is different from all mentioned papers as we focus both on the availability of public childcare as well as on the degree of subsidization. Furthermore, as we take a clear public finance perspective, we have to go further than just capturing the

[5]Eckhoff Andresen and Havnes (2019) is a rare example of a similar estimate. They refine the typical back-of-the-envelope calculation of the static effect of childcare subsidies on income tax revenues and estimate the impact of (subsidized) childcare availability on annual earnings. They find that an increase of NOK 66,000 (USD 8,000) in annual earnings and an average tax rate on the additionally earned income of 14%. However, their perspective is static, whereas we consider the dynamic impact of childcare subsidies on earnings and hence paid income taxes over the life cycle.

aggregate labor supply responses to counterfactual policies. We model rich heterogeneity in family structures as well as preferences to pin down which mothers are marginal or inframarginal for a given policy scenario. This approach allows us to accurately quantify the net fiscal effects of the distribution of behavioral responses to the policies.

In terms of the public finance literature, many papers have emphasized that the implied effects on labor supply provide a rationale for subsidizing childcare (see, e.g., Domeij and Klein 2013, Ho and Pavoni 2020, Bastani, Blomquist, and Micheletto 2020). The goal of this paper is to make this argument more operational from an applied point of view and quantify the size of the implied fiscal externality of childcare subsidies. The paper is also related to Colas, Findeisen, and Sachs (2021), who study the same question for college subsidies but with different trade-offs.

1.3 Background, data, and stylized facts

We now turn to the institutional background of public childcare in Germany and introduce our data source and sample. Furthermore, we present a number of stylized facts on maternal employment and public childcare enrollment that motivate our analysis.

1.3.1 Institutional background in Germany

Childcare facilities. Three different types of childcare institutions can be distinguished in Germany: First, children below the age of 3 are taken care of in nurseries (day care centers). Approximately around the time when children turn 3, they enter kindergarten and stay there until they start school at age 6. Lastly, during school age, children may attend after school care centers in the afternoon.[6] The distinction between these institutions matters because childcare fees differ by the attended institution and, therefore, by child age.[7] Contrary to, e.g., the US, however, the quality of care provided at these childcare institutions does not depend on the price. Strict regulation in terms of caretaker qualification and child-to-caretaker ratios yields a homogeneous level of quality across Germany. Furthermore, 95% of childcare institutions are either operated by municipalities or by non-profit organizations.[8]

[6] Only a small share of elementary schools operates with mandatory full-day schooling (Klemm 2014).
[7] See Appendix A3.1 for a detailed description of the determinants of childcare fees.
[8] See Authoring Group Educational Reporting (2018).

Recent public childcare expansion. While all three types of childcare institutions have been continuously present throughout Germany since the early 1990s, their use and prevalence have changed substantially in the past thirty years. Public childcare supply for children below 3 (nursery slots) remained scarce until 2005. In that year, the German government committed to creating 230,000 additional nursery slots by October 2010, increasing the childcare coverage for children aged 0 – 2 from just 5 to 17 slots per 100 children. Focusing policy efforts further on nurseries, a legal entitlement to a public childcare slot for all children aged 1+ was introduced in August 2013.[9] This reform effectively ended the rationing of public childcare.[10]

1.3.2 Data

For our analysis, we focus on German females aged between 20 and 65 who are currently not in education and form a household with a full-time working partner.[11] We track this group over the time span 2012 to 2017 in a representative longitudinal survey data set, the German Socio-Economic Panel (GSOEP). The GSOEP is an unbalanced household panel that has been running since 1984. Its scope is comparable to the US Panel Study of Income Dynamics, as it provides annual socio-economic and demographic information on the household and individual level.[12]

1.3.3 Stylized facts

Effect of child birth on maternal labor supply. Persistent effects of parenthood on maternal labor market outcomes – also called child penalties – are a well-established fact in the recent literature for a growing number of countries (see, e.g., Kleven, Landais, and Søgaard 2019, Angelov, Johansson, and Lindahl 2016, as well as Kleven, Landais, Posch, et al. 2019).

[9]Related family policy laws: 1996 - Reform of the §SGB VIII, 2005 - Tagesbetreuungsausbaugesetz, 2008 - Kinderfoerderungsgesetz.

[10]See Appendix A2.3 for a discussion of the remaining gap between childcare demand and childcare enrollment for 0 – 2 year olds.

[11]In line with our modelling framework described in Section 1.4, we furthermore limit the sample to females who have at most three children and gave birth (if any) only between ages 20 and 40 (85.26% of observations). See Appendix A1 for additional details on the sample.

[12]Source: FDZ-SOEP (2019). See Goebel et al. (2019) for a detailed description.

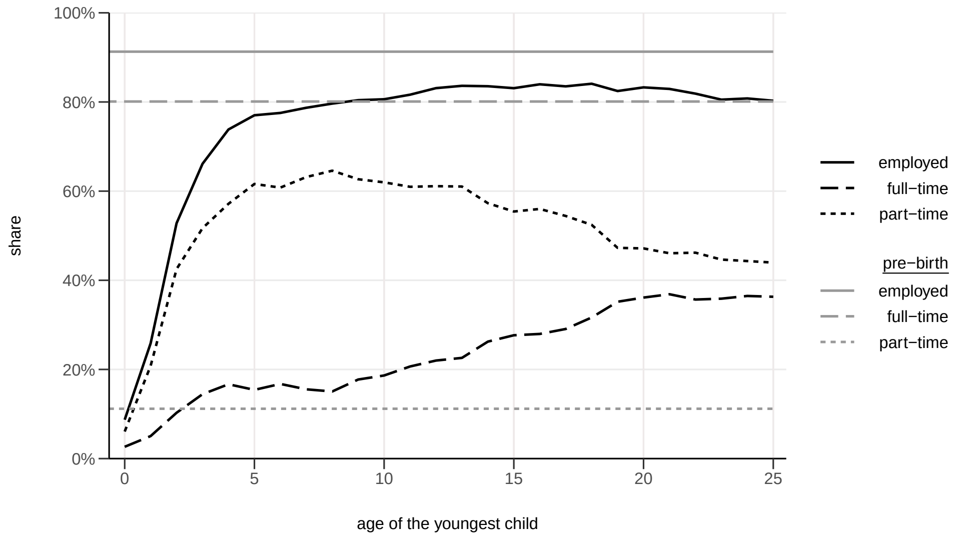


Figure 1.1: Maternal employment by the age of the youngest child

Notes: Full-time work corresponds to 40h/week, part-time corresponds to 20h/week. Maternity leave is treated as non-employment. Sample: females aged 20 to 65 that are not in education and live with a full-time working partner, conditional on having at least one child. Source: 2000 to 2017 GSOEP, FDZ-SOEP (2019).

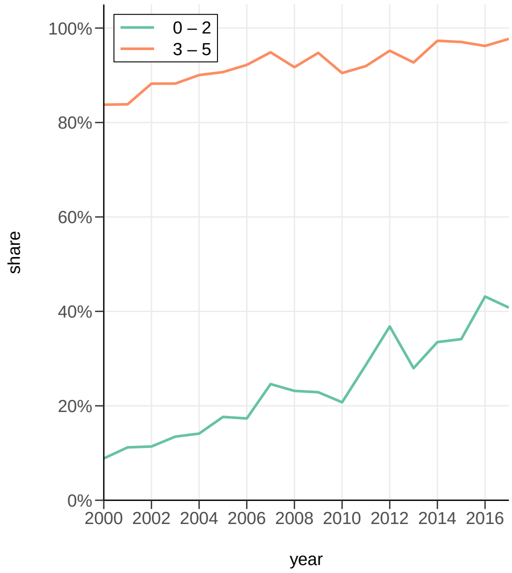
This also holds for Germany where childbirth has a substantial and sustained effect on maternal employment.[13] Figure 1.1 illustrates this effect in the GSOEP data. It compares maternal pre-birth employment rates to their post-birth development depending on the age of the youngest child. The average pre-birth employment rate of future mothers (illustrated by the horizontal lines) is above 90%, of which 80% work full-time ($\approx$ 40 hours per week). In the first year after childbirth, the employment rate drops by more than 80 percentage points.[14]

As children grow older, the share of mothers working full-time rises gradually. While it reaches 15% by the end of statutory maternity leave at age 3, from then on it only increases slowly to just about 20% when the youngest child is 10 years old. Afterwards, the full-time share continues to rise gradually, but stays below 40%.

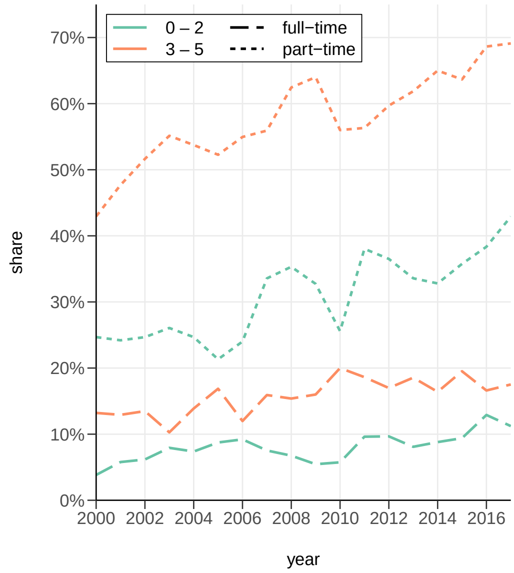
[13]Kleven, Landais, Posch, et al. (2019) use the event study approach proposed by Kleven, Landais, and Søgaard (2019) and data from 6 countries (US, UK, Denmark, Sweden, Austria and Germany) to show that long-run earnings penalties range from 21% in Denmark to 61% in Germany.

[14]Note that until the third birthday of the child mothers are entitled to maternity leave. During such a maternity related temporary 'non-employment'-spell mothers receive financial compensation up to 12 months and have the right to return to their pre-birth job. 77.86% of mothers whose youngest child is aged 0 to 1 state that they are on maternity leave.

Turning to part-time, prior to first birth, only 12% of future mothers work with reduced hours. This share increases sharply in the first year after childbirth and reaches 62% with a 5-year-old. At this point, total employment reaches almost 78% and barely increases afterwards. What changes is the composition: As children grow older, more mothers switch from part-time to full-time.



(a) Public childcare enrollment



(b) Employment of mothers

Figure 1.2: Development of public childcare enrollment and maternal employment

Notes: Enrollment is binary in the sense that it is not conditional on a minimum number of hours. Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, conditional on having a child aged 0 – 2 or 3 – 5. Source: 2000 to 2017 GSOEP, FDZ-SOEP (2019).

Reform effects on childcare enrollment and maternal labor supply. Figure 1.2a plots childcare enrollment shares for different child age brackets across recent years. In 2000, more than 80% of 3 – 5 year old children were already attending childcare, while the share was only 9% for 0 – 2 year olds. Over the years during which public childcare for 0 – 2 year olds was substantially expanded, this enrollment share has increased to more than 40% in 2017. Overall, childcare attendance of 0 – 2 year olds has increased by more than 30 percentage points over the last 17 years. Enrollment of 3 – 5 year old children has also increased, especially during the early 2000s, reaching 97% in 2017.

To relate this pattern of increased childcare attendance to employment rates, Figure 1.2b plots maternal full-time and part-time shares by the age of the youngest child.

During the 2000 to 2017 time span, when childcare enrollment of children aged 0 – 2 increased substantially, mothers of children in this age bracket have been increasing their labor force participation both along the extensive and the intensive margin. Part-time rates among these mothers have increased by more than 15 percentage points, while full-time rates have also increased slightly. Looking at mothers with 3 – 5 year olds, part-time employment shares have risen substantially from 45 to more than 65% between 2000 and 2017. During the same period, the full-time share has increased by around 5 percentage points.

1.4 Model

We now present our dynamic model of households faced with labor supply and childcare choices. A household is composed of two adults with up to three children. Households' decision making is unitary and forward looking. The unit time period is 3 years. Marriages are formed at the age of 20 and are stable. Both spouses are of the same age, retire at 65, and have a remaining lifespan of 15 years after retirement. Fertility follows an exogenous stochastic process, which captures the substantial empirical heterogeneity in family composition and the age of parents at first birth.

Households with young children make two decisions each period: how to provide care for the children and how much labor to supply. Regarding childcare, they decide between the female spouse caring for the children at home, which we call 'domestic childcare', and externally provided childcare. The latter can either be informal childcare by, e.g., grandparents, or the use of public childcare services.[15] Labor supply choices are discrete: The female spouse can work full-time, part-time, or choose not to participate, while the male spouse is assumed to always work full-time.[16] An important dynamic component of our framework comes from the positive impact of current working hours on the expected growth rate of (potential) wages. Hence, career breaks imply wage penalties.

A distinct feature of our model is the large amount of heterogeneity. Households differ in education, which is an important component in the stochastic wage and fertility

[15]The introduction of informal childcare is motivated by the fact that we observe some mothers who work more hours than they buy public childcare for. See Appendix A2 for a discussion of how public childcare hours correspond to working hours in the data.

[16]Close to 90% of fathers of children below 9 work full-time in our sample. Therefore we rule out that fathers provide domestic childcare during working hours.

processes. Besides education, female wages, male wages, and family composition, households are heterogeneous in three further dimensions: their preference for domestic childcare, their taste for the leisure of the female spouse, and their access to free informal childcare. The latter three are unobservable to the econometrician. As we argue below, accounting for this unobserved heterogeneity is key to capture the large heterogeneity in childcare and labor supply choices of households.

1.4.1 Children

We model fertility and family composition in the following way: We allow children to be born, one by one, to mothers between the ages of 20 and 40. After the first birth, subsequent siblings can only be born in one or two 3-year intervals, i.e., all age gaps between children of a family can only be 3 or 6 years. The fertility process is stochastic and is determined by the education of the mother, the age of the mother, and the presence of older siblings.

For our model purposes, the child age ranges that are relevant are $(0-2)$, $(3-5)$, $(6-8)$, and $(9+)$. We denote K a 4-element vector indicating the presence of a child in these age brackets. For example, a family whose composition is represented by the vector $K = (0, 1, 1, 0)$ has two children, the youngest aged between 3 and 5 and the eldest aged between 6 and 8. By assumption, each of the first three elements of K can only be 0 or 1 since only one child can be born in each period. Transitions between different values of K are governed by fertility events and the ageing of the household's children. Finally, we assume that households cannot have more than three children.[17]

1.4.2 Preferences

Households value female leisure time L, household consumption c, and domestic childcare dcc. Household consumption is made comparable across different household sizes k by applying a square root equivalence scale. Even when the female spouse works full-time, she has a number $\overline{L}$ of leisure hours and can devote $\overline{dcc}$ hours to domestic childcare. Hence, we should think of dcc and L, respectively, as time with children and leisure during normal working hours.

[17] Only 4.99% of households have more than three children in our data.

Preferences are reflected in the following instantaneous utility function:

$$u(c, L, dcc) = \left(1 - \mathcal{G}(g, K)\right)\left((1-\alpha)\frac{\left(\frac{c}{\sqrt{k}}\right)^{1-\gamma_c} - 1}{1-\gamma_c} + \alpha\frac{\left(L + \overline{L}\right)^{1-\gamma_L} - 1}{1-\gamma_L}\right) + \tag{1.1}$$

$$\mathcal{G}(g, K)\left(\frac{\left(dcc + \overline{dcc}\right)^{1-\gamma_{dcc}} - 1}{1-\gamma_{dcc}}\right),$$

where the three CRRA coefficients γ_c, γ_L and γ_{dcc} as well as $\overline{L}$ and $\overline{dcc}$ are homogeneous across all households.

Preference heterogeneity. Households' preferences are heterogeneous in two dimensions: α and g. α represents the relative taste for female leisure over consumption. The parameter g is the preference for domestic childcare relative to leisure and consumption. For now, the reader should simply think of them as preferences for time with children and leisure, respectively. In Section 1.4.5 we discuss how these parameters can be interpreted in a more general way.

We allow the taste for domestic childcare to vary with the age of the child by scaling g via $\mathcal{G}$. Based upon the child-age vector K, the functional form of $\mathcal{G}$ corresponds to

$$\mathcal{G}(g, K) = \begin{cases} g & \text{if youngest child's age} \in [0, 3], \\ g \cdot \kappa & \text{if youngest child's age} \in [3, 9). \end{cases}$$

This allows us to capture the sharp difference in public childcare enrollment between 0 – 2 year olds and 3 – 5 year olds, as previously illustrated in Figure 1.2a.

1.4.3 Constraints

Childcare hours constraints and childcare expenditures. We now describe the time constraint for childcare provision. For each $j = 1, 2, 3$ relating to the age ranges (0 – 2), (3 – 5), and (6 – 8), a child needs an age-specific number of hours of childcare within normal working hours, $\bar{t}_j$. Normal working hours are 40 hours per week. In the first and second child age categories, the child needs care all of the time, whereas in the third category, the child needs care in the non-school hours only since she is enrolled in compulsory primary school already. There are three ways in which parents can fulfill the childcare need: maternal time, which we denote dcc and refer to as domestic childcare, informal (other) childcare denoted oth by, e.g., grandparents, or public (market) childcare services, i.e., in a nursery, denoted mcc.

Informal childcare is free and only available to some households. The variable oth ranges between 0 and 40 hours a week. If available, households always prefer to use oth hours of costless informal childcare over mcc hours of costly public childcare. In that sense, oth is a reduced form for whether informal childcare is available and considered equally good as public childcare. We provide a longer discussion about the interpretation and role of oth in Section 1.4.5.

Public childcare is always available at a fee, normalized to full-time use, which depends on the age j of the child, the family structure K, and the household gross income y:

$$p\,(j, K, y)\,.$$

For a given amount of domestic childcare and informal childcare use, the resulting amount of public childcare necessary for a child of age j is thus given by:

$$mcc(j) = \max\left\{0, \bar{t}_j - dcc - oth\right\} \tag{1.2}$$

and the sum of public childcare of all children in the household:

$$Tcc(K) = \sum_{j=1}^{3} K(j) \cdot mcc(j),$$

where $K(j)$ is the j-th element of the vector K indicating if a child of age j currently lives in the household.

The household expenditure for public childcare of all its children is given by:

$$Ecc(K, y) = \sum_{j=1}^{3} K(j) \cdot mcc(j) \cdot p(j, K, y).$$

This equation clearly shows that public childcare implies higher expenditure the more children a household has because childcare fees have to be paid for all children. This contrasts domestic and informal childcare, where not more time is needed to look after one more child. To put it more formally, we assume that taking care of J children at home requires one hour of domestic or informal childcare but J hours of public childcare.

Parental time constraint. At each age t, the household has to choose between female labor supply (lm_t), female leisure (L_t), and the provision of domestic childcare (dcc_t). Total available time is normalized to 1, so the time constraint is written as:

$$lm_t + L_t + dcc_t = 1. \tag{1.3}$$

Budget constraint. We abstract from borrowing and saving to keep the state space tractable despite the large amount of heterogeneity. In that sense, the budget constraints are static and given by:

$$c_t + Ecc(K_t, y_t) = y_t - T(y_t), \tag{1.4}$$

where

$$y_t = 40 \cdot \left(w_{m,t}\left(w_{m,t-1}\right) + lm_t \cdot w_{f,t}\left(w_{f,t-1}, lm_{t-1}\right)\right). \tag{1.5}$$

$T(\cdot)$ captures the tax and transfer system and $lm_t \in \{0, 0.5, 1\}$ represents non-participation, part-time, and full-time work, respectively. $w_{m,t}(w_{m,t-1})$ is the male wage that follows an exogenous Markov process. The female wage $w_{f,t}\left(w_{f,t-1}, lm_{t-1}\right)$ also follows a Markov process. Importantly, it is a function of past labor supply. This is an important dynamic component of the budget constraint: Current female labor supply decisions do not only affect current earnings but also future wages and therefore earnings.

Wage growth differs across age and across the wage distribution. Furthermore, the expected wage penalties for working lower or no hours also differ depending on the level of the wage as well as on age. Despite the Markov property, the wage process is thereby able to capture the key dynamics of more complex human capital accumulation frameworks (such as in, e.g., Blundell, Costa Dias, et al. 2016), as we demonstrate in Section 1.5.3.

Finally, once the spouses retire, they get a fraction $\mathcal{B}$ of their last period's full-time earnings potential as retirement benefits.

1.4.4 Dynamic decision problem

We summarize all heterogeneity by a seven-dimensional state space vector:

$$\Omega_t = (t, w_{m,t}, w_{f,t}, K_t, educ, g, oth, \alpha).$$

At each age t, the household has to choose female labor supply (lm_t) and the use of domestic childcare (dcc_t), which also imply the choices of consumption (c_t), female leisure (L_t), and the use of public childcare (mcc_t). The labor supply decision is discrete between full-time, part-time, and non-participation. The domestic childcare decision is close to continuous with decisions in steps of 5 hours from 0 to 40 hours. The three constraints that the household faces are the need for childcare of current children (equation (1.2)), the time constraint for the female spouse (equation (1.3)), and the budget constraint (equation (1.4)).

The full dynamic household problem is defined for a given state space vector Ω_t as:

$$V(\Omega_t) = \max_{lm_t, dcc_t} u(c_t, L_t, dcc_t|\Omega_t) + \beta\mathbb{E}[V(\Omega_{t+1}|\Omega_t, lm_t)], \tag{1.6}$$

subject to the constraints (1.2), (1.3), and (1.4). This problem can be broken down into an intratemporal choice and an intertemporal choice. The intratemporal decision reflects the optimal time allocation between leisure and domestic childcare for a given female labor supply choice lm_t. The intertemporal choice is to choose the optimal labor supply, lm_t, given the conditional optimal choice of domestic childcare $dcc_t(lm_t)$.

Intratemporal decision problem. The optimal intratemporal choice consists in solving the following static problem for a given discrete female labor supply decision lm_t and the state space vector Ω_t:

$$u^*(\Omega_t, lm_t) = \max_{dcc_t} u(c_t, L_t, dcc_t|\Omega_t, lm_t), \tag{1.7}$$

subject to the childcare constraint (1.2), the time constraint (1.3), and the budget constraint (1.4).

Intertemporal decision problem. The optimal intertemporal choice consists in maximizing lifetime utility $V(\Omega_t)$ by choosing female labor supply lm_t:

$$V(\Omega_t) = \max_{lm_t} u^*(\Omega_t, lm_t) + \beta V(\Omega_{t+1}|\Omega_t, lm_t) \tag{1.8}$$

for given dynamics of the states $\left(\Omega_{t+1}|\Omega_t, lm_t\right)$. The model is solved by backward induction from retirement. We assume that during retirement all income is used as consumption and the entire time endowed per period is used as leisure since, by construction, there are no children to be taken care of.

1.4.5 Unobserved heterogeneity

We now provide a more thorough discussion of the role that the unobserved heterogeneity parameters (g, oth, α) play. First, we allow for heterogeneity in leisure preferences α to account for the sizeable variation in labor supply conditional on wages. Such heterogeneity in leisure preferences (or equivalently, disutility of work) is a common component in structural models to be able to match hours worked decisions (such as in, e.g., Blundell, Costa Dias, et al. 2016).

A more important and more unique feature of our model is the assumption about heterogeneity in g and oth. The heterogeneity in these two parameters is necessary to account for the heterogeneity in childcare decisions conditional on observables that we observe in the data.

While we have introduced g as a preference parameter, we think of g in a more general sense as a reduced form that reflects a number of different aspects: Heterogeneity in g can capture i) the true preference heterogeneity for spending time with the child. It can also capture ii) heterogeneity in how much parents (dis)like their child being in daycare or informal childcare, e.g., due to social norms, positive (negative) peer effects, or (mis-)trust in the quality of the childcare institutions. Furthermore, it may also reflect iii) the fixed (utility) cost of bringing children to daycare, e.g., of a car ride, as we do not model geography and distance.[18]

Finally, we turn to informal childcare oth. A specific value of 10 hours, for example, does not only reflect the availability of such an amount of informal childcare. It also reflects that it is desired by the household in the sense that the household prefers 10 hours of free informal childcare over 10 hours of paid public childcare. Hence, oth is a reduced form that – besides availability – also reflects preference heterogeneity: Some households may want the grandparents to take care of their children, others may not.

The distribution of (g, oth, α) conditional on observables is key to capturing the observed behavior of households. Furthermore, it allows us to predict how the behavior of households changes if policies change.

[18] See Appendix A2.3 for corresponding evidence on the reasons why parents do not send their children to public childcare.

1.5 Estimation

Our estimation can be decomposed into two broad parts. First, we estimate and calibrate various parameters without using the explicit structure of the model: We set policy parameters such as tax rates, childcare fees, but also our assumptions on childcare needs in Section 1.5.1. Then, we estimate the fertility process in Section 1.5.2 and the wage process in Section 1.5.3.

In a second step, we quantify the remaining parameters by using the explicit structure of the model.[19] First, we set the homogeneous preference parameters. Second, we estimate the distribution of heterogeneous preference parameters by maximum likelihood. The homogeneous preference parameters are set in a way such that the estimated model delivers labor supply elasticities that are consistent with quasi-experimental evidence.

1.5.1 Childcare need and government policies

In this section, we first calibrate the childcare need of the different age groups. We also calibrate the costs of public childcare and estimate the childcare fee schedule. Second, we quantify the government policies that are required as exogenous inputs for our model.

Table 1.1: Weekly childcare need within normal working hours

Children's age interval	0 – 2	3 – 5	6 – 8	9+
Hours of childcare needed ($\bar{t}_j$)	40	40	15	0
Minimum public childcare norm	0	20	0	0

Childcare need. Table 1.1 summarizes the assumed childcare need for children of different ages: If a child is younger than 6, the childcare need is set to 40 hours per week, i.e., 100% of the time. To account for the fact that almost all 3 – 5 year olds attend kindergarten at least half-days (see Figure 1.2a), we impose that 20 of the 40 hours required

[19]For the estimation, we furthermore operationalize the large state space as follows: To capture the age range from 20 to 80, we set up $t = 20$ 3-year model periods. Heterogeneity in male and female wages is captured by 5 and 10 gridpoints, respectively, education by 2 different levels, and the family structure K as introduced in Section 1.4.1 requires 18 state space points. The unobserved heterogeneity in g and α is captured by 20 gridpoints each, while 17 gridpoints are sufficient for oth.

for this age interval have to be covered by public childcare. Half-day public childcare attendance in this age range has become close to a social norm. For children aged 6 – 8, the need reduces to 15 hours per week because these children attend compulsory schooling for 25 hours per week. This yields for each child age j, the age-specific weekly hours of childcare needed, $\bar{t}_j$.

Public childcare cost structures. We approximate the cost structure of public childcare institutions by assuming the costs to be linear in the number of children. This abstracts from any possible non-linearities driven by, e.g., capacity constraints, but provides a reasonable average value for public spending per child. We use the values in Table 1.2, which are provided by the German Statistical Office.

Table 1.2: Average government spending per child for 40h/week of public childcare

Children's age interval	0 – 2	3 – 5	6 – 8
Annual cost	€11,919	€7,983	€6,780

Notes: See Statistisches Bundesamt (2012), converted to 2017 prices.

Childcare fees. We use data from the 2013, 2015, and 2017 GSOEP waves, which contains information on public childcare hours per day and monthly fees paid.[20] We normalize the monthly fees by the reported daily public childcare hours to extract the monthly fee of full-time public childcare, defined by an attendance of 8 hours per day or 40 hours per week. For this purpose, we assume linearity of childcare fees in hours.

Given that we also observe a fraction of households paying zero fees, we estimate a Tobit model of childcare fees as a function of gross household income, which we also interact with the number of siblings (see Appendix A3.2 for details). Figure 1.3 shows the estimated fee schedule. Childcare fees are slightly increasing in household income (between 2% and 3% at the margin) and decrease with the number of siblings. Furthermore, fees are higher for younger children.

[20] In terms of the sample construction, this estimation is based on the same sample as laid out in Section 1.3.2.

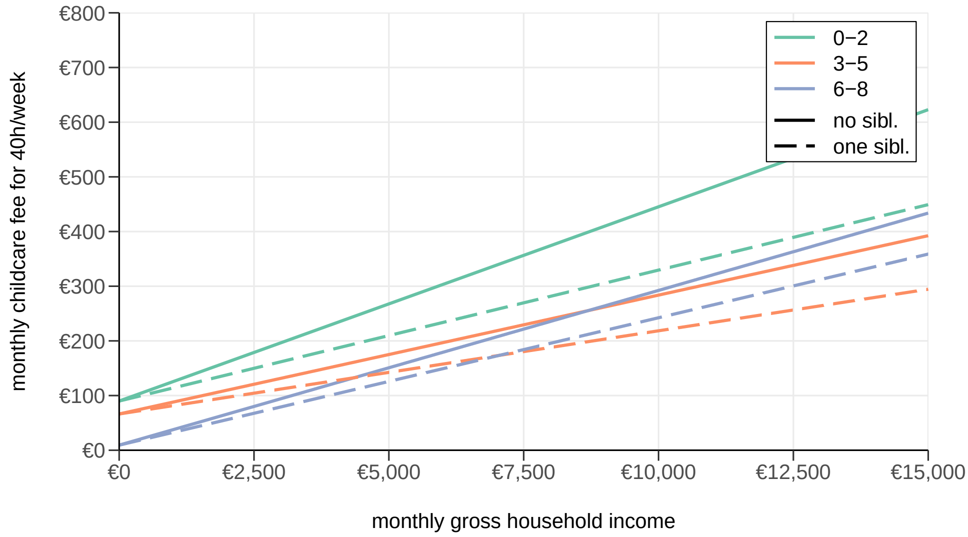


Figure 1.3: Estimated childcare fee schedule

Notes: See Appendix A3.2 for details on the underlying Tobit regression. All values are adjusted to 2017 price levels. Source: 2013, 2015, 2017 GSOEP, FDZ-SOEP (2019).

Taxes. We use the Matlab implementation of the German tax and transfer code provided by Bick, Brüggemann, et al. (2019) to map gross to net income and calculate tax revenues.[21] The implementation is based on the annual OECD "Taxing Wages" reports and takes into account federal income taxes as well as social security contributions, cash benefits, and standard deductions. Aside from a precise implementation of the non-linearities of the tax code, it includes joint taxation of couples as well as child benefits for each child in the household. Marginal tax rates faced by women differ with their spouses' income and child allowances reduce the taxable income of the household.

Pensions. We approximate the German pension system by assuming that households receive 40% of both partners' last period's potential gross full-time earnings throughout retirement. The pension share, $\mathcal{B}$, is therefore set to 0.4, which matches the replacement rate reported in OECD (2017).

Interest rate. We set the interest rate of the government to 3% per 3-year model period, which corresponds approximately to 1% per annum.

[21]We calculate tax revenues the sum of income tax payments, social security contributions for public sickness and care insurance, and solidarity surcharge payments.

1.5.2 Estimation of the fertility process

As introduced in Section 1.4.1, children are assumed to be born one at a time in any 3-year model period to mothers aged 20 to 40. We also restrict households to have at most three children. The determinants of fertility are the age and education of the mother and the number and ages of children already present in the family. The transition probability between family composition K and family composition K' faced by a household aged t, with an education level *educ* captures the (deterministic) ageing of existing children and the fertility hazard over the next period. Our estimate of the birth rate for this household is simply the sample average of birth events conditional on $(t, educ, K)$.

To make sure that we can also identify the less frequent fertility transition probabilities robustly, we compute them on an alternative larger data set, the German Microcensus. Specifically, we use the 2014 and 2018 Microcensus waves and focus on births taking place from 2012 to 2017.[22]

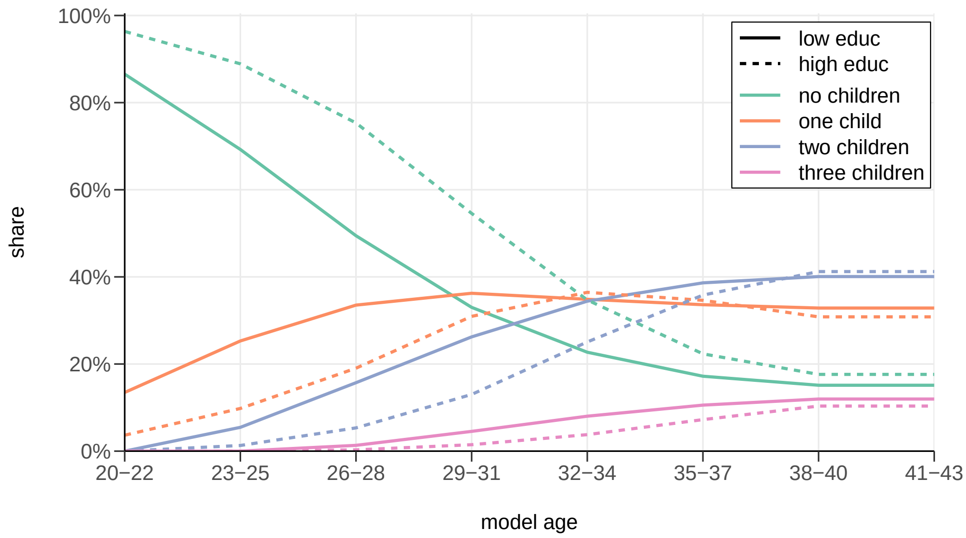


Figure 1.4: Family composition as implied by the fertility process

Notes: 'low educ' corresponds to no A-level, 'high educ' corresponds to having obtained an A-level. Sample restrictions as laid out in Section 1.3.2. Sources: FDZ-StABL (2020a), FDZ-StABL (2020b).

[22]We select the sample from the Microcensus data using the same restrictions as for our GSOEP survey data (see Section 1.3.2). Sources: FDZ-StABL (2020a), FDZ-StABL (2020b). This yields us 71,165 observations of households aged 20 to 40.

Figure 1.4 illustrates our estimates of the evolution of shares of families with zero to three children over the age of the mother and by education level, referring to having obtained an A-level or not. In terms of completed fertility, i.e., number of children at age 41, the figures are similar in both education groups: about 45% of households have two children, about 30% (respectively 10%) have one (respectively three) child(ren) and about 15% of households remain childless. The timing of births, however, differs markedly between education levels, with most curves for the high education group exhibiting a lag of three years, i.e., one model period, relative to the low education group. Only the share of families with three children grows nearly simultaneously for both groups. By age 34 (respectively 37) for the low (respectively high) education group, the majority of households have completed their fertility. Half the households have had at least one child by age 26 (respectively 29) in the low (respectively high) education group.

1.5.3 Estimation of the wage process

Using the 2000 to 2017 GSOEP data, we observe monthly gross labor income as well as contracted working hours.[23] This allows us to directly compute hourly wages for every individual that is working. For females who choose not to work, on the other hand, we do not observe any labor income and therefore, we impute their *potential* gross hourly wages using a selection corrected wage model (see Appendix A3.3 for details).

We then estimate the following equation for the wage process of females:

$$\begin{aligned} log(w_{f,it}) = \alpha + \beta_1 log(w_{f,it-1}) + \beta_2 \mathbb{1}\{lm_{it-1} = NP\} + \\ \beta_3 \mathbb{1}\{lm_{it-1} = PT\} + \beta_4 educ_i + \mathcal{A}(t) + \varepsilon_{it}^{w_f}, \end{aligned} \tag{1.9}$$

where $\mathbb{1}\{lm_{it-1} = NP\}$ and $\mathbb{1}\{lm_{it-1} = PT\}$ are dummy variables that indicate whether a woman i was either not working or working part-time in period $t-1$. The coefficients β_2 and β_3 are of particular interest for our analysis since they measure the dynamic wage penalty from working less than full-time. β_4 captures the wage increase due to having obtained an A-level and $\mathcal{A}(t)$ is a third-order polynomial in age. Note that the implied wage process is a Markov process, where the individual wage is drawn from a log-normal distribution that depends on the previous wage, previous employment decision, age, and education. The estimated coefficients are shown in Appendix-Table A.3.

[23] We extend the sample for the wage process estimation back until 2000 to ensure that we can robustly capture the key dynamics with a sufficient number of observations. Otherwise, we use exactly the same sample restrictions as described in Section 1.3.2.

The implied age-wage profiles follow a hump-shaped profile, which is consistent with the literature (see Appendix-Figure A.6). The estimated wage penalties for working part-time or not working instead of working full-time are substantial and amount to 5.5% and 16.5% per 3-year model period.[24]

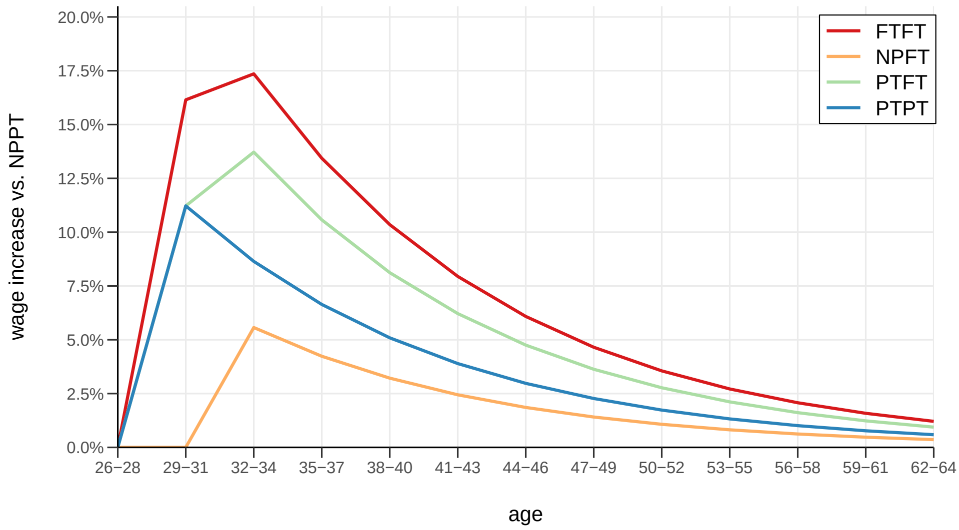


Figure 1.5: Illustration of the wage process dynamics

Notes: Relative increase in wages of different labor supply patterns, always compared to not working at age 26 – 28 and working part-time at age 29 – 31. NP, PT, and FT denote not working, part-time work, and full-time work, respectively. PTFT denotes part-time work at age 26 – 28 and full-time work at age 29 – 31. All other patterns are defined analogous. Simulations based on female wage process estimates from Appendix-Table A.3.

In Figure 1.5, we illustrate our wage process by looking at the benefits of increasing labor supply relative to a typical labor supply pattern of mothers. Specifically, we consider a mother that conceives her child at 26. The benchmark is that she does not work while the child is 0 – 2, works part-time when the child is 3 – 5, and works full-time afterwards. The graph illustrates the dynamic wage gains that the mother would obtain if she increased her labor supply. The blue line shows the case if the mother already starts working part-time when the child is 0 – 2. The red line shows the case if the mother switches to full-time work both when the child is 0 – 2 and also when the child is 3 – 5.

Aside from the substantial wage gains from increasing labor supply, the graph clearly illustrates that the potential wage gains are quite persistent. This will play a key role for

[24]Based on the estimates in Appendix-Table A.3, transformed into percent changes.

the dynamic fiscal effects that we present later: If public childcare allows a mother to increase her labor supply, this does not only affect her tax payments while the child is young, but also her tax payments in the future.[25]

Finally, we also estimate the male wage process in a similar fashion (see Appendix-Table A.3 for details). Since we only consider males that always work full-time, the wage equation does not contain part-time or non-employment penalties.

1.5.4 Calibration of homogeneous preference parameters

In line with Blundell, Costa Dias, et al. (2016), we set the discount factor β to 0.94 per 3-year model period.[26] We calibrate the homogeneous parameters of the utility function in equation (1.1) to match, jointly with the estimated distribution of unobserved heterogeneity, data moments of labor supply and public childcare take-up as well as benchmark labor supply elasticities from the literature. We discuss this further in Section 1.6, where we present the model fit.

Table 1.3: Calibrated homogeneous model parameters

parameter	β	γ_c	γ_L	$\overline{L}$	γ_{dcc}	$\overline{dcc}$	κ
value	0.94	1	2	1	1.125	4	0.075

Table 1.3 provides an overview of all calibrated parameters. We use a log specification for consumption ($\gamma_c = 1$) and set the CRRA coefficient on leisure (γ_L) to 2 and on domestic childcare (γ_{dcc}) to 1.125. The leisure endowment $\overline{L}$ is calibrated to 1 and the domestic childcare endowment $\overline{dcc}$ will be set to 4, if children below 9 are present. Furthermore, the shifter for the preference for domestic childcare κ is 0.075. It implies that households have a much stronger preference to spend time with children below age 3 compared to 3

[25] To give the reader an idea about magnitudes, assume that a mother has an hourly wage of €20 at age 26. If she decided to work part-time instead of not working when the child is 0 – 2, her earnings would increase by €20,000 each year in the first three years. Due to the positive effect on future wages, the increase in the net-present value of earnings, however, is a bit more than €100,000. The dynamic wage effects make up around 40% of the overall effect in this simple example. Due to the joint taxation system in Germany, the earnings of most secondary earners are taxed at a rate of 50% or more at the margin, so that this translates into an increase in tax revenue by €50,000 or more.

[26] Corresponding to an annual discount factor of 0.98.

to 8 year old children. This allows us to capture social norms in a reduced form way, as many women spend the first years after birth out of the job caring for their children.[27]

1.5.5 Maximum likelihood estimation of heterogeneous preferences

In our model, the preference for domestic childcare g, the availability of informal childcare oth, and the taste for female leisure α are driving forces of female labor supply and childcare decisions during parenthood. Their unobserved joint distribution determines how female labor supply and therefore women's career paths react to changes in childcare policies. In this subsection, we present our estimation to identify this joint distribution of permanent unobserved heterogeneity.

We first introduce our methodological approach by explaining the components of the likelihood function as well as our assumptions. Then we outline the identification of the unobserved heterogeneity and discuss our selection of constant characteristics x on which we condition the distribution of unobserved heterogeneity. Finally, we present the results of the maximum likelihood estimation of the joint distribution of unobserved heterogeneity.

Methodology. Our data comprises of observations for two model periods of female labor supply and total public childcare take-up of the household, which we denote by the vectors $(lm^n, Tcc^n) = (lm_1^n, lm_2^n, Tcc_1^n, Tcc_2^n)$. We omit the household index n in most of the equations below for ease of notation.

The time-varying state variables are summarized in the vector s and are also observed over two model periods. s includes the wages of the male and female spouse (w_m and w_f), the family composition K, and age t.[28] The unobserved heterogeneity $h = (g, oth, \alpha)$ and the constant characteristics x, on the other hand, are assumed to be constant. Given s and h, our model predicts female labor supply and total public childcare for each period p in a deterministic manner.

Let us denote these predicted choices as $\widehat{lm_p}$ and $\widehat{Tcc_p}$. The likelihood of the unobservable 'true' choices (lm_p, Tcc_p) being predicted for a household with characteristics (s_p, h) is thus as shown in equation (1.10) below. The terminology 'true' choices is used

[27]In the 2016 wave of the German General Social Survey around 40% of respondents agree with the statement "A small child is bound to suffer if his or her mother goes out to work." Source: GESIS (2017).

[28]s contains all observed states of the model besides the female education level $educ$, which is contained in x.

to distinguish these from the choices observed in the data, which include measurement errors.

$$\ell(lm_p, Tcc_p|s_p, h) = \begin{cases} 1 & \textit{iff} \quad \widetilde{lm_p}(s_p, h) = lm_p \text{ and } \widetilde{Tcc_p}(s_p, h) = Tcc_p, \\ 0 & \text{otherwise.} \end{cases} \tag{1.10}$$

We assume that the wages of both spouses, w_m and w_f, as well as total public childcare Tcc are observed with measurement error, but the discrete labor supply choices lm are observed without error. Furthermore, we assume that all measurement errors are independent of each other. We denote observed wages and total public childcare as $(\widetilde{w}, \widetilde{Tcc})$, in contrast to the 'true' measurement error-free quantities (w, Tcc).

For each period p and each spouse q, we assume that the distribution of the observed wage conditional on the 'true' wage, denoted $\ell(\widetilde{w_{p,q}}|w_{p,q})$, is such that the absolute value of the measurement error in log-wages follows a type II extreme value distribution.[29] Similarly, the conditional distribution of observed total public childcare $\ell(\widetilde{Tcc_p}|Tcc_p)$ given the 'true' amount of total public childcare Tcc_p is such that the absolute value of the error is also distributed as a type II extreme value distribution for each period p.[30]

$$\ell(\widetilde{w_{p,q}}|w_{p,q}) = \mathcal{F}(\log(\widetilde{w_{p,q}}) - \log(w_{p,q})), \tag{1.11}$$
$$\text{where } |\epsilon_{p,q}| = |\log(\widetilde{w_{p,q}}) - \log(w_{p,q})| \sim EV_{\text{II}},$$

$$\ell(\widetilde{Tcc_p}|Tcc_p) = \mathcal{F}(\widetilde{Tcc_p} - Tcc_p), \text{ where } |u_p| = |\widetilde{Tcc_p} - Tcc_p| \sim EV_{\text{II}}, \tag{1.12}$$

where $p = 1, 2$ denotes the time period, $q = m, f$ denotes the spouse, and $(\epsilon_{p,q}, u_p)$ denote the measurement errors in wages and total public childcare, respectively. We omit the conditioning on the full set of time-varying characteristics s in (1.11) and (1.12), as well as in (1.13) below, to ease the notation, but these expressions should still be understood as conditional on it.

[29] We use a generalized extreme value distribution with $\sigma = 0.026$ and $\xi = 0.5$ for the measurement error in wages. This distribution matches the one estimated in Blundell, Costa Dias, et al. (2016) in terms of its mean measurement error and the measurement error at the 90th and 95th percentile.

[30] We use a generalized extreme value distribution with $\sigma = 1.05$ and $\xi = 0.5$ for the measurement error in public childcare. This distribution implies a mean measurement error of 2.5h/week as well as 5.6h/week and 8.7h/week at the 90th and 95th percentile.

The likelihood of observing the choices of a household in period p can then be written as:

$$\ell(lm_p, \widetilde{Tcc_p}|\widetilde{w_p}, h) = \int\int\int \ell(lm_p, Tcc_p|w_p, h) \cdot \\ \mathcal{F}(\epsilon_{p,m}) \cdot \mathcal{F}(\epsilon_{p,f}) \cdot F(u_p)\, d\epsilon_{p,m} d\epsilon_{p,f} du_p. \tag{1.13}$$

The likelihood of an individual household trajectory given the full set of time-varying characteristics $\widetilde{s} = (\widetilde{s_1}, \widetilde{s_2})$, which includes $\widetilde{w}$, and unobserved heterogeneity h is thus:

$$\ell(lm, \widetilde{Tcc}|\widetilde{s}, h) = \prod_{p=1}^{2} \ell(lm_p, \widetilde{Tcc_p}|\widetilde{s_p}, h). \tag{1.14}$$

Our object of interest is the joint distribution of unobserved heterogeneity $\ell(h|x)$, which we aim to recover from the observed household choices. We estimate this distribution conditional on a set of constant household characteristics, denoted x, whose choice we discuss later. The likelihood of observing a household's sequence of choices $(lm, \widetilde{Tcc})$ conditional on observed characteristics is given by the following expression:

$$\ell(lm, \widetilde{Tcc}|\widetilde{s}, x) = \int_{\boldsymbol{h}} \ell(lm, \widetilde{Tcc}|\widetilde{s}, h) \cdot \ell(h|x)\, dh. \tag{1.15}$$

Finally, our sample likelihood is the product of all individual likelihood contributions of the N households in our data:

$$\mathcal{L} = \prod_{n=1}^{N} \ell(lm^n, \widetilde{Tcc^n}|\widetilde{s^n}, x^n). \tag{1.16}$$

Joint distribution of unobserved heterogeneity. Zooming into the joint distribution of unobserved heterogeneity $\ell(h|x)$, it is the product of the marginal distributions of g, oth, and α, which we assume to be independent conditional on constant characteristics x:

$$\ell(\underbrace{g, oth, \alpha}_{=h}|x) = \ell^g(g|\boldsymbol{x}^g) \cdot \ell^{oth}(oth|\boldsymbol{x}^{oth}) \cdot \ell^{\alpha}(\alpha|\boldsymbol{x}^{\alpha}),$$

where $\boldsymbol{x}^g, \boldsymbol{x}^{oth}$, and $\boldsymbol{x}^{\alpha}$ are subsets of x that are allowed to have some overlap. Such overlap creates correlations between the marginal distributions ℓ^g, ℓ^{oth}, and ℓ^{α} without assuming an explicit correlational structure.

We assume the underlying data-generating process of each type of heterogeneity, $het \in \{g, oth, \alpha\}$, to have the following functional form:

$$het^{\mu} = \gamma^{het} + \boldsymbol{x}^{het}\boldsymbol{\beta}^{het} + u^{het}, \tag{1.17}$$

where $\boldsymbol{x}^{het} \subseteq x$ denotes the vector of constant characteristics used in the estimation of the heterogeneity type het. We assume that the error term u^{het} follows a normal distribution with mean zero and standard deviation σ^{het}. Furthermore, we assume that the three errors u^{g}, u^{oth}, and u^{α} are mutually independent. Hence, het^{μ} is normally distributed conditional on $\boldsymbol{x}^{het}$ with a covariate-dependent mean $\gamma^{het} + \boldsymbol{x}^{het}\boldsymbol{\beta}^{het}$ and a covariate-independent standard deviation σ^{het}:

$$het^{\mu}|\boldsymbol{x}^{het} \sim \mathcal{N}\left(\gamma^{het} + \boldsymbol{x}^{het}\boldsymbol{\beta}^{het}, \left(\sigma^{het}\right)^{2}\right). \tag{1.18}$$

Each dimension of heterogeneity, g, oth, and α, is defined on the closed interval $[0, 1]$ as set up in Section 1.4. Therefore, we truncate the normal distribution of $het^{\mu}|\boldsymbol{x}^{het}$ at 0 and 1. The parameter μ simply allows for a wider set of functional forms and is set to $\frac{1}{3}$ for α and g and to 1 for oth.

Our maximum likelihood procedure will consequently estimate the parameters (γ^{g}, $\boldsymbol{\beta}^{g}$, σ^{g}, γ^{oth}, $\boldsymbol{\beta}^{oth}$, σ^{oth}, γ^{α}, $\boldsymbol{\beta}^{\alpha}$, σ^{α}) which maximize the sample likelihood function given in equation (1.16).[31]

Sample. We conduct the maximum likelihood estimation with data from the GSOEP covering 2012 to 2017 (see Section 1.3.2 for details). For every household, we convert the data from the six years into two corresponding model periods. Further details on this process can be found in Appendix A1 and summary statistics on the sample are presented in Appendix-Table A.1.

We furthermore restrict the sample to females who have at least one child of any age. This leaves us with an estimation sample of 2,178 households. 1,073 of these households face some childcare need in at least one of the two periods, as a child below age 9 lives in the household in the respective period. The other half of the sample does not face a childcare need in either period as their children are aged 9 or older. Nonetheless,

[31] The number of estimated parameters in the likelihood function depends on the number of constant characteristics $\boldsymbol{x}^{het}$.

the subsample with older children plays an important role in the identification of the unobserved heterogeneity, as we discuss next.

Identification. In the absence of a formal proof, we provide an intuition for the identification of the time-invariant parameters that govern the joint distribution of unobserved heterogeneity. The identification is conditional on the calibrated and reduced form regression inputs (Sections 1.5.1, 1.5.2, 1.5.3), the homogeneous preference parameters (Section 1.5.4), and the previously described assumptions for our maximum likelihood procedure. The distribution of $h = (g, oth, \alpha)$ will be jointly and set identified. There are three interacting ingredients that identify the time-invariant unobserved heterogeneity: i) cross-sectional variation in choices conditional on the same observed states, ii) the longitudinal dimension of our panel data, iii) using data not only from households with small children, but also from those with older children. The following paragraphs describe the three ingredients in more detail.

First, we observe households making different choices conditional on the same observed states s and constant characteristics x. Within our model, these differences in choices are therefore driven by differences in unobserved heterogeneity h. For illustration purposes, consider the example of a household with a single child aged 0 – 2 and a part-time working mother that buys 20 hours of public childcare. From this household's choices in isolation, oth is identified to be ≤ 0.5 (≤ 20 hours), as otherwise the household would buy less public childcare. Nonetheless, oth is only set identified: For a given preference for leisure α, the above choices could result from different combinations of g and oth. A low preference for domestic childcare g relative to leisure α would be consistent with oth close to 20 hours, i.e., the mother consumes leisure and does not provide much domestic childcare. On the contrary, a high preference for domestic childcare g relative to leisure α would be consistent with oth close to 0 hours, i.e., the mother spends a lot of time on domestic childcare and little on leisure.

Now, let us consider variation in the two choices which helps to identify the distribution of unobserved heterogeneity: i) A higher amount of public childcare bought implies a higher preference for leisure, lower preference for domestic childcare and decreases the upper limit of the amount of informal childcare. ii) A decrease in the amount of public childcare bought implies that the household's informal childcare use oth is strictly positive because otherwise the household would be unable to cover the childcare need while the mother works part-time. iii) If the mother were to work full-time, that would imply a lower preference for leisure, a lower preference for domestic childcare, and would point-identify oth at 20 hours. (iv) If the mother would be not working, that would reflect a higher preference for leisure without necessarily affecting g and oth as the household still consumes 20 hours of public childcare.

Turning to the second ingredient, using panel data is crucial for two reasons: i) The longitudinal dimension of the data and the associated temporal variation strongly facilitates identification because it allows to disentangle temporary shocks from the time-invariant unobserved heterogeneity. ii) Changes in family composition over time also affect which dimension of heterogeneity matters in which period: Consider a household in which a child below 9 is present in one period but not in the other, i.e., either a new child is born in the second period or the youngest child is between 6 and 8 in the first period. Then, the preference for leisure (α) helps to explain the choices in both periods, whereas the preferences for domestic childcare (g) and the availability of informal childcare (oth) help to identify the choices while a child that requires childcare is present. In addition, deterministic changes in the family composition, i.e., when at least one child between 0 and 8 is present in both periods, also facilitate the joint identification of g, oth, and α.

Third, the estimation sample also includes households who have children without childcare need (age 9 and above) in both periods. For these, the only unobserved heterogeneity that matters is the preference for leisure α, which explains the variation in their labor market choices conditional on wages and other observed characteristics. Hence, this group adds significantly to the identification of the distribution of α, independently of g and oth.

The combination of all three just described ingredients allows us to credibly identify the joint distribution of g, oth, and α.

Constant characteristics x. The fact that we allow the joint distribution of unobserved heterogeneity h to differ by constant characteristics x is important in multiple ways: The characteristics in x allow us to capture that subgroups in our data may have very different preferences, thereby improving the capability of our model to predict the behavioral patterns in the data. In addition, we select constant characteristics x to address the initial conditions problem, i.e., that the time-invariant joint distribution of unobserved heterogeneity might have affected the initial values of our time-varying state variables. We include all covariates in x as indicator variables. Thereby, they act as mean shifters in the data-generating process of each type of heterogeneity as introduced in equation (1.17). Summary statistics of the variables in x can be found in Appendix-Table A.1.

First, we estimate the distribution of the preference for domestic childcare ℓ^g conditional on indicator variables for living in East Germany and for being Catholic at age 20. Both variables are expected to affect the distribution as the prevalent social norms regarding domestic childcare are potentially very different in these populations. Additionally, we include an indicator for having primarily worked in a 'demanding occupation', i.e., an occupation with a high share of interactive non-routine tasks.[32] This last constant characteristic is motivated by Adda, Dustmann, and Stevens (2017), who have shown that women select themselves out of analytical jobs if they prefer to spend time with their children.

Second, we estimate the distribution of the availability and preference for informal childcare ℓ^{oth} separately for East and West Germany. Social norms about childcare differ significantly between both regions, which likely affects the availability of grandparental childcare.[33] Therefore and for computational reasons, we only estimate the mean of ℓ^{oth} for East Germany and fix the standard deviation to 1, as this allows for greater numerical stability in the face of very little variation.[34] For households living in West Germany, we do estimate the standard deviation and additionally include maternal education and whether the household lives in an urban area in $\boldsymbol{x}^{oth}$.

Third, we estimate the distribution of leisure preferences ℓ^α conditional on maternal education and a 'demanding occupation'. Both variables are included to control for the potential initial conditions problem in wages.

Results. The estimated coefficients from our maximum likelihood estimation (MLE) are shown in Table 1.4.[35] Figure 1.6 shows the implied cumulative distribution functions of g, oth, and α.

First, considering the preference for domestic childcare (g), Table 1.4 shows that women who live in former East Germany have lower preferences for domestic childcare than those in former West Germany, as reflected by $\beta_{\text{east}} < 0$. Additionally, if the mother was Catholic at age 20, we observe a higher preference for domestic childcare, in line

[32] Specifically, we code an occupation as a 'demanding occupation' if the share of interactive non-routine tasks is greater than one-third. We use the task classification of 3-digit occupations by Dengler, Matthes, and Paulus (2014).

[33] See, e.g., Hank, Tillmann, and Wagner (2001) for a discussion on differences in institutional childcare and childcare norms.

[34] The implied distributions are effectively identical for other values of the standard deviation. They can be characterized as close to a corner solution, which can be represented by a number of different mean and standard deviation combinations in a very similar fashion.

[35] See Appendix A3.5 for details on the optimization routine and Appendix A3.6 for details on the sensitivity of the estimates.

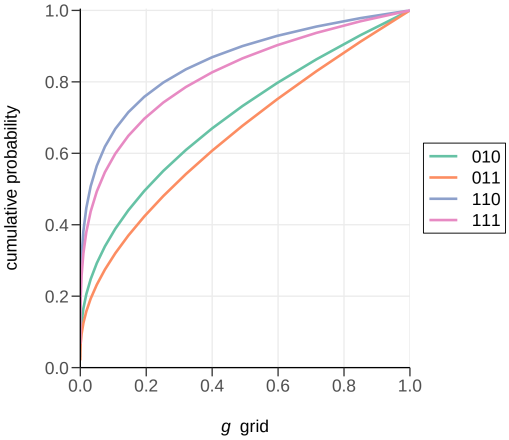
with the prior that Catholics have more conservative views regarding childcare. This is also visible in Figure 1.6a: The cumulative distribution function of Catholic mothers living in West Germany is the orange line. It features the lowest mass at low values of g, i.e., on a low preference for domestic childcare. Overall, the distinction between East and West Germany is the more important factor and both distributions for West Germany first-order stochastically dominate the East distributions.

Table 1.4: Maximum likelihood estimates

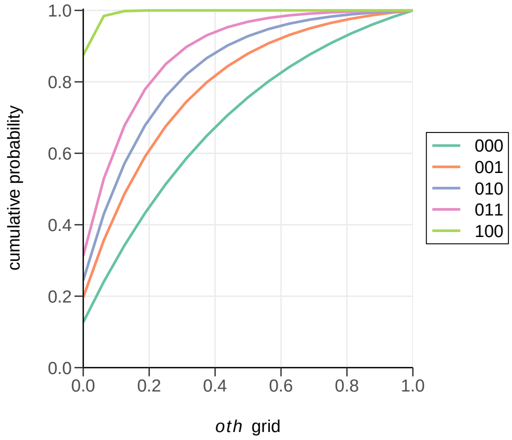
	domestic childcare (g)	avail. informal childcare (oth^{west})	avail. informal childcare (oth^{east})	leisure (α)
γ	1.37	−7.12	−33.25	1.63
β_{east}	−2.25			
$\beta_{\text{demanding occupation}}$	0.01			−0.22
β_{catholic}	0.58			
$\beta_{\text{high education}}$		−10.44		−0.28
β_{urban}		−6.27		
σ	1.00	2.00	$1.00^{\dagger}$	0.55

Notes: 'east' indicates having lived in former East Germany at some point; 'demanding occupation' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine; 'catholic' indicates being Catholic at age 20; 'high education' indicates having obtained at least an A-level; 'urban' indicates living predominantly in an urban area. See Appendix A3.6 for an illustration of the sensitivity of the estimates. † fixed to 1 for computational reasons.

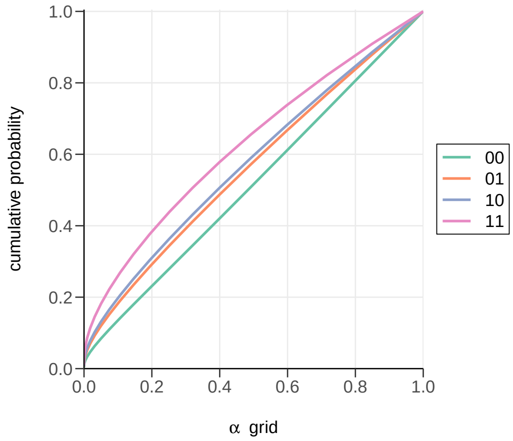
Second, we find large differences in the distribution of the availability of informal childcare (oth) between East and West Germany. As visible in Figure 1.6b, only very few households in East Germany rely on informal childcare (light green line). Generally, the large differences between East and West Germany point to cultural differences in the use (and availability) of grandparental and other informal childcare. Focusing on West Germany, the coefficients in Table 1.4 show that lower educated mothers as well as those not living in urban areas have a higher availability of informal childcare. Hence, as also illustrated in Figure 1.6b, households with highly educated mothers living in urban areas rely only little on informal childcare.



(a) domestic childcare preference
covariates: east, demanding occupation, catholic



(b) informal childcare availability
covariates: east, high education, urban



(c) leisure preference
covariates: demanding occupation, high education

Figure 1.6: Marginal cumulative distribution functions of the unobserved heterogeneity

Notes: The legend of each subfigure indicates if the respective indicator – in the same order as the covariates listed below the subfigure – is 0 or 1. In case of the preferences for domestic childcare, we omit the plots for *no* demanding occupation to facilitate the illustration because the estimated difference by demanding occupation is very small. 'east' indicates having lived in former East Germany at some point; 'demanding occupation' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine; 'catholic' indicates being Catholic at age 20; 'high education' indicates having obtained at least an A-level; 'urban' indicates living predominantly in an urban area.

Third, we turn to the marginal distribution of the preference for leisure (α), which we estimate conditional on maternal education and occupational characteristics. Both covariates matter almost equally, as shown in Table 1.4: being highly educated or having primarily worked in a demanding occupation both imply a lower preference for leisure. The resulting cumulative distribution functions in Figure 1.6c reflect this as well. Those who are highly educated and have worked primarily in a demanding occupation have the highest mass at low values of α.

1.6 Model Fit

Based on the just presented estimated joint distribution of unobserved heterogeneity and the calibrated parameters of the utility function, we now turn to the model fit. Specifically, we evaluate the ability of the estimated structural model to match data moments for the two choices, female labor supply lm and total use of public childcare Tcc, as well as empirical estimates of the participation elasticity and the Hicksian (compensated) elasticity of total hours from Chetty, Guren, et al. (2011).

Table 1.5: Model fit for labor supply

	Children 0 – 2			Children 3 – 5		
	NP	PT	FT	NP	PT	FT
Model	0.50	0.40	0.10	0.16	0.61	0.23
Data	0.55	0.40	0.05	0.17	0.65	0.18
	Children 6 – 8			**Children 9+**		
	NP	PT	FT	NP	PT	FT
Model	0.15	0.62	0.23	0.13	0.63	0.24
Data	0.13	0.68	0.19	0.16	0.54	0.30

Notes: PT and FT denote the female working part-time and full-time, respectively. NP denotes not working. Sample as defined in Section 1.5.5. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

First, Table 1.5 shows the model and data moments for female labor supply by the age of the youngest child. Overall, we are able to achieve a good fit of the labor supply patterns of mothers conditional on child age. In particular, the model matches the observed increase in participation and hours once the youngest child turns 3. Our model is also able to match the labor supply pattern of mothers with completed fertility, i.e., children aged 9 or older. Splitting up the fit by the hourly wage of the mother and father in Appendix-Figures A.8a and A.8b affirms that the conclusions above also hold across the wage distribution.

Next, we evaluate the fit with respect to the total public childcare take-up of households by the age of their youngest child. Families with the same age of the youngest child might face a different total childcare need as they might or might not have older children. To take this into account, we normalize each household's total public childcare take-up by the household's total childcare need. This yields the share of childcare need that a household covers with public childcare, which we denote as $m(Tcc)$.[36]

Table 1.6: Model fit for total public childcare take-up

	Children 0 – 2		
	$m(Tcc) \leq \overline{m}_1$	$\overline{m}_1 < m(Tcc) \leq \overline{m}_2$	$\overline{m}_2 < m(Tcc)$
Model	0.46	0.34	0.20
Data	0.44	0.40	0.16
	Children 3 – 5		
	$m(Tcc) \leq \overline{m}_1$	$\overline{m}_1 < m(Tcc) \leq \overline{m}_2$	$\overline{m}_2 < m(Tcc)$
Model	0.00	0.53	0.47
Data	0.02	0.52	0.46
	Children 6 – 8		
	$m(Tcc) \leq \overline{m}_1$	$\overline{m}_1 < m(Tcc) \leq \overline{m}_2$	$\overline{m}_2 < m(Tcc)$
Model	0.33	0.17	0.50
Data	0.29	0.23	0.48

Notes: $m(Tcc)$ denotes the share of total childcare need that is covered through public childcare. $\overline{m}_1$, respectively $\overline{m}_2$, indicates that the household would cover a share of 33%, respectively 75%. Sample as defined in Section 1.5.5. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

[36] For example, consider a household with two children, the first being between 0 and 2 and the second between 3 and 5. Their total childcare need is 80 hours per week. If the older child goes to Kindergarten for 30 hours per week, while the younger child attends nursery for 30 hours per week, the household will cover 80% of the total childcare need by public childcare ($\frac{60}{80} = 0.8$).

Table 1.6 summarizes three public childcare take-up moments by the youngest child's age: i) the share of households who cover less than 33% of their total childcare need with public childcare ($m(Tcc) \leq \overline{m}_1$), ii) those who cover between 33% and 75% with public childcare ($\overline{m}_1 < m(Tcc) \leq \overline{m}_2$), and iii) those who cover more than 75% with public childcare ($\overline{m}_2 < m(Tcc)$). In summary, we achieve a good fit for all child age brackets. The fit is even almost perfect for children between 3 and 5. Again, splitting up the fit by the hourly wage of the mother and father in Appendix-Figures A.9a and A.9b affirms that we match the overall dynamics across the wage distribution.

Third, we evaluate two labor supply elasticities: the participation elasticity and the Hicksian (compensated) elasticity of total working hours. Getting the female labor supply responses right is crucial for the credibility of our policy experiments in Section 1.7. To obtain the model counterparts for the empirical estimates, we simulate a permanent increase in the gross female wage rate by 1% and calculate the labor supply responses along the participation and total hours margin. The resulting participation elasticity is 0.16, which is within the range of estimates for women in the quasi-experimental literature.[37] In terms of the Hicksian (compensated) elasticity of total working hours, our model prediction is 0.32. This estimate is smaller than the meta-study aggregate of 0.59 reported in Chetty, Guren, et al. (2011), but the aggregate stems from a wide range of estimates of which some are well in line with our results.[38]

1.7 Policy Experiments

We now use the results from the empirical estimation of our structural model to simulate different childcare policy reforms. Our estimates of the tastes for domestic childcare and leisure as well as access to informal childcare allow us to predict the distribution of responses to counterfactual policy experiments. For each counterfactual policy reform, we are able to simulate which mothers will adjust their labor supply or public childcare take-up and how much these behavioral adjustments will contribute to government revenue. We can also quantify these effects across the income distribution and assess the extent of redistribution afforded by the progressivity of the childcare fee schedule.

We start by evaluating the net fiscal effects of changes in childcare policies. The idea is that some (or all) of the initial cost of subsidizing public childcare is offset by additional

[37] The range reported in Chetty, Guren, et al. (2013) for (non-single) women is 0.15 to 0.3.

[38] See Chetty (2012) for details on the full range.

revenue from income tax payments of mothers who decide to work more as a result of the policy ('marginal' mothers). Some of this additional income tax revenue will be perceived in the current period. Thereafter, some more revenue will be generated in the next periods since our model includes a positive return on labor supply in terms of earnings potential and a positive impact of earnings potential on the incentives to work.

We will evaluate two policies: First, the large public childcare expansion for 0 – 2 year old children in the late 2000s. Second, a permanent increase in childcare subsidies from the current levels for all child age groups. We consider both untargeted subsidy increases as well as work-contingent and full-time-contingent subsidies.

To disentangle the fiscal effects, it is useful to separate the following three contributions: First, *direct effects* are fiscal effects that arise from behavioral adjustments in response to changes in public childcare availability or fees, while the household has children in an age range affected by the policy. Second, *anticipation effects* capture fiscal effects from behavioral adjustments in response to changes in public childcare availability or fees while the household has no children in a policy affected age range, but may have in the future. Third, *dynamic wage effects* contain fiscal effects through higher wages and behavioral adjustments in response to higher wages, where higher wages are the result of past behavioral changes. In our policy simulations, we decompose the total fiscal effect into the dynamic wage effect and the direct plus anticipation effect to work out the importance of the former.

For both policies, we distinguish two ways to answer our postulated question on the degree of self-financing:

i) We focus on the amount of subsidy spent and income taxes received in the model period of the policy introduction and call this perspective *Impact period*. This perspective includes both direct as well as anticipation effects, but no dynamic wage effects (as there are no past behavioral changes yet).
ii) We simulate the remaining life cycle starting from the observed age, wages, and family composition of the household and derive the discounted net fiscal effect. As opposed to i), this accounts for dynamic wage effects and we refer to this perspective as *All periods*. Note that this perspective includes the effects from the *Impact period* and also contains additional direct and anticipation effects from childbirths after the model period of the policy introduction.

We then turn to examining the redistributive impact of the variations in childcare fees with household income by comparing the childcare fee schedule to the income tax schedule. Building on the thought experiment of Okun's leaky bucket, we quantify the marginal excess burden of both schedules: For each Euro taken from above-median income households, how much reaches those with below-median income and how much is lost due to lower labor supply incentives? Furthermore, we map out how the marginal excess burdens are driven by the labor supply elasticities implied by our estimated model and thereby illustrate the economic mechanisms at play. Finally, we discuss potential policy conclusions and present a possible reform of the childcare fee schedule.

1.7.1 Public childcare expansion for 0 – 2 year olds

As laid out in Section 1.3.1, the mid-2000s were characterized by substantial rationing of public public childcare slots for 0 – 2 year old children in former West Germany.[39] A large expansion policy was started in 2005 and effectively ended the rationing by approximately 2012.[40] To evaluate this policy, we limit the sample to West Germany and introduce in our model that a share of households does not have access to any public childcare for 0 – 2 year olds. Comparing this counterfactual to the non-rationed observed state of the economy therefore allows us to calculate the self-financing degree of the public childcare expansion.

However, one potential issue with this approach is that we may overestimate additional tax revenues for the following reason: The estimated (low) levels of informal childcare availability allow only very few mothers of 0 – 2 year olds to work in the rationed counterfactual. A comparison to the non-rationed baseline would therefore show large increases in participation rates and tax revenues. Aggregate data from 2005, however, shows employment rates for mothers of 0 – 2 year old children of 30%, despite very low public childcare enrollment rates of about 7.5% (see Appendix-Figures A.2a and A.2b in the Appendix).

This implies that, when public provision of childcare was rationed, a sizeable share of mothers of 0 – 2 year olds were relying on informal childcare, as no private market for childcare existed.[41] When more public childcare slots became available as rationing came to an end, many mothers started sending their children to childcare facilities instead of relying on informal care arrangements. These mothers therefore did not contribute

[39] There was no rationing in former East Germany due to persistent institutional structures from the socialist era, see Section 1.3.1 for details.

[40] See Appendix A2.3 for a discussion of the remaining gap between childcare demand and childcare enrollment for 0 – 2 year olds.

[41] See Section 1.3.1 for details on the German childcare institutions.

additional tax revenue since they were already working, but still increased their take-up of subsidized public childcare. This leads to a lower degree of self-financing of the policy compared to the case if we were to assume that only our estimated (low) levels of informal childcare were available.

However, the switch from informal to public childcare just described is a transition that goes beyond the capabilities of our model, which is best suited to capture the current institutional environment.[42] Nevertheless, we believe that it is insightful to study the ending of rationing with a slightly modified version of our model. This perspective allows us to contrast the effects of expanding access to public childcare with the effects of changes in the current fee schedule, which we focus on in Sections 1.7.2 and 1.7.3.

Therefore, we add the following component to our model: Under rationing, each mother has access to a fallback option for childcare, additionally to the estimated oth distribution. This fallback childcare is free and uniformly distributed, but is fully crowded out once public childcare is available.[43] Thereby, it allows mothers to work in the rationed environment, but it does not play a role for the decision making in the non-rationed environment.

Within this modified framework, we calibrate the availability of fallback childcare to 17.5h/week and the share of households without access to public childcare to 85%. This allows us to closely match the 2005 employment and childcare enrollment rates.[44] Adding the fallback childcare component therefore represents a minimally invasive extension to be able to capture the observed transition between modes of childcare from a heavily rationed to a non-rationed environment. The simplistic extension thereby achieves our main goal of avoiding an overestimation of the fiscal effect of the expansion due to an overestimation of the labor supply effects. Furthermore, the implied increases in the employment rate (+15pp) and public childcare use (+33pp) from the rationed to the non-rationed scenario are also consistent with quasi-experimental evidence: We find that about 45% of mothers who start using public childcare also start to work. This matches

[42]Specifically, this transition violates two core assumptions of our model: i) informal childcare availability is time-constant, and ii) households always prefer informal childcare over public childcare.

[43]We think of it as an additional availability of, e.g., grandparents in the face of no other alternative mode of childcare. However, public childcare is preferred to this fallback option at all levels of income or grandparents themselves may have scaled back their involvement in childcare as a viable alternative became available.

[44]Our modified model predicts an employment rate of 30% for mothers with children aged 0 – 2 and childcare enrollment of 0 – 2 year olds of 6%, well in line with the historical targets of 30% and 7.5% respectively.

well with quasi-experimental evidence from Bauernschuster and Schlotter (2015) for an earlier reform.[45]

Table 1.7: Self-financing degree of the public childcare expansion

	total	female hourly wage		
		≤ €15	€20	≥ €25
Impact period	71.1%	38.2%	82.3%	170.1%
All periods	103.3%	63.9%	120.9%	212.5%

Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiment: ending rationing of public childcare for 0 – 2 year olds.

Results. The main results on the self-financing degree of the public childcare expansion are presented in the first column of Table 1.7. For the *Impact period*, the increased tax revenue from mothers increasing their labor supply due to the availability of public childcare already makes up for 71.1% of the increased spending on childcare subsidies. Considering *All periods*, the self-financing degree increases to 103.3%, implying that the public childcare expansion for 0 – 2 year olds was fully self-financing when taking into account the effects from future periods. Given that 55% of mothers increased their public childcare usage without reacting on the labor supply margin, this implies that the average mother who started to work paid taxes that are around twice the average subsidy paid per childcare slot.

Splitting up the effects by initial female wages, the additional columns in Table 1.7 reveal substantial effect heterogeneity across the wage distribution. Focusing on mothers with wages of ≤ €15 per hour, the expansion of childcare was only 38.2% self-financing in the *Impact period*. In comparison, it was already 82.3% self-financing for those who earn around €20 per hour and even 170.1% self-financing for mothers with wages above €25 per hour. These effects increase to 63.9%, 120.9%, and 212.5%, respectively, for *All periods*,

[45]Going from the rationed to the non-rationed scenario, participation of mothers of 0 – 2 year olds increases from 30% to 45%, while public childcare take-up increases from 6% to 39%. Bauernschuster and Schlotter (2015) study an earlier expansion of public childcare for 3-year-olds in Germany and find that 35% of the increased public childcare take-up stems from mothers taking up work.

illustrating that even for those earning a wage of around €20, the increased government revenue from taxes recoups more than 100% of the spending on subsidies.

In Figure 1.7b we present a decomposition based on the three types of effects introduced at the beginning of this section, namely the direct and anticipation effect versus the dynamic wage effect. Based on the *All periods* perspective, we capture the dynamic wage effect as the share of the total self-financing degree that can be attributed to changes in the wage distribution. Throughout the wage distribution, Figure 1.7b illustrates that the dynamic wage effect makes up around one-third of the total net fiscal effect of the public childcare expansion. Overall, the childcare expansion only becomes fully self-financing once dynamic wage effects are taken into account.

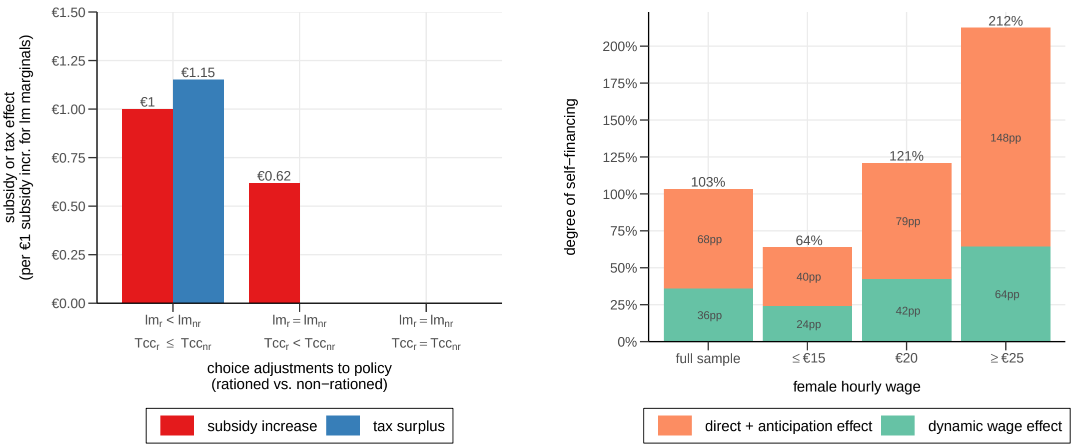
These results also underscore that expanding public childcare generates high fiscal returns throughout the wage distribution and that the effects are not just driven by high wage individuals. This affirms that our main conclusions on the fiscal effects of the expansion are unlikely to be driven by the specific assumptions of our model extension to incorporate rationing.[46]

To better understand the underlying drivers of the self-financing rates, Figure 1.7a illustrates the contributions of different groups of individuals to the 71.1% *Impact period* self-financing degree. While some mothers do not respond to the policy, others change labor supply and use of public childcare. We refer to the group who changes their labor supply as 'marginal' mothers. Those mothers who use public childcare and thereby consume subsidies but do not change their labor supply due to the reform are referred to as 'inframarginal'.

The bars in Figure 1.7a show the magnitude of the increase in childcare subsidy spending and income tax revenue for each subgroup, normalized in relation to a one Euro subsidy increase for those who are marginal with respect to their labor supply (red bar on the left). Focusing on the marginal mothers (blue bar, $lm_r < lm_{nr}$), the policy generated €1.15 in tax surplus per Euro spent on subsidies, which makes it clearly self-financing for this subgroup even when focusing only on the *Impact period.* However, the overall self-financing degree is negatively impacted by the subgroup which only increases their public childcare take-up, but does not increase their labor supply ($lm_r = lm_{nr}$, $mcc_r < mcc_{nr}$). For every Euro spent on subsidies for the marginals, the government has to

[46] The assumption of fallback childcare to be uniformly distributed affects who is marginal or inframarginal w.r.t. the labor supply decision. Even with a distribution of fallback childcare skewed to high wage mothers, which would make them less likely to be marginal, the policy would still generate high fiscal returns.

spend 62 cents on subsidies for the inframarginals, which leads to the *Impact period* self-financing degree of 71.1%.



(a) *Impact period* (b) *All periods*

Figure 1.7: Results decomposition for the public childcare expansion

Notes: Decomposition of the effects from Table 1.7. lm_r and Tcc_r denote the labor supply and public childcare decision in the rationing scenario. lm_{nr} and Tcc_{nr} denote the labor supply and public childcare decision in the non-rationing scenario. *All periods* decomposed into: i) *direct effects + anticipation effects*, fiscal effects from behavioral adjustments in response to changes in the public childcare availability or fees while the household has children/may have children in the future in a policy affected age range. ii) *dynamic wage effects*, fiscal effects from higher wages and behavioral adjustments in response to higher wages, where higher wages are the result of past behavioral changes. Decomposition by initial wages at policy introduction. Values within the bars in Figure 1.7b denote percentage point contributions.

1.7.2 Childcare subsidy increases

After evaluating the expansion of public childcare slots for a given degree of subsidization of the unit price of childcare, we now turn to quantifying the fiscal externalities of small increases in the subsidy. This exercise investigates to which degree self-financing attribute of childcare subsidies holds for increases from the current level, keeping in mind that in the status quo childcare is already heavily subsidized (see Section 1.5.1).

Untargeted childcare subsidy increase. We first study an untargeted increase in the subsidization of the hourly childcare fee that is equivalent to increasing monthly subsidies for full-time public childcare (40h/week) by €50. Exemplified for a household with only a

0 – 2 year old and gross earnings of €4,000, this translates into a 21.5% decrease in the hourly childcare fee from €1.4 to €1.1.[47]

Table 1.8: Self-financing degree of childcare subsidy increases

		female hourly wage		
	total	≤ €15	€20	≥ €25
(a) *untargeted +€50*				
Impact period	4.1%	3.4%	5.4%	4.3%
All periods	5.9%	5.3%	7.1%	6.2%
(b) *work-contingent +€50*				
Impact period	8.7%	7.4%	11.9%	8.0%
All periods	12.4%	11.0%	15.4%	11.6%
(c) *full-time-contingent +€50*				
Impact period	33.5%	29.6%	40.4%	33.6%
All periods	50.1%	48.3%	54.0%	47.2%

Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiments: increase in the subsidization of the hourly childcare fee that is equivalent to increasing monthly subsidies for full-time public childcare (40h/week) by €50.

Panel (a) in Table 1.8 shows that for such an untargeted increase in childcare subsidies increased tax revenues only make up for 5.9% (4.1%) of the increased subsidy spending considering *All periods* (*Impact period*). We find the fiscal effects to be similar across both time perspectives for all three wage levels and illustrate in Appendix-Figure A.11a that the dynamic effects contribute approximately one-third.[48]

[47] Plugging in €4,000 and 40h of childcare need for a 0 – 2 year old into the childcare fee schedule from Appendix-Table A.2.

[48] Appendix-Figure A.11 illustrates the importance of the dynamic wage effects for the self-financing degrees of the three policy experiments that increase childcare subsidies. Comparing these contributions across the wage distribution, they are decreasing in wages for all experiments, i.e., they are always highest for ≤ 15 (up to almost 50%) and lowest for ≥ 25. This highlights that while low wage individuals do not generate high immediate tax revenues from working when their children are young, their dynamic returns to experience generate substantial additional future tax revenue.

The underlying driver for the low self-financing rate is the very high share of inframarginal mothers, who receive a windfall gain through higher subsidies but whose labor supply is unaffected.[49] Decomposing the *Impact period* effect by the choice margins in Appendix-Figure 1.7a, we find that for every Euro spent on marginal mothers, the government has to pay €40.64 to inframarginal mothers. This is barely counterbalanced by €1.68 in additional tax revenue from marginal mothers who change their labor supply. The small and untargeted increase in the subsidy is therefore self-financing for those who are induced to increase their labor supply, but the substantial spending on inframarginals makes the policy very costly. The latter is also the main difference to the results for the public childcare expansion in Section 1.7.1: Figure 1.7a showed a rather low spending on inframarginals for the ending of rationing, which was a fully self-financing policy. Appendix-Figure A.10a, on the other hand, illustrates that the spending on inframarginals outweighs the spending on marginals by more than a factor 40 for an untargeted increase in subsidies from the current level.

Work-contingent childcare subsidy increase. Second, we study the effect of targeting a similar increase in the subsidization of the hourly childcare fee to households in which the mother works part-time or full-time.[50] The targeting works through two angles: i) Since the additional subsidy is not available to non-working mothers, the amount spent on mothers who are inframarginal in their labor supply decision decreases. ii) The work-contingent policy increases the incentives for mothers to enter the labor market through a reduction in the costs of taking up employment. Both angles contribute to a higher self-financing degree.

Panel (b) in Table 1.8 shows that in the *Impact period* 8.7% of childcare subsidies are refinanced through additional income tax revenues. This number increases to 12.4%, when we simulate household behavior over *All periods*, with dynamic effects contributing again around one-third (see Appendix-Figure A.11b). Splitting up the effect by female wage levels, the €20 category stands out slightly, with effects that are 3 to 4 percentage points higher than the other two wage levels.

The higher self-financing rate compared to the untargeted subsidy increase is mostly driven by lower subsidy spending on inframarginals. Appendix-Figure A.10b shows that the additional subsidy spending on inframarginals is reduced to €23.14 per Euro spent on

[49] For every marginal mother, the government has to pay increased subsidies to more than 300 inframarginal mothers.

[50] The increase is again equivalent to increasing monthly subsidies for full-time public childcare by €50).

marginals. Simultaneously, increased labor supply incentives also play an important role, as the share of marginal mothers increases and the tax surplus generated by them for every Euro in subsidies rises to €2.10. Despite the high self-financing degree of ≥ 200% for the marginals, the spending on the inframarginals still outweighs the overall tax surplus substantially.

Full-time-contingent childcare subsidy increase. Third, we study an increase in the subsidization of the hourly childcare fee targeted to households in which the mother works full-time.[51] This policy effectively cuts out all households who may increase their use of public childcare to consume additional leisure and creates additional incentives to work full-time.

Panel (c) in Table 1.8 clearly shows the effectiveness of the targeting: The degree of self-financing increases to 34.1% when we consider the *Impact period* and even further to 50.1% for *All periods*. Comparing the effects across wage levels, the effects are largest at the medium wage level, reaching up to 54%, and the relative contribution of the dynamic effect is again around one-third (see Appendix-Figure A.11c).

In comparison to the previous two policies, Appendix-Figure A.10c shows that the full-time work contingency drastically decreases the additional subsidies paid to those who do not respond in terms of labor supply. For every Euro in additional subsidies spent on marginal mothers, the government has to spend just €4.92 on inframarginal mothers. Furthermore, the marginal mothers again generate a sizeable tax surplus of €1.99 per Euro of additional subsidies consumed by them. The lower spending on inframarginals together with the strong labor supply incentives make a full-time-contingent subsidy increase therefore a much less costly policy.

1.7.3 Redistribution through childcare fees

The existing childcare fee schedule is increasing in household income (see Figure 1.3). This implies that subsidies are falling with income and that the childcare fee schedule effectively redistributes within the childcare using population. Based on this observation, a number of policy relevant questions arise: How progressive should a childcare fee schedule be? Is it too progressive in Germany? Or should it rather be more progressive since the current schedule still implies substantial subsidies for high-income households? These are generally thorny normative questions that require to trade off the utility of

[51]The increase is again equivalent to increasing monthly subsidies for full-time public childcare by €50).

high and low-income households. For such purposes, it is typically required to specify a social welfare function. A common alternative is to work with social marginal welfare weights (Saez and Stantcheva 2016), which capture how much the government values one more (marginal) Euro in the hands of different income groups. However, the choice of these welfare weights is again a normative question.

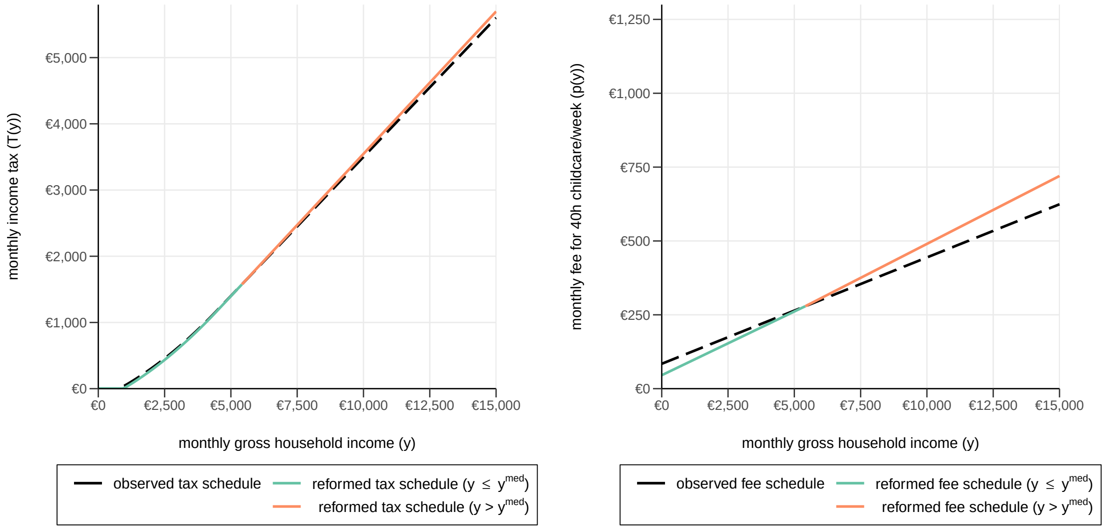
We circumvent the normative decision of how to weigh the utility of high and low-income households by making use of the so-called 'inverse-optimum' approach of optimal tax theory (Bourguignon and Spadaro 2012, Lorenz and Sachs 2016, Lockwood and Weinzierl 2016, Jacobs, Jongen, and Zoutman 2017). This allows us to quantify how society currently trades off equity and efficiency through the tax schedule. We then ask whether the childcare fee schedule should be more or less progressive if the same weights on equity (the distribution of the pie) and efficiency (the size of the pie) were applied.

Most of the inverse-optimum literature focuses on rather simple static models and can therefore consider fully non-linear tax reforms in the spirit of Mirrlees (1971) to quantify the welfare weights for all income levels. In our large dynamic model, such a granular analysis is not tractable. We therefore focus on a more coarse measure: We consider redistribution from above to below-median income households and quantify the relation of the average inverse-optimum weights between these two groups.

Progressive tax reform. For the quantification, we consider a simple parametric tax reform. We adjust the observed income tax schedule $T(y)$ in the following way leading to the reformed tax schedule $T^*(y)$:

$$T^*(y) = \begin{cases} T(y) + \tau_1^T \left(y - y^{med}\right) & \textit{for } y > y^{med} \\ T(y) - \tau_2^T(\tau_1^T) \left(y^{med} - y\right) & \textit{for } y \leq y^{med} \end{cases} \tag{1.19}$$

We fix τ_1^T, the tax increase for above-median income households ($y > y^{med}$), to 0.01 and choose τ_2^T, the tax decrease for below-median income households ($y \leq y^{med}$), such that the reform is government-budget neutral. We implement this parametric reform as a one-off reform: The reformed tax schedule is only applied in the *Impact period* and households are fully aware that the observed tax schedule will be applied in all future periods. To account for the dynamic effects of behavioral adjustments in response to the reform, we conduct all simulations until the end of the life cycle (*All periods* perspective). The resulting reformed income tax schedule is illustrated in Figure 1.8a.



(a) *Tax reform* (b) *Childcare fee reform*

Figure 1.8: Illustration of budget-neutral reforms to quantify the marginal excess burden

Notes: The income tax schedule representation in Figure 1.8a abstracts for illustration purposes from social security contributions and child benefits/allowances. For the childcare fee schedule in Figure 1.8b, we focus, for illustration purposes, on the case with a 0 – 2 year old and no siblings. The budget-neutral parameter combinations are $\tau_1^T = 0.01$ and $\tau_2^T = 0.00744$ as well as $\tau_1^c = 0.01$ and $\tau_2^c = 0.00714$.

We then calculate the marginal excess burden of the tax schedule based on the reform. It quantifies the relation of the average inverse-optimum weights between both groups in the following intuitive fashion: For each Euro that the government takes from above-median income households, how much reaches those with below-median income? In the absence of behavioral responses, the full Euro would reach below-median income households. However, individual decisions are endogenous with respect to the reform. As the reformed tax schedule is steeper throughout (see Figure 1.8a), this implies lower labor supply incentives for all households. Consequently, households reduce their labor supply which lowers their tax payments at the margin. This leads to the marginal excess burden.

For the tax schedule, we find that the marginal excess burden is 31.3 cents per Euro of revenue raised. Hence, for each Euro the reform takes from above-median income households, 68.7 cents reach the below-median income households and 31.3 cents are lost due to lower labor supply incentives. Note that this also tells us that society values 1 Euro in the hands of above-median income households as much as 68.7 cents in the hands of below-median income households. The average inverse-optimum weight of those with above-median income is 31.3% smaller than the average inverse-optimum weight of those with below-median income. Or phrased differently, redistribution from above to below-median income households is considered desirable if at least 68.7 cents reach the below-median income households.

Progressive childcare fee reform. We then apply the same type of reform to the childcare fee schedule, i.e., we adjust the observed childcare fee schedule $p(y)$ in the following way leading to the reformed childcare fee schedule $p^*(y)$:

$$p^*(y) = \begin{cases} p(y) + \tau_1^c \left(y - y^{med}\right) & \textit{for } y > y^{med} \\ p(y) - \tau_2^c(\tau_1^c) \left(y^{med} - y\right) & \textit{for } y \leq y^{med} \end{cases} \tag{1.20}$$

We again set $\tau_1^c = 0.01$ and choose τ_2^c such that the reform is government-budget neutral. The resulting reformed childcare fee schedule, which is illustrated in Figure 1.8b, is also steeper throughout. Marginal childcare fees increase for all income groups, which lowers the incentives to work just as the tax reform did. However, note that Figure 1.8b also shows that childcare fees per hour (holding income constant) decrease for below-median income and increase above. As will become clear below, this implies more complex behavioral reactions than the tax reform.

Based on the childcare fee reform, we find a marginal excess burden of 21.2 cents per Euro of revenue raised for the childcare fee schedule. This corresponds to 78.8 cents reaching below-median income households for each Euro the reform takes from above-median income households. Therefore, increasing redistribution through the childcare fee schedule would imply efficiency cost that are one-third lower than what society is willing to pay for redistribution through the income tax schedule. Applying the inverse-optimum weights found for the tax schedule would yield the policy recommendation that the childcare fee schedule should be made more progressive.

Economic mechanisms. To illustrate the underlying mechanics that drive the differences in the marginal excess burdens, we start with a look at the *Impact period* perspective (different to the just described results). This perspective allows us to carefully decompose the effects by the changes in labor supply choices in Table 1.9, similar to Figures 1.7a and A.10. Afterwards, we show that accounting for the dynamic effects, as done above, is crucial as altered labor supply choices have sizeable long-run effects via their persistent impact on the wage distribution.

We start with the tax schedule, for which we find an *Impact period* marginal excess burden of 20.5 cents, i.e., 20.5 cents are lost when redistributing an additional Euro through the tax system. This can be decomposed into 7.9 cents (38.5% of the marginal excess burden) originating from labor supply reductions of below-median income households and 12.5 cents (61%) due to labor supply reductions of above-median income households. Changes in the take-up of public childcare do not contribute to the marginal excess burden at all.

Table 1.9: Decomposition of the *Impact period* marginal excess burden

	tax schedule (marg. excess burden: 0.205)		childcare fee schedule (marg. excess burden: 0.172)	
	$lm \downarrow$	$lm \uparrow$	$lm \downarrow$	$lm \uparrow$
HH income < median	0.079	0.000	0.013	−0.005
	[38.5%]	[0.0%]	[7.6%]	[−3.2%]
HH income ≥ median	0.125	0.000	0.118	0.000
	[61.0%]	[0.0%]	[68.8%]	[0.0%]
	unchanged lm		unchanged lm	
Contribution of Tcc changes	0.001		0.046	
	[0.5%]		[26.8%]	

Notes: Behavioral reactions in response to budget-neutral reforms as set up in equations (1.19) and (1.20) ($\tau_1^T = 0.01$, $\tau_2^T = 0.00744$, $\tau_1^c = 0.01$, $\tau_2^c = 0.00714$), but simulated only for the model period of the reform (*Impact period*). Values in parentheses denote the contribution to the marginal excess burden.

Second, focusing on the childcare fee schedule, we find an *Impact period* marginal excess burden of 17.2 cents. Of these 17.2 cents, only 7.6% are due to lower labor supply of below-median income households (1.3 cents). Lower labor supply of the high-income households, on the other hand, makes up 68.8%, adding up to 11.8 cents. Additionally, some below-median income households increase their labor supply, which translates into a reduction of the marginal excess burden by −3.2%. The remaining component of the marginal excess burden (4.6 cents, 26.8%) can be attributed to changes in Tcc that are not accompanied by changes in labor supply.

To understand the intuition behind these results, first note that both reforms have in common that they increase the effective marginal tax rate on labor income: The tax reform increases the marginal tax rate $T'(y)$ and the childcare fee reform increases the marginal hourly price of childcare $p'(y)$ (see Figure 1.8). The increase in the effective marginal tax rate for a given amount of public childcare translates into decreased labor supply incentives for all income levels in both reforms. This effect in isolation leads to labor supply reductions and therefore losses in tax revenue. In the case of the tax schedule, it is the sole driver of the marginal excess burden, as illustrated in Table 1.9.

For the childcare fee schedule, however, there is an additional effect that plays an important role for the size of the marginal excess burden: the reform also changes the absolute hourly price of childcare $p(y)$, which also affects the decision of how to allocate time between work, leisure, and domestic childcare. Due to the reform, childcare fees (at a given income level) decrease for households with below-median income, while they increase for households with above-median income.

Our results in Table 1.9 show that the contribution of changed labor supply from below-median income households is much lower for the childcare fee reform compared to the tax reform (0.7 cents vs. 7.9 cents in Table 1.9). This illustrates that the two opposing labor supply effects for below-median income households almost cancel each other out. The revenue lost due to labor supply changes of households with above-median income, on the other hand, is very similar in both cases (around 12 cents): changes in the absolute price of childcare per hour do not play much of a role for above-median income households.[52] In summary, this shows that below-median income households are more likely to be marginal in their decision to work with respect to a change in the absolute price of childcare $p(y)$ than above-median households.

A final important difference between both schedules is that decreasing the absolute price of childcare per hour also affects the leisure vs. domestic childcare trade-off. Leisure becomes more expensive relative to domestic childcare for above-median income households and vice versa for below-median income households. As public childcare is heavily subsidized, any change its take-up also affects the government budget and thereby the efficiency cost of redistribution. Aggregating over above and below-median income, we observe that poorer households increase their Tcc take-up to consume leisure more than richer households decrease it, respectively. This yields the additional 4.4 cents in

[52]To think about an example, compare (high income) female doctors to (low income) female nurses: Doctors are unlikely to stop to use childcare and stop to work because the absolute price of childcare per hour increases. In contrast, a non-negligible share of nurses who did not work previous to the reform considers to start using childcare and work.

marginal excess burden from households with unchanged labor supply behavior, shown at the bottom of Table 1.9.

Dynamic effects. Based on these decompositions, we now turn to the role of the dynamic wage effects for the marginal excess burden. Recall that the marginal excess burdens including dynamic effects are substantially larger than the *Impact period* ones shown in Table 1.9: 31.3 cents vs. 20.5 cents for the tax schedule and 21.2 cents vs. 17.2 cents for the childcare fee schedule. This highlights the importance of accounting for dynamic wage effects when evaluating the efficiency cost of redistribution of both schedules.

Comparing the size of the dynamic effect between both schedules, it is substantially larger for the tax schedule. This difference can be traced back to the role that labor supply reductions play for the *Impact period* marginal excess burden, as these directly drive the dynamic effects via the wage process. Changes in labor supply account for 100% of the *Impact period* marginal excess burden in the tax schedule, but only for 73% in the childcare fee schedule. The remaining marginal excess burden contribution from households with unchanged labor supply (via Tcc changes), on the other hand, does not imply any dynamic effects.

Furthermore, Table 1.9 illustrates that especially lower labor supply from below-median households is an important contributor in the case of the tax schedule, but not in the childcare fee schedule. As the dynamic wage effects are especially large for low-income households,[53] this is an additional contributor to the large dynamic component of the marginal excess burden of the tax schedule.

Policy implications. Based on the results of the differing marginal excess burdens in the tax schedule and the childcare fee schedule, we can derive two alternative policy implications: First, a simple conclusion is that society has a weaker desire to redistribute from above-median to below-median income in the childcare-using population relative to the general population. In that case, it would be optimal to have a lower marginal excess burden in the childcare fee schedule.

Second, if the government has the same desire for redistribution in the childcare-using population as it has in the general population, then the childcare fee schedule should be more progressive. One potential reform would be to increase the progressivity of the childcare fee schedule using the simple parametric approach from equation (1.20). In

[53] See the effect decompositions from the previous sections in Figures 1.7b and A.11.

Figure 1.9, we implement such a reform that increases the income dependency of the hourly price of public childcare. Specifically, we increase the income dependency at the margin from around 3% to around 10% by setting $\tau_1^c = 0.07$ and redistribute the additional revenue via the corresponding budget-neutral τ_2^c.[54] The childcare fee schedule becomes considerably steeper, i.e., prices become considerably more progressive.[55] Nevertheless, the reformed schedule still implies a lower marginal excess burden of 29.6 cents than the tax schedule (31.3 cents). Hence, applying the inverse-optimum weights found for the tax schedule would suggest to implement this reform.

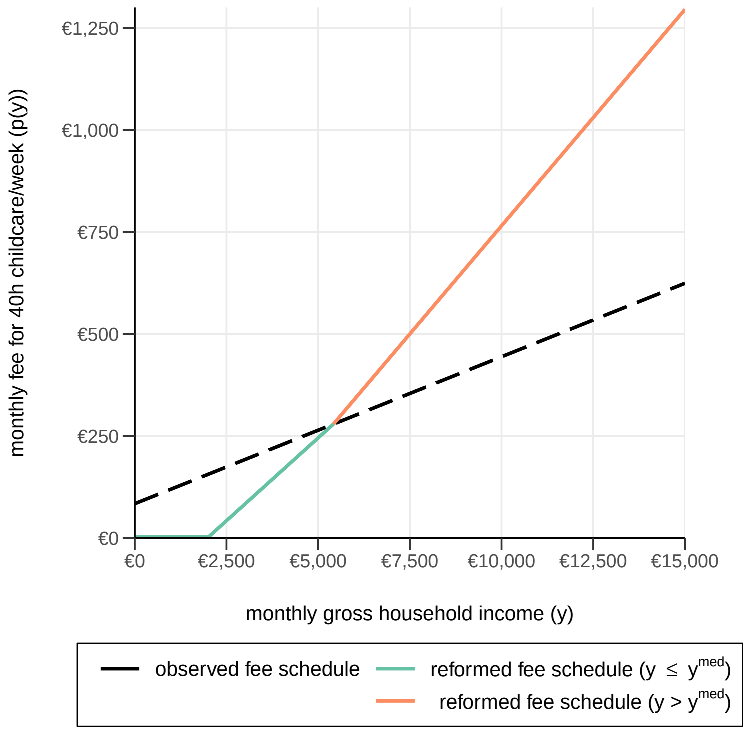


Figure 1.9: Potential reform of the childcare fee schedule

Notes: Parametric budget-neutral reform of the childcare fee schedule as set up in equation (1.20) with $\tau_1^c = 0.07$ and $\tau_2^c = 0.04518$. For illustration purposes, we focus on the case with a 0 – 2 year old and no siblings.

[54] Under the observed childcare fee schedule, one marginal Euro in monthly gross household income increases hourly childcare fees by around 3 cents (see Appendix-Table A.2). With $\tau_1^c = 0.07$, one marginal Euro in monthly gross household income for above-median income households increases hourly childcare fees by around 10 cents.

[55] We do not allow for negative childcare fees and therefore cap them below at 0. We also cap them at the top to the actual cost of a slot (see Table 1.2).

1.8 Conclusion

In this paper, we develop a dynamic structural model of unitary households' decision making to study universal childcare programs in Germany. The two key endogenous choices in the model are maternal labor supply and how to provide care for young children. To account for heterogeneity in these choices, we allow for rich observed and unobserved heterogeneity: Households differ with respect to their education, wages, and timing and spacing of up to three children. Furthermore, they also differ in their preference for domestic childcare, their taste for maternal leisure, and their access to free informal childcare by, e.g., grandparents. We estimate the model by maximum likelihood with German panel data and show that we fit empirical moments and predict reasonable labor supply elasticities.

We use our model to simulate counterfactual childcare policies. We start by studying the effects of a recent public childcare expansion for 0 – 2 year olds. We find that the introduction of more publicly provided slots was completely self-financing through the impact on maternal life cycle earnings and tax revenue.

We then turn to evaluating the current childcare fee schedule. Our results also indicate that the degree of subsidization is likely too high: a further marginal increase in the subsidies is only 6% self-financing. Such an increase would primarily be a transfer to families that are inframarginal in their public childcare and labor supply decisions. The numbers are different, however, if the increase in subsidies was contingent on mothers working full-time: such an increase would be 50% self-financing.

Finally, we turn to the redistribution that is entailed in the progressive childcare fee schedule. We illustrate that the current high subsidies for high-income families can barely be grounded on efficiency rationales. We show this by comparing the childcare fee schedule to the income tax schedule in terms of each schedules' efficiency cost of redistribution. We find that making the childcare fee schedule more progressive would imply efficiency costs that are one-third lower than what society currently pays for redistribution through the tax schedule. This implies that if society would like to solve the trade-off between equity and efficiency for the childcare-using population similarly to how it does it for the general population, childcare fees should be made more progressive. The current system would, however, be optimal if societal preferences for redistribution were significantly lower in the childcare-using population.

One important aspect our analysis abstracts from are the effects of public childcare on children's development. Incorporating these would likely increase the self-financing degree of the different counterfactuals we consider, as Felfe and Lalive (2018) and Cornelissen et al. (2018) document positive effects of public childcare on child development for Germany. Furthermore, as Cornelissen et al. (2018) show that the gains are especially large for children from disadvantaged backgrounds, these results strengthen our suggestion to make childcare subsidies more progressive.

Appendix A1 Details on the Sample Construction

In addition to the sample restrictions described in Section 1.3.2, we condition on observing the following covariates for every female: hourly wage if working, hourly wage of the partner, education (A-level or not), religion (Catholic at age 20), state of residence, predominantly living in an urban or rural area, demanding occupation (having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine). Furthermore, we drop households that are either in the top 1% or bottom 1% of the male or female wage distribution to avoid distortions.

Data processing for the maximum likelihood estimation Starting out with six waves of GSOEP data (2012 – 2017), we only keep households that are observed at least twice within this time frame. Then we allocate all children into the corresponding model child age brackets (see Section 1.4) and the household into the corresponding child-age structure K. Next, we assign each observation to a 3-year-spanning model period, ensuring that these line up with the evolution of the child-age structure K across time. Finally, we average all household variables of interest within the assigned model periods and only keep households with complete information for two model periods.

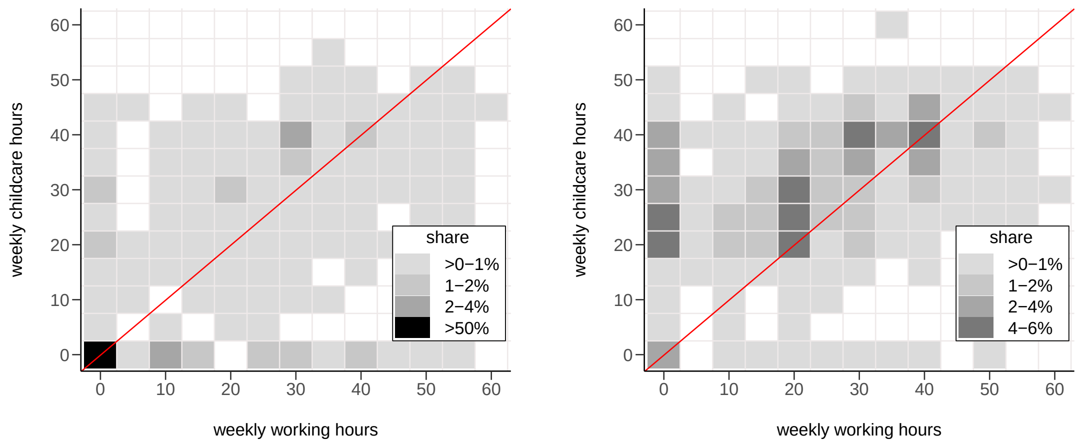
Table A.1: Summary statistics for the MLE sample

	mothers of 0 – 8 year olds	mothers of 9+ year olds
age	36.10	50.53
male wage (w_m)	€23.58	€23.95
female wage (w_f)	€17.31	€16.17
share high education	51%	32%
number of children aged 0 – 8	1.28	
age of youngest child	3.40	
share living in former East	20%	27%
share demanding occupation	43%	35%
share catholic	29%	32%
share urban	64%	60%
N	1,073	1,105

Notes: 'east' indicates having lived in former East Germany at some point; 'demanding occupation' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine; 'catholic' indicates being Catholic at age 20; 'high education' indicates having obtained at least an A-level; 'urban' indicates living predominantly in an urban area. Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner. Source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

Appendix A2 Additional Stylized Facts

A2.1 Childcare hours vs. working hours



(a) Children aged 0 – 2

(b) Children aged 3 – 5

Figure A.1: Maternal working hours vs. public childcare hours

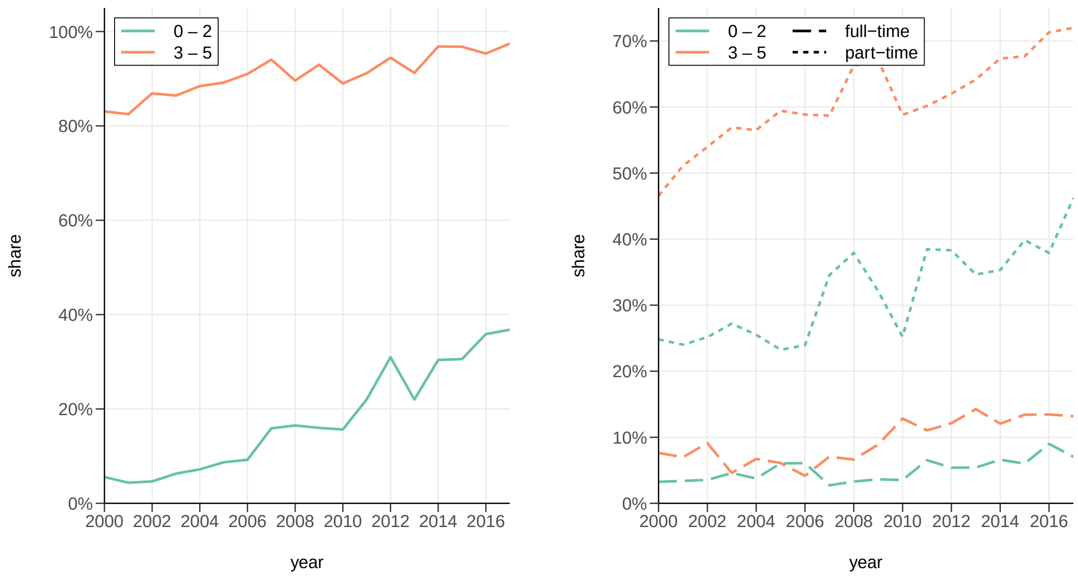
Notes: Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, conditional on having a child aged 0 – 2 for Figure A.1a or 3 – 5 for Figure A.1b. Source: 2009 to 2017 GSOEP, FDZ-SOEP (2019).

Concerning the joint decision making of mothers in terms of the use of public childcare and labor supply, Figures A.1a and A.1b plot the weekly working hours of the mother against the number of hours the child spends in childcare.

As Figure A.1a shows, more than half of the mothers of 0 – 2 year olds neither work nor send their children to public childcare. Of those who do work, not everybody sends their child to public childcare at the same time (mass at zero public childcare hours and positive working hours). Furthermore, we observe some children in public childcare whose mothers are not working at all.

For the 3 – 5 year olds in Figure A.1b, the picture looks different. Most children attend public childcare and many mothers work to some degree. Nevertheless, similar to the 0 – 2 year olds, a share of mothers also does not work while their children are in public childcare (mass at zero working hours). Focusing on the mass close to the red 45-degree line, we observe that with increasing weekly working hours, children also spend more time in public childcare.

A2.2 Developments in West Germany



(a) Public childcare enrollment (b) Employment of mothers

Figure A.2: Development of public childcare enrollment and maternal employment in West Germany

Notes: Enrollment is binary in the sense that it is not conditional on a minimum number of hours. Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, conditional on having a child aged 0 – 2 or 3 – 5. Source: 2000 to 2017 GSOEP living in West Germany, FDZ-SOEP (2019).

A2.3 Childcare demand and enrollment of 0 – 2 year olds

Rationing of public childcare slots for 0 – 2 year olds was formally ended with the introduction of a legal claim to a slot for every 1+ year old in 2013. Nevertheless, recent government reports still show a sizeable gap between public childcare demand and enrollment (BMFSFJ 2016, BMFSFJ 2020). In 2015, for example, this gap amounted to 18.9 (11.7) percentage points for 1 to 2 (2 to 3) year olds, with 35.8% (61.3%) of children enrolled in public childcare and 54.7% (73%) of parents reporting demand for public childcare. Multiplying these values by the total number of children in the specific age range yields a shortage of around 300,000 nursery slots, which has been reported in multiple policy briefs and newspaper articles over the years (Geis 2018, Geis-Thöne 2020, Kulms and Nehls 2018).

There are two possible underlying explanations for this gap: First, it may reflect rationing, i.e., the complete absence of slots for a sizeable number of children. This would imply that parents are not enforcing their legal claims by taking local authorities to court. Second, the gap may reflect the choice of some parents not to take up the public childcare slots that they are offered. Reasons for this may be, for example, personal preferences on childcare quality or the location of the facility. In the eyes of our model, the second explanation would be captured by our estimated preference for domestic childcare.

To understand the role that both explanations play for the gap between reported demand and observed take-up, we start by reconstructing the numbers from the government reports within a single data set, the DJI-KiBS panel.[56] This is different from the approach taken in the reports, which do use the same survey data to calculate the demand for public childcare, but then rely on administrative childcare enrollment data and birth records to calculate take-up. Instead, we calculate both demand and enrollment from the survey data to ensure that we accurately capture the gap within the sample at hand. In terms of the sample restrictions, we focus on children below age 3 and limit the analysis to married and cohabiting couples (as for the estimation).[57]

[56] The DJI-KiBS panel is a large-scale survey of parents of children in different age brackets, focusing on institutional care of children in Germany. It is run by the German Youth Institute (DJI) since 2012 and includes more than 35,000 observations of children under age 18 per year. For our analysis, we use waves 3 and 4 (corresponding to the years 2014 and 2015), as for those years the questionaire included very detailed questions on the reasons for the lack of public childcare use/conditions under which public childcare would be used. Source: FDZ-DJI (2017).

[57] Additionally, we do not include children below 1, for whom public childcare demand and enrollment is virtually non-existent.

Based solely on the DJI-KiBS data, we find that 9.39% of parents report a gap between their demand and enrollment. However, only 3.18% of parents report that they did not use public childcare because they were not offered any slots. A larger share of 6.21% was offered a slot, but did not use public childcare for other reasons, which we explore in more detail below. This illustrates that within our population of interest, rationing is likely to be substantially lower than in the reports and policy briefs mentioned above.

With the rich DJI-KiBS data, we can additionally split the data into subgroups by the amount of hours of public childcare that parents demand. Specifically, we find that of the 3.18% who were not offered a slot, two-thirds (2.09%) did demand more than 20 hours of public childcare per week, while one-third (1.09%) stated 20 hours or less. Of the 6.21% who were offered a slot but decided not to enroll their children, around half (3.07%) demanded more than 20 hours per week, while the other half (3.14%) demanded $\leq$ 20 hours. With this differentiation of the scope of the demand at hand, we now turn to a more detailed investigation of the reasons that parents state for not using public childcare.

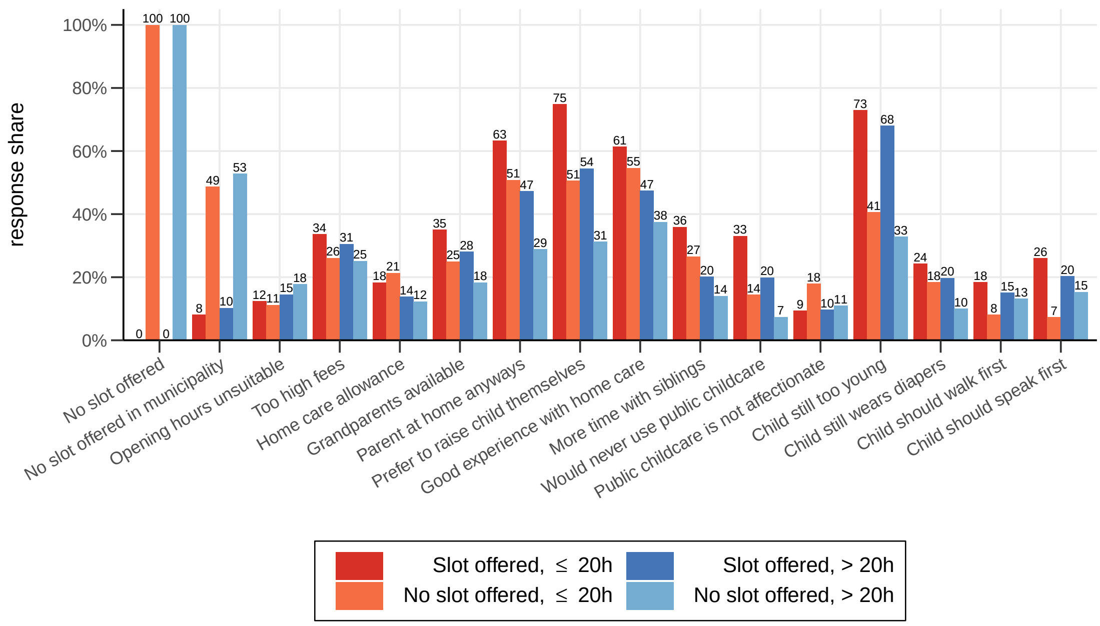


Figure A.3: Reasons for not using public childcare

Notes: Children aged 1 or 2 with cohabiting parents in 2014 or 2015, for whom parents report higher public childcare demand than enrollment. Source: FDZ-DJI (2017).

In Figure A.3, we present the responses to a multiple choice question on these reasons. More than half of the parents who were offered a slot, but did not use it (darker colors), responded that they thought their child was too young or preferred to raise their child themselves, given the public childcare at offer. Of those parents that were not offered a slot (lighter colors), around half responded that there is no public childcare facility close to where they live. Many parents also responded that they were home themselves anyway or have had good experiences with providing childcare at home.

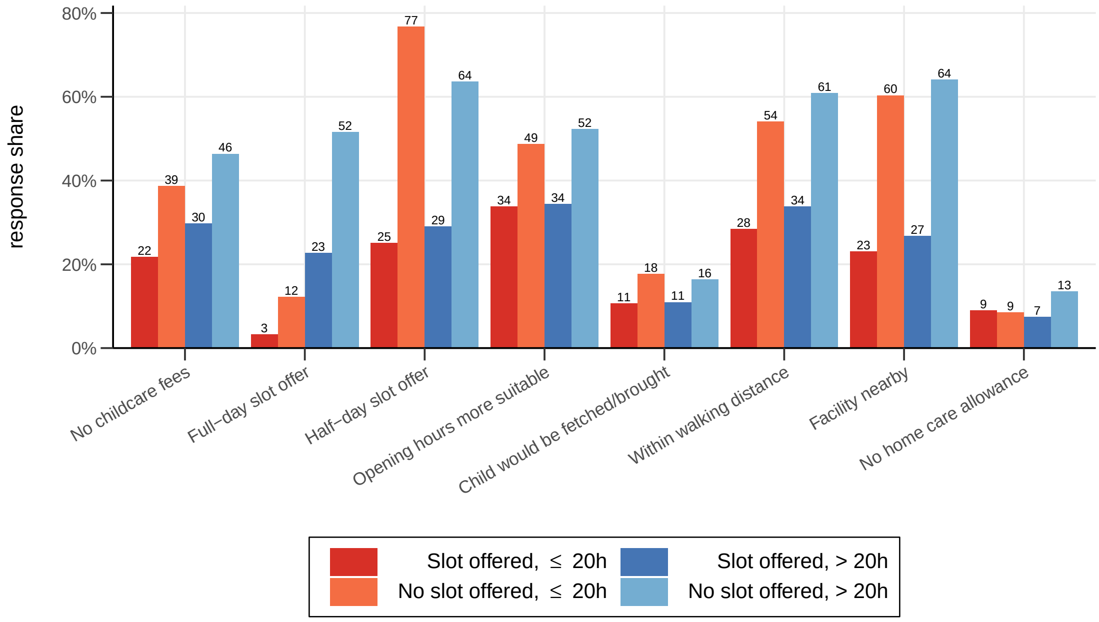


Figure A.4: Conditions under which public childcare would be used

Notes: Children aged 1 or 2 with cohabiting parents in 2014 or 2015, for whom parents report higher public childcare demand than enrollment. Source: FDZ-DJI (2017).

In the interviews, parents were also asked under which conditions they would be using public childcare. Figure A.4 presents a first set of responses to this question. Many parents, also those that were offered a slot, responded that they would have chosen to use public childcare if there had been an available half-day slot, if the opening hours were more suitable, or if the facility had been within walking distance/nearby.

Figure A.5 presents a second set of responses to the same question, in this case specifically related to the quality of care. Among the group of parents, who did get a slot offer, but did not take it, the group size and the number of caretakers are reported to be the most relevant factors.

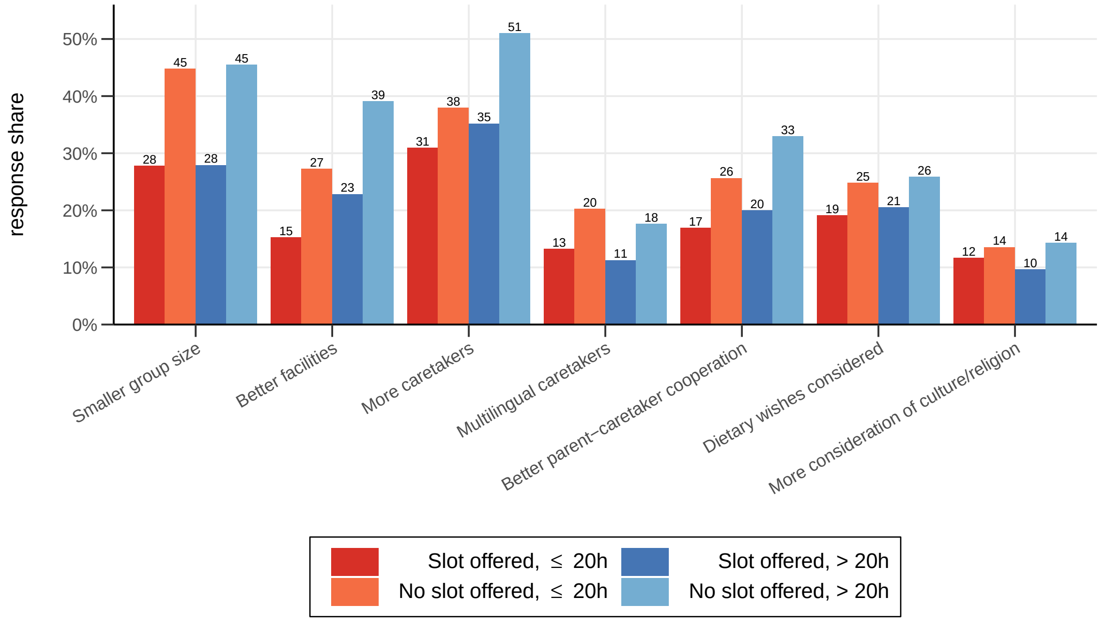


Figure A.5: Conditions under which public childcare would be used (quality dimensions)

Notes: Children aged 1 or 2 with cohabiting parents in 2014 or 2015, for whom parents report higher public childcare demand than enrollment. Source: FDZ-DJI (2017).

In summary, this detailed look at unsatisfied public childcare demand reveals that in cases where a slot was offered, the reasons for non-enrollment are mostly related to quality preferences and locational convenience of the facility. This implies that the majority of the stated gap between demand and enrollment is due to a higher preference for domestic childcare vs. the alternative of public childcare at a given quality level/location.

And even in cases where no slot was offered, parents also state a number of other reasons why they are not using public childcare. This illustrates that even if this subgroup were offered slots, their take-up would likely also fall short of their stated demand, as, e.g., distance to the facility plays an important role for the decision to enroll. A plausible alternative interpretation of the 'no slot' responses in light of these reasons would also be that some parents answer that they did not get a slot offered if they only got a slot in the neighboring municipality, but not very close to their home.

Both of these interpretations underline that the true scope of rationing is likely even below the 3.18% that we find in the DJI-KiBS data. While there is a sizeable amount of unsatisfied demand for public childcare for 0 – 2 year olds in Germany despite the legal claim, this is mostly due to slots not meeting the preferences of parents and therefore parents prefer to use domestic childcare instead.

Appendix A3 Additional Details on the Estimation

A3.1 Institutional background on the determinants of childcare fees

Child age. One of the important determinants of childcare fees in Germany is the age of the child. This is due to the fact that children of different ages visit different childcare institutions, as laid out in Section 1.3.1. Fees are usually higher for younger children since the costs of operating nurseries are higher than those of kindergartens.

Regional variation. Childcare fees in Germany differ further on a regional level because of two reasons: First, the fee schedules are set discretionary on a municipality level. Second, the German constitution ensures that the federal states bear the political responsibility for their respective education systems. Since public childcare is part of the education system, different federal states have implemented different regulations concerning the fee schedules.

Further determinants of childcare fees. Despite their autonomy, different states define in their legislation vastly similar determinants of childcare fees besides child age:[58]

1. *Household income:* In 11 out of 16 states the household income has to be used as a determinant and in two additional states it can be used.[59]
2. *Number of children in the household:* In 12 out of 16 states, childcare fees are determined conditional on the number of children in the household. Furthermore, in one additional state it can be used optionally as a determinant and two further states condition on the number of children in the household that attend nursery or kindergarten.

[58] See Authoring Group Educational Reporting (2018): *Education in Germany 2018*, Section C2, p. 70–71 and Table C2-15web.

[59] Note that different municipalities differ in their definitions of household income: First, municipalities differ in using the net or gross household income. Second, they also condition on household income of different years (current year or previous years).

A3.2 Estimation of the childcare fee schedule

We use the following linear model to estimate the childcare fee schedule reflected in the structural model by $p(j, K, y)$ separately for each child age bracket j:

$$\begin{aligned} p_{nt} = \; & \alpha + \beta_1 y_{nt} + \beta_2 \left(y_{nt} \times \mathbb{1}\left\{\text{one sibling with age} < 17 \text{ in HH}\right\}_{nt} \right) \\ & + \beta_3 \left(y_{nt} \times \mathbb{1}\left\{\text{two siblings with age} < 17 \text{ in HH}\right\}_{nt} \right) + \epsilon_{nt}. \end{aligned} \tag{A.1}$$

The dependent variable p_{nt} is the monthly fee that household n would pay for full-time childcare (40h/week) for a child aged j in year t. The interaction terms of gross household income with indicators for the number of siblings capture discounts granted to families with multiple children.[60] Our empirical model thereby closely reflects the current childcare fee schedule regulation as laid out in the previous section.

We estimate equation (A.1) as a Tobit regression with censoring at €0 and €725, the lowest and highest observed monthly childcare payments in our data.[61] We abstract from regional variation to keep the state space of the structural model tractable.

Results. The results of the Tobit regressions are summarized in Table A.2. Monthly childcare fees increase significantly in gross household income for all age brackets. Average fees are estimated to be highest for the youngest children, who require the most intensive care. The presence of siblings implies a significant reduction of the income gradient for 0 – 2 and 3 – 5 year olds, decreasing it by more than half if two siblings live in the household.

[60] As we are mainly interested in predicting childcare fees, we only include covariates that are in line with the institutional setup described above. The stand-alone sibling dummies are not included as they do not add any explanatory power.

[61] We observe a number of households not paying any fees for a positive amount of childcare hours. Furthermore, we cap the fees to the maximum observed value to ensure that the rescaling to full-time equivalent fees does not yield unreasonably high values.

Table A.2: Tobit estimation of the childcare fee schedule

Dependent variable:
Monthly fee for public childcare attendance of 40h/week

	child age		
	0 – 2	3 – 5	6 – 8
gross HH income	0.036	0.022	0.028
	(0.0037)	(0.0015)	(0.0043)
gross HH income × 1 sibling in HH	−0.012	−0.0065	−0.0050
	(0.0028)	(0.0012)	(0.0035)
gross HH income × 2 siblings in HH	−0.022	−0.010	−0.0052
	(0.0040)	(0.0015)	(0.0041)
constant	90.0	66.3	9.34
	(18.3)	(6.38)	(16.2)
N	362	1,950	626

Notes: Sample: Children attending public childcare for whom childcare fees and childcare hours are observed. Tobit regressions with censoring at €0 and €725. All values are adjusted to 2017 price levels. Source: 2013, 2015, 2017 GSOEP, FDZ-SOEP (2019).

A3.3 Imputation of potential wages for non-working females

For the imputation of potential wages of non-working females, we use the following static wage model:

$$log(w_{f,it}) = \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}, \tag{A.2}$$

where $w_{f,it}$ is the wage of female i in period t and $\mathbf{X}$ contains the following Mincer-type covariates: linear and quadratic terms for age, full-time work experience, and part-time work experience. Furthermore, we include indicators for different education levels, namely an indicator for a lower track school degree and vocational training, an indicator for an A-level, and an indicator for a university degree. Additionally, $\mathbf{X}$ also contains the number of children below age 5, the overall number of children, an urban indicator, an indicator for living in former East Germany, and a full set of year indicators.

Wages are only observed if a woman works (participation$_{it}$ = 1), which is determined by:

$$\mathbf{Z}_{it}\boldsymbol{\zeta} + v_{it} > 0, \tag{A.3}$$

where $\mathbf{Z}$ contains $\mathbf{X}$ along with a set of exclusion restrictions. Following Bargain, Orsini, and Peichl (2014) and in line with our model, we use as exclusion restrictions indicators for the presence of $0-2$, $3-5$, $6-8$, $9-17$, or 18+ year old children in the household. Furthermore, we include the husband's gross wage quintile and the net household income if the female chooses not to work.

In line with the selection correction procedure proposed by Semykina and Wooldridge (2010), we run a Probit version of equation (A.3) for each time period. In these, we also include the individual specific means of all covariates in $\mathbf{Z}$ across 2000 to 2017, denoted by $\overline{\mathbf{Z}}$:

$$Pr(\text{participation}_i = 1) = \Phi\left(\mathbf{Z}_i\boldsymbol{\zeta} + \overline{\mathbf{Z}}_i\boldsymbol{\xi}\right). \tag{A.4}$$

After estimating (A.4) for each year, we obtain the inverse Mills ratios λ_{it}, which we then use as control functions in the selection corrected version of the wage equation (A.2):

$$log(w_{f,it}) = \mathbf{X}_{it}\boldsymbol{\rho} + \overline{\mathbf{Z}}_i\boldsymbol{\xi} + \gamma\lambda_{it} + u_{it}. \tag{A.5}$$

With the estimated coefficients $\boldsymbol{\rho}$ and $\boldsymbol{\xi}$ at hand, we impute the wages of the non-working females.

A3.4 Wage process estimation results

Table A.3: Estimation of female and male wage dynamics

	$log(w_{f,it})$	$log(w_{m,it})$
age_{it}	0.020	0.010
	(0.013)	(0.0093)
age_{it}^2	−0.0026	−0.0018
	(0.0017)	(0.0012)
age_{it}^3	0.000082	0.000059
	(0.000067)	(0.000048)
$educ_i$	0.076	0.027
	(0.0052)	(0.0036)
NP_{it-1}	−0.18	
	(0.0065)	
PT_{it-1}	−0.057	
	(0.0057)	
$log(w_{f,it-1})$	0.75	
	(0.0057)	
$log(w_{m,it-1})$		0.91
		(0.0040)
constant	0.70	0.30
	(0.032)	(0.023)
σ^2	0.28	0.20
	(0.0017)	(0.0012)

Notes: See Section 1.5.3 for the regression setup. 'educ' indicates having obtained at least an A-level, NP and PT denote not working and working part-time, respectively. Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, non-participation wages imputed as described in Appendix A3.3. Source: 2000 to 2017 GSOEP, FDZ-SOEP (2019).

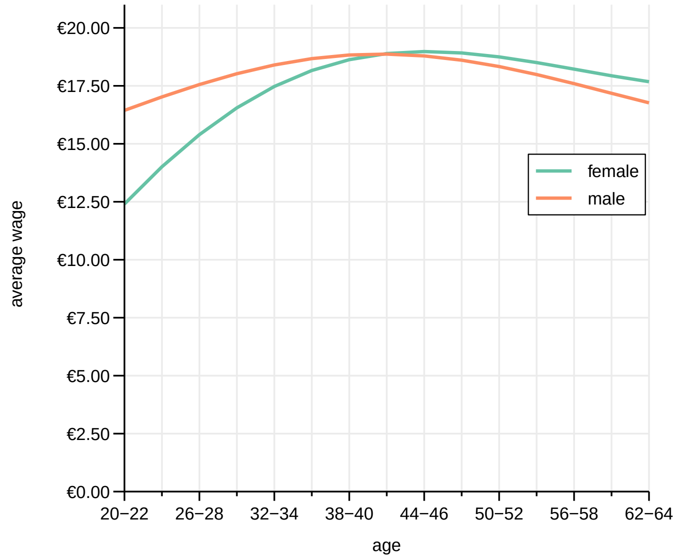


Figure A.6: Wage dynamics across the life cycle

Notes: Simulated average wages starting with the observed wage distributions for individuals without an A-Level aged 20 – 22 and assuming continuous full-time work for the remaining life cycle. Based on wage process estimates from Table A.3.

A3.5 Global optimization routine

To solve the optimization problem numerically, we use the basin-hopping algorithm in combination with a Matlab built-in minimization routine for constrained target functions (*fmincon*). The basin-hopping algorithm is a stochastic global optimisation algorithm used in various fields (Chemistry, Applied Mathematics, ...), which was first introduced by Wales and Doye (1997).[62]

Intuitively, the procedure works as follows: We set an (arbitrary) initial starting point and solve for a (possibly local) minimum given the specified constraints on the parameters using *fmincon*. As we do not know the shape of the multidimensional objective function, we cannot be sure to have found the global minimum. To increase the likelihood of finding the global minimum, the basin-hopping algorithm then applies a random perturbation to the parameters of the previously found (potentially local) minimum and restarts the minimization routine *fmincon* at the perturbed parameters. The basin-hopping algorithm then compares the new minimum to the previous one and records the point with the lowest target function value as a candidate for the global minimum. The algorithm repeats the procedure, always keeping track of the point that yielded the lowest target function value, until either a predetermined number of iterations has been completed or the global minimum candidate did not change for a predetermined number of iterations. Trading off runtime and precision, we set the number of iterations in the basin-hopping algorithm to 1,000.

Detailed description of the basin-hopping algorithm. As described above, the basin-hopping algorithm is a stochastic global optimization algorithm. Starting from an arbitrary prespecified initial parameter combination, it employs a minimization routine (in our case Matlab's *fmincon*) to find a (possibly local) first minimum. This first minimum is the first candidate for the global minimum. Using the parameter combination of the minimum found after the first minimization, the algorithm then applies a random perturbation to these parameters (which we will call 'taking a *step*' from now on) and restarts the minimization routine. If the routine returns a new minimum whose function value is lower than the current global minimum candidate, this parameter combination becomes the current global minimum candidate. The *step*-taking, which is discussed in detail below, is then repeated for either a prespecified number of times or terminated if

[62] Our Matlab implementation of the basin-hopping algorithm follows the SciPy Python implementation (Virtanen et al. 2019).

the global minimum candidate has not changed for a prespecified number of *steps*. After termination, the final global minimum candidate is returned as the global solution to the minimization problem.

The *step*-taking procedure is the key mechanism of the algorithm to search the multidimensional target function for minima. To be able to escape the basins of attraction of local minima, the following adaptive procedure is used to conduct the *step*-taking:[63]

Starting with an initial step-size – which we set to 0.2 –, each of the parameters of the first minimum is separately perturbed by a random shock whose size is at most equal to the (initial) step-size. This implies that the perturbed version of each parameter is within a ±0.2 interval around the respective parameter of the first minimum. After running the minimization routine with the perturbed parameters as the starting point, the newly found (possibly local) minimum is compared to the first minimum. If the newly found minimum is i) lower in function value than the first minimum or ii) its function value is larger, but close to the first minimum's function value, the newly found minimum is *accepted*.[64]

Whenever a newly found minimum is *accepted*, it becomes the starting point for the next and all future *steps*. By allowing minima to be *accepted* despite not being new global minimum candidates, the algorithm retains some flexibility to explore the target function in the proximity of the current global minimum candidate.

The procedure furthermore includes an adaptive adjustment of the search radius (as determined by the size of the perturbations, i.e., the *step*-size). After every n-th *step*, the algorithm compares the rate at which *steps* are *accepted* to a target rate.[65] If the *acceptance* rate is below the target rate, this implies that the perturbations are too large, i.e., the proximity of local minima is not sufficiently explored. Consequently, the step-size is adjusted downwards by 10%, narrowing the search radius. If the *acceptance* rate is above the target rate, this implies that the perturbations are too small, i.e., the algorithm is likely stuck in a basin of attraction. Consequently, the step-size is adjusted upwards

[63] A *basin of attraction* is the set of possible starting points of a minimization routine that leads to the same minimum (Nusse and Yorke 1996).

[64] The determination of what is 'close' is based on a Metropolis-Hastings criterion, see the SciPy implementation for details (Virtanen et al. 2019). A 'temperature' parameter governs the acceptance probability, which should be comparable to the separation in function value between local minima. We set this parameter to 30 as we typically observe separations of this size between the local minima of our target function.

[65] We set n to 5, i.e., we adjust the *step*-size after every fifth *step* and the target rate to 0.5, i.e., 50% of steps should be accepted.

by 10%, widening the search radius to be able to escape the current basin of attraction and explore the target function outside of it.

The two just described components of the *step*-taking procedure, i) which minima to *accept*, and ii) how to adjust the step-size, allow the basin-hopping algorithm to explore a large range of possible directions in terms of the parameters.

A3.6 Sensitivity of estimated structural parameters

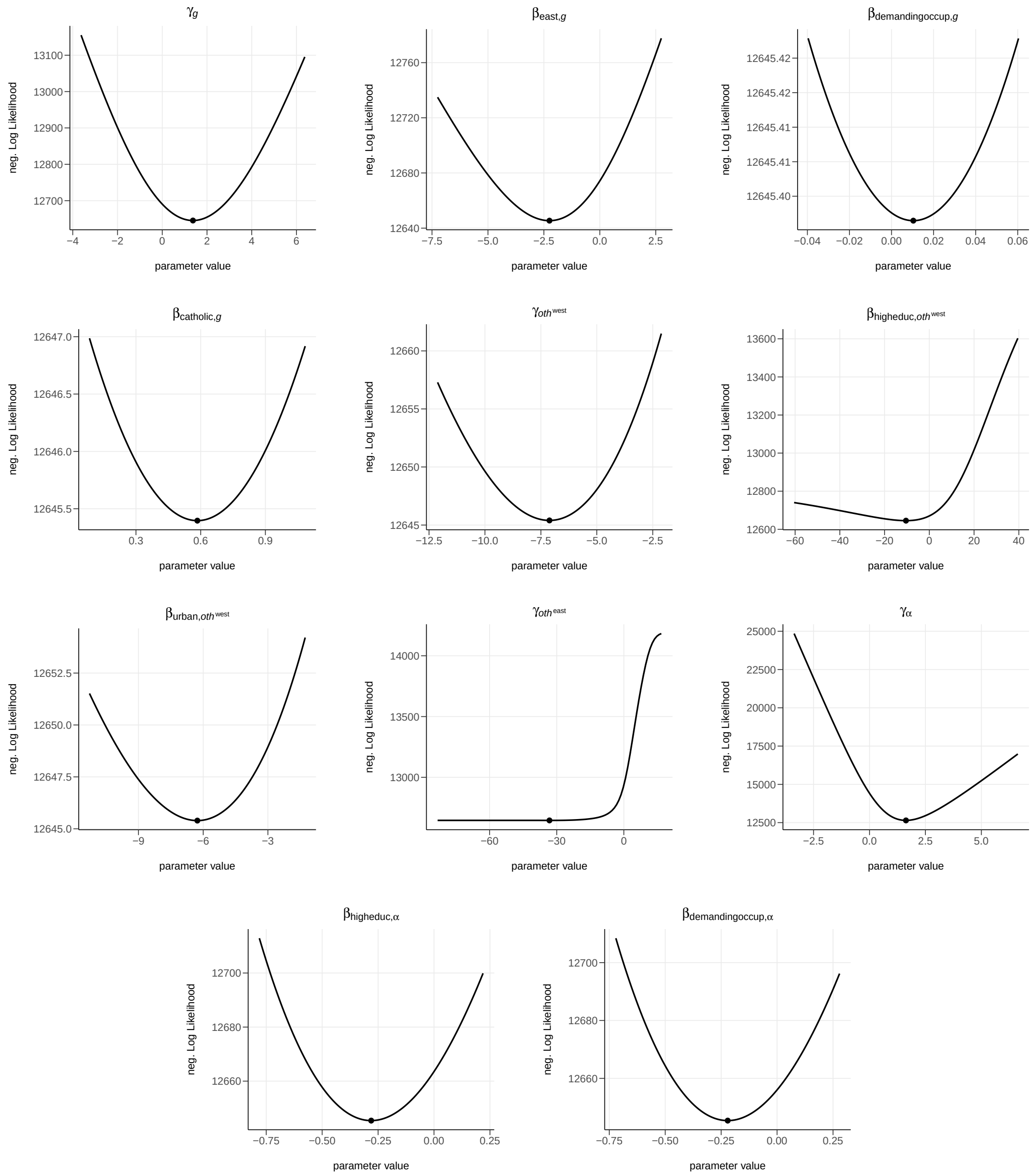


Figure A.7: Sensitivity of estimated structural parameters

Notes: Illustration of changes in the negative Log-Likelihood in response to small deviations of each parameter from its estimated value (labelled by black dot), keeping all other parameters at their point estimates. We constrain the estimation of σ_{oth} within the interval [0,2] to facilitate the computation, but the resulting CDF (as displayed in Figure 1.6) remains virtually identical if we allow for a larger interval. 'demanding occup' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine. 'high educ' indicates having obtained at least an A-level.

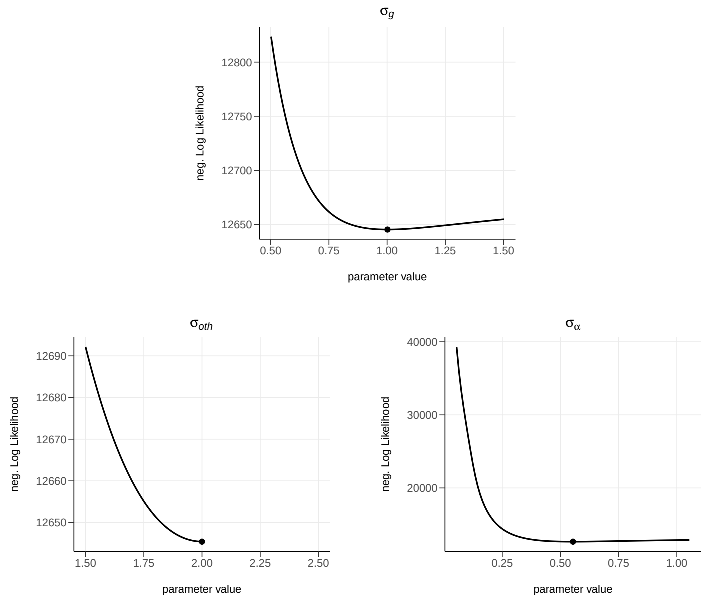
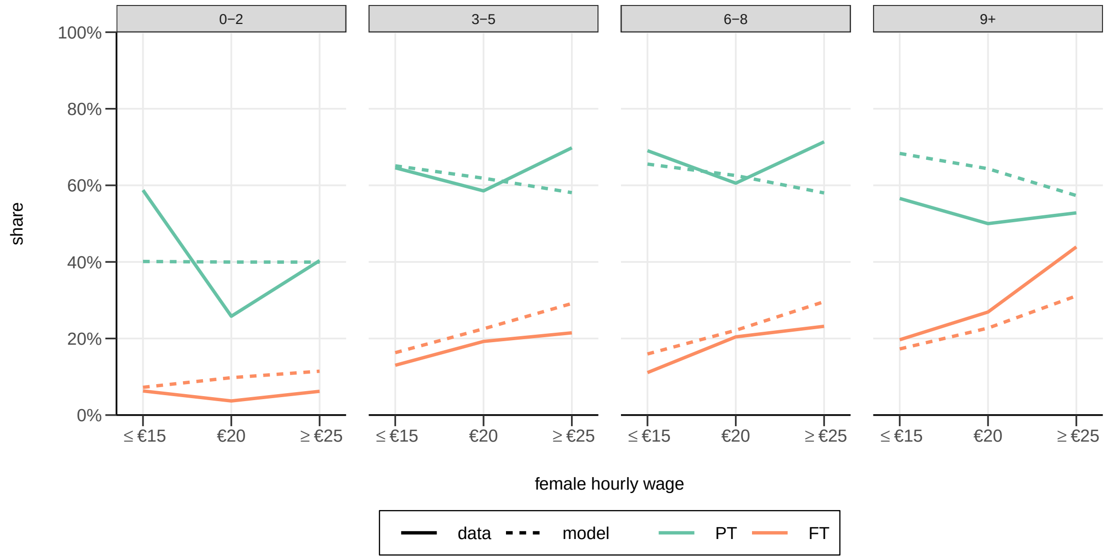


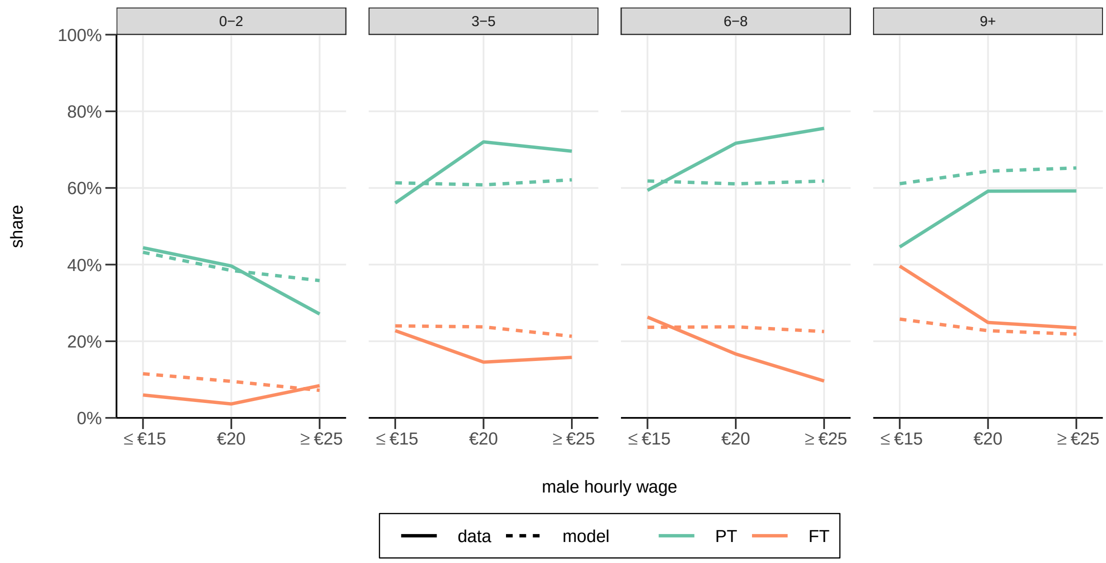
Figure A.7 (cont.): Sensitivity of estimated structural parameters

Notes: Illustration of changes in the negative Log-Likelihood in response to small deviations of each parameter from its estimated value (labelled by black dot), keeping all other parameters at their point estimates. We constrain the estimation of σ_{oth} within the interval [0,2] to facilitate the computation, but the resulting CDF (as displayed in Figure 1.6) remains virtually identical if we allow for a larger interval. ‘demanding occup’ indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine. ‘high education’ indicates having obtained at least an A-level.

Appendix A4 Additional Model Fit Illustrations



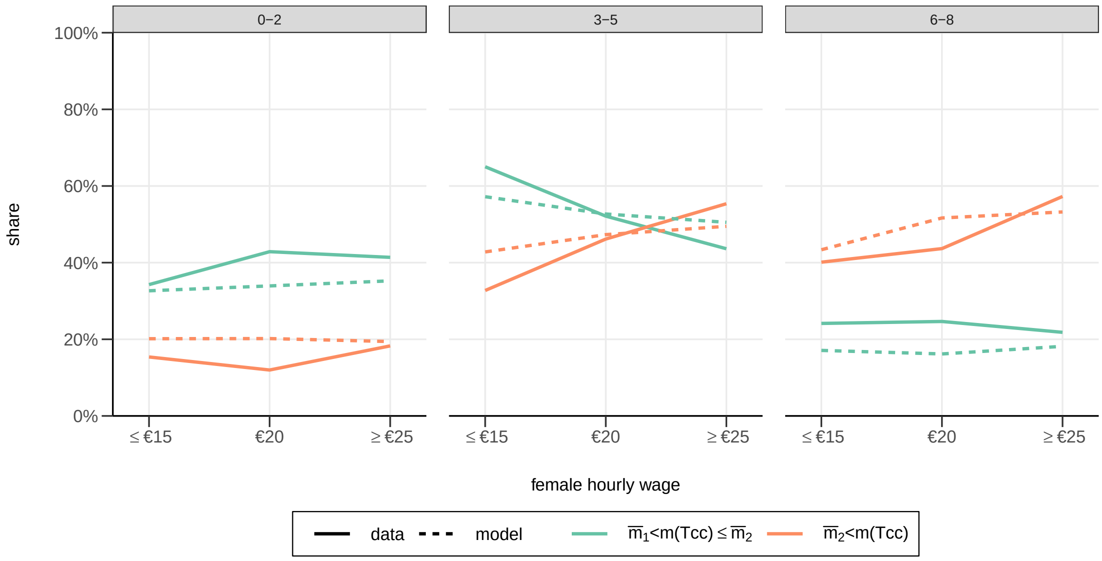
(a) Model fit for labor supply by female wages



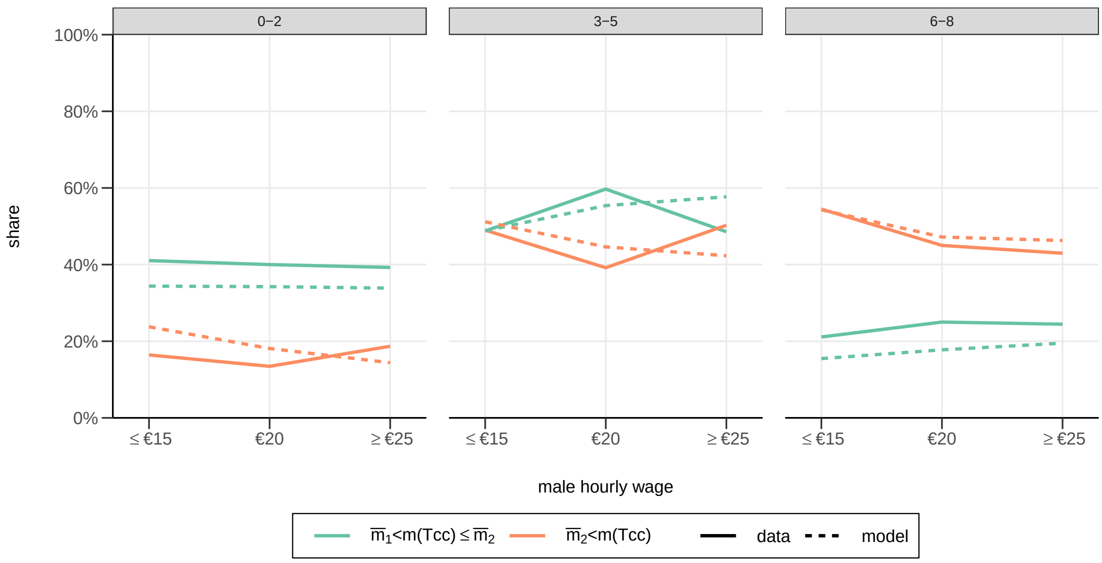
(b) Model fit for labor supply by male wages

Figure A.8: Model fit for labor supply by male and female wages

Notes: PT and FT denote the female working part-time and full-time, respectively. Sample as defined in Section 1.5.5. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).



(a) Model fit for public childcare take-up by female wages



(b) Model fit for public childcare take-up by male wages

Figure A.9: Model fit for public childcare take-up by male and female wages

Notes: $m(Tcc)$ denotes the share of total childcare need that is covered through public childcare. $\overline{m}_1$, respectively $\overline{m}_2$, indicates that the household would cover a share of 33%, respectively 75%. Sample as defined in Section 1.5.5. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

Appendix A5 Additional Results

Table A.4: Detailed results for the public childcare expansion

		female hourly wage		
	total	≤ €15	€20	≥ €25
(a) *Impact period*				
tax surplus	1,017.68	259.42	488.84	269.41
subsidy increase	1,431.18	678.64	594.18	158.36
self-financing	71.1%	38.2%	82.3%	170.1%
(b) *All periods*				
tax surplus	2,105.87	625.95	1,014.90	465.02
subsidy increase	2,037.79	979.43	839.51	218.85
self-financing	103.3%	63.9%	120.9%	212.5%

Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiment: ending rationing of public childcare for 0 – 2 year olds.

Table A.5: Detailed results for an untargeted childcare subsidy increase

		female hourly wage		
	total	≤ €15	€20	≥ €25
(a) *Impact period*				
tax surplus	42.36	21.60	15.46	5.30
subsidy increase	1,044.43	632.90	288.08	123.45
self-financing	4.1%	3.4%	5.4%	4.3%
(b) *All periods*				
tax surplus	111.20	59.00	40.40	11.80
subsidy increase	1,877.38	1,120.13	565.39	191.86
self-financing	5.9%	5.3%	7.1%	6.2%

Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiment: increase in the subsidization of the hourly childcare fee that is equivalent to increasing monthly subsidies for full-time public childcare (40h/week) by €50.

Table A.6: Detailed results for a work-contingent childcare subsidy increase

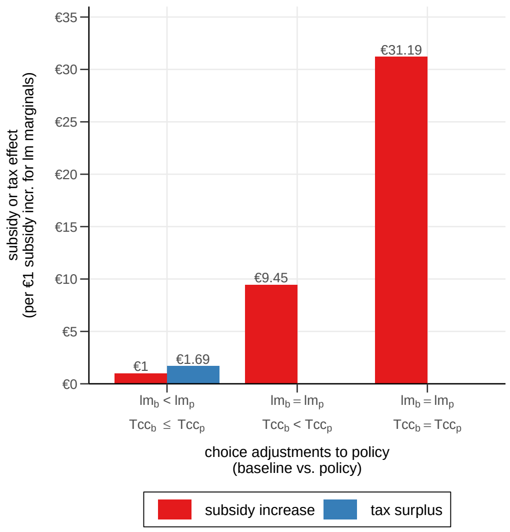
		female hourly wage		
	total	≤ €15	€20	≥ €25
(a) *Impact period*				
tax surplus	79.16	40.49	29.74	8.94
subsidy increase	911.36	549.68	250.16	111.52
self-financing	8.7%	7.4%	11.9%	8.0%
(b) *All periods*				
tax surplus	204.03	107.61	76.45	19.98
subsidy increase	1,649.07	980.35	496.04	172.68
self-financing	12.4%	11.0%	15.4%	11.6%

Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiment: work-contingent increase in the subsidization of the hourly childcare fee that is equivalent to increasing monthly subsidies for full-time public childcare (40h/week) by €50.

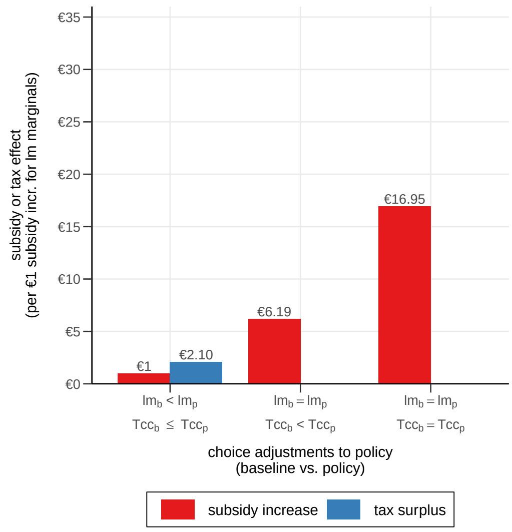
Table A.7: Detailed results for a full-time-contingent childcare subsidy increase

		female hourly wage		
	total	≤ €15	€20	≥ €25
(a) *Impact period*				
tax surplus	96.37	46.15	35.45	14.78
subsidy increase	287.72	156.03	87.75	43.94
self-financing	33.5%	29.6%	40.4%	33.6%
(b) *All periods*				
tax surplus	262.37	134.66	95.55	32.15
subsidy increase	523.80	278.70	177.05	68.06
self-financing	50.1%	48.3%	54.0%	47.2%

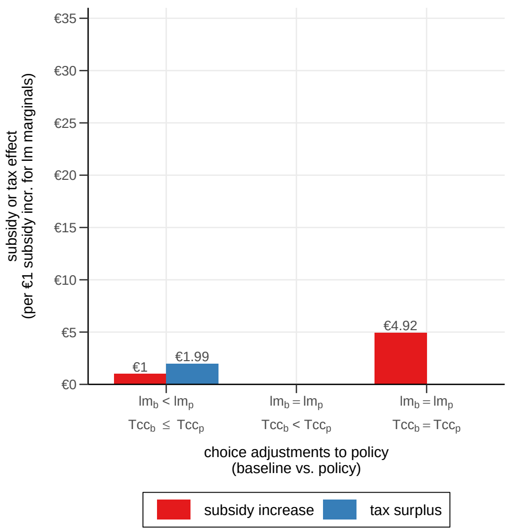
Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiment: full-time-contingent increase in the subsidization of the hourly childcare fee that is equivalent to increasing monthly subsidies for full-time public childcare (40h/week) by €50.



(a) *untargeted +€50*



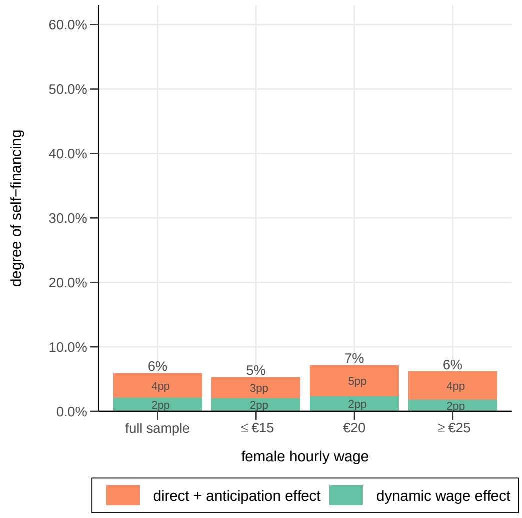
(b) *work-contingent +€50*



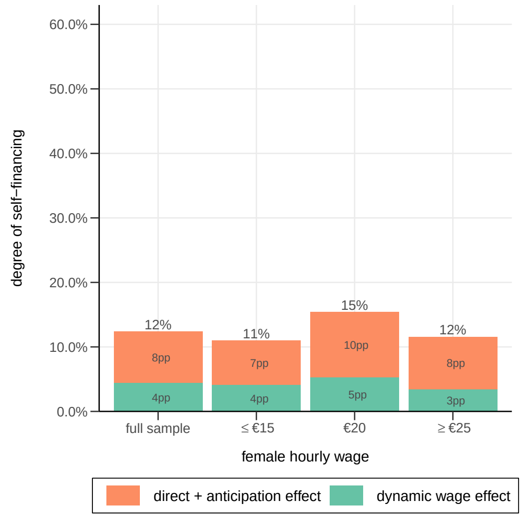
(c) *full-time-contingent +€50*

Figure A.10: *Impact period* effect decomposition for childcare subsidy increases

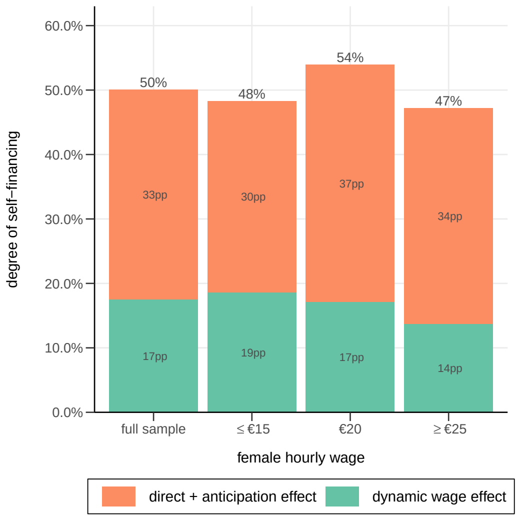
Notes: *Impact period* effects from Table 1.8 decomposed by subsidy increase and tax surplus, normalized by sample size. lm_b and Tcc_b denote the labor supply and public childcare decision in the baseline scenario. lm_p and Tcc_p denote the labor supply and public childcare decision in the respective policy scenario.



(a) *untargeted +€50*



(b) *work-contingent +€50*



(c) *full-time-contingent +€50*

Figure A.11: *All periods* effect decomposition for childcare subsidy increases

Notes: *All periods* effects from Table 1.8 decomposed into two main components: i) *direct effects + anticipation effects*, fiscal effects from behavioral adjustments in response to changes in the public childcare availability or fees while the household has children/may have children in the future in a policy affected age range. ii) *dynamic wage effects*, fiscal effects from higher wages and behavioral adjustments in response to higher wages, where higher wages are the result of past behavioral changes. Decomposition by initial wages at policy introduction. Values within the bars denote percentage point contributions.

Chapter 2

Women's Early Careers and Childcare Policy

Abstract: Early career years for both men and women are characterized by a fast accumulation of human capital. However, despite comparable working hours, the wages of young men quickly outgrow those of young women, even after accounting for observable differences. One potential reason is that employers take the cost of (long) maternity leave into account when making promotion decisions. In this paper, I first present novel descriptive evidence on the gender gap in early career promotion rates using administrative data from Germany. Second, I exploit quasi-random variation induced by a staggered expansion to identify the causal effect of public childcare availability on women's pre-birth promotion probabilities. I find that a 10 percentage point increase in public childcare availability improves the promotion probability of young women by 1.7 percentage points, which translates into a 10% increase. This result is consistent with public childcare effectively reducing leave-related costs for employers and thereby boosting women's pre-birth career prospects through a reduction of statistical discrimination.

2.1 Introduction

The gender wage gap is a powerful force shaping income differences between women and men across all ages, making it a key contributor to lower economic independence of women.[1] While differences in human capital accumulation are an important factor among middle age workers, they do not play a role for the early career gender wage gap. Previous research has rather pointed to differences in observable characteristics such as selection into occupations. A less studied, but potentially also important dimension is the degree to which expected future labor supply patterns affect promotion decisions, in particular for young women. Focusing on intra-firm interactions of employees with their employers allows to zoom into this matter and can furthermore shed new light on the broader effects of family policy on women's labor market outcomes.

In this paper, I first use German administrative social security data to map out patterns in promotion rates of women prior to first birth, i.e., *potential* mothers, and compare them to young men. I then use reform-driven variation in public childcare availability to study the role of the anticipation of future labor supply for the early career progress of women.[2] Specifically, I evaluate whether a large-scale expansion of public childcare, a family policy aimed at mothers, has impacted the promotion probabilities of young women *before* they become mothers.

The key channels in question are the following: On the employer side, firms take into account that a sizeable share of women temporarily drops out of the labor force around childbirth, imposing organizational and replacement costs. This gives employers an incentive to engage in statistical discrimination with respect to promotions, if the costs of dropping out are increasing in the position of a worker in the firm. On the employee side, young females, who have to take a break around childbirth, may have a lower incentive to engage in high on-the-job effort if they cannot benefit much from career advances. Through a reduction of the average time out of the labor market due to childbirth, childcare policies have the potential to reduce statistical discrimination and encourage young women to pursue more ambitious career paths. By asking whether public childcare provision has affected the labor market outcomes of *potential* mothers, I can therefore investigate the role of expected future labor supply for early career progress.

[1]As documented, for example, by Goldin (2014) and Blau and Kahn (2017). Tinios, Bettio, and Betti (2015) illustrate the long-lasting effects by relating the gender wage gap to the gender gap in pensions.

[2]This channel has been previously highlighted in Kiessling et al. (2019) and Kuziemko et al. (2018).

I identify the causal effect of public childcare on young women's labor market outcomes by exploiting idiosyncratic variation in the staggered expansion of public childcare for children aged 1+ from 2006 to 2014 in former West Germany.[3] During these years, public childcare availability increased substantially throughout West Germany, but the timing of the expansion varied substantially across counties. Thereby, the reform generated temporal and spatial variation in public childcare availability at the county $\times$ year level that can be exploited for identification. This variation has previously been used by, e.g., Bauernschuster, Hener, and Rainer (2016) in a generalized difference-in-differences strategy similar to the one that I will employ.

I capture career progress by relying on a wage-based measure of promotions as the main outcome of interest. As wage growth early in the life cycle is especially driven by large infrequent positive wage shocks and these are often also associated with increases in responsibility, such a definition of promotions captures the key career dynamics that I am interested in. Therefore, I leverage high-quality administrative data from the German Federal Employment Agency to define promotions as a 10 percentage point increase in wage growth over the average wage growth of employees with the same educational attainment at the same firm.

In terms of descriptive evidence on the gender gap in early career dynamics, I document large differences in promotion rates between occupations. After controlling for selection into occupations and firms, I find that young men are promoted at significantly higher rates than young women, which is in line with findings from Bronson and Thoursie (2020) for Sweden. The estimates of the gender gap in promotions range from -1.6 to -2.8 percentage points, which translates into young women being promoted 10% to 20% less often.

Turning to the effects of public childcare, I provide novel evidence that an increase in the availability of public childcare has a positive and significant causal effect on the promotion probability of young female workers. A 10 percentage point increase in the childcare coverage rate increases their promotion rate by approximately 1.7 percentage points, which translates into a 10% increase. While the effect is not driven by differential selection into occupations, I do document substantial heterogeneity by firm size, namely, that the effects are driven by larger firms with $\geq$ 50 employees. In line with the literature, I find that the timing of first birth is unaffected. Furthermore, I do not find an effect of public

[3] I focus on West Germany as the public childcare system in the East was already substantially more developed prior to the reform, because it had persisted from the socialist GDR (Hank, Tillmann, and Wagner 2001).

childcare availability on wage growth for young men. This implies that the results are not driven by promotions of young women instead of young men into leadership positions. Finally, I show that my results are robust to a number of alternative specifications in terms of the sample and my definition of promotions.

With the observed response to the policy variation as a lens into the employer-employee interaction, the results show that expected future labor supply plays an important role for promotion decisions. This underlines that statistical discrimination is a plausible driver of the early career gender gap in promotion rates. Additionally, the results show that childcare policy has an important impact that goes beyond the well-documented effect of increasing mothers' labor market attachment: My findings are consistent with public childcare effectively reducing leave-related costs for employers and thereby boosting women's pre-birth career prospects through a reduction of statistical discrimination. From a cost-benefit perspective, the indirect effects on *potential* mothers provide an additional source of revenue for the government, reducing the net fiscal cost of such childcare policies.

The remainder of the paper is structured as follows: I first provide an overview of the literature in Section 2.2. Section 2.3 introduces the data and presents a number of descriptives on the early career gender gap in promotion rates in Germany. Section 2.4 introduces a stylized model of promotion decisions and statistical discrimination to sketch out the role that public childcare availability may play. In Section 2.5 I lay out the institutional background on public childcare in Germany as well as the details of the specific reform. In Section 2.6 I set up my empirical strategy and discuss the identifying variation. Section 2.7 presents my main results along with a number of robustness checks and Section 2.8 concludes.

2.2 Related Literature

The following short review of the literature on the gender wage gap highlights its most important determinants and explanatory factors. It will be necessary to credibly account for these to isolate my effect of interest, i.e., the role of anticipated future labor supply. Furthermore, I review previous evidence on the effects of family policy on women's labor market outcomes.

Evidence on the gender wage gap. The gender wage gap is a thoroughly documented feature of wage distributions around the world, with recent estimates ranging from 3.5%

to 32.5% across the OECD.[4] Its determinants have been studied extensively and one of the most dominant explanatory factors for the gender wage gap is the arrival of children and related reductions in labor supply.[5] These labor supply reductions after childbirth lead to a depreciation of human capital for mothers, but also play an important role in shaping the expectations of and about younger women (who are not *yet* mothers), on which I will focus in this paper.

The other important determinant of the gender wage gap is differential selection into fields of study, occupations, and firms. Women are underrepresented among graduates of STEM and economics subjects, which have been shown to yield long-lasting labor market payoffs over other subjects (Francesconi and Parey 2018, Kirkeboen, Leuven, and Mogstad 2016). Upon entering the labor market, women have also been shown to select into more child-friendly occupations (Felfe 2012, Hotz, Johansson, and Karimi 2017) with linear earnings, more flexible hours (Goldin 2014, Zucco 2019), and lower skill requirements as well as a lower share of non-routine tasks (Adda, Dustmann, and Stevens 2017, Gelblum 2020, Peto and Reizer 2021). I will therefore investigate potential heterogeneity in my results with respect to occupations and focus on within-firm comparisons by benchmarking individual wage growth with the average wage growth at each firm.

Zooming into the establishment level, the gender wage gap has also been documented in within-firm wage rates, with women being offered lower entry wages (Liu 2016), fewer women making their way up the organizational hierarchy (Folke and Rickne 2020a, Giannetti and T. Y. Wang 2020) and even being paid less once they make it into management positions (Geiler and Renneboog 2015). Potential explanations include gender differences in bargaining behavior, risk preferences, negotiation style, as well as self-assessment.[6] However, aside from such behavioral or preference differences between men and women, discriminatory employer behavior also plays an important role.

Such discrimination can be split in two broad categories: i) taste-based discrimination and sexual harassment and ii) statistical discrimination (Bohren et al. 2019). Taste-based

[4]Defined as the difference between median earnings of women relative to median earnings of men for full-time employees. OECD average: 12.8%, Germany: 15.3% (OECD 2021a).

[5]See Blau and Kahn (2017) and Altonji and Blank (1999) for summaries of the literature. Whether in event study frameworks (Angelov, Johansson, and Lindahl 2016, Kleven, Landais, Posch, et al. 2019, Lucifora, Meurs, and Villar 2017), studies using quasi-experimental policy variation (Costa Dias, Joyce, and Parodi 2021, Lundborg, Plug, and Rasmussen 2017) or structural life cycle models (Adda, Dustmann, and Stevens 2017, Blundell, Costa Dias, et al. 2016), the robust finding is that women who drop out after giving birth experience a significant and persistent decrease in their earnings.

[6]See Card, Cardoso, and Kline (2016), Kiessling et al. (2019), Roussille (2021), Cortes et al. (2021), and Exley and Kessler (2019) for investigations of these explanations.

discrimination has been shown to contribute to gender wage inequality by making it more costly for women to pursue, become, and remain in leadership positions (Folke and Rickne 2020b, Folke, Rickne, et al. 2020). In this paper, I will focus on the second dimension, statistical discrimination, as my policy variation affects this component. Statistical discrimination implies in my context that women are promoted at lower rates than men, because employers extrapolate the drop-out probability and associated cost from the population average to each individual. This implies that individual female workers are affected by the average behavior of all female workers.

Previous studies have furthermore shown that the hazard of dropping out depresses women's labor market entry wages and explains a sizeable share of the early career gender gap in wage growth.[7] Statistical discrimination therefore leads to women climbing the within-firm job ladder substantially slower than their male coworkers, starting from the very beginning on. The large disparities observable later in the life cycle are partially a consequence of this early divergence (Bayer and Kuhn 2020, Manning and Swaffield 2008). As my policy variation in public childcare availability potentially changes the costs associated with a female worker dropping out, I can use it to collect evidence on statistical discrimination and how policies may affect its scope.

The effects of family policy on women's labor market outcomes. Turning to the family policy literature, previous work has shown that reforms can improve women's labor market outcomes across a large number of dimensions, reducing the gender gap in most cases.[8] Focusing on the effects of childcare policy, there is a large literature studying its effects on women's outcomes. Especially broad is the evidence for the effects of public childcare availability on maternal labor supply, where studies find positive effects in most countries.[9] For the reform I focus on, the 2006 to 2014 expansion of public childcare in Germany, K.-U. Müller and Wrohlich (2020) document increases in labor market participation and hours worked. The earlier reentry of mothers also led them to be promoted at higher rates (Chhaochharia et al. 2021). This specific reform has furthermore been studied in terms of fertility, with positive effects on the number of children, but no effect on the timing of first birth (Bauernschuster, Hener, and Rainer 2016). The broadly

[7] See Neumark and Vaccaro (2020), Xiao (2020), Bronson and Thoursie (2020), Bagger, Lesner, and Vejlin (2019) and Biewen and Seifert (2018) for different examples of this effect.

[8] See Doran, Bartel, and Waldfogel (2018), Laun and Wallenius (2021), Olivetti and Petrongolo (2017), Kleven, Landais, Posch, et al. (2020) for policy evaluations in this regard.

[9] Literature examples include Nollenberger and Rodríguez-Planas (2015), Carta and Rizzica (2018), and Cascio, Haider, and Nielsen (2015). Norway is a notable exception, where Havnes and Mogstad (2011) find no effects.

documented increases in labor supply motivate my hypothesis that public childcare decreases the expected costs of female employees dropping out by decreasing the average time that mothers spend out of the labor force.

Focusing on whether family policy affects statistical discrimination, this paper is closest to the following ones, which evaluate other policy reforms in this regard: J. Jessen, R. Jessen, and Kluve (2019) study a policy that switched firms' payments for mandatory maternity leave compensation from direct payments to employees who dropped out to a flat fee only dependent on the size of the workforce (and independent of its gender composition). This reform reduced the costs directly tied to dropping out and led to significant increases in the wages of female employees in the affected firms. Looking at a different family policy dimension, Thomas (2020) evaluates the introduction of a mandatory maternity leave policy, i.e., an increase in the drop-out length, and finds that it decreased the pre-birth promotion probability by almost one-third. In a similar spirit, Ginja, Karimi, and Xiao (2020) study the expansion of parental leave in Sweden and find that it lowered hiring probabilities for women and depressed their entry wages.

2.3 Data and Descriptives

2.3.1 Sample

Data sources and construction. My main data source is the Sample of Integrated Labor Market Biographies (SIAB), which is provided by the Federal Employment Agency of Germany.[10] It is a two percent random sample drawn from the universe of all German employees, excluding the public sector and the self-employed. In terms of scope, the SIAB contains the full set of employment spells that occurred between 1975 and 2017. This allows me to reliably track young individuals from the very start of their working life. As the SIAB also contains firm identifiers for every employment spell, I can furthermore merge a number of administratively collected employer characteristics from the Establishment History Panel (BHP) to the data and track individuals' movements between firms.[11]

[10]This study uses the weakly anonymous Sample of Integrated Labour Market Biographies (Years 1975-2017). Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access. Source: FDZ-IAB (2021b). See Antoni et al. (2019) for a detailed description.

[11]The weakly anonymous Establishment History Panel 1975-2017 was used in the same fashion as the SIAB. Source: FDZ-IAB (2021a). See Schmucker et al. (2018) for a detailed description.

In terms of the variables available, the SIAB includes a number of demographics such as age, education, nationality, along with detailed employment spell characteristics such as occupational codes and part-time status.[12] Crucially, these also include individual daily wages, which are calculated from the per-spell wages divided by the number of calendar days in each spell and converted into real terms. The underlying source of this wage information are mandatory employer reports linked to social security payments. They are therefore, contrary to survey-based wage measures, not subject to subjective reporting and related measurement errors. A drawback of the data is, however, that the wage data is top-censored at the social security contribution limit.

From the linked establishment data, I retrieve information on firm size, 2-digit industry codes as well as the mean firm-specific wage of all full-time employees by education level.[13] The latter variable is computed by the data provider on the universe of labor market biographies, for which top-censored wages are imputed to capture the mean establishment wage more accurately.

Lastly, the SIAB contains county identifiers for the place of residence of every individual. Using these identifiers, I merge a number of county characteristics to the individual employment data. The county level data stems from the INKAR database and a federal database of regional statistics.[14] The most important information from these sources is my measure of public childcare availability: the local childcare coverage rate. It is defined as the number of 0 – 2 year olds per county who are enrolled in childcare divided by the total number of 0 – 2 year olds per county.

With the complete data set at hand, I leverage the employment histories to construct a number of additional variables, such as work experience as well as the labor market entry age. Furthermore, I employ the procedure proposed by D. Müller and Strauch (2017) to identify births for female individuals to be able to focus my analysis on career development prior to first childbirth.[15]

[12]As 2-digit occupations I use 14 occupational segments based on the German Classification of Occupations 2010 (Matthes, Meinken, and Neuhauser 2015).

[13]For 2-digit industries I use 13 categories based on the NACE Rev. 2 classification (with the following aggregations: D+E, G+H, K+L, M+N, P+Q, and R+S+T).

[14]INKAR is a database of regional development indicators provided by the Federal Institute for Research on Building, Urban Affairs, and Spatial Development (BBSR Bonn 2021). The federal database of regional statistics is provided by the German Federal Statistical Office (FDZ-StABL 2021).

[15]The procedure works as follows: German labor regulation mandates women to stop working at least six weeks prior to the projected birth date. At this point, employers file a special notice to the social security administration indicating the entry into mandatory maternity leave. Using the occurrence and dates of these notices, it is possible to recover the timing of first birth for most women who were working prior to that. See D. Müller and Strauch (2017) for details.

While the data quality of the SIAB is overall very high, there are nevertheless two variables that require imputation: First, I follow Fitzenberger and Seidlitz (2020) in correcting the part-time indicator prior to 2012, as I have to rely on this indicator to be able to focus on the wage development of full-time working individuals.[16] Second, I apply the same imputation procedure that is used to construct the mean wage at the firm level to impute individual wages above the social security contribution limit.[17] This allows me to include wage developments above the contribution limit in my analysis. Finally, I convert the SIAB spell data to a panel by using each individuals dominant spell, i.e., the one that covers the most months for every year.

Sample definition. Building on the data set described above, I proceed to construct my main sample. I include the years 2006 to 2014 and focus on West Germany, as this is the time span and location during which I have reform-driven variation in public childcare availability. I exclude individuals who are in education and only focus on full-time workers, because for these I can reasonably assume that they work similar hours, which then allows me to compare wages across workers.[18] As I am interested in within-firm career dynamics, I restrict the sample to consecutive full-time workers who stay at the same firm.[19] This allows me to benchmark their wage growth vs. their coworkers.

Prior to first birth, the share of women working full-time is very comparable to that of young men (≈ 80%) and they only start to substantially differ when children enter the picture, as illustrated in Figure B.1. Therefore, I would ideally want to limit the sample to women prior to first birth. For the analysis of the effects of public childcare, however, the timing of first birth may be directly affected itself. To avoid endogenous sample selection on an outcome, I therefore restrict the sample to the age range between the pre-reform

[16]Before 2012, the indicator had no relevance for social security payments and there is evidence that firms were substantially underreporting switches from full-time to part-time. In 2012, an update of the reporting procedure increased the reliability of this variable substantially, serving as the basis for the imputation procedure proposed by Fitzenberger and Seidlitz (2020).

[17]Following the procedure laid out in Section 8.1 in Schmucker et al. (2018).

[18]As I only observe a part-time indicator, I cannot accurately compute hourly wages for part-time workers, because hours worked in part-time differ substantially. Therefore, I focus on full-time workers, for whom hours worked have a much lower variance (see Figure 3.1 in Chapter 3).

[19]This strategy abstracts from switching employers as another important potential way to realize wage growth and focuses on within-firm dynamics. It captures the majority of individuals in the sample age range (≈ 80%, see Appendix-Table B.1), but there may be differential selection into switching the employer based on unobservables.

median entry age of non-college graduates (age 22) and the median age at first birth of college graduates (age 32) instead of conditioning on prior to first birth.[20]

Table 2.1: Summary statistics

	women	men
age	27.55	28.00
non-german	8.16%	11.80%
college degree	22.32%	20.00%
experience in full-time (years)	4.72	5.01
daily earnings (log EUR)	4.40	4.56
age at labor market entry	23.30	23.67
age at first birth	29.71	-
occupations		
goods production	7.17%	48.96%
personal services	37.50%	12.80%
business services	47.04%	20.46%
IT and science	1.98%	4.45%
other servcies	6.31%	13.34%
occupational characteristics		
occupation with demanding tasks	30.12%	21.51%
occupation with demanding know-how	20.34%	20.96%
individuals	57,630	80,758
individuals × year	176,538	260,499

Notes: 'demanding tasks' denotes a share of analytic non-routine tasks $\geq \frac{1}{3}$) and 'demanding know-how' requirement denotes that specialist or expert knowledge is required. Sample: individuals between ages 22 and 32 (median labor market entry age without college and median age at first birth with college) working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

Summary statistics. Turning to the summary statistics of the sample in Table 2.1, I end up with 57,630 women in 176,538 annual observations and 80,758 men in 260,499 annual observations. With the county level analysis in mind, this implies approximately 180

[20] I determine this age range using data from the years 2000 to 2005 to avoid contamination through effects of the reforms.

women per each of the 325 West German counties. In terms of the core demographics (age, nationality, education), women and men are very comparable. The women in my sample have slightly less full-time experience and lower daily wages then men, but both enter the labor market at a similar age.

Substantial differences are, however, visible between occupations: 85% of women work either in personal or business services occupations, while only 33% of men do so. Approximately half of men are employed in jobs related to goods production, but only 7% of women. IT and science occupations make up only a small share for both genders (2% and 4%), while other services (safety, logistics, cleaning) account for 13% of men and 6% of women. If promotion probabilities differ across occupations, then the segregation of men and women in terms of occupations will impact the observed gender difference in promotions.

To compare wage developments across occupational characteristics, I add two additional occupational categorizations: First, I code an occupation as consisting of a demanding task set, if more than one-third of its tasks are analytic non-routine tasks.[21] Second, I code an occupation as requiring demanding know-how, if the occupational code submitted by the employer indicates that the job requires specialist or expert skills. In my sample, 30% of women and 22% of men work in occupations with demanding tasks and around 20% of each gender work in jobs requiring demanding know-how.

Comparing Table 2.1 to Appendix-Table B.1, which contains all employees within my target age and year range without conditioning on consecutive full-time work at the same employer, allows to judge the selectivity of the sample. While the sample is positively selected in terms of full-time experience and wages, especially for women, the remaining characteristics are very comparable. Acknowledging this positive selection, descriptive comparisons across genders have to be interpreted with some caution. Nevertheless, the full-time restricted sample is the relevant one to investigate within-firm career progression early in the life cycle across genders, as it rules out that gender differences are driven by differences in labor supply.

2.3.2 Measuring within-firm career progress via promotions

To compare within-firm career dynamics consistently across different firms, I require a comparable and reliable measure of promotions. Turning to the literature, a number of

[21] I use the occupational task classification into analytic non-routine, interactive non-routine, cognitive routine, manual routine, and manual non-routine provided by Dengler, Matthes, and Paulus (2014).

alternatives have been proposed: i) self-reported promotion, ii) changes in job titles or job assignment, iii) changes in managerial responsibility (supervision of employees), or iv) wage-based measures.[22]

Given my high-quality data set on wages, I will employ the latter to construct an objectively quantifiable promotion measure for all employees in my sample. Defining promotions as large upward moves within the wage hierarchy of each firm allows me to broadly analyze within-firm career dynamics. This definition will focus the analysis on large upward moves accompanied by significant wage increases and abstracts from horizontal moves or job title changes without substantial changes in compensation. Furthermore, it does not rely on subjective assessments of career dynamics, for which previous research has shown substantial differences between women and men (Exley and Kessler 2019). Additionally, wage-based promotions do capture career moves that do not involve additional managerial responsibility (such as among specialists). Finally, previous research has shown that such large wage growth episodes are highly correlated with the other measures of promotions mentioned above (Bronson and Thoursie 2020).

To make a wage-based promotion measure comparable across firms, it is important to separate large wage increases for individuals from firm-wide wage increases.[23] To accomplish this, I follow Bronson and Thoursie (2020) and leverage the scope of my administrative data to benchmark individual wage growth against the mean wage growth at the firm. More specifically, I construct 'relative wage growth' as the difference between an individual's wage growth and the mean wage growth of full-time workers with the same education level at the same firm. This approach eliminates any firm $\times$ education specific wage dynamics from the individual wage dynamics, such as, e.g., the effects of collective bargaining agreements. Furthermore, it allows for a comparable classification of promotions across firms by only measuring wage growth in excess of the mean wage growth at each firm.

Figure 2.1 presents the resulting 'relative wage growth' distribution.[24] For my measure of promotions, I binarize the distribution through a split along the 10% margin. Thereby, I

[22]Self-reported promotions have been used in McCue (1996) and Kosteas (2011) for example. See Klaauw and Dias da Silva (2011) for an example of coding changes in (standardized) job titles or associated job assignments as promotions. Biewen and Seifert (2018) use changes in managerial responsibility, e.g., the headcount of subordinates, to construct a promotion measure.

[23]Appendix-Figure B.2 illustrates the distribution of annual real wage growth in my sample. While most wage growth is incremental, there are occasional large positive shocks, as illustrated by the right tail of the distribution.

[24]See Appendix-Figure B.2 for a histogram of non-benchmarked real wage growth in comparison. The benchmarking notably increases the centering of the distribution around 0.

follow Bronson and Thoursie (2020) again, who show that the 10% threshold is well suited to separate lower (incremental) growth episodes from large upward moves. Visualized by the red vertical line in Figure 2.1, this split creates an easily interpretable and comparable wage-based measure of promotions.[25]

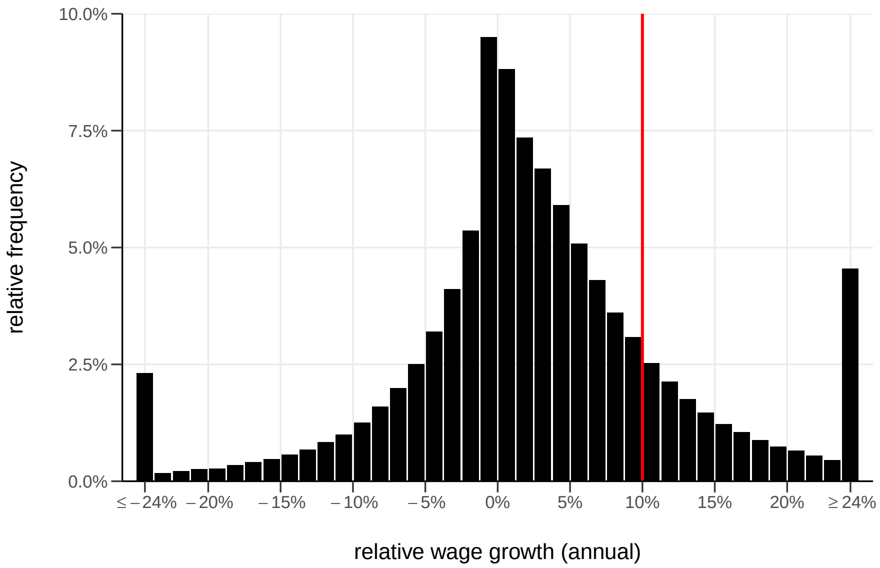


Figure 2.1: Relative wage growth distribution

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level, all values in real terms. The vertical red line visualizes the promotion definition threshold at 10% relative wage growth. Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

Promotions defined as such can be characterized as follows in my sample: They occur with low frequency, as on average only 17.76% of individuals are promoted every year. The associated wage growth is large in magnitude, with wage growth from promotion episodes accounting for 58% of total wage growth. Furthermore, promotions lead to persistent wage growth, as 82% of individuals retain or exceed the wage level achieved during a promotion episode five years later. In summary: Promotions are rare, but they do matter substantially for career progress.

[25]I also show robustness of my results to alternative values for the cutoff in Section 2.7.4.

2.3.3 Gender differences in promotion rates

Building on the measure of promotions just introduced, I now turn to descriptive evidence on gender differences in promotion rates. Aside from an aggregate comparison across genders, I take into account the previously described differences between women and men in sorting across occupations. My aim for these descriptives is to highlight patterns in promotion rates among young workers who have just entered the labor market, i.e., for whom there is no gender difference in labor market experience that may drive such patterns for older workers.

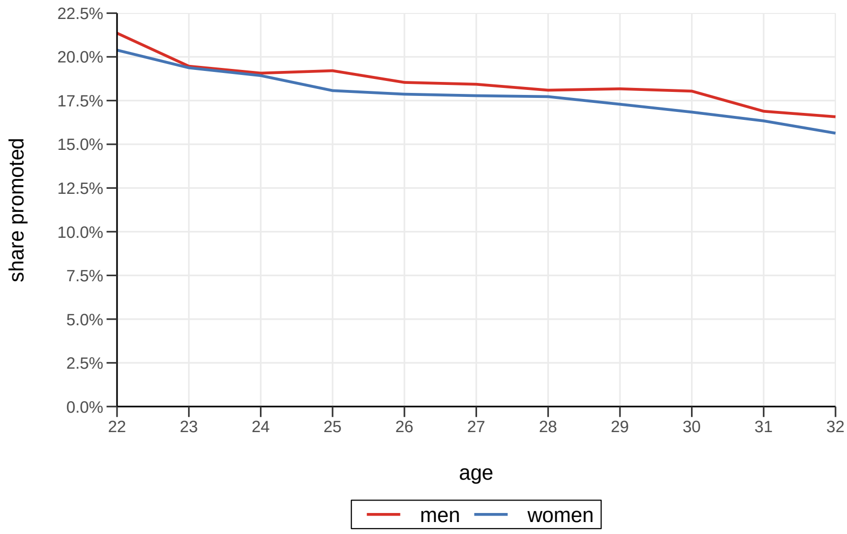


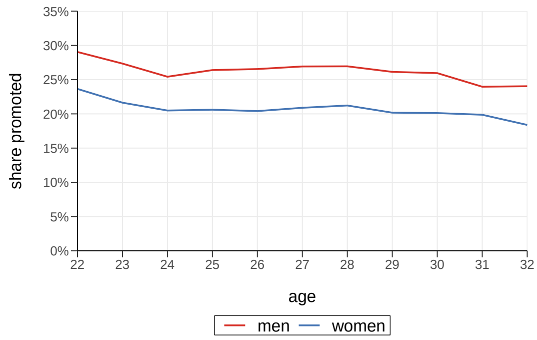
Figure 2.2: Promotion rates by age and gender

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

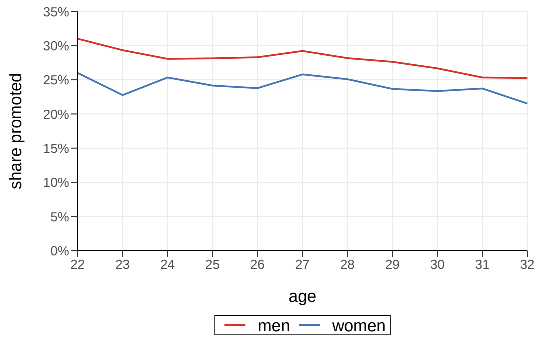
I start with a descriptive comparison of promotion rates across age and gender in Figure 2.2. Promotion rates are decreasing in age (from approximately 21% at age 22 to around 16% at age 32), which is in line with promotion opportunities becoming rarer as individuals climb the career ladder.[26] Comparing promotion rates across genders, I (surprisingly) do not find clear differences between women and men, despite men having

[26]Given the higher wage level of experienced workers, the same relative wage increase implies a larger increase in absolute terms. Furthermore, there may be fewer positions available to be promoted into the higher an individual has climbed the career ladder already.

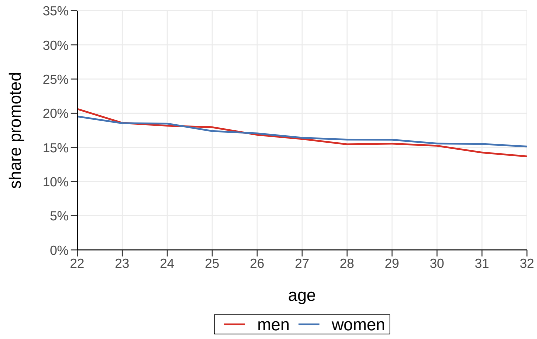
slightly higher rates almost across the entire age range. This may, however, be simply the result of the differences in the observables discussed above, if, for example, the women in my sample tend to work in occupations with higher average promotion rates. To investigate this, I turn to splits by occupational characteristics next.



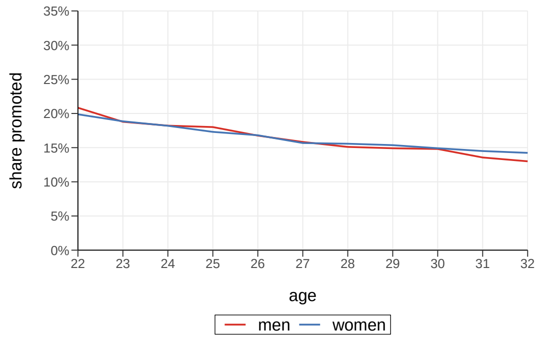
(a) demanding tasks



(b) demanding know-how



(c) not demanding tasks



(d) no demanding know-how

Figure 2.3: Promotion rates by age, gender, and occupational characteristics

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Subsamples with demanding tasks (share of analytic non-routine tasks ≥ $\frac{1}{3}$) or demanding know-how requirement (specialist or expert knowledge required). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

In Figure 2.3 I report promotion rates across ages for the subsamples of individuals working in jobs with demanding tasks or that require demanding know-how. Contrary to Figure 2.2, these two figures both show a clear gender gap. Focusing first on jobs composed of demanding tasks in Figure 2.3a, men are promoted on average 25.7% of the time, compared to only 20.4% for women. Across the entire age range 22 to 32, a steady 5 percentage point gap is visible. Shifting the focus to occupations that require demanding know-how in Figure 2.3b, the gender gap is smaller (around 3 percentage points), but also persistent across the age range. In summary, I find that when comparing women and men in jobs with similarly demanding task content or required know-how, young men are promoted at substantially higher rates than their female peers.

Turning to jobs without demanding tasks or no expert know-how required, there are no gender gaps visible in Figures 2.3c and 2.3d, but promotions rates are substantially lower (by 5 to 10 percentage points). The positive selection of women (e.g., higher share working in jobs with demanding tasks) is therefore a clear driver of the gender parity in aggregate promotion rates in Figure 2.2.

To investigate this selection further, I split the sample by occupations in Appendix-Figure B.3, as this is where the summary statistics in Table 2.1 also showed large differences. Overall, the heterogeneity in promotion rates across occupations illustrates that there are substantial differences in wage dynamics by occupation. Focusing on personal and business services in Appendix-Figures B.3b and B.3c, where 85% of women and one-third of men work, I document sizable gender gaps. The gaps range from 2 to 5 percentage points for personal services and from 4 to 7 percentage points for business services. Furthermore, despite being below the promotion rate of men in business services, the average promotion rate of women in business services at 19.31% is more than a full percentage point higher than the population average male promotion rate at 18.17%. This illustrates the positive selection of women into occupations with above average promotion rates, yielding the aggregate parity in Figure 2.2 despite clear occupation-specific gender gaps. Examples of such occupations with above average promotion rates as well as a high female share are jobs business organization and strategy, in insurance and financial services as well as in law consulting and marketing.[27]

However, the documented gender gaps within occupations may also be rooted in selection. The sample of males working in the predominantly female sectors (business and personal services) may be positively selected, i.e., only the most able men work in these sectors and those get promoted at higher rates.

Therefore, I now turn to investigating the gender gap in promotion rates in a more structured way using a regression framework in which I condition on the complete set of observables. More specifically, I regress my binary promotion indicator on an indicator for women, a set of controls, and firm as well as year fixed effects. This allows me to account for selection on observables and eliminate time-invariant firm heterogeneity.

[27] These occupations account for 15.5% of females, but for just 7.5% of males. Women in these occupations are promoted at a rate of 22% (vs. 18.17% average male promotion rate).

The results can illustrate the scope of the gender wage gap, but cannot pin down the causal effect of female gender on promotions, due to, e.g., selection on unobservable individual characteristics.

Table 2.2: The gender gap in promotions

	promotion (relative wage growth ≥ 10%) – mean: 0.1776 –			
	(1)	(2)	(3)	(4)
women	-0.0167^{***}	-0.0159^{***}	-0.0285^{***}	-0.0159^{***}
	(0.0025)	(0.0025)	(0.0027)	(0.0025)
human capital controls	✓	✓	✓	✓
occupational characteristics		✓	✓	✓
occupation/industry FE			occupation	industry
year + firm FE	✓	✓	✓	✓
N	402,674	402,674	402,674	402,674

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Covariates included: human capital (age, age sq., non-German, education, FT experience, FT experience sq.), occupational characteristics (demanding tasks, demanding know-how), occupation and industry FE (2-digit). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

Table 2.2 presents the results for these panel data regressions, where the mean of the dependent variable, the promotion indicator, is 17.76% across all columns. The coefficient on the women indicator captures the gender gap in promotion rates conditional on selection on the observable characteristics included in each column.

Starting in column (1), I include a set of controls capturing differences in human capital: a second-order polynomial in age and full-time work experience along with nationality and education. Given these controls, I find a gender gap of −1.67 percentage points, corresponding to women being promoted approximately 10% less often than men. Taking into account differences in occupational characteristics (task composition/required know-how) in column (2), the gap remains similar at −1.59 percentage points. However, when adding granular controls for different occupations in column (3), the gender gap in promotions increases substantially to −2.76 percentage points. In turn, this implies that women are promoted 20% less often than men. Controlling alternatively for 2-digit industry codes in column (4) yields similar coefficients as in columns (1) and (2). Keeping in mind that wage increases from promotion episodes are very persistent, the large gender gaps in promotion rates in Table 2.2 translate into substantially lower wage trajectories for women over their life cycle.

To ensure that these results are not driven by women that gave birth in their early 20s and returned to work quickly, I re-estimate all descriptive results while restricting the sample of women to prior to first birth in Appendix B2.[28] The graphs and tables based on the restricted sample confirm all findings from above.

With this descriptive evidence at hand, I will next illustrate a plausible underlying mechanism in a lightweight, but intuition-rich model. I then also use the model to introduce how childcare policy can affect young women's career prospects and contribute to reducing the just described gender gap.

2.4 Theoretical Considerations on Women's Career Dynamics

Given the descriptive evidence in the previous section, I introduce a model of gender-specific career dynamics that can explain the observed gender gap in promotion rates. This model builds on the framework laid out in detail in Bronson and Thoursie (2020). It provides a clear formulation of firms' decision making on promotions and generates the key prediction that early in their careers, women are promoted less frequently than men. I then extend the model to capture how the availability of public childcare may affect the firms' promotion decisions. Finally, I discuss alternative explanations for how childcare may affect women's early career dynamics.

2.4.1 Benchmark model

Building on the model originally developed by Gibbons and Waldman (1999), Bronson and Thoursie (2020) introduce a simplified life cycle model, which carries a rich intuition

[28] I take a conservative approach in terms of first births and restrict the sample to prior to the first long employment interruption, as I can only reliably identify first birth if women are working before.

for early career developments. I lay out the core components of this model here for ease of exposition and to illustrate where public childcare comes into play.[29]

General environment. The economy consists of a continuum of identical firms, who can freely enter into production, and a continuum of heterogeneous workers. All agents are risk neutral, have full information, and do not discount the future. Workers i differ by gender (male and female) and with respect to their innate ability θ_i. Furthermore, they accumulate labor market experience x_{it} linearly with every period they work. Together, these two components characterize the period-specific *effective* ability of each worker: $\eta_{it} = \theta_i f(x_{it})$, with $f'(\cdot) > 0$ and $f(0) = 0$. The time horizon of the model spans the full working life and consists of four periods that are illustrated in Figure 2.4.

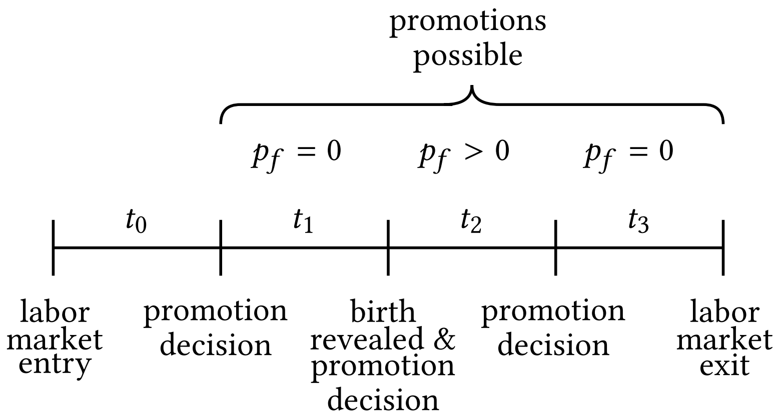


Figure 2.4: Model structure and timing

Notes: Taken with slight modifications from Figure 12 in Bronson and Thoursie (2020), p. 54.

The periods are: t_0, the labor market entry period; t_1, the pre-childbearing period; t_2, the period with potential childbearing; t_3, the post-childbearing period. At the beginning of period t_2, a share p_f of women gives birth and stops working temporarily while taking parental leave until the end of t_2. In all other periods, p_f is set to 0.

Production. Each firms' production consists of $J + 1$ jobs who become more demanding in j and for which labor is the only input. A worker i in job j in period t has the production function $y_{ijt} = d_j + c_j \eta_{it}$, where $c_j, d_j > 0$ and $c_{j+1} > c_j$, $d_{j+1} < d_j$. Output therefore

[29] For the original proofs and an extensive discussion of the model, see Bronson and Thoursie (2020), Section 6, and Appendix B.

consists of an ability-independent job-specific component d_j and an ability-dependent job-specific component $c_j\eta_{it}$. Furthermore, as $d_{j+1} < d_j$, the ability-independent component is the largest for low-level jobs, while output depends increasingly on *effective* ability for more demanding jobs ($c_{j+1} > c_j$). As workers accumulate labor market experience, their *effective* ability increases and they become more productive at more demanding jobs.

In periods when workers do not work, but remain attached to their firm via parental leave, the production function becomes $y_{ijt} = -k_j$ with $k_{j+1} > k_j > 0$. k_j captures that firms have to hire and train temporary replacements and are faced with coordination and knowledge transfer costs when employees go on leave. These costs are increasing in the complexity of the job. This incorporates that a manager or a worker conducting complex tasks is more costly to temporarily replace than a low-level worker with no hierarchical power or simple stand-alone tasks. Firms cannot avoid or decrease these costs through lay-offs or demotions, which lines up with German parental leave regulation that allows leave-takers to return to their old jobs.[30]

Promotions & equilibrium. Firms optimize the efficiency of their output through the allocation of workers to jobs and make zero profits. Workers are paid a per-period wage that corresponds to their output $w_{ijt} = y_{ijt}$. At the beginning of the life cycle in period t_0, it is most productive to hire all workers into job $j = 0$, as they do not have any work experience yet. From period t_1 on, firms make promotion decisions at the beginning of each period, but after observing potential childbirths (see Figure 2.4). These promotion decisions can be expressed in terms of *effective* ability thresholds $\bar{\eta}^j$ as follows: $\bar{\eta}^j$ solves $d_j + c_j\eta_{it} = d_{j-1} + c_{j-1}\eta_{it}$, assuming that c_j and d_j are such that $\bar{\eta}^{j-1} < \bar{\eta}^j < \ldots < \bar{\eta}^J$. If $\eta_{it} > \bar{\eta}^j$, then worker i is more productive in job j than in job $j-1$ and it is efficient to promote that worker into job j. In terms of the timing of promotions, individuals with the highest innate ability θ_i will be promoted first, as their *effective* ability exceeds the threshold $\bar{\eta}^j$ the earliest (even with only little work experience). For male workers, for whom there is no uncertainty in terms of labor market participation, this implies the following *effective* ability threshold for promotions:

$$\bar{\eta}^j = \frac{d_{j-1} - d_j}{c_j - c_{j-1}}. \tag{2.1}$$

[30]Mothers are entitled to job retention in terms of pay and responsibility, if they return with the same number of hours within three years.

Early life cycle dynamics for women. For female workers, on the other hand, firms take into account that they may incur cost k_j in period t_2, if childbirth occurs. In period t_3 and in period t_2 if no childbirth occurs, i.e., in cases where all uncertainty with respect to fertility is resolved, female workers are promoted like men according to (2.1). In period t_2, however, firms will not promote women that have just given birth, as k_j is increasing in j. Turning to period t_1, firms know that a share p_f of women will leave the labor force in period t_2. The firms do not know specifically which female workers will give birth and as workers are unable to credibly signal whether they intend to have children. Therefore, firms incorporate the uncertainty of childbearing into their promotion decision making concerning all women. The additional cost component k_j yields the following *effective* ability threshold for promotions of women in period t_1:

$$\bar{\eta}^* = \bar{\eta}^1 + p_f \frac{k_1 - k_0}{c_1 - c_0}. \tag{2.2}$$

Given that $p_f > 0$, $k_1 > k_0$, and $c_1 > c_0$, the additional additive component $p_f \frac{k_1 - k_0}{c_1 - c_0}$ is strictly positive. This implies that $\bar{\eta}^* > \bar{\eta}^1$, i.e., the t_1 promotion threshold is higher for women and therefore fewer women are promoted in period t_1. In other words, women face a gender penalty in terms of their pre-childbearing promotion probabilities.

This main result from Bronson and Thoursie (2020) now serves as the basis for the following extension of their model to incorporate the role of public childcare.

2.4.2 Extension with public childcare

Bronson and Thoursie (2020) summarize in k_j all costs to the employer when a worker is on parental leave and only allow these costs to vary by job level j. While this abstraction is reasonable given a fixed policy environment, I will now allow for childcare policy interventions to affect k_j. This extension is motivated by evidence that childcare policy did change the labor market participation rates of mothers, allowing them to return to the labor market earlier (K.-U. Müller and Wrohlich 2020). Specifically, I introduce heterogeneity in k_j by the availability of public childcare in the following fashion (derivation in Appendix B3):

$$\bar{\eta}^* = \bar{\eta}^1 + p_f \left[p_{cc} \frac{k_1^{cc} - k_0^{cc}}{c_1 - c_0} + (1 - p_{cc}) \frac{k_1^{ncc} - k_0^{ncc}}{c_1 - c_0} \right]. \tag{2.3}$$

With p_{cc} as the probability that a female worker has access to public childcare, the costs of parental leave are now heterogeneous by the availability of public childcare. In this framework, I assume that public childcare (denoted by the superscript *cc*) lowers the leave-taking cost for employers compared to the situation without public childcare (denoted by the superscript *ncc*), i.e., $k_j^{cc} < k_j^{ncc}$. This can be seen as a reduced form approach to incorporating the evidence referenced above, namely that public childcare induces female workers to take shorter parental leave and return to the labor market earlier. Even in cases of high labor market participation rates of mothers without it, public childcare can be seen as a factor that increases the reliability and flexibility of non-domestic childcare, allowing mothers to return to their original jobs (in terms of hours worked/responsibility) more quickly.

Given this extension, I can now investigate how an increase in the availability of public childcare affects the *effective* ability threshold for promotions of female workers in period t_1. To help with the exposition, I assume that $k_{j+1} > k_j$ can be captured by $k_{j+1} = \tau_{j+1} \cdot k_j$ with $\tau_j > 1$.

$$
\begin{aligned}
\bar{\eta}^* &= \bar{\eta}^1 + p_f \left[p_{cc} \frac{k_1^{cc} - k_0^{cc}}{c_1 - c_0} + (1 - p_{cc}) \frac{k_1^{ncc} - k_0^{ncc}}{c_1 - c_0} \right] \\
&= \bar{\eta}^1 + \frac{p_f}{c_1 - c_0} \left[p_{cc} \cdot (k_1^{cc} - k_0^{cc}) + (1 - p_{cc}) \cdot (k_1^{ncc} - k_0^{ncc}) \right] \\
\frac{\partial \bar{\eta}^*}{\partial p_{cc}} &= \frac{p_f}{c_1 - c_0} \left[(k_1^{cc} - k_0^{cc}) - (k_1^{ncc} - k_0^{ncc}) \right] \\
&= \frac{p_f}{c_1 - c_0} \left[(k_1^{cc} - k_1^{ncc}) + (k_0^{ncc} - k_0^{cc}) \right] \\
&= \frac{p_f}{c_1 - c_0} \left[\tau_1 \cdot (k_0^{cc} - k_0^{ncc}) + (k_0^{ncc} - k_0^{cc}) \right] \\
&= \frac{p_f}{c_1 - c_0} \left[-\tau_1 \cdot (k_0^{ncc} - k_0^{cc}) + (k_0^{ncc} - k_0^{cc}) \right] \\
&= \frac{p_f}{c_1 - c_0} (k_0^{ncc} - k_0^{cc}) \left[1 - \tau_1 \right]
\end{aligned}
$$

With $p_f > 0$, $c_1 > c_0$, $k_j^{ncc} > k_j^{cc}$, and $\tau_j > 1$, I can conclude that $\frac{\partial \bar{\eta}^*}{\partial p_{cc}} < 0$. In other words, as childcare availability increases, the *effective* ability promotion threshold for female workers in period t_1 decreases, which leads to more women being promoted. This model prediction of a positive relationship between childcare availability and women's promotion probabilities is what I will take to the data in the following sections.

2.4.3 Alternative explanation

Aside from the mechanics laid out above, another mechanism for how public childcare availability may affect young women's promotion probabilities is the following: As forward-looking agents, young women may anticipate that they can return to work earlier as public childcare becomes available. This increases their incentive to accumulate human capital and exert high effort on the job, as they are able to benefit from their experience/moves up the career ladder during a longer time in the labor market. Such higher effort may increase their chances of being promoted, leading to a positive relationship of public childcare availability and promotions as well.[31]

2.5 Background on Family Policy in Germany

The current family policy landscape in Germany consists of three main components: i) child benefits/allowances and joint taxation, ii) paid maternity leave and employment protection, and iii) subsidized public childcare. Using these three components, German policy makers attempt to shape the labor market and fertility incentives for women to achieve specific policy goals. Since the mid 1990s, one of the main goals has been to increase the compatibility of sustained career building with having a family.[32]

Focusing on the labor market incentives prior to first birth, the status quo of the first two components looks as follows: In terms of the tax and transfer system, joint taxation creates labor supply disincentives for spouses that earn less than their partners, which are women in many cases.[33] In terms of maternity leave, mothers are entitled to employment protection for up to three years and income contingent benefits for up to

[31] One strategy to test this alternative explanation vs. the mechanics in the model (which focus on costs to the employers) would be to look at the effect of public childcare availability on the promotion probabilities of women that choose to remain childless. These women will not respond with higher on-the-job-effort to public childcare availability, as such does not matter for their decision making. If their promotion probabilities would nevertheless be affected, that would be evidence for the mechanics introduced in the model. As it is impossible to reliably identify women who choose to remain childless in my data, I am unable to conduct this test and the findings will have to be interpreted with the alternative explanation in mind.

[32] See Bundesministerium für Familie, Senioren, Frauen und Jugend (2004) for the draft of the federal daycare expansion law ("Tagesbetreuungsausbaugesetz"), which includes a discussion of the policy makers motives.

[33] The German joint taxation scheme works as follows: The incomes of both spouses are summed up, divided by two and the tax on that hypothetical income is calculated. The couple then pays double that tax. Given the progressive German tax code, this creates higher (lower) marginal tax rates for the lower (higher) earning spouse that than under individual taxation (Bick, Brüggemann, et al. 2019).

one year. As benefits amount to 65% of pre-birth earnings and are only capped at €1,800, this creates incentives to increase earnings prior to giving birth. Importantly, these first two family policy components affect the entire female population in a similar manner, i.e., there is little to no regional heterogeneity.

2.5.1 The German childcare market

Turning to subsidized public childcare, there are two separate types of institutions: early childcare (ECC) facilities cater to 0 – 2 year olds and kindergartens cater to ages 3 to 6. For both types, the German childcare market consists predominantly of facilities operated by municipalities or non-profit organizations. Only a small share of 3% are operated by for-profit organizations, including some that are integrated in large firms (Statistisches Bundesamt 2014).

In terms of fees, parents pay on average a monthly fee of €250 for a public ECC slot, with the fee schedule being progressive in household income and regressive in the number of siblings. As the actual costs for a slot amount to more than €1,000 per month, ECC care is substantially subsidized (Stern et al. 2015). The quality of childcare in these facilities is heavily regulated and thereby largely homogeneous. These regulations concern almost all aspects of operation, from opening hours to group size, staff-to-child ratios, and staff qualifications. The strict regulations are a likely reason for the small size of the private market (Bauernschuster, Hener, and Rainer 2016, Felfe and Lalive 2018).

Regarding the supply of slots, there is historically and still contemporaneously quite a lot of regional heterogeneity. The main dimension of this heterogeneity is between former West and East Germany. As the socialist government in the East supported and sometimes even mandated the use of public childcare for all age ranges, East Germany had a substantially higher supply and uptake of slots than West Germany at reunification.[34] In the mid 1990s, more than 50% of children below 3 were attending childcare in the East, while only 1.2% did so in the West. Aside from the significant East-West divide, the local nature of the providers also gives rise to regional heterogeneity in the supply of ECC slots. The entire scope of this heterogeneity is illustrated in Figure 2.5, which plots the county distribution of the childcare coverage rate (number of childcare slots / headcount of children) for 2002 and 2014. For both years separately, the childcare coverage rates are

[34]The socialist government in the former GDR (East Germany) had set up a broad public childcare system to allow mothers to work. This system persisted after reunification of Germany. See Hank, Tillmann, and Wagner (2001) for details on the differences in the childcare systems between West and East Germany after reunification and Hüsken (2010) for a detailed discussion of the situation in the mid 2000s.

significantly higher in East Germany than in the West. Furthermore, a large degree of regional heterogeneity between counties is also visible, especially in West Germany.

2.5.2 Recent public childcare reforms

The other source of heterogeneity that Figure 2.5 clearly illustrates and that I will focus on now is reform-induced temporal variation in the supply of ECC slots. For this discussion, I will focus on West Germany, as the following reforms aimed to end the rationing of public childcare in the West, while there was little to no rationing in the East.

Since the mid 1990s, national governments from varying parties have introduced a large amount of legislature targeted at increasing the number of public childcare slots. Starting in 1996 with the "Kinder- und Jugendhilfegesetz", a legal entitlement to a slot for all children aged 3 and above was introduced. To meet the demand and comply with the entitlement, the number of kindergarten slots was substantially expanded, resulting in almost full coverage for 3+ year olds by the early 2000s (Bauernschuster and Schlotter 2015).

There was, however, only a very limited supply of ECC slots, i.e., for 0 – 2 year olds, available at that time, with on average just 5 slots per 100 children in 2002 (Bien, Rauschenbach, and Riedel 2006). This situation of severe rationing of slots with long waiting lists and lotteries implied that many parents (mostly mothers) had to provide childcare domestically and could not choose to return to the labor market early (Wrohlich 2008). In recognition of the substantial externalities of this situation, the federal government took legislative action in the form of multiple reforms of the laws governing the provision of public childcare for 0 – 2 year olds:

- In 2005, the federal daycare expansion law ("Tagesbetreuungsausbaugesetz") came into effect, in which the federal government committed itself to increase the supply of ECC slots by 230,000 by 2010, amounting to an increase to 17 slots per 100 children. To achieve this coverage, the law explicitly called for a demand-oriented expansion of slots.
- In 2007, federal and local government representatives convened to an ECC summit ("Krippengipfel") and committed to further increase the childcare coverage rate for ECC to 35% by 2013.
- In 2008, the federal government passed the law to promote children ("Kinderförderungsgesetz"), establishing a step-wise introduction of a legal claim to a

public ECC slot for children aged 1 to 2. From October 1st, 2010 on, working parents were eligible and from August 1st, 2013, all parents became eligible. To meet the demand arising from the legal claim, the law committed counties to increase supply sufficiently. The counties then also became liable for the legal claim, forcing them to either pay for substitute care arrangements or to compensate foregone earnings due to domestic childcare provision if they could not offer parents an ECC slot.

All these legislative efforts were accompanied by substantial public investment programs, which provided the funding for the targeted expansion. Importantly, the additional funding for public childcare did not crowd out spending on other family policy tools such as child benefits/allowances and maternity leave benefits (Bauernschuster, Hener, and Rainer 2016).

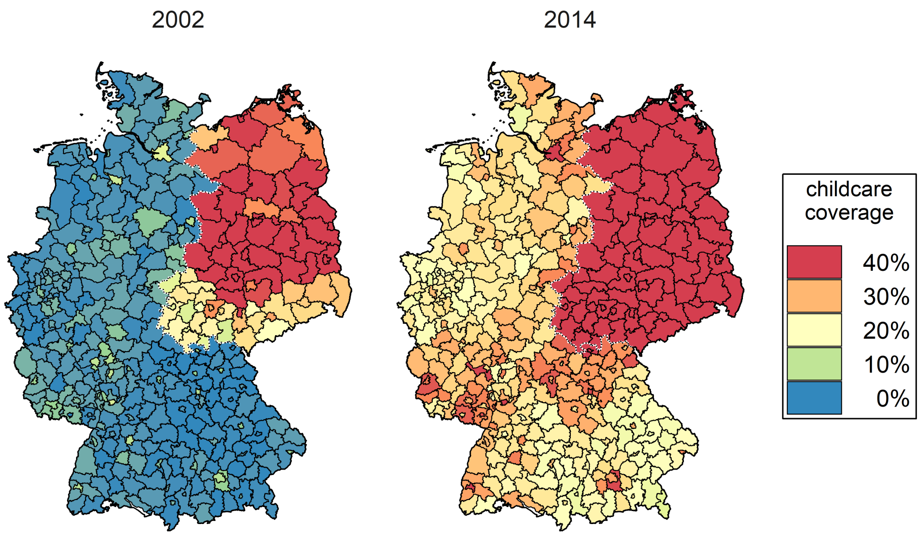


Figure 2.5: Childcare coverage rates across counties and time

Notes: Childcare coverage: number of children aged 0 – 2 enrolled in public childcare divided by total number of children aged 0 – 2 in a county, > 40% collapsed into 40% bin. Sample: German counties. Source: BBSR Bonn (2021).

Returning to Figure 2.5, it becomes clear from the 2002 levels that almost all West German counties had to drastically increase their childcare coverage rates to meet the targets. With the 2002 average childcare coverage rate in the West at just 2.2%, local authorities throughout the country responded to the reforms by substantially increasing the supply of ECC slots, reaching 26.2% by 2014. Figure 2.5, however, also illustrates the regional heterogeneity in the extent to which counties expanded their childcare coverage. Comparing the developments in the very north for example, some counties reached over 30% by 2014, while others, sometimes even neighboring counties, only increased coverage to around 20%, starting out from very similar 2002 levels. This is a first indication of regional heterogeneity in the response to the reforms, which generates the temporal and spatial variation that I will exploit in my identification strategy. Appendix-Figure B.8 illustrates the heterogeneity in the expansion in more detail by replicating the maps from Figure 2.5 for all years from 2006 to 2014.

In 2014, after the introduction of the legal claim, the reform-induced expansion was concluded, as illustrated in Appendix-Figure B.7. Therefore, I will focus on the time span 2006 to 2014 for the remainder of the paper. After all, the national ECC childcare coverage rate of 26.2% in 2014 fell short of the 35% goal, but the reforms were successful in increasing the rate by more than factor 10.

2.6 Empirical Strategy

I employ a generalized difference-in-differences framework similar to the one in Akerman, Gaarder, and Mogstad (2015) to estimate the causal effects of interest. Specifically, I regress the labor market outcome y_{ict} of individual i living in county c in year t on the main explanatory variable cc_{ct}, a set of controls as well as a number of fixed effects:

$$\mathrm{y}_{ict} = \alpha + \beta\ \mathrm{cc}_{ct} + \mathbf{X}_{ict}\boldsymbol{\gamma}_{\mathrm{x}} + \mathbf{Z}_{ct}\boldsymbol{\gamma}_{\mathrm{z}} + \delta_t + \delta_c + \delta_c \times t + \epsilon_{ict}, \tag{2.4}$$

where cc_{ct} is the childcare coverage rate, defined as childcare enrollment of $0-2$ year olds divided by the headcount of $0-2$ year olds in county c in year t. The two vectors of controls capture structural differences between individuals/counties and are defined as follows (see Section 2.6.2 for rationale): $\mathbf{X}_{ict}$ denotes a set of time-varying individual level controls, including a second-order polynomial in age, a college dummy, a non-German dummy, as well as a second-order polynomial in full-time work experience. $\mathbf{Z}_{ct}$ denotes a set of time-varying county level controls, including log population density, the population share of women aged 20 to 40, the employment rate of women, the fertility rate, log GDP per capita, the share of firms with $\geq$ 50 employees and a set of sector dummies. The fixed effects are year fixed effects δ_t, county fixed effects δ_c, and linear county time trends $\delta_c \times t$.

As the county fixed effects capture all time-constant differences between counties, they effectively control for all permanent unobservable heterogeneity in this regard that may be correlated with the childcare coverage rate. Furthermore, the year fixed effects absorb common time shocks and the linear county time trends capture the 2006 – 2014 county-specific trend in the outcome variable. Identification is therefore only based on within-county deviations from the respective linear time trends and conditional on the controls included. This is precisely the quasi-experimental variation induced by the reform that I aim to exploit and discuss in detail next.

2.6.1 Reform-induced variation in public childcare availability

As described in Section 2.5, the German authorities enacted a number of public childcare reforms between 2005 and 2008. These led to heterogeneous childcare expansion responses across counties, creating the spatial and temporal variation in childcare coverage rates that I will exploit for identification. This strategy has been used in a similar fashion to study the effects of this reform on fertility in Bauernschuster, Hener, and Rainer (2016) and on maternal labor supply in K.-U. Müller and Wrohlich (2020). To understand the source of the heterogeneity in the county level responses, I now map out the administrative procedure of creating new childcare slots during this time as well as the relevant constraints.

Federal and state level funding for the childcare expansion was allocated to the county level, mainly based on the current and projected headcount of children below age 3. On the county level, youth welfare offices ("Jugendämter") were responsible to allocate the funding to municipalities and other local childcare providers. This year-by-year allocation of funding was based on the present supply of ECC, the (projected) demand of ECC, and the quality of the expansion plans that providers submitted (Felfe and Lalive 2018).

The two main obstacles for expansions were shortages of suitable construction sites and shortages of qualified ECC staff. As ECC quality regulation requires at least 2.5 square meters of space per child, heterogeneity in the scarcity of building ground implied that some municipalities were quickly able to apply for funding, while others took longer to find suitable space. In terms of scarcity of staff, Germany as a whole was lacking around 45,000 ECC workers, with only 20,000 graduating each year. Together with the childcare sector paying very low wages, especially compared to the high cost of living in inner cities, this made it difficult for providers to quickly put together reliable proposals. Aside from these two sources of obstacles, additional variation was induced by the funding allocation procedure in cases of oversubscription. In such cases, i.e., when the youth welfare offices received more valid applications than funding was available, they resorted to lotteries and waiting lists to determine who would receive funding in the current year and who in the following year (Felfe and Lalive 2018).

The application process therefore generated spatial and temporal variation from i) heterogeneity in the availability/municipal approval of suitable construction sites, ii) heterogeneity in the availability of qualified staff, iii) idiosyncratic timing of the county level approval in cases of excess demand. Additional variation in the approval probability stems from idiosyncratic errors in local demand and population projections by the youth welfare offices, as documented by Hüsken (2010) and Hüsken (2011). Furthermore, heterogeneity in the experience with the complex funding and regulation system for public childcare both on the applicants' as well as the approvers' side, led to additional heterogeneity in the timing of the expansion.

Taken together, these sources of variation indicate that the growth in childcare coverage rates was not just driven by the key predictors of demand for ECC slots, but also by local supply shocks due to idiosyncrasies in the process of creating new slots. These ECC supply shocks are arguably orthogonal to other (unobserved) determinants of women's early career labor market outcomes, allowing me to identify the causal effects of cc_{ct} on the outcomes of interest.

To illustrate the just described variation in cc_{ct}, I plot the deviations of the year-on-year childcare coverage growth rates from the 2006 – 2014 county-specific average growth rate for three selected years in Figure 2.6. Positive values indicate that counties expanded their childcare coverage rate by an above average rate in a given year, with 100% implying an expansion twice as large as the 2006 – 2014 average growth rate. Negative values congruently indicate below average year-on-year growth rates. If a county had expanded its childcare coverage at a steady pace from 2006 to 2014, it would show up with values around zero for all years in Figure 2.6.

Comparing the West German counties across the three years depicted in Figure 2.6, the regional heterogeneity in the timing of ECC supply shocks becomes apparent. There are some regional clusters, for example in the very west, where most counties have expanded ECC rather late (negative values in 2006 – 2007, high positive values in 2013 – 2014). However, even within this cluster, there is substantial heterogeneity in the timing of the ECC expansion. This is, for example, visible for many neighboring counties who are shown with large positive values (colored in red) at different points in time. Appendix-Figure B.9 confirms this assertion further, by illustrating the heterogeneity in the expansion timing for all years from 2006 to 2014.

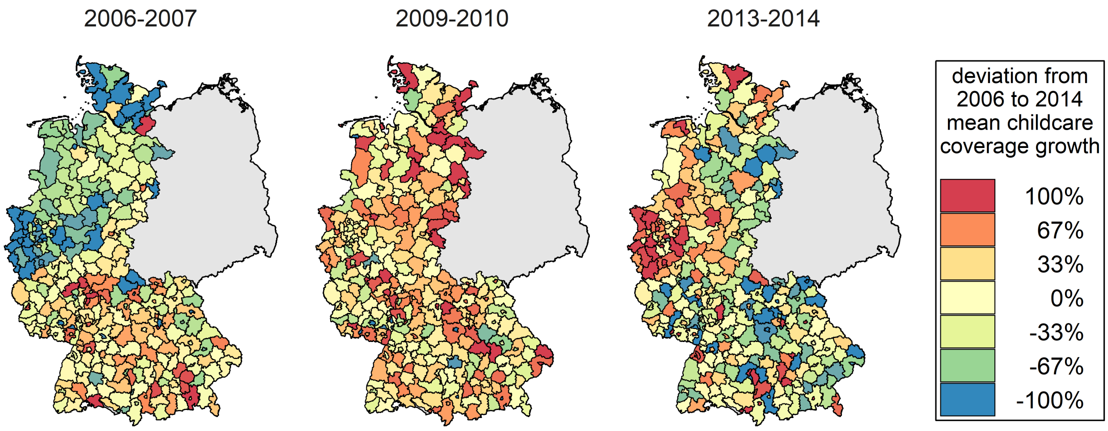


Figure 2.6: Year-on-year variation in childcare coverage growth rates

Notes: Childcare coverage: number of children aged 0 – 2 enrolled in public childcare divided by total number of children aged 0 – 2 in a county. Positive values indicate that a given year-on-year growth rate exceeds the 2006 – 2014 county level mean growth rate in childcare coverage, negative values vice versa. Values above/below +/-100% collapsed into the closest bin. Sample: West German counties. Source: BBSR Bonn (2021).

2.6.2 Assessment of the identification strategy

The just described exogenous reform variation has been used in the literature to identify the causal effects of public childcare on, e.g., fertility and maternal labor supply (Bauernschuster, Hener, and Rainer 2016, K.-U. Müller and Wrohlich 2020). I now turn to evaluating the key identification threat for my specific application: The timing of the childcare expansion might be related to underlying trends in women's labor market outcomes across counties, which would imply a violation of the common trends assumption.

To investigate this threat, I first check whether the timing of the expansion may be related to differences in pre-reform childcare coverage trends. Specifically, I compare the pre-reform developments of childcare coverage rates in counties that expanded mostly in the first half of the reform period (*early expanders*) with counties that expanded mostly in the second half of the reform period (*late expanders*). Appendix-Figure B.10 shows that the childcare coverage rates for both groups were low but very comparable in all years for which data is available prior to the reform. The development of women's promotion rates in early and late-expanding counties is also very similar during the pre-reform time span 2000 to 2005, as illustrated in Figure B.11.[35]

As the early vs. late expanders perspective may, however, be affected by compositional differences, I turn next to a more structured investigation of whether the expansion speed was related to key correlates of women's early career progression. Such relationships may arise for any of the following plausible reasons:

- Counties may have selected into expanding childcare early depending on the underlying time trends in labor market outcomes, i.e., to counteract negative trends in labor market outcomes via becoming more attractive for highly qualified women.
- Counties may have selected into expanding childcare early depending on the gains in women's labor market outcomes from public childcare, i.e., those expanded quickly where women were likely to be the most responsive to the availability of public childcare.
- The reforms were explicitly focused on a demand-oriented expansion of childcare. Wealthier counties may have faced a less severe shortage of slots initially and therefore the expansion speed may have been inversely related to the economic situation of counties.
- Parental preferences towards childcare and maternal employment may differ across counties and may have resulted in differential political pressure to expand childcare quickly.

Timing of the childcare expansion. To examine these relationships, I investigate whether the timing of the childcare expansion is related to baseline county characteristics that may be correlated with women's labor market outcomes. Specifically, I follow

[35] There is, however, no reform effect visible in Appendix-Figure B.10. This discrepancy to the results shown in Section 2.7, where I do find an effect of childcare coverage on promotions, is likely due to compositional differences between early and late expanding counties. The generalized difference-in-differences specification (2.4) does account for such differences in, e.g., the educational or occupational composition across counties.

Akerman, Gaarder, and Mogstad (2015) and run the following regression:

$$\Delta \text{cc}_{ct} = \delta_t + \delta_c + [\delta_t \times \mathbf{W}_c]' \boldsymbol{\psi}_t + \epsilon_{ct}, \tag{2.5}$$

where $\Delta \text{cc}_{ct} = \text{cc}_{ct} - \text{cc}_{c,t-1}$, while δ_t and δ_c are again year and county fixed effects. In $\mathbf{W}_c$ I include county characteristics from 2005 (pre-reform) that capture county demography, the economic situation of the county, the industry structure in the county, and the political preferences within the county. The resulting coefficients $\boldsymbol{\psi}_t$ show the relationship between each of the baseline covariates and the size of the expansion per year, with the county fixed effects absorbing the differences in the absolute size of the expansion.

More specifically on the covariates, I capture baseline demographic differences by including log population size, log population density (to proxy for urbanization), the population share of females aged 20 to 40, the fertility rate as well as the share of females graduating from the highest high-school track. Differences in local economic circumstances are taken into account via log GDP per capita, the unemployment rate, the employment rate of women, the employee share of women, and the 2003 to 2005 pre-expansion per-period growth rates in employment and earnings. Furthermore, I include the log of the number of firms, the share of firms with $\geq$ 50 employees, as well as the employment shares for manufacturing, construction, trade/hospitality, financial services, and public/health/education to check whether the baseline industry composition impacted the expansion timing. Lastly, I capture county differences in political preferences by including vote shares from the 2005 federal election for the conservative parties as well as the center-left and the green party.[36]

Appendix-Figure B.12 plots the estimation results of equation (2.5) for $\boldsymbol{\psi}_t$ and the accompanying 95% confidence intervals. The results show that the timing of the expansion is unrelated to both the baseline industry structure as well as the baseline political preferences. In terms of demographics and the economic situation, the expansion timing is also unrelated to all but two of the baseline county characteristics. The notable exceptions are population density as well as the employment rate of women: More densely populated counties have slowed down the expansion towards the end of the expansion period relative to less densely populated counties. A high employment rate of

[36]I use the secondary vote share in the federal elections on September 15th, 2005. Conservative parties are CDU and CSU, the main opposition in this election. The center-left and the green party are SPD and Bündnis 90/Die Grünen, the incumbent parties in this election.

women, on the other hand, appears to have induced counties to expand childcare earlier than others.

To account for these two potential confounders, I will include them as time-varying characteristics in my main specification. Along with the other county level covariates in equation (2.4), they account for systematic differences in the demand and supply of ECC, as such may drive the speed of the expansion. I additionally include in Appendix B6 results for all regressions without the set of time-varying county controls to document that the effects are not driven by the inclusion of controls.

Finally, I also investigate whether the reforms have affected migration patterns between counties differentially in early and late expanding counties. This tackles the concern that parents may migrate to counties with a high supply of ECC. Appendix-Figure B.13 compares the migration developments between counties across the pre-reform and reform period. Both groups of counties display very similar developments, suggesting that the reform has not led to substantial migration from late to early expanding counties.

2.7 Results

Employing the estimation strategy laid out in the previous section, I now turn to my results on the causal effect of public childcare provision on young women's wage growth and promotion probabilities. My hypothesis based on the model from Section 2.4 is that public childcare has a positive impact on women's promotion rates, as it decreases the *effective* ability promotion threshold for women. I explore heterogeneity of the effects with respect to firm size, as an important proxy for the organizational structure of employers, and occupational characteristics, as these mattered substantially in my earlier discussion of the descriptives. Furthermore, I look at a number of other outcomes capturing how women may have adjusted their behavior in response to the policy to investigate potential channels. Finally, I also investigate whether public childcare provision affects the promotion rates of young men, which should be unaffected based on the model, and conduct a number of robustness checks.

2.7.1 Main results

Table 2.3 presents the results on the effect of public childcare provision on three different wage growth measures. I include year and county fixed effects as well as linear county time trends in all columns, as set up in Section 2.6, along with a set of covariates that always includes the time-varying individual and county characteristics discussed before. Additionally, I control for 2-digit occupations or industries in columns denoted by *ico* and *ics*, respectively. First, in column (1), the dependent variable is raw individual wage growth. The effect of public childcare is positive, but not significantly different from 0. The same holds for column (2), where I use relative wage growth, i.e., wage growth in excess of the average wage growth at the firm, on the left-hand side. Given that a large share of wage growth occurs in rare, but large wage shocks, I focus in columns (3) to (5) on my main outcome variable of interest, the binary promotion indicator. It is equal to 1, if an individual's relative wage growth is greater or equal to 10%.

Table 2.3: The effect of public childcare on wage growth and promotions

	wage growth	rel. wage growth	promotion (relative wage growth ≥ 10%)		
	(1)	(2)	(3)	(4)	(5)
childcare coverage	0.0380	0.0370	0.1668^{*}	0.1691^{**}	0.1671^{*}
	(0.0397)	(0.0431)	(0.0868)	(0.0862)	(0.0863)
covariate set	*ico*	*ico*	*ic*	*ico*	*ics*
year & county FE	✓	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓	✓
mean dependent var.	0.0332	0.0269	0.1781	0.1781	0.1781
N	176,538	176,538	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Covariate sets: *i* (age, age sq., nongerman, college, FT experience, FT experience sq.), *c* (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), *o* (2-digit occupations), *s* (2-digit industries). See Appendix-Table B.6 for results without covariates. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

For the causal effect of public childcare on promotions, I find in columns (3) to (5) that a 10 percentage point increase in the childcare coverage rate increases the promotion probability of young women ceteris paribus by around 1.7 percentage points. Given the mean promotion rate of 17.81%, this translates into a 10% increase in the promotion probability of young women. The coefficients remain very similar when including occupation or industry fixed effects. All of them are statistically significant at least at the 10% level, with the effect that includes occupation fixed effects being significant at the 5% level as well.

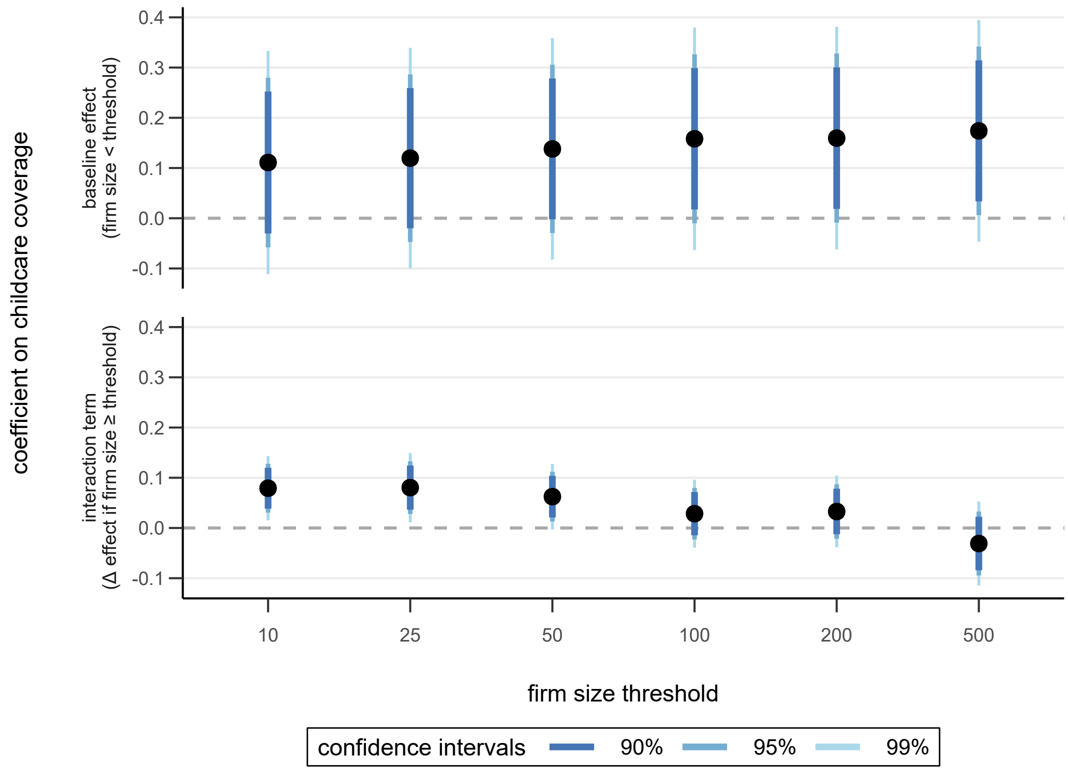


Figure 2.7: Heterogeneity of the effect of public childcare on promotions by firm size

Notes: Different thresholds for binary firm size indicators across columns, e.g., = 1 if number of full-time employees ≥ 10 (column 1), = 0 otherwise. Specifications including individual, county, and occupation controls, see Appendix-Table B.3 for full details. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Confidence intervals based on robust standard errors clustered at the county level.

Effect heterogeneity. I investigate next whether the effects are heterogeneous by the size of the firm, as small and large firms may differ substantially in their internal promotion decision making processes. Specifically, I include interaction terms of binary firm size indicators with the childcare coverage rate and present the results in Figure 2.7. The binary indicators differentiate at different thresholds between smaller and larger firms by the number of full-time employees. The baseline effects in the top panel therefore capture the effect of public childcare on promotions for firms smaller than the threshold, i.e., with < 10 employees in the first column. The interaction terms in the bottom panel accordingly capture the difference in the effect size for firms larger than the threshold vs. firms smaller than the threshold, i.e., ≥ 10 employees vs. < 10 in the first column.

Looking at the pattern in the top panel, I find that for firms with < 50 employees public childcare does not significantly improve young women's promotion rates. It does so, however, for larger firms, as indicated by the positive and significant interaction terms in the bottom panel (first three columns). Starting in the fourth column with the firm size threshold equal to 100, the pattern switches to a small and insignificant interaction term, while the baseline coefficients in the top panel turn significant. In other words, the effect of public childcare on promotions does not differ anymore between firms with < 100 and ≥ 100 employees (neither for the larger firm size thresholds 200 and 500). Instead, the baseline coefficients in the top panel are close to the estimates in Table 2.3.

This pattern illustrates that the positive effects of public childcare availability on promotion rates of young women are driven by firms with ≥ 50 employees. A potential explanation is that employers with larger workforces are more attentive to policy changes, as they have more instances of maternity leave each year and a larger organizational structure to process relevant information. Alternatively, larger firms may also have a greater number of positions available for mothers to return to. This might induce mothers to return quickly, if made possible through access to public childcare, from which their employers benefit. As promotions are approximately equally frequent in small and large firms, this pattern is not driven by small firms just offering fewer promotions in general.[37]

Additionally, I investigate whether the effect of the availability of public childcare on promotions is heterogeneous with respect to occupational characteristics. In columns (1) and (2) of Table 2.4, I interact the occupational characteristics used before, namely demanding task content (high load on analytic non-routine tasks) and demanding know-how required (specialist or expert knowledge), with the childcare coverage rate. The coefficient on the childcare coverage rate therefore becomes the effect of public childcare on women working in non-demanding tasks in column (1), which turns out to be positive and slightly larger than the effect from Table 2.3. A 10 percentage point increase in the childcare coverage rate implies on average a 1.856 percentage point increase in the promotion probability of women working in non-demanding tasks (significant at the 5% level).

[37]If any, promotions are slightly more frequent in small firms. The average promotion rate in firms with < 50 employees is 18.62% and 17.30% in firms with ≥ 50 employees.

Table 2.4: Heterogeneity of the effect of public childcare on promotions by occupational characteristics and education

	promotion (relative wage growth $\geq$ 10%) – mean: 0.1781 –		
	(1)	(2)	(3)
childcare coverage	0.1856**	0.1735**	0.1639*
	(0.0859)	(0.0875)	(0.0870)
demanding tasks	0.0288***		
	(0.0056)		
childcare coverage × demanding tasks	−0.0547**		
	(0.0237)		
demanding know-how		0.0396***	
		(0.0063)	
childcare coverage × demanding know-how		−0.0177	
		(0.0306)	
college			0.0859***
			(0.0070)
childcare coverage × college			0.0284
			(0.0349)
covariate set	*ico*	*ico*	*ico*
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
N	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. 'demanding tasks' defined as occupations with a share of analytic non-routine tasks $\geq \frac{1}{3}$), 'demanding know-how' as occupations with specialist or expert knowledge required. Covariate sets: i (age, age sq., nongerman, college, FT experience, FT experience sq.), c (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), o (occupation segments, 14 catg.), s (industry segments, 13 catg.). See Appendix-Table B.8 for results without covariates. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Women working in occupations with demanding tasks have a slightly higher baseline promotion rate (positive and significant, but small coefficient), but a significantly lower effect of public childcare on promotions than those with non-demanding tasks. This indicates that women working in non-demanding occupations benefit more from public childcare. Career-oriented women working in demanding occupations may have been using other childcare arrangements before and therefore the reform affected their return behavior less. For women in less demanding occupations, on the other hand, the availability of affordable public childcare may be more important in determining their labor market participation prospects as mothers.[38]

Nevertheless, I still find that women working on demanding tasks benefit from public childcare, but the effect of a 10 percentage point increase in the childcare coverage rate is approximately 0.5 percentage points lower than for those with non-demanding tasks. Turning to my second occupational characteristic, the know-how requirement of the job in column (2), mostly confirms the results from column (1). The baseline estimate of 0.1735 is close to the one from Table 2.3, significant at the 5% level, and the effect of the interaction term is also negative, but small and indistinguishable from zero.

I furthermore include an interaction with my college indicator in column (3) of Table 2.4. However, I do not find evidence for heterogeneity between women with or without a college degree, and the baseline effect in the first row remains very similar to the ones before. This emphasizes that the differences in column (1) are not driven by education, but rather by job characteristics.

2.7.2 Investigation of potential channels

To evaluate the plausibility of the channels laid out in Section 2.4, namely a decrease in statistical discrimination or an increase in employee effort, I next investigate a number of other potentially affected outcomes. Instead of promotions, I examine the effects of public childcare availability on the timing of first birth and on the labor supply of young women, the two key components of my sample selection (continuous full-time work). Furthermore, I investigate whether public childcare encourages women to take up occupations with a more demanding task composition, higher know-how requirements, or induced them to switch occupations or industries. This allows me to investigate

[38]This explanation is in line with the results in K.-U. Müller and Wrohlich (2020), who find that the public childcare expansion has increased participation the most among medium skilled mothers.

whether such occupational adjustments or switches, i.e., responses by the employees, are an important driver of my results.

Table 2.5: The effect of public childcare on fertility and labor supply

	first birth next period	first interruption next period	consecutive full-time
	(1)	(2)	(3)
childcare coverage	−0.0453	−0.0367	0.1476**
	(0.0357)	(0.0530)	(0.0702)
covariate set	*ico*	*ico*	*ico*
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
mean dependent var.	0.0377	0.0764	0.6963
N	294,569	269,893	269,893

Notes: Binary indicators for first birth or first long employment interruption (≥ 1 year) next period, conditional on observing either at some point. Covariate sets: i (age, age sq., nongerman, college, FT experience, FT experience sq.), c (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), o (occupation segments, 14 catg.), s (industry segments, 13 catg.). See Appendix-Table B.9 for results without covariates. Sample: working women between ages 22 and 32, not out of the labor force in previous period, prior to first birth (column (1)) or prior to first long employment interruption (≥ 1 year, column (2)). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Fertility and labor supply. I start out by looking at the effects on the timing of first birth in Table 2.5, columns (1) and (2). The dependent variable is a binary indicator for first birth in the next period in column (1) and I include all individuals prior to first birth in the sample (not conditioning on consecutive full-time at the same employer). In column (2), I use a binary indicator for the first long employment interruption instead (not working for a full year) as the more conservative alternative, because I can only reliably identify first births for women working prior to giving birth. For both outcomes, 'first birth next period' and 'first long employment interruption next period', I find negative, but insignificant effects of public childcare availability. Consistent with these results, I find no significant effects on age at first birth/age at first employment interruption in Table B.4. I conclude, in line with Bauernschuster, Hener, and Rainer (2016), that public childcare availability does not affect the timing of first birth and therefore does not affect my sample selection endogenously in this regard.

Turning to labor supply, I use again the larger sample of all women in the labor market prior to their first long employment interruption to test whether public childcare affects the propensity to observe an individual working full-time consecutively. In column (3) of Table 2.5, I find a positive and significant effect of public childcare availability on the consecutive full-time indicator, but it is quantitatively small. A 10 percentage point increase in the childcare coverage rate implies a 1.476 percentage point increase in the share working consecutively full-time, which is a 2% increase given the 69.63% mean of the dependent variable. There is therefore a small effect of endogenous sample selection, which may point to some women exerting more effort prior to first birth, in line with the alternative channel sketched out in Section 2.4.3.

Table 2.6: The effect of public childcare on occupational choices

	demanding tasks	demanding know-how	switched occupation	switched industry
	(1)	(2)	(3)	(4)
childcare coverage	−0.0319	−0.0174	−0.0285	−0.0197
	(0.0669)	(0.1250)	(0.0812)	(0.0226)
covariate set	*ico*	*ico*	*ico*	*ics*
year & county FE	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓
mean dependent var.	0.3012	0.2034	0.0662	0.0099
N	176,538	176,538	176,538	176,538

Notes: 'demanding tasks' defined as occupations with a share of analytic non-routine tasks $\geq \frac{1}{3}$), 'demanding know-how' as occupations with specialist or expert knowledge required, 'switched occupation/industry' as a binary indicator for different 2-digit occupational/industry code in prev. period. Covariate sets: i (age, age sq., nongerman, college, FT experience, FT experience sq.), c (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), o (occupation segments, 14 catg.), s (industry segments, 13 catg.). See Appendix-Table B.10 for results without covariates. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Occupational characteristics and switching occupations. To understand whether public childcare availability affects women's selection into specific occupations, I use my indicators for occupational characteristics as dependent variables in Table 2.6. I find that public childcare affects neither the propensity to work in occupations with demanding tasks in column (1) nor with demanding know-how requirements in column (2), with both effects close to and indistinguishable from zero. In line with these results, I furthermore find no effects on occupation or industry switches in columns (3) and (4) of Table 2.6. This indicates that public childcare availability does not affect the occupational composition to the degree to which I measure it.

Table 2.7: The effect of public childcare on male wage growth and promotions

	wage growth	rel. wage growth	promotion (relative wage growth $\geq$ 10%)		
	(1)	(2)	(3)	(4)	(5)
childcare coverage	−0.0052	0.0204	0.0899	0.0862	0.0912
	(0.0341)	(0.0344)	(0.0716)	(0.0712)	(0.0717)
covariate set	*ico*	*ico*	*ic*	*ico*	*ics*
year & county FE	✓	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓	✓
mean dependent var.	0.0372	0.0311	0.1817	0.1817	0.1817
N	260,499	260,499	260,499	260,499	260,499

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Covariate sets: *i* (age, age sq., nongerman, college, FT experience, FT experience sq.), *c* (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), *o* (occupation segments, 14 catg.), *s* (industry segments, 13 catg.). See Appendix-Table B.11 for results without covariates. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

2.7.3 Results for men

After focusing on the effects of public childcare on women's labor market outcomes, I now turn to results on men. As men do not drop out after having children, the model in Section 2.4 does not predict that their promotion rates would be affected by the reform.

Table 2.7 replicates Table 2.3 for the male part of my sample. In columns (1) and (2), I start again with raw and coworker-benchmarked wage growth as the dependent variables. I find in both columns small and insignificant effects of public childcare on the wage growth rates of men. Turning to columns (3) to (5) with promotions as my key outcome of interest, I find no evidence for an effect of public childcare on the promotion rates of men, whether I include occupation/industry controls or not. This result can be seen as evidence for the hypothesis that the effect for women is not driven by promotions into a fixed set of leadership positions, as then I should see a negative effect on men if women instead of men are promoted. Rather, the effect appears to be driven by women being promoted into jobs in which they are most productive for the firm and these are higher paid jobs. Alternatively, as my sample focuses on young workers only, the additional promotions for young females could also be crowding out older workers from moving into leadership positions.

2.7.4 Robustness to alternative sample and promotion definitions

To judge the robustness of my findings, I run a number of alternative specifications that address different potential weaknesses in my approach. First, I investigate whether the effects are driven by young mothers (who give birth early and return within my sample age range 22 to 32), as K.-U. Müller and Wrohlich (2020) did show that the reforms increased the labor market participation of mothers. To address this concern, I re-estimate my effects using a restricted sample of non-mothers. Second, I use a number of alternative promotion definitions to test whether my results are sensitive to the specific construction of my measure. Third, I expand and contract the sample age range to check whether my results are specific to the selected age range.

Subsample of women prior to first employment interruption. While I already restrict my sample with the target to capture pre-first birth labor market behavior, I may still include some mothers who gave birth in their early 20s and have returned already before age 32. Therefore, I re-estimate my main results in Table 2.8 only with women prior to their first long employment interruption, a conservative measure for prior to first birth as I can only determine first birth reliably if a woman is working before. This approach avoids that the results are driven by the increased labor market participation of mothers returning due to public childcare, but is potentially problematic if the timing of first birth is itself an endogenous outcome. However, as shown previously in Table 2.5, I find no evidence for an effect of public childcare on the timing first birth.

Table 2.8: The effect of public childcare on wage growth and promotions (subsample prior to first employment interruption)

	wage growth	rel. wage growth	promotion (relative wage growth ≥ 10%)		
	(1)	(2)	(3)	(4)	(5)
childcare coverage	0.0469	0.0551	0.1733*	0.1743*	0.1729*
	(0.0387)	(0.0415)	(0.0973)	(0.0965)	(0.0967)
covariate set	*ico*	*ico*	*ic*	*ico*	*ics*
year & county FE	✓	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓	✓
mean dependent var.	0.0397	0.0334	0.1767	0.1767	0.1767
N	155,100	155,100	155,100	155,100	155,100

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Covariate sets: *i* (age, age sq., nongerman, college, FT experience, FT experience sq.), *c* (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), *o* (occupation segments, 14 catg.), *s* (industry segments, 13 catg.). Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm, prior to first long employment interruption (≥ 1 year). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

In columns (1) and (2) of Table 2.8, I again find no effects on raw wage growth or wage growth benchmarked at the firm level. Columns (3) to (5) furthermore illustrate that my previous findings for the effect of public childcare provision on the promotion probability of young women do hold up: I again find a significant positive effect ranging from 0.1733 to 0.1743, an approximately 10% increase in the promotion rate for a 10 percentage point increase in the childcare coverage rate. These results therefore confirm that the main findings are not driven by mothers, but that they do originate in pre-birth labor market interactions between employers and female employees.

Alternative promotion definitions. To test whether my results are sensitive to the specific definition of promotions, I use a number of alternatives in Table 2.9. In column (1), I define a promotion as a raw wage growth rate of 10% or more, i.e., not benchmarked vs. the average firm wage growth. I find that a 10 percentage point increase in the childcare coverage rate significantly increases the alternatively defined promotion rate by 1.452 percentage points. The effect is slightly lower than the 1.691 percentage points in Table 2.3, but very much in line with my previous results. This is therefore reassuring evidence that my results are not driven by the benchmarking of individual wage growth with the average firm wage growth.

Table 2.9: The effect of public childcare on promotions
(alternative promotion definitions)

	wage growth $\geq$ 10%	relative wage growth $\geq$ 5%	relative wage growth $\geq$ 15%
	(1)	(2)	(3)
childcare coverage	0.1452*	0.2135*	0.0622
	(0.0822)	(0.1129)	(0.0642)
covariate set	*ico*	*ico*	*ico*
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
mean dependent var.	0.1652	0.3380	0.1009
N	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Covariate sets: *i* (age, age sq., nongerman, college, FT experience, FT experience sq.), *c* (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), *o* (occupation segments, 14 catg.), *s* (industry segments, 13 catg.). Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

In columns (2) and (3), I adjust the promotion threshold in relative wage growth to 5% and 15% instead of 10%. For 5% in column (2), this almost doubles the mean promotion rate in the sample to 33.8%, making promotions substantially less rare. In line with my previous results, I find a positive and significant effect of public childcare provision on the promotion probability of young women. The size of the coefficient at 0.2135 is larger than before, but smaller relative to the mean promotion rate at 6.3% for a 10 percentage point increase in the childcare coverage rate. For 15% as the promotion threshold in column (3), the mean promotion rate in the sample drops to just 10.09%. The coefficient on the childcare coverage rate continues to be positive, but is substantially smaller and I cannot distinguish it from zero any more. This disparity could be driven by the fact that promotions are rarer under this stricter definition and are more likely to be tied to higher (managerial) positions. As argued in Section 2.7.3, where I discussed that I do not find negative effects for men, this may indicate again that the results are driven by medium-level promotions instead, which are not captured by the more restrictive definition anymore.

Alternative sample age ranges. Lastly, I use two alternative age ranges, namely expanding the age range to 20 – 34 and contracting it to 24 – 30. In column (1) in Appendix-Table B.5, I show that my results are robust to increasing the age range to 20 – 34. The coefficient of interest drops to 0.1353, but remains positive and significantly different from zero. If I restrict the sample age range to a tighter interval (24 – 30) in column (2), the effects are positive again, but become smaller (0.1132) and the standard error increases. The implied sample size reduction from the compressed age range therefore appears to decrease the precision of my estimates, rendering the coefficient insignificant.

2.7.5 Alternative generalized difference-in-differences specifications

A recently emerged strand of the econometrics literature[39] has pointed out that generalized difference-in-differences models as the one I employ (equation (2.4)) are subject to the following concern: It has been shown that the coefficient of interest (β in my case) in such models can be expressed as the weighted sum of the difference-in-differences coefficients of all 2×2 group-time pairs. The weights of these separate difference-in-

[39] For example, Goodman-Bacon (2018), Chaisemartin and D'Haultfœuille (2020), Callaway and Sant'Anna (2020), and Imai and Kim (2021).

differences coefficients sum up to 1, but some of them can be negative if treatment effects are heterogeneous over time or across groups. This stems from the fact that in these 2×2 group-time pairs, the previously treated groups are used as 'control groups', because their treatment status does not change. The treatment effects in the 'control group' are therefore differenced out by the difference-in-differences design, which can result in negative weights (Chaisemartin and D'Haultfœuille 2020). These negative weights can yield heavily biased estimates, which even may have the wrong sign.

While this young strand of the literature has proposed a number of new estimators that are not subject to the just described concern, these estimators are so far limited to specific cases of generalized difference-in-differences models. They cover mostly only simple cases of binary treatments and even more flexible variants currently only work with discrete treatments. Given that I have a continuous treatment (childcare coverage rate), these newly proposed estimators unfortunately do not cover my use case yet.[40] Therefore, I cannot provide a fully comparable robustness check of my estimates.

Nevertheless, to take into account that the effect of public childcare on promotions may be heterogeneous, I show in Table 2.10 results using two of the newly proposed estimators with different discretized versions of my treatment. The specifications in columns (1) to (3) use the estimator proposed in Chaisemartin and D'Haultfœuille (2020) (CDH2020), which is robust to heterogeneous treatment effects, while in columns (4) to (6) I use the estimator from Chaisemartin and D'Haultfœuille (2021) (CDH2021), which is robust to heterogeneous and dynamic treatment effects. For both estimators, it is necessary to introduce thresholds for continuous treatment variables to separate counties where treatment does not change (change in childcare coverage rate < threshold) from counties where treatment does change (change in childcare coverage rate ≥ threshold). The idea for the threshold values is that they should be large enough such that there are counties for whom the childcare coverage rate changes by less than the threshold, but small enough such that a change below the threshold is unlikely to affect promotion probabilities.

Turning to the results in Table 2.10, I find positive effects throughout all six columns, i.e., regardless of the estimator and the discretization threshold. This result is reassuring because it shows that potentially negative weights in my primary approach (specification (2.4)) do not appear to flip the sign of my coefficients of interest. The size of the effects, however, varies substantially across estimators and discretization thresholds.

[40] A very recent working paper by Callaway, Goodman-Bacon, and Sant'Anna (2021) makes an effort to close this gap.

Table 2.10: The effect of public childcare on promotions
(alternative generalized difference-in-differences specifications)

	promotion (relative wage growth ≥ 10%) – mean: 0.1781 –					
	CDH2020			CDH2021		
	(1)	(2)	(3)	(4)	(5)	(6)
childcare coverage	0.3317	0.1796	0.0580	0.0109	0.0091	0.0023
	(0.2876)	(0.2838)	(0.2535)	(0.0162)	(0.0115)	(0.0095)
dynamic effect in $t + 1$				0.0460	0.0368	0.0042
				(0.0388)	(0.0240)	(0.0198)
discretization threshold	0.02	0.025	0.03	0.02	0.025	0.03
covariate set	-	-	-	-	-	-
year & county FE	✓	✓	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓	✓	✓
N	80,264	76,929	70,679	18,070	21,109	24,554
$N_{switchers}$	43,644	33,183	25,253	8,527	8,563	8,342

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. CDH2020 columns use the estimator proposed in Chaisemartin and D'Haultfœuille (2020) (robust to heterogeneous treatment effects), while CDH2021 columns use the estimator proposed in Chaisemartin and D'Haultfœuille (2021) (robust to heterogeneous and dynamic treatment effects). Both estimator require a threshold to identify units with no treatment status change (below threshold change in childcare coverage), referenced as discretization threshold in the table. $N_{switchers}$ denotes the number of observations with a change in treatment status. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Bootstrapped standard errors clustered at the county level in parentheses (1,000 reps), significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Using the approach from CDH2020, the effect in column (2) is of similar size as the ones that I presented earlier, but a smaller discretization threshold in column (1) almost doubles the effect and a larger threshold in column (3) reduces the effect by two-thirds. Using CDH2021, the effects are overall substantially smaller and I find some evidence for dynamic treatment effects, with large effects in the period after the initial treatment. However, in terms of significance, none of the coefficients in Table 2.10 is significantly different from zero.

Overall, these results are indicative that the potential issue of negative weights does not flip the sign of my effects, but it is important to keep in mind that these alternative estimates are not only based on other estimators, but also on different samples. The samples that can be used with these estimators are highly dependent on the specific discretization threshold and procedure, implying different samples throughout Table 2.10. Furthermore, due to the lower number of observations, I am not able to include the set of covariates from the baseline results in these regressions.[41] After all, these results are therefore not fully comparable to my main results, but they do shed light on the main concern of negative weights for my generalized difference-in-differences specification (2.4).

2.8 Conclusion

In this paper, I use rich social security data from Germany to map out the early career dynamics of women vs. men and the role that childcare policy can play to improve the career trajectories of women. Following Bronson and Thoursie (2020), I leverage the combination of high-quality individual and firm level wage data to construct a wage-based measure of promotions that is subject to little measurement error. This allows me to reliably compare the promotion patterns of women and men across a multitude of heterogeneous firms.

I show that gender wage gaps are not just driven by differences in labor supply around childbirth, but originate at the very beginning of the working life. Prior to giving birth, young women are promoted at significantly lower rates than their male peers, contributing early on to the divergent wage trajectories of these two groups. This observation is consistent with employers engaging in statistical discrimination, i.e., promoting young women at lower rates than men, because the employers expect to incur higher maternity leave related costs from women in higher positions.

To test this hypothesis, I exploit a staggered expansion of public childcare for below 3 year olds for identification. This expansion has potentially reduced the maternity leave related costs for employers and thereby the extent of statistical discrimination, as it provided mothers with a reliable and affordable mode of care.

[41] The number of observations is especially decreasing at the county level, because only counties whose childcare coverage rate did only change below the discretization threshold can be used as control units. This results in bootstrap samples for whom the number of covariates exceeds the sample size, given that year fixed effects, county fixed effects, and linear county time trends are always included in line with my identification strategy.

Using a generalized difference-in-differences framework, I find a positive and significant causal effect of public childcare availability on the promotion probability of young women. A 10 percentage point increase in public childcare availability increases the promotion probability by 1.7 percentage points. This translates into a 10% increase in young women's promotion rates and corresponds to, taking the number at face value, more than 50% of the gender gap in promotions. I interpret these results as evidence consistent with i) statistical discrimination playing a role for the early career dynamics of women and ii) that childcare policy can reduce the scope of statistical discrimination.

Furthermore, I find that the effects are driven by larger firms. Plausible reasons include that larger firms are better informed about childcare policies or have more jobs available for mothers to work in effectively, e.g., part-time with suitable hours. Looking at men, I find no evidence for a crowding out of promotions for young male workers. This suggests that the additional promotions for young females are either into specialist positions or crowd out older workers from leadership positions. Finally, I show that public childcare does not lead to delays in fertility, i.e., my results are not driven by postponement of first birth, and I also demonstrate that the results hold up for a number of different specifications.

In light of Chaisemartin and D'Haultfœuille (2020)'s recent criticism of generalized difference-in-differences designs like the one I employ, I provide evidence that the sign of my effects holds up using more robust newly developed estimators. The effects are, however, not significant anymore, but as these new estimators currently require discretized treatments, they are not fully comparable to my specification with a continuous treatment.

In terms of evidence on statistical discrimination, further research is needed to determine its full scope. While my results point towards it playing an important role, a structural approach such as in Neumark and Vaccaro (2020) or Xiao (2020) would be an avenue to quantify its extent and compare it to the effects of dropping out of the labor market around childbirth.

All in all, I show that expected future labor supply patterns play an important role for early career promotion decisions. Through this channel, I show that family policies targeted at mothers can also have a substantial effect on women prior to even becoming mothers. This is an important insight to take into account when evaluating the effectiveness or the net fiscal cost of such policies, as such early career gains can lift women into higher overall wage trajectories.

Appendix B1 Additional Summary Statistics

Table B.1: Summary statistics (extended sample)

	women	men
age	27.52	27.67
non-german	9.99%	13.44%
college degree	23.93%	22.12%
experience in full-time (years)	3.76	4.15
daily earnings (log EUR)	4.00	4.30
share switched employer	21.29%	20.53%
age at labor market entry	23.46	23.96
age at first birth	28.85	-
occupations		
goods production	7.03%	45.26%
personal services	38.71%	15.04%
business services	44.66%	20.67%
IT and science	1.72%	4.10%
other servcies	7.88%	14.93%
occupational characteristics		
occupation with demanding tasks (high load analytic non-routine)	27.56%	21.45%
occupation with demanding know-how (specialist/expert knowlegde req.)	19.89%	21.02%
individuals	84,898	104,616
individuals × year	348,405	412,888

Notes: Sample: working individuals between ages 22 and 32, working part-time or full-time, not out of the labor force in previous period. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

Appendix B2 Additional Descriptives

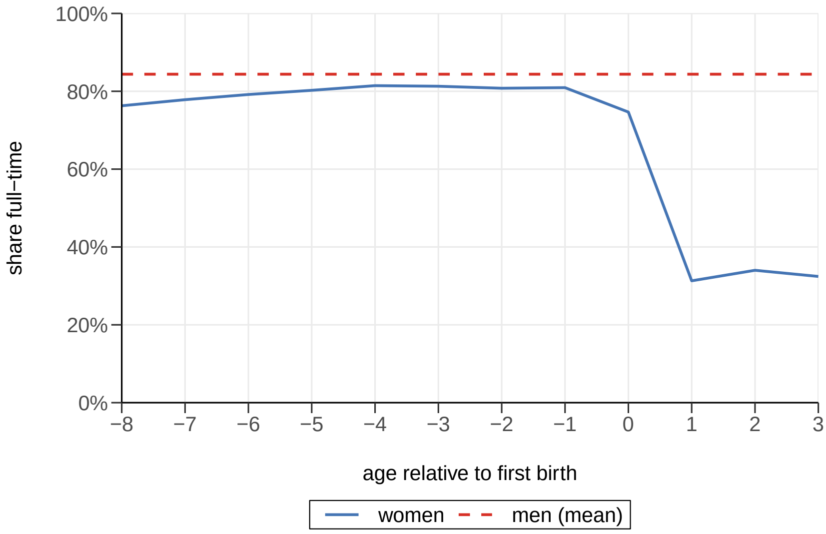


Figure B.1: Women's full-time share prior to first birth

Notes: Sample: individuals between ages 22 and 32, not out of the labor force in previous period and women conditional on observing first birth. men: sample mean full-time share. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

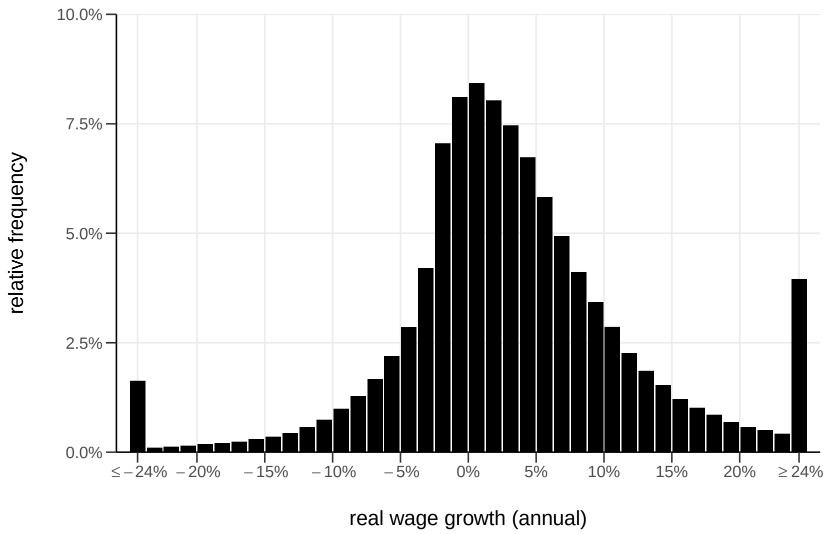
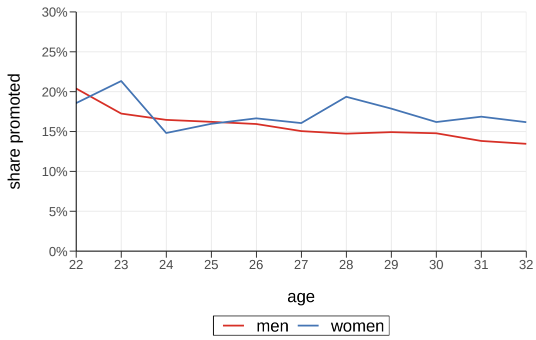
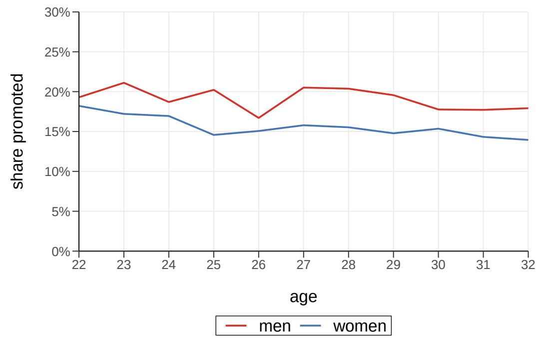


Figure B.2: Wage growth distribution

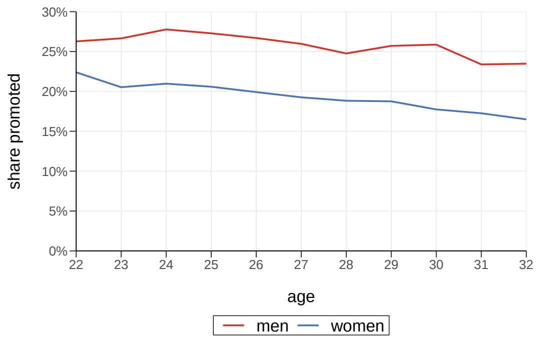
Notes: All values in real terms. Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).



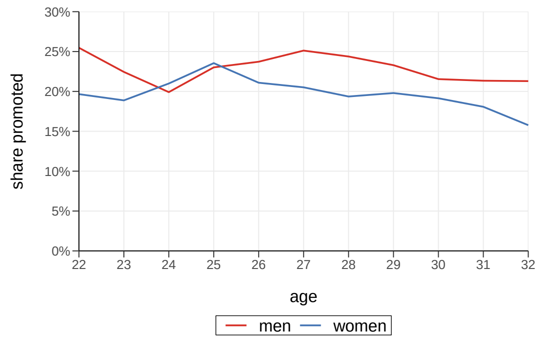
(a) goods production



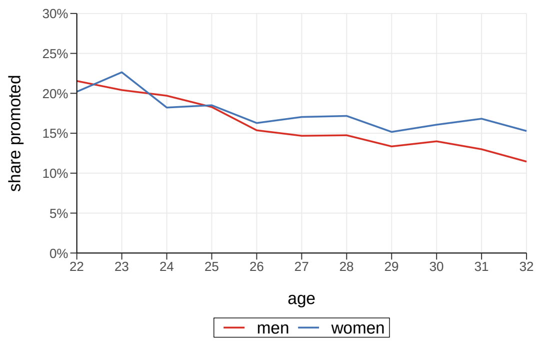
(b) personal services



(c) business services



(d) IT and science



(e) other services

Figure B.3: Promotion rates by age, gender, and occupational sector

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

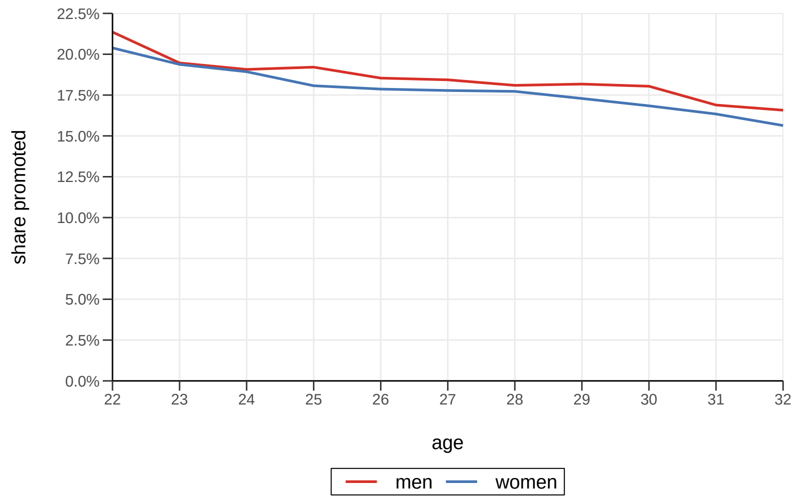
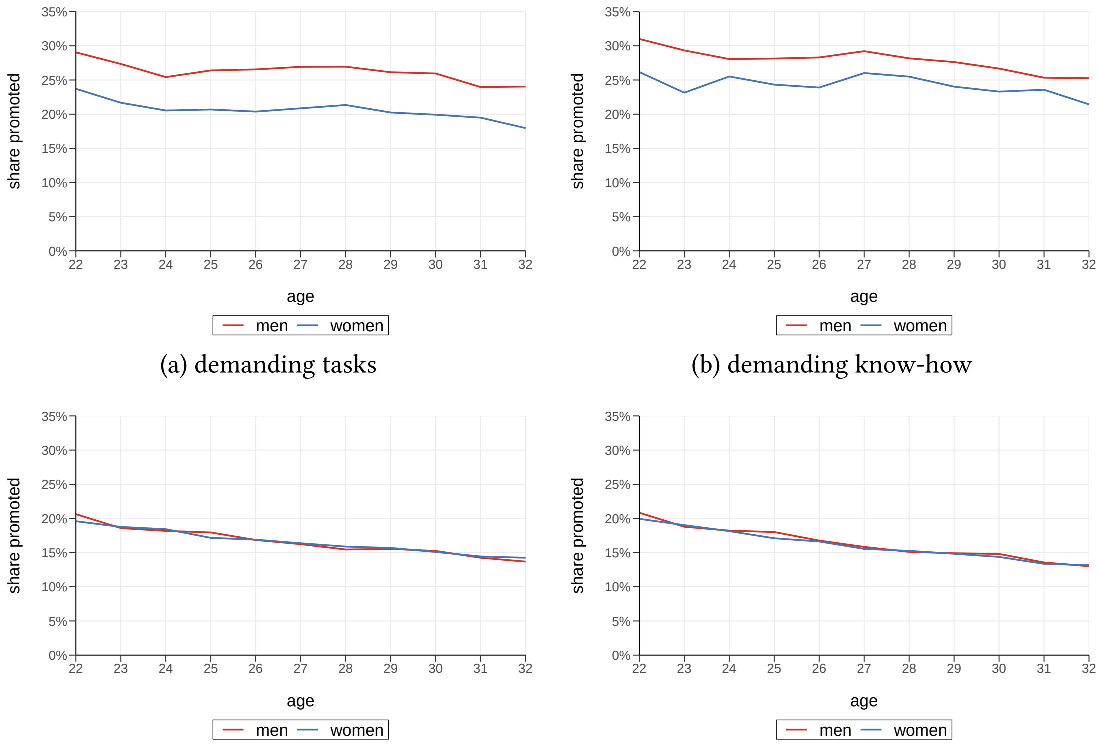


Figure B.4: Promotion rates by age and gender
(prior to first employment interruption)

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm, women prior to first long employment interruption (≥ 1 year). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

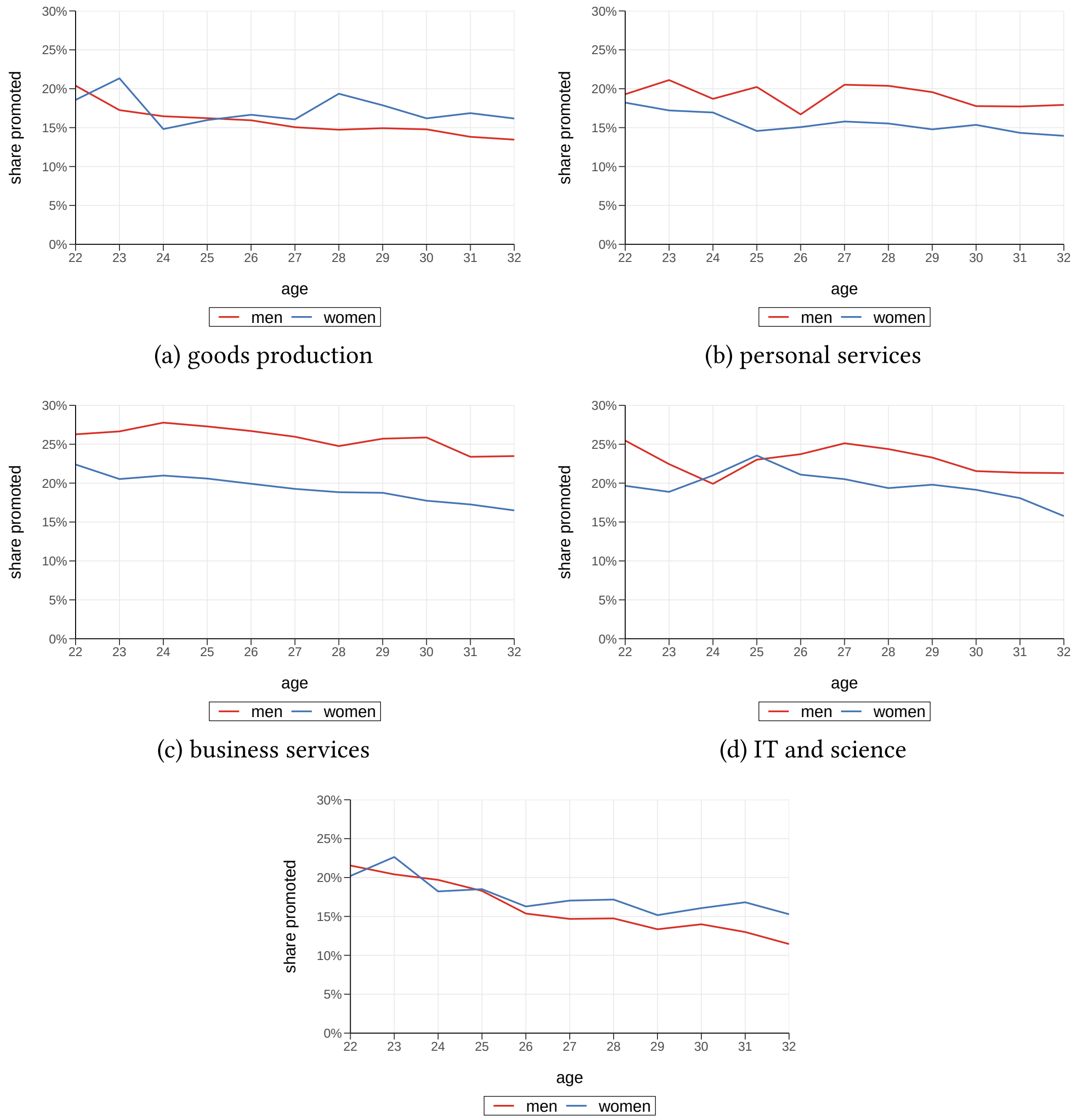


(a) demanding tasks (b) demanding know-how

(c) no demanding tasks (d) no demanding know-how

Figure B.5: Promotion rates by age, gender, and occupational characteristics (prior to first employment interruption)

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Subsamples with demanding tasks (share of analytic non-routine tasks ≥ $\frac{1}{3}$) or demanding know-how requirement (specialist or expert knowledge required) Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm, women prior to first long employment interruption (≥ 1 year). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).



(a) goods production
(b) personal services
(c) business services
(d) IT and science
(e) other services

Figure B.6: Promotion rates by age, gender, and occupational sector (prior to first employment interruption)

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm, women prior to first long employment interruption (≥ 1 year). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

Table B.2: The gender gap in promotions (prior to first employment interruption)

	promotion (relative wage growth $\geq$ 10%) – mean: 0.1772 –			
	(1)	(2)	(3)	(4)
women	-0.0193^{***}	-0.0185^{***}	-0.0316^{***}	-0.0185^{***}
	(0.0026)	(0.0026)	(0.0028)	(0.0026)
human capital	✓	✓	✓	✓
occupational characteristics		✓	✓	✓
occupation/industry FE			occupation	industry
year + firm FE	✓	✓	✓	✓
N	382,585	382,585	382,585	382,585

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Covariates included: human capital (age, age sq., nongerman, education, FT experience, FT experience sq.), occupational characteristics (demanding tasks, demanding know-how), occupation and industry FE (2-digit). Sample: individuals between ages 22 and 32, working full-time in consecutive periods at the same firm, women prior to first long employment interruption ($\geq$ 1 year). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

Appendix B3 Derivations for the Model Extension

In all periods, it is efficient to promote a male worker from job $j-1$ to job j if his effective ability exceeds the threshold value $\bar{\eta}^j$ (equation (2.1)), which solves the following equality:

$$\begin{aligned} d_{j-1} + c_{j-1}\eta_{it} &= d_j + c_j\eta_{it} \\ \eta_{it} &= \frac{d_{j-1} - d_j}{c_j - c_{j-1}} \equiv \bar{\eta}^1 \end{aligned}$$

For women in period t_3, the firm's decision problem is equivalent to the one for men, as there is no uncertainty with respect to childbirth and incurring leave-related cost k_j. The same holds for women who remain childless in period t_2, as for them all uncertainty with respect to childbirth is resolved. Women who give birth in period t_2 are never promoted, as employers would only incur higher leave-taking costs $k_{j+1} > k_j$.

In period t_1, however, firms take into account the possibility that female workers may drop out in t_2 when making promotion decisions, as they cannot fire or demote them prior to leave-taking. Promotions in t_1 are possible from $j = 0$ to $j = 1$, as all workers are initially hired into job $j = 0$. Let the expected output of a woman who is not promoted in t_1 be denoted as $V_1(j = 0)$ and as $V_1(j = 1)$, if she is promoted to job $j = 1$ in t_1:

$$\begin{aligned} V_1(j=0) &= d_0 + c_0\eta_{it} + p_f \cdot \left[p_{cc} \cdot (-k_0^{cc}) + (1-p_{cc}) \cdot (-k_0^{ncc})\right] + (1-p_f) \cdot V_2^* \\ V_1(j=1) &= d_1 + c_1\eta_{it} + p_f \cdot \left[p_{cc} \cdot (-k_1^{cc}) + (1-p_{cc}) \cdot (-k_1^{ncc})\right] + (1-p_f) \cdot V_2^* \end{aligned}$$

V_2^* denotes the expected output if a woman remains childless in t_2, which is identical in both cases as it is independent of V_1. Promotion decisions for women in t_1 are therefore based on the threshold value $\bar{\eta}^*$ (equation (2.2)), which solves the following equality:

$$\begin{aligned} d_0 + c_0\eta_{it} + p_f \cdot \left[p_{cc} \cdot (-k_0^{cc}) + (1-p_{cc}) \cdot (-k_0^{ncc})\right] &= \\ d_1 + c_1\eta_{it} + p_f \cdot \left[p_{cc} \cdot (-k_1^{cc}) + (1-p_{cc}) \cdot (-k_1^{ncc})\right] &. \end{aligned}$$

Rearranging the terms yields:

$$\begin{aligned}
\eta_{it}(c_1 - c_0) &= d_0 - d_1 + p_f \cdot \left[p_{cc} \cdot (k_1^{cc} - k_0^{cc}) + (1 - p_{cc}) \cdot (k_1^{ncc} - k_0^{ncc})\right] \\
\eta_{it} &= \frac{d_0 - d_1}{c_1 - c_0} + p_f \cdot \left[p_{cc} \cdot \frac{k_1^{cc} - k_0^{cc}}{c_1 - c_0} + (1 - p_{cc}) \cdot \frac{k_1^{ncc} - k_0^{ncc}}{c_1 - c_0}\right] \\
\eta_{it} &= \bar{\eta}^1 + p_f \cdot \left[p_{cc} \cdot \frac{k_1^{cc} - k_0^{cc}}{c_1 - c_0} + (1 - p_{cc}) \cdot \frac{k_1^{ncc} - k_0^{ncc}}{c_1 - c_0}\right] \equiv \bar{\eta}^*
\end{aligned}$$

Appendix B4 Additional Figures on the Public Childcare Expansion

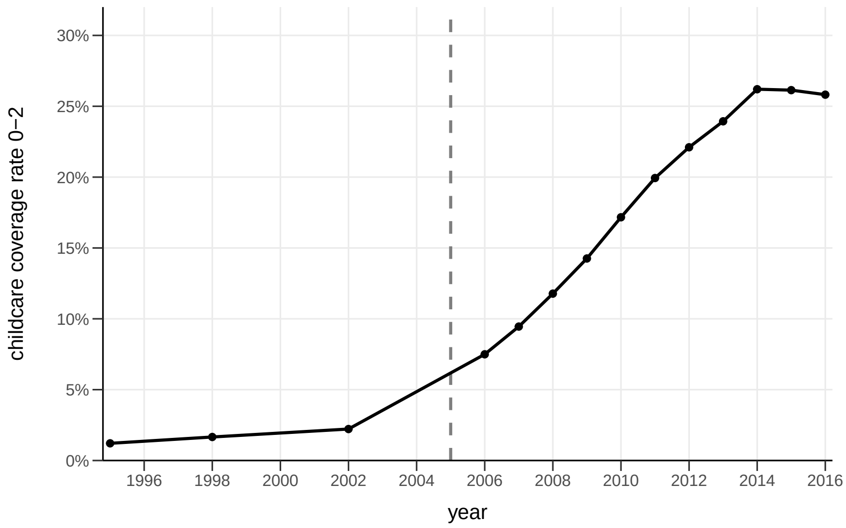


Figure B.7: Development of the childcare coverage rate

Notes: Childcare coverage: number of children aged 0 – 2 enrolled in public childcare divided by total number of children aged 0 – 2 in a county. Sample: West German counties. Source: BBSR Bonn (2021), values only available for years indicated by dots.

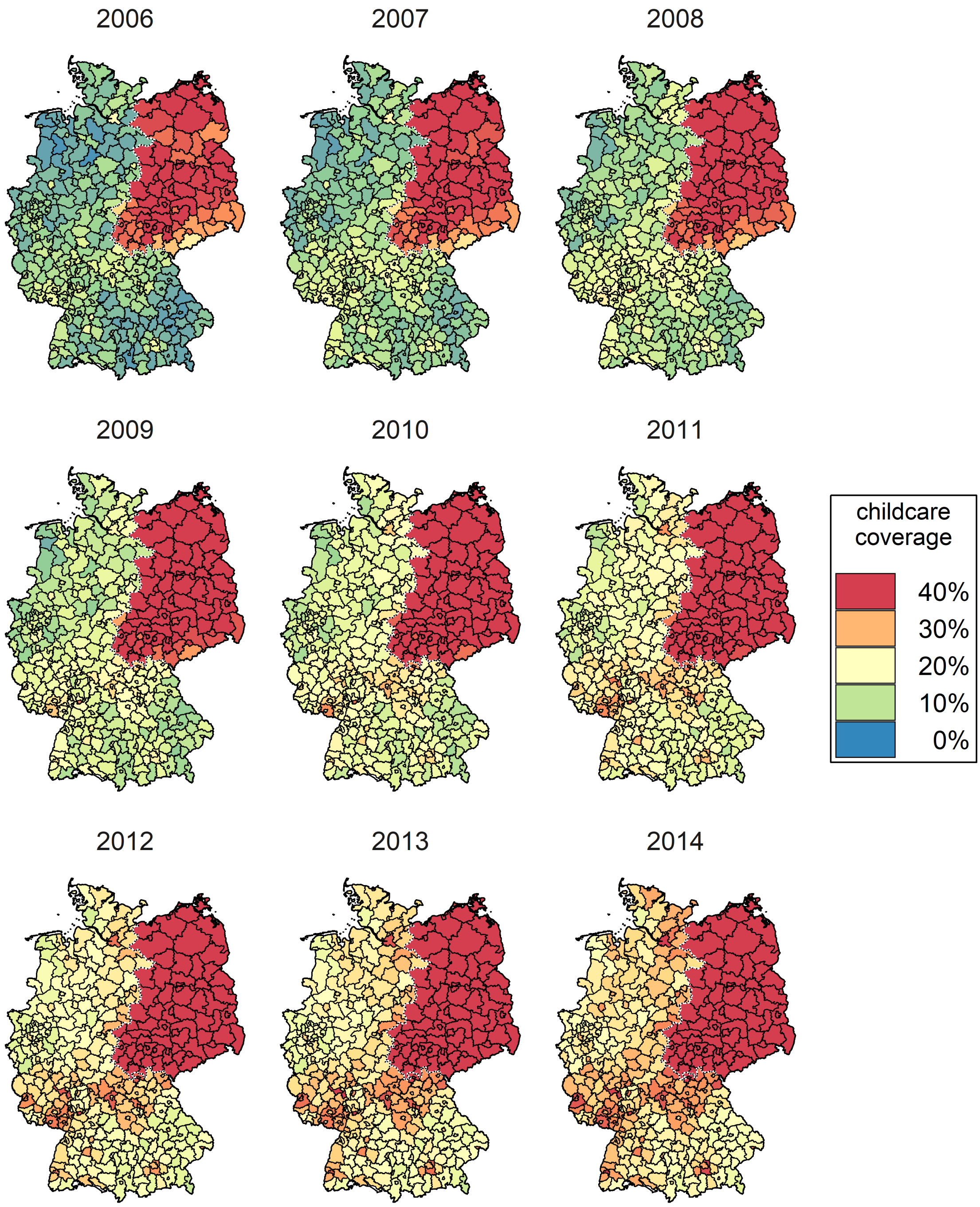


Figure B.8: Childcare coverage rates across counties and time

Notes: Childcare coverage: number of children aged 0 – 2 enrolled in public childcare divided by total number of children aged 0 – 2 in a county, > 40 collapsed into 40% bin. Sample: German counties. Source: BBSR Bonn (2021).

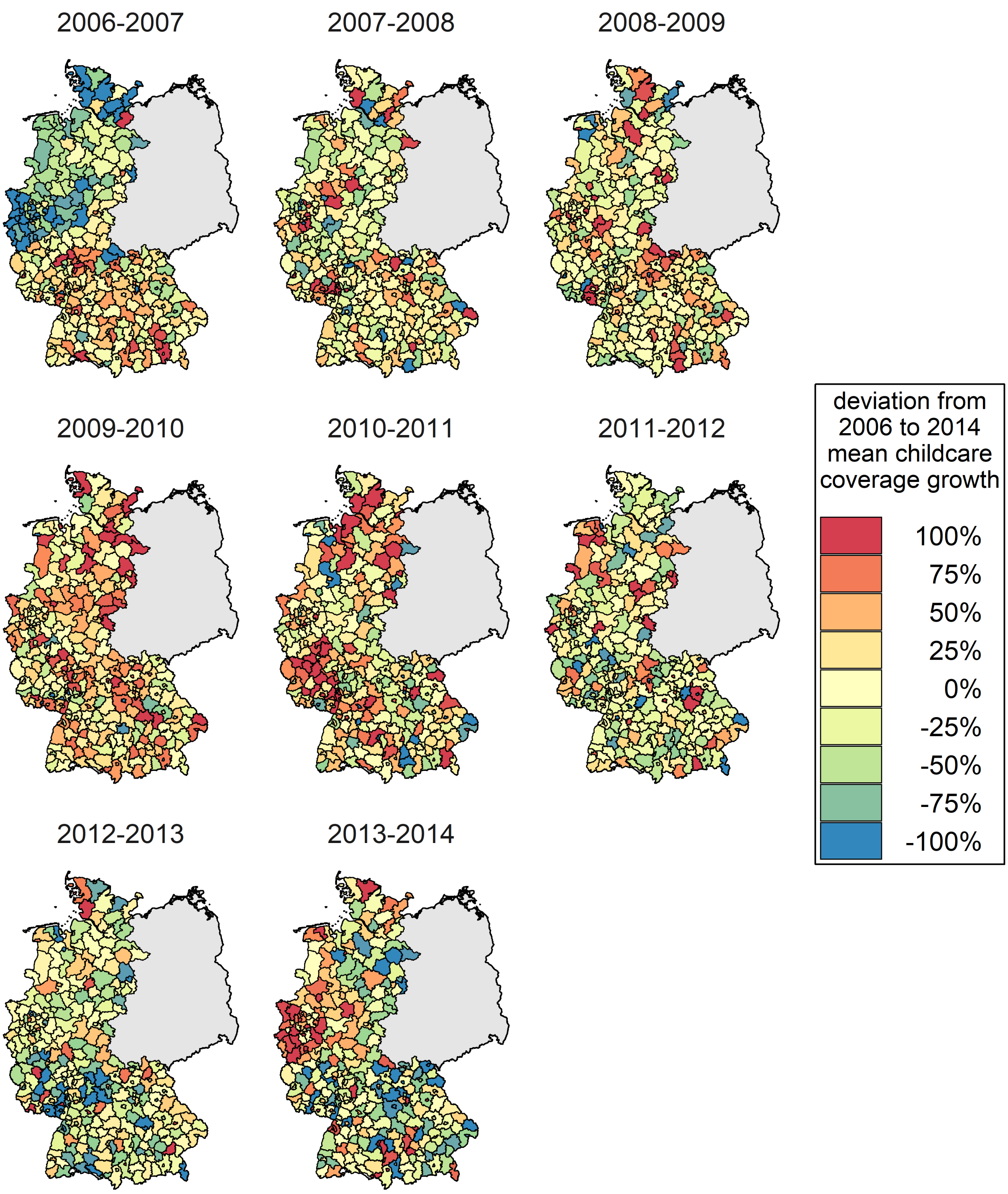


Figure B.9: Year-on-year variation in childcare coverage growth rates

Notes: Childcare coverage: number of children aged 0 – 2 enrolled in public childcare divided by total number of children aged 0 – 2 in a county. Positive values indicate that a given year-on-year growth rate exceeds the 2006 – 2014 county level mean growth rate in childcare coverage, negative values vice versa. Values above/below +/-100% collapsed into the closest bin. Sample: West German counties. Source: BBSR Bonn (2021).

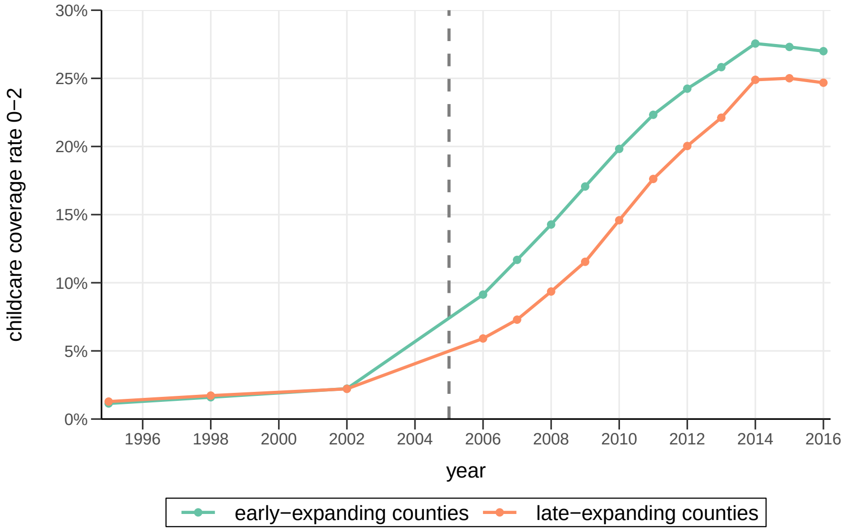


Figure B.10: Childcare coverage rates over time by childcare expansion speed

Notes: Childcare coverage: number of children aged 0 – 2 enrolled in public childcare divided by total number of children aged 0 – 2 in a county. Classification of counties into early-expanding (≥ 50% of 2006 – 2014 childcare coverage growth until 2009) and late-expanding (< 50% of childcare coverage growth until 2009). Sample: West German counties. Source: BBSR Bonn (2021).

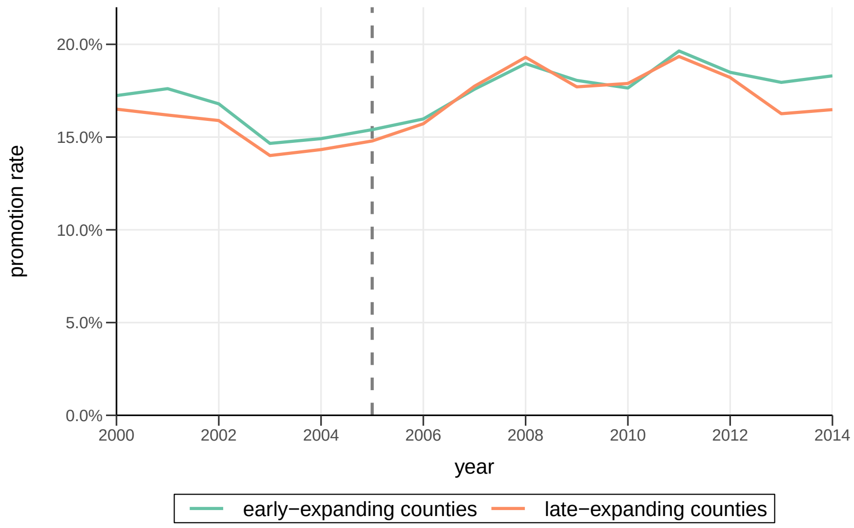


Figure B.11: Promotion rates over time by childcare expansion speed

Notes: Promotions: relative wage growth ≥ 10% (wage growth benchmarked at firm × education level). Classification of counties into early-expanding (≥ 50% of 2006 – 2014 childcare coverage growth until 2009) and late-expanding (< 50% of childcare coverage growth until 2009). Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b).

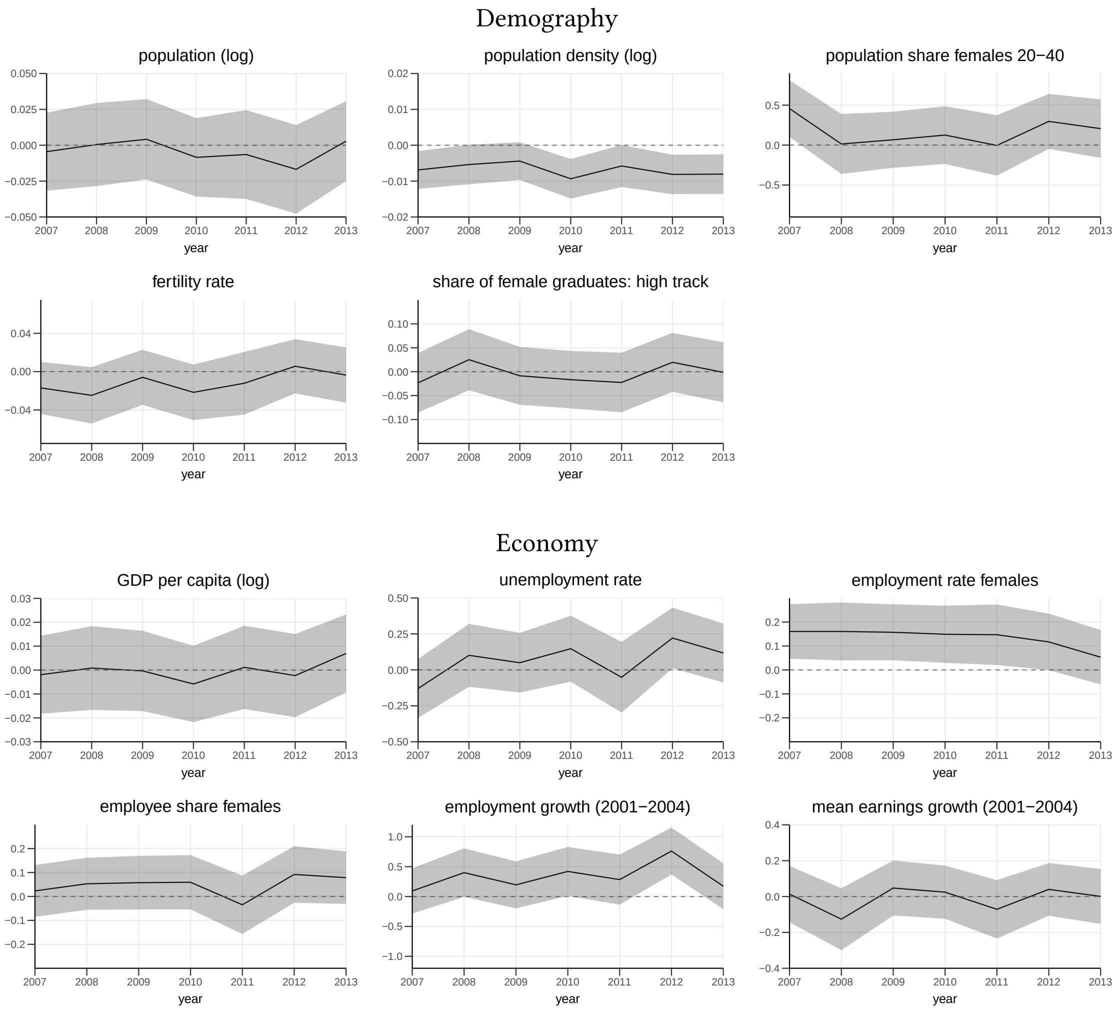
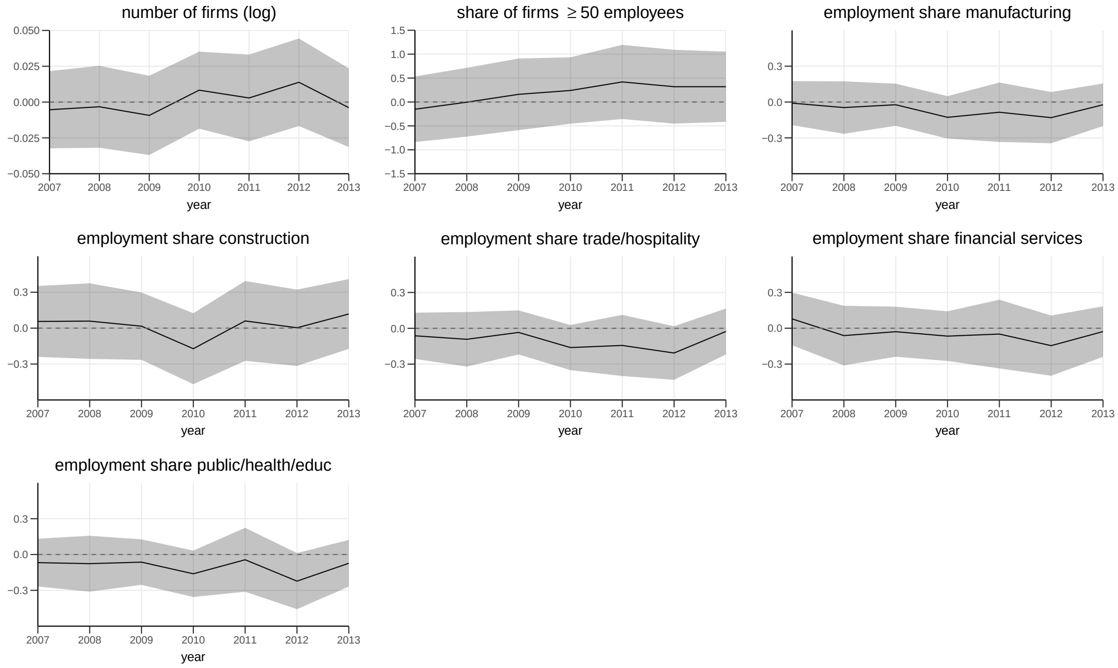


Figure B.12: Relation of the expansion timing and 2005 county characteristics

Notes: Estimates from equation (2.5) for the vector $\boldsymbol{\psi}_t$ for every t, reference category is 2014 in all cases. Shaded in grey are the associated robust 95% confidence intervals. Sample: West German counties. Source: BBSR Bonn (2021) and (FDZ-StABL 2021).

Industry structure



Political preferences

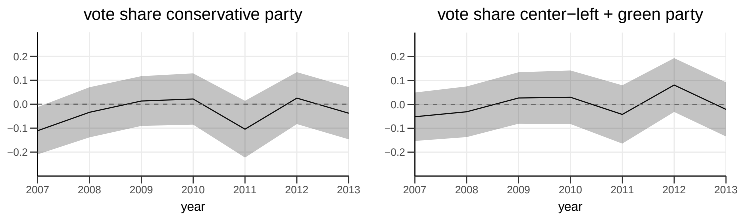


Figure B.12 (cont.): Relation of the expansion timing and 2005 county characteristics

Notes: Estimates from equation (2.5) for the vector $\boldsymbol{\psi}_t$ for every t, reference category is 2014 in all cases. Shaded in grey are the associated robust 95% confidence intervals. Sample: West German counties. Source: BBSR Bonn (2021) and (FDZ-StABL 2021).

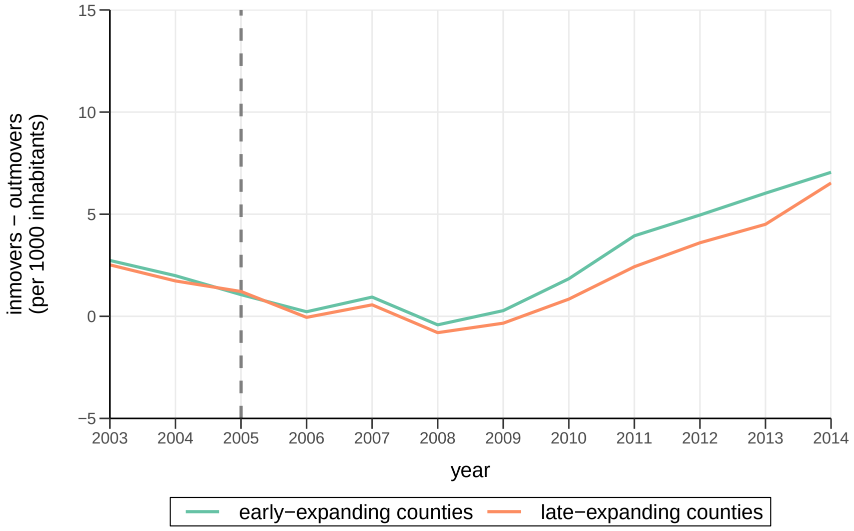


Figure B.13: Migration patterns by childcare expansion speed

Notes: inmovers – outmovers: number of individuals who moved into the county minus number of individuals who moved out of the county. Classification of counties into early-expanding (≥ 50% of 2006 – 2014 childcare coverage growth until 2009) and late-expanding (< 50% of childcare coverage growth until 2009). Sample: West German counties. Source: BBSR Bonn (2021).

Appendix B5 Additional Results Tables

Table B.3: Heterogeneity of the effect of public childcare on promotions by firm size

	promotion (relative wage growth ≥ 10%) – mean: 0.1781 –					
indicator threshold	≥ 10	≥ 25	≥ 50	≥ 100	≥ 200	≥ 500
	(1)	(2)	(3)	(4)	(5)	(6)
childcare coverage	0.1109	0.1196	0.1380	0.1582*	0.1595*	0.1739**
	(0.0861)	(0.0851)	(0.0854)	(0.0858)	(0.0859)	(0.0856)
firm size indicator	−0.0194***	−0.0298***	−0.0295***	−0.0229***	−0.0221***	−0.0046
	(0.0050)	(0.0048)	(0.0045)	(0.0047)	(0.0050)	(0.0064)
childcare coverage × firm size indicator	0.0792***	0.0803***	0.0624**	0.0284	0.0328	0.0310
	(0.0248)	(0.0267)	(0.0253)	(0.0262)	(0.0277)	(0.0325)
covariate set	*ico*	*ico*	*ico*	*ico*	*ico*	*ico*
year & county FE	✓	✓	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓	✓	✓
N	176,538	176,538	176,538	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Different thresholds for binary firm size indicators across columns, e.g., = 1 if number of full-time employees ≥ 10 (column (1)), = 0 otherwise. Covariate sets: i (age, age sq., nongerman, college, FT experience, FT experience sq.), c (log pop. density, population share females 20–40, employment rate females, fertility rate, log GDPPC), o (occupation segments, 14 catg.), s (industry segments, 13 catg.). See Appendix-Table B.7 for results without covariates. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.4: The effect of public childcare on age at first birth and age at first employment interruption

	age at first birth	age at first employment interruption
	(1)	(2)
childcare coverage	0.3861	−0.7723
	(2.8584)	(2.1863)
covariate set	$i^{\dagger}c$	$i^{\dagger}c$
year & county FE	✓	✓
linear county trends	✓	✓
mean dependent var.	28.98	28.85
N	8,996	15,405

Notes: Covariate sets: $i^{\dagger}$ (nongerman, college), c (log pop. density, population share females 20–40, employment rate females, fertility rate, log GDPPC). Sample: working women between ages 22 and 32, not out of the labor force in previous period, for whom first birth or first long employment interruption (≥ 1 year) is observed. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.5: The effect of public childcare on promotions (alternative sample age ranges)

	age range 20 – 34	age range 24 – 30
	(1)	(2)
childcare coverage	0.1353*	0.1132
	(0.0794)	(0.1006)
covariate set	*ico*	*ico*
year & county FE	✓	✓
linear county trends	✓	✓
mean dependent var.	0.1764	0.1786
N	213,433	123,853

Notes: Dependent variable in all columns: binary promotion indicator, i.e., relative wage growth $\geq$ 10% (wage growth benchmarked at firm $\times$ education level). Covariate sets: *i* (age, age sq., nongerman, college, FT experience, FT experience sq.), *c* (log pop. density, population share females aged 20 – 40, employment rate females, fertility rate, log GDPPC), *o* (occupation segments, 14 catg.). Sample: women between ages 20 (24) and 34 (30), working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Appendix B6 Results without Covariates

Table B.6: The effect of public childcare on wage growth and promotions (specification without covariates)

	wage growth	relative wage growth	promotion (relative wage growth ≥ 10%)
	(1)	(2)	(3)
childcare coverage	0.0189	0.0261	0.1516
	(0.0392)	(0.0433)	(0.0930)
covariate set	-	-	-
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
mean dependent var.	0.0332	0.0269	0.1781
N	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.7: Heterogeneity of the effect of public childcare on promotions by firm size (specification without covariates)

	promotion (relative wage growth ≥ 10%) – mean: 0.1781 –					
indicator threshold	≥ 10	≥ 25	≥ 50	≥ 100	≥ 200	≥ 500
	(1)	(2)	(3)	(4)	(5)	(6)
childcare coverage	0.0697	0.0913	0.1138	0.1353	0.1375	0.1533^{*}
	(0.0918)	(0.0910)	(0.0916)	(0.0926)	(0.0925)	(0.0917)
firm size indicator	−0.0043	-0.0146^{***}	-0.0156^{***}	-0.0116^{**}	-0.0131^{**}	0.0029
	(0.0052)	(0.0052)	(0.0050)	(0.0051)	(0.0054)	(0.0071)
childcare coverage × firm size indicator	0.1132^{***}	0.1012^{***}	0.0767^{***}	0.0419	0.0497^{*}	−0.0098
	(0.0256)	(0.0277)	(0.0273)	(0.0277)	(0.0300)	(0.0370)
covariate set	-	-	-	-	-	-
year & county FE	✓	✓	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓	✓	✓
N	176,538	176,538	176,538	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Different thresholds for binary firm size indicators across columns, e.g., = 1 if number of full-time employees ≥ 10 (column (1)), = 0 otherwise. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.8: Heterogeneity of the effect of public childcare on promotions by occupational characteristics and education (specification without covariates)

	promotion (relative wage growth ≥ 10%) – mean: 0.1781 –		
	(1)	(2)	(3)
childcare coverage	0.1428	0.1394	0.1580*
	(0.0927)	(0.0970)	(0.0914)
demanding tasks	0.0249***		
	(0.0053)		
childcare coverage × demanding tasks	0.0341		
	(0.0256)		
demanding know-how		0.0557***	
		(0.0066)	
childcare coverage × demanding know-how		0.0656*	
		(0.0357)	
college			0.1184***
			(0.0072)
childcare coverage × college			0.0243
			(0.0365)
covariate set	-	-	-
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
N	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. 'demanding tasks' defined as occupations with a share of analytic non-routine tasks ≥ $\frac{1}{3}$), 'demanding know-how' as occupations with specialist or expert knowledge required. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.9: The effect of public childcare on fertility and labor supply (specification without covariates)

	first birth next period	first interruption next period	consecutive full-time
	(1)	(2)	(3)
childcare coverage	−0.0454	−0.0308	0.0991
	(0.0349)	(0.0539)	(0.0912)
covariate set	-	-	-
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
mean dependent var.	0.0377	0.0764	0.6963
N	294,569	269,893	269,893

Notes: Binary indicators for first birth or first long employment interruption ($\geq$ 1 year) next period, conditional on observing either at some point. Sample: working women between ages 22 and 32, not out of the labor force in previous period, prior to first birth (column (1)) or prior to first long employment interruption ($\geq$ 1 year, column (2)). Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.10: The effect of public childcare on occupations (specification without covariates)

	demanding tasks	demanding know-how	switched occupation	switched industry
	(1)	(2)	(3)	(4)
childcare coverage	−0.0340	−0.0215	−0.0310	−0.0256
	(0.0874)	(0.1454)	(0.0787)	(0.0223)
covariate set	-	-	-	-
year & county FE	✓	✓	✓	✓
linear county trends	✓	✓	✓	✓
mean dependent var.	0.3012	0.2034	0.0368	0.0049
N	176,538	176,538	176,538	176,538

Notes: ‘demanding tasks’ defined as occupations with a share of analytic non-routine tasks $\geq \frac{1}{3}$), ‘demanding know-how’ as occupations with specialist or expert knowledge required, ‘switched occupation/industry’ as a binary indicator for different 2-digit occupational/industry code in prev. period. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Table B.11: The effect of public childcare on male wage growth and promotions (specification without covariates)

	wage growth	relative wage growth	promotion (relative wage growth ≥ 10%)
	(1)	(2)	(3)
childcare coverage	−0.0171	0.0109	0.0762
	(0.0344)	(0.0343)	(0.0700)
covariate set	-	-	-
year & county FE	✓	✓	✓
linear county trends	✓	✓	✓
mean dependent var.	0.0332	0.0269	0.1781
N	176,538	176,538	176,538

Notes: 'relative wage growth' is defined as individual wage growth minus the firm level mean wage growth of full-time employees with the same education level. Sample: women between ages 22 and 32, working full-time in consecutive periods at the same firm. Years 2006 – 2014, West Germany. Source: FDZ-IAB (2021b). Robust standard errors clustered at the county level in parentheses, significance at the 10%, 5%, and 1% level denoted by *, ** and ***.

Chapter 3

Part-time Wage Penalties across the Working Hours Distribution*

Abstract: We investigate heterogeneity of wages and wage growth rates across different working hours to quantify the implications of splitting work equally among spouses. Based on a combination of German administrative and survey data, we account for endogenous selection into specific hours by exploiting reforms of the tax system. We find substantial heterogeneity in part-time wage penalties, ranging from −2% to −18% compared to full-time. The heterogeneity in wage growth penalties is similar, but less severe. Both penalties are not linearly decreasing in working hours. High penalties for working high part-time hours suggest that splitting work equally would imply sizeable wage losses.

*This chapter is based on joint work with Ulrich Schneider.

3.1 Introduction

The labor market participation of women has significantly increased in recent decades. Nevertheless, a large fraction of women, especially mothers, continues to work fewer hours than their male counterparts. This high share of part-time employment is an important contributor to the gender wage gap and to child related earning losses for women, as, for example, documented by Kleven, Landais, Posch, et al. (2019).[1] A more equal distribution of working hours among spouses seems to be a potential path to reduce such gender differences. To understand the earnings implications of this potential path, it is crucial to understand how various working hours choices affect wage trajectories.

In this paper, we provide novel evidence on part-time penalties in wage levels and wage growth rates, which vary substantially across different working hours choices. Furthermore, we highlight the role that part-time penalties play for the incentives of how to split work among spouses and thereby inform the policy debate on gender equality in the labor market.

For our analysis, we leverage a combination of rich administrative data on earnings with survey data on hours to construct a high-quality panel of hourly wages for females in Germany. To control for selection into different working hours, we exploit numerous reforms of the German tax and transfer system along with conventional exclusion restrictions. Aside from full-time (> 34 hours), we focus on three common part-time working hours choices: working around 25% of the week (≤ 16 hours), working around 50% (> 16 to ≤ 24 hours), working around 75% (> 24 to ≤ 34 hours). Each of these part-time working hours choices accounts for around one-third of the overall part-time share, which itself amounts to around 50% for a large part of the life cycle.[2]

Our selection-corrected estimates of the part-time penalty in hourly wages show substantial heterogeneity, which is missed when treating part-time as a uniform choice: Using a uniform part-time indicator, we find penalties ranging from −6% to −10% across specifications. Splitting part-time into the three categories lined out above results in a much larger heterogeneity, with penalties ranging from −2% to −18%. The heterogeneity is especially driven by large penalties for working low hours in part-time (≤ 16h), which are more than twice as large as the ones for higher hours. Our findings of high penalties

[1]Throughout the OECD 17.9% of mothers with at least one child aged 0 – 14 work part-time (OECD 2021b). See, e.g., Bick, Brüggemann, et al. (2019) and Goldin (2014) for discussions of recent developments. See also Blau and Kahn (2017) on the relation of part-time work to the gender wage gap.

[2]As illustrated in Appendix-Figure C.1.

for working few hours are in line with earlier studies that relied solely on survey data (see, e.g., Gallego Granados 2019, and Paul 2016).

In addition to the literature, we also show that working high part-time hours (> 24 to ≤ 34h) does not only imply sizeable penalties as well, but can even carry higher ones than medium part-time hours (> 16 to ≤ 24h). This is the case for occupations and industries in which such high hours are uncommon.

In terms of part-time penalties in yearly wage growth rates, we estimate selection-corrected penalties between -1 and -1.4 percentage points for a uniform part-time indicator. Using our differentiated part-time working hours choices, we again document sizeable heterogeneity with penalties ranging from -0.7 and -1.9 percentage points. Compared to the average wage growth rate in full-time of 3.30%, these penalties amount to up to -50%. In terms of heterogeneity, the penalties are again the largest for working only few hours and decrease with increasing working hours, but not in a linear fashion. Wage growth penalties remain significant even for working many hours within part-time, which highlights that there are hours-independent elements in firms' cost functions.

All together, these results inform important policy questions: They illustrate the effects of splitting work equally among partners vs. the typical lopsided allocation. Thus, our results contribute to a better understanding of the division of labor among spouses and the related old-age poverty risk for women.[3] Under the assumption that our results for women also apply for men, we can conduct a simplified back-of-the-envelope calculation. A switch from the typical allocation of one partner working 100% and the other 50% to an equal share of 75% each would imply the following: In terms of hourly wages, both partners would face a -7% to -10% wage penalty, instead of just one partner facing a -2% to -6% penalty. In terms of wage dynamics, both partners would lose out on 0.7 to 1.3 percentage points in wage growth, instead of just one partner losing out on 0.9 to 1.4 percentage points. These results imply that couples would incur both a net income loss and potentially move to a lower wage growth trajectory as a consequence. Part-time penalties, therefore, create a sizeable disincentive to split work more equally among spouses.

In terms of the literature, we add to a large body that analyzes the causal impact of working part-time on wages and wage growth rates (see, e.g., Jones and Long 1979, Ermisch and Wright 1993, Aaronson and French 2004, and Paul 2016, among others). We improve the typical approach that concentrates on a uniform part-time indicator to

[3]See Tinios, Bettio, and Betti (2015) for an extensive investigation of how labor supply differences lead to the gender gap in pensions in the EU.

estimate the difference in wages and wage growth rates compared to full-time work by considering various part-time working hours choices.

We also discuss how our results relate to common explanations for part-time penalties. From the firms' side, these are especially per-person costs related to training, recruitment, as well as coordination. From the employees' side, the penalties might stem from slower human capital accumulation or differences in task content. Our findings for working few or medium hours are well in line with these explanations. However, we also find increasing wage penalties between medium and high part-time hours. This result is in contrast to the hypothesized cost structure of firms and the human capital theory. We show that this result is driven by occupations and industries in which high hours in part-time are uncommon. Thus, a potential explanation is that the incorporation of high part-time hours in the work flow might be especially costly in some occupations/industries and firms pass on these costs to the workers.

The remainder of the paper is structured as follows: We first discuss the related literature on part-time penalties in Section 3.2. Afterwards, we introduce the data and sample in Section 3.3. Section 3.4 lays out our empirical strategy to estimate part-time penalties. Sections 3.5 and 3.6 present the effects of working part-time on hourly wage levels and wage growth rates, respectively. In Section 3.7 we relate our findings to common explanations for part-time penalties. Finally, Section 3.8 concludes.

3.2 Related Literature

The analysis of part-time penalties is an active research area that has yielded a large set of estimates for various countries. The natural starting point for such analyses are models of hourly wages that condition on a rich set of individual characteristics. With early work going back to Jones and Long (1979), wage penalties of up to −25% have been documented for working part-time instead of full-time.[4] Many of these studies, however, point out that a large part of these penalties diminishes substantially after controlling for wage level differences across occupations and industries, job and firm characteristics, as well as

[4]See, e.g., Preston and Yu (2015) for Australia, Mumford and Smith (2009) for the UK, Jepsen et al. (2005) for Belgium, Fernández-Kranz and Rodríguez-Planas (2011) for Spain, Bardasi and Gornick (2008) for US, UK, Germany, Italy. Sweden is a notable exception, where Bardasi and Gornick (2008) find a small part-time premium. For an extensive overview of different estimates see Table 1 in Schrenker (2020).

contract types.[5] This highlights the large role that segregation in terms of jobs and firms plays, with part-time workers being often employed in low-paying occupations. Given our administrative data with detailed occupational and industry codes, we are able to take this segregation into account. The large cross-country heterogeneity in the estimates referenced above also demonstrates that local institutional characteristics and cultural factors are important for the extent of the part-time penalty. Thus, country-specific studies on part-time wage penalties should only be carefully generalized.

In addition to the effects on hourly wages, part-time work has also been shown to substantially affect wage growth trajectories (see, e.g., Fouarge and Muffels 2009, and Connolly and Gregory 2010). In other words, part-time work leads to a productivity penalty in future periods, which is a prominent feature in many structural life cycle models. Blundell, Costa Dias, et al. (2016) estimate, for example, for the UK that part-time work contributes only around one-eighth as much to human capital growth as full-time, while Schneider (2017) finds a 50% difference for Germany. These studies also underline the importance of investigating the wage effects of working part-time both at the level as well as the growth rate. In this paper, we therefore leverage the longitudinal dimension of our data to investigate the effects of different part-time hours at both margins.

Correcting for selection into part-time. Asides from segregation into different jobs and firms, the second important dimension of selectivity is endogenous selection into working part-time. If specific groups of individuals, e.g., in terms of preferences, constraints, or ability, are more likely to work part-time than others, then the observed differences vs. full-time wages will at least partially reflect this selection and not just the causal effect of working part-time. Adda, Dustmann, and Stevens (2017) show, for example, that early career choices can be linked to fertility preferences, illustrating that women with high desired fertility self-select into occupations in which part-time is more prevalent. Models that only condition on observables are therefore unlikely to identify the causal effect of working part-time on wages. To overcome this identification challenge, a number of different approaches have been brought forward to correct estimates for selection into part-time. We discuss the three most common ones: individual fixed effects, simultaneous hours choice and wage equation estimations, and Heckman selection correction models.

[5] See for occupations/industries: Preston and Yu (2015), Bardasi and Gornick (2008), Jepsen et al. (2005); for firm characteristics: Mumford and Smith (2009); for contract types: Fernández-Kranz and Rodríguez-Planas (2011).

For panel data sets, individual fixed effects have been employed to eliminate potential omitted variable bias driven by time-constant unobserved heterogeneity. These models identify the penalty only from individuals who are observed working different hours at different points in time. Typical estimates of this approach range from a −10% to −12% part-time penalty for Spain (Fernández-Kranz and Rodríguez-Planas 2011) and the UK (Connolly and Gregory 2008), a negligible penalty for the US (Hirsch 2005), to a small part-time premium for Australia (Booth and Wood 2008).

A second strand of the literature relies on the joint estimation of working hours choice and wage equations to account for the endogeneity of selection into part-time. These strategies use exclusion restrictions from institutional labor market regulations such as social security limits or family composition/household characteristics to separate the working hours decision from the wage equation.[6] The results are comparable to the ones from the previously discussed approaches, with penalties of up to −9% for Germany and no significant penalties for the US.

Lastly, corrections in the spirit of Heckman selection models (Heckman 1979) are also employed to correct for the endogeneity of the part-time choice.[7] Such applications start by estimating the probability of working in a specific working hours category. The obtained inverse Mills ratios are then used as control functions and plugged into separate wage equations for each working hours category.[8] To break the endogeneity, most of this work relies on the presence and age of children as well as marital status as exclusion restrictions. While there has been some discussion on the credibility of these exclusion restrictions and the reliability of the estimates,[9] the results using this approach indicate that positive selection into full-time is an important driver of part-time penalties (see, e.g., Mulligan and Rubinstein 2008).

Building on this work, Costa Dias, Joyce, and Parodi (2021) and Eisenhauer et al. (2020) set up similar control function approaches, but estimate two staggered selection equations: one for selection into employment, and one for selection into working full-time instead of part-time. By including the inverse Mills ratios from both selection

[6] See, e.g., Aaronson and French (2004) and Paul (2016) for applications exploiting labor market regulations and E. Wolf (2002) for the use of household characteristics.

[7] This approach has been popular in the literature. See for example, Ermisch and Wright (1993), Hardoy and Schøne (2006), Manning and Petrongolo (2008), Bardasi and Gornick (2008), Mulligan and Rubinstein (2008), Matteazzi, Pailhé, and Solaz (2014), Schrenker (2020).

[8] We use the term 'control functions' throughout this paper as introduced and defined in Heckman and Robb (1985, 1986).

[9] See, e.g., Manning and Petrongolo (2008) and Fernández-Kranz and Rodríguez-Planas (2011) for such discussions.

equations in a single wage model, they are able to directly identify the causal effects of working part-time on wages. As exclusion restrictions, they leverage variation in the tax code over years that affects the incentive to work either part-time or full-time differently. Focusing on wage growth rates, both find that working part-time results in substantially lower wage trajectories (Costa Dias, Joyce, and Parodi 2021, for the UK, Eisenhauer et al. 2020, for Germany). In this paper, we build on their work and construct control functions for four different working hours choices. For our exclusion restrictions, we also exploit that reforms of the tax and transfer system have had different effects on these choices.

Non-homogeneous part-time penalties. While there is a large literature on part-time penalties in general, only a few papers study heterogeneity in part-time penalties. Notable exceptions include Gallego Granados (2019) and Goldin (2014). Gallego Granados (2019) studies heterogeneity in part-time wages across the wage distribution and finds sizeable and persistent penalties for the lowest wage quintile, but not for higher ones. Goldin (2014) documents that firms disproportionately penalize short working hours. Two other closely related examples, which work purely with survey data, are E. Wolf (2002) and Paul (2016). E. Wolf (2002) finds a large penalty for part-time jobs with few hours per week (< 20h). Similarly, Paul (2016) documents hourly wage penalties for working very few hours (< 15h), but not for 16 to 34 hours, and negative wage growth effects for both categories. We broaden this analysis to a third part-time category that allows us to differentiate between full-time and close to full-time work, e.g., 75%. We extend the work further by using a unique combination of rich administrative data with survey data. This allows us to reduce the scope for biases originating from measurement error in earnings.

3.3 Data and Descriptives

To credibly investigate the scope of part-time penalties, we require a high-quality panel data set of hourly wages. Focusing on Germany, a country with a high prevalence of part-time work especially among mothers,[10] no large administrative panel data set exists that entails data on hourly wages.[11] Therefore, we leverage the combination of the following two data sources: First, we collect information on contracted working hours and a broad

[10] As shown in Appendix-Figure C.1.

[11] A very recent counterexample is Dustmann et al. (2020) who are able to construct a short six year panel from source data of the Federal Employment Agency's Statistics Department.

set of background variables from a longitudinal survey study, the German National Education Panel Study (NEPS[12]). The NEPS is a representative study covering adults born between 1944 and 1986 that live in private households and contains working hours information across the complete set of employment spells of every individual. Second, we link this data to administrative data from the German Federal Employment Agency's Integrated Employment Biographies (IEB[13]) that contains information on employment spells including daily earnings, but only a coarse part-time indicator.[14]

This combination of data sources yields us a panel of hourly wages that only contains minimal measurement error on the earnings side, as the earnings data stems directly from official social security reporting channels. There are, however, a number of difficulties arising when combining these two data sources: First, the sample is limited to individuals who consent to their data being matched with their social security records and for whom this match can be established (73.63% of all NEPS participants). Furthermore, we need to match NEPS and IEB spells, where the latter only contain episodes that are subject to social security contributions (i.e., no self-employment or public employment). To avoid mismatching working hours to earnings, we condition our sample on a match of the part-time indicator in the IEB data with the working hours recorded in the NEPS, as illustrated in Table 3.1 below.[15] Finally, the IEB earnings data is top-censored at the social security contribution limit, which concerns approximately the top ten percent of earnings. To be able to include these top earners in our analysis as well, we leverage information on firm-specific earnings levels to impute the censored values.[16]

[12]This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Adults. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS has been carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. Source: FDZ-LIfBi (2019). See 2011 for an introduction to the NEPS.

[13]This study uses survey data of the National Educational Panel Study (NEPS), Starting Cohort 6 (SC6) linked to administrative data of the IAB (NEPS-SC6-ADIAB 7518). Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently by remote data access. Source: FDZ-IAB (2020). See Bachbauer and C. Wolf (2020) for a detailed description.

[14]We use the term 'daily earnings' to allow for a clear differentiation with hourly wages, but the label 'daily wages' would also be appropriate as only earnings from the respective employment spell are accounted for.

[15]We previously apply the correction proposed by Fitzenberger and Seidlitz (2020) to account for the 2012 structural break in the reporting procedure regarding the part-time indicator.

[16]We use information on the average daily earnings at each firm computed on the universe of employees to apply the imputation procedure laid out in Section 8.1 in Schmucker et al. (2018).

Sample. For our final sample, we additionally apply the following restrictions: We focus on prime-working age individuals (aged 20 to 54) observed between 1975 and 2017, who are not in education. As the prevalence of part-time is very low among German men, we follow previous literature and limit our analysis to women.[17] Furthermore, we use the weights provided for use with the NEPS throughout our analysis to account for the sampling procedure and convert all monetary values to 2015 prices.

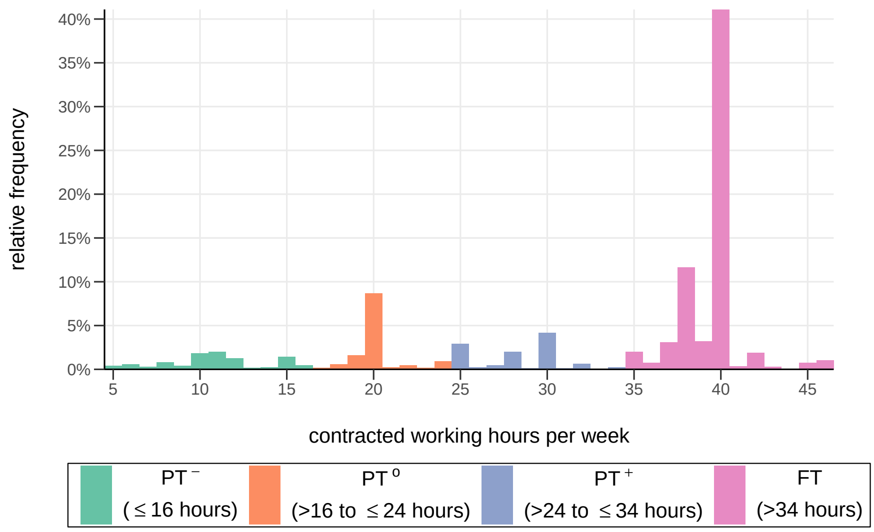


Figure 3.1: Histogram of women's working hours

Notes: Data collapsed into 1h bins. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Figure 3.1 presents a histogram of the contracted hours in our sample. While the clear majority of observations is centered around the common full-time range 35 to 40, the histogram illustrates the substantial heterogeneity in part-time hours. We bin the NEPS data into the following four categories, which we will rely on for our entire analysis: PT^- denotes working 1 to 2 days per week (≤ 16 hours), PT° denotes working around 3 days a week (> 16 to ≤ 24 hours), PT^+ denotes working around 4 days per week (> 24 to ≤ 34 hours), and FT denotes working full-time (> 34 hours). This categorization allows us to differentiate between three common part-time hours choices, that can be also framed more loosely as less than 50% (PT^-), around 50% (PT°) and more than 50% (PT^+).

[17] This is in line with the previous literature using German data, such as E. Wolf (2002), Paul (2016), Gallego Granados (2019), and Schrenker (2020).

Based on these categories, we can investigate the extent to which part-time penalties are heterogeneous across the hours distribution, while still keeping a sufficient number of observations in each bin.

Matching of working hours data. To judge the match quality of our two data sources, Table 3.1 contrasts the binned NEPS hours data with the binary IEB part-time indicator. The high overlap between the NEPS and the IEB classification (68.72% for part-time, 90.93% for full-time) illustrates that our spell matching procedure works reasonably well. Of the 93,689 annual observations for which we can match both data sources, we keep those for whom the IEB data and the NEPS data agree on part-time or full-time status. Furthermore, we additionally keep PT^+ observations that are classified as full-time in the IEB data, as this is the threshold category between part-time and full-time.[18] This leaves us with a total of 78,896 annual wage observations, which stem from 5,673 individual women. For our investigation of wage growth rates, the sample is smaller (N = 64,299) as we require consecutive wage observations. Furthermore, we condition our wage growth sample on not switching the working hours bin (96.2% of observations) to be able to separate wage growth penalties from hourly wage penalties. Additionally, we observe 10,998 unemployment spells, which we use in the selection equation for employment.

Summary statistics. Women in our sample earn an average hourly wage of €11.50 and experience an average wage growth rate of 2.83%, as shown in the summary statistics in Appendix-Table C.1. The average age of 36.28 corresponds well to the midpoint of our age range 20 – 54 and around half of the observations are with children. To be able to control for heterogeneous family backgrounds, we follow Costa Dias, Joyce, and Parodi (2021) and construct binary background measures based on median splits of two principal components from the following covariates: an indicator for working at age 18, indicators for college education of the mother and the father, the number of siblings, the number of older siblings and an indicator capturing whether the woman was living with both her parents at age 15.

[18]There is no clear hours threshold for the IEB classification into part-time and full-time. Neither employers, nor employees face any consequences when a part-time worker is labeled as full-time and employers often use previous year's forms. Therefore, we do not drop the mismatching observations in the threshold category PT^+. A robustness check without these observations reported in Appendix C5 confirms all of our main results. Furthermore, see Appendix-Table C.1 for summary statistics comparing the matched to the unmatched sample. The substantially lower wages for the unmatched sample are an indicator that for these we are mismatching full-time NEPS hours to part-time IEB earnings of, e.g., secondary jobs.

Table 3.1: Comparison of working hours in NEPS and IEB data

NEPS data	IEB data		
	PT indicator	FT indicator	Total
PT^-	8,400	912	9,312
(≤ 16 hours)	21.65%	1.66%	9.94%
PT°	10,292	1,745	12,037
(> 16 to ≤ 24 hours)	26.52%	3.18%	12.85%
PT^+	7,971	2,319	10,290
(> 24 to ≤ 34 hours)	20.54%	4.22%	10.98%
FT	12,136	49,914	62,050
(> 34 hours)	31.28%	90.93%	66.23%
Total	38,799	54,890	93,689
	100.00%	100.00%	100.00%

Notes: NEPS data categorization based on contracted working hours per week, IEB data are indicators directly reported by employers. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Aside from the individual characteristics, we are interested in investigating the role that different job characteristics play for part-time wage penalties. To do so, we construct five additional measures, whose prevalence across the four working hours choices is presented in Table 3.2.

The first measure regards the task composition of occupations: We code occupations as consisting of 'demanding tasks' if more than one-third of their typical tasks can be classified as analytic non-routine.[19] We complement this measure with a variable that focuses on the skill level required for the respective position. If the occupational code submitted by the employer indicates that specialist or expert skills are necessary, then such occupations are coded as 'demanding know-how' occupations. Table 3.2 reports that 'demanding tasks' and 'demanding know-how' are the least prevalent in PT^- and the prevalence of both measures increases in the part-time hours worked.

[19]We use the occupational task classification into analytic non-routine, interactive non-routine, cognitive routine, manual routine, and manual non-routine provided by Dengler, Matthes, and Paulus (2014).

Table 3.2: Job characteristics across working hours choices

	PT^-	PT°	PT^+	FT
demanding tasks	14.61%	23.18%	25.85%	24.85%
demanding know-how	10.48%	14.29%	19.41%	17.27%
fixed-term contract	13.88%	16.29%	16.02%	11.30%
common PT^+ occupation	20.03%	18.77%	34.35%	28.49%
common PT^+ industry	40.67%	29.10%	51.12%	38.03%

Notes: Hours categorization (weekly): PT^- denotes ≤ 16 hours, PT° denotes > 16 to ≤ 24 hours, PT^+ denotes > 24 to ≤ 34 hours, FT denotes > 34 hours. Covariate definitions: 'demanding tasks' indicates that at least one-third of the tasks entailed in an individuals occupation can be classified as analytic non-routine, 'demanding know-how' indicates that a specialist or expert skill level is necessary for a job. 'common PT^+ occupation/industry' denotes the prevalence of PT^+ relative to PT° is above its median value across all occupations/industries. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Another dimension, which has been highlighted in the literature, concerns fixed-term vs. permanent contracts. As Table 3.2 shows, fixed-term contracts are slightly more prevalent in part-time than in full-time jobs.

Finally, we introduce two measures to understand how common high part-time hours are within an occupation or industry. The first measure is constructed on the basis of the 3-digit occupational code. For each occupation, we calculate the share $\frac{PT^+}{PT^\circ + PT^+}$. If this share is above its median across all occupations, we label the occupation as a 'common PT^+ occupation'. Therefore, the indicator captures whether PT^+ is a commonly observed working hours choice in a specific occupation. It proxies high supply and simultaneous high take-up of PT^+, which we then contrast with occupations where PT^+ is an uncommon choice.[20] For our industry-based measure, we use the 3-digit industry code instead of the occupational code to construct the analogous indicator 'common PT^+ industry'.

[20] As the classification takes place at the occupation level, this allows for values below 50% for PT^+ in Table 3.2.

3.4 Empirical Strategy

To estimate the causal effect of working part-time on wages, we set up a regression model such as the one presented in equation (3.1). We use two different dependent variables: The first is the log wage ($log(w_{it})$), which we use to estimate whether hourly wage levels differ between working hours choices. Our second dependent variable of interest is the growth rate of hourly wages in log points ($\Delta log(w_{it}) = log(w_{it+1}) - log(w_{it})$), which we use to estimate how different working hours affect wage trajectories.

Our starting point is the following specification:

$$y_{it} = \alpha + \beta \mathrm{PT}_{it} + \boldsymbol{X}_{it}\gamma + \delta_t + \epsilon_{it}, \tag{3.1}$$

where y_{it} is one of the two labor market outcomes discussed above for woman i at time t. PT_{it} is a binary indicator for working part-time instead of full-time, corresponding to working less or equal to 34 hours per week. $\boldsymbol{X}_{it}$ contains a set of common controls in wage regressions, namely second-order polynomials in age as well as in part-time and full-time experience, along with a college degree indicator, an indicator for residing in former East Germany and an indicator for non-German citizenship. Furthermore, $\boldsymbol{X}_{it}$ also contains tenure, firm size, as well as 2-digit occupation and industry fixed effects.[21] δ_t captures a full set of year fixed effects. The main coefficient of interest is β as it measures whether women working part-time earn lower wages or experience lower wage growth than in full-time.

As we are not only interested in the aggregate penalty attached to working part-time, we expand equation (3.1) as follows:

$$y_{it} = \alpha + \beta_1 \mathrm{PT}^{-}_{it} + \beta_2 \mathrm{PT}^{\circ}_{it} + \beta_3 \mathrm{PT}^{+}_{it} + \boldsymbol{X}_{it}\gamma + \delta_t + \epsilon_{it}, \tag{3.2}$$

where PT^{-}_{it} is a binary indicator for working around 25% of the week (≤ 16 hours), while PT°_{it} and PT^{+}_{it} are equivalently constructed for around 50% (> 16 to ≤ 24 hours) and around 75% (> 24 to ≤ 34 hours), respectively. The coefficients of interest in equation (3.2) are β_1, β_2, and β_3, capturing the penalties for working different part-time hours vs. working full-time.

[21]For occupations we use 14 occupational segments based on the German Classification of Occupations 2010 (Matthes, Meinken, and Neuhauser 2015) and for industries we use 13 categories based on the NACE Rev. 2 classification (with the following aggregations: D+E, G+H, K+L, M+N, P+Q, and R+S+T).

To eliminate potential omitted variable bias from time-invariant unobserved factors, we extend (3.1) and (3.2) by the inclusion of individual fixed effects μ_i for some specifications.

3.4.1 Accounting for selection

The key concern for specifications (3.1) and (3.2) is endogenous selection into working specific hours. Additionally, we only observe wages for those who work, which is likely to be a selected sample of the population.

To account for the selection into working and for the selection into specific working hours, we employ a selection correction model that builds on the seminal work by Heckman (1979). More specifically, we set up an extended control function approach, as recently applied in Costa Dias, Joyce, and Parodi (2021). The key component in this approach is to exploit tax reforms that induced changes in the incentives to work different hours for the main exclusion restrictions. We construct these changes by extending the detailed tax code implementation from Bick, Brüggemann, et al. (2019), which captures all year-on-year changes in the German tax and transfer system. To better capture the decision to work at all, our implementation additionally covers changes in unemployment benefits.[22]

For the construction of the exclusion restrictions, we follow Costa Dias, Joyce, and Parodi (2021) closely: First, we regress female wages on full sets of interactions between age and year indicators with education.[23] Based on the estimated coefficients, we predict female wages and calculate gross household income for each of the working hours choices. Using predicted wages allows us to separate the exogenous variation induced by the tax reforms from potentially endogenous variation in observed wages. Furthermore, we only include women's earnings to avoid that correlations with spousal earnings contaminate our measure.

Next, we apply the tax and transfer code to calculate disposable household income for each working hours choice for every period. Finally, we regress our simulated disposable

[22]Our implementation covers variation in the tax code for the last 35 years of the sample, i.e., between 1983 to 2017 (93% of observations). The variation in our main exclusion restriction is therefore limited to this time frame.

[23]To be more precise, our dependent variables are interactions of indicators for each age between 20 and 54 with our college indicator and with interactions of indicators for all years 1975 to 2017 with our college indicator.

income on a set of family demographics to net out any aggregate effects. An overview of these different steps is provided in Appendix-Figure C.3.

This procedure leaves us with a set of residuals that captures how reforms of the tax and transfer system have affected the disposable household income for different working hours choices. Thus, our exclusion restrictions capture the reform-induced variation in the incentives to work different hours across the years (as, e.g., in Blundell, Duncan, and Meghir 1998, and Aaronson and French 2004).

Selection into employment. With these exclusion restrictions at hand, we start by constructing a control function for employment, i.e., a Heckman selection correction model, as follows:

$$\Pr\left(\text{employment}_{it} = 1 | \cdot\right) = \Phi\left(\alpha^e + \boldsymbol{Z}^e_{it}\boldsymbol{\beta}^e + \tilde{\boldsymbol{X}}_{it}\boldsymbol{\gamma}^e\right), \qquad (3.3)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. $\boldsymbol{Z}^e_{it}$ includes residualised simulated disposable income from the choice of not working as described above as the main exclusion restriction. As additional exclusion restrictions for the decision to participate in the labor market, we include a maternity indicator and the number of children aged 0 – 3 in $\boldsymbol{Z}^e_{it}$.[24] $\tilde{\boldsymbol{X}}_{it}$ contains a number of control variables, namely a third-order polynomial in age, indicators for college degree, non-German citizenship, marital status, and whether the household lives in former East Germany. Furthermore, we include the two family background measures described in Section 3.3 and their interaction.[25] After estimating specification (3.3), we obtain the corresponding inverse Mills ratio (λ^e) to use it as a control function for the selection into employment.

Selection into full-time vs. uniform part-time. Making use of λ^e, we construct a second control function to account for the endogeneity of the choice to work full-time vs. part-time in general ($\leq$ 34h) by estimating the following model:

$$\Pr\left(\text{FT}_{it} = 1 | \cdot\right) = \Phi\left(\alpha^{ft} + \boldsymbol{Z}^{ft}_{it}\boldsymbol{\beta}^{ft} + \tilde{\boldsymbol{X}}_{it}\boldsymbol{\gamma}^{ft} + \eta^{ft}\lambda^e_{it}\right) \quad \text{if employment}_{it} = 1. \qquad (3.4)$$

[24] Employment protection, i.e., the right of employees to return to their pre-birth jobs, lasts until the child to whom the employment interruption is related turns three for the majority of our sample time span (since 1992).

[25] These binary background measures are based on median splits of two principal components from the following covariates: an indicator for working at age 18, indicators for college education of the mother and the father, the number of siblings, the number of older siblings and an indicator capturing whether the women was living with both parents at age 15.

$\boldsymbol{Z}_{it}^{ft}$ contains two components based on the reform-induced variation in the tax and transfer system: First, the residualised simulated disposable income from working part-time (20h) and second, the increment in residualised simulated disposable income between working part-time and working full-time (40h). These two components of $\boldsymbol{Z}_{it}^{ft}$ summarize the effects of changes in the tax and transfer system on the incentives to work part-time or full-time. Additionally, we include a set of exclusion restrictions in $\boldsymbol{Z}_{it}^{ft}$ that are related to time constraints: a maternity indicator, the age of the youngest and oldest child, and the number of children in the age brackets $0-6$, $7-14$, $15-18$, and $19-25$. In $\tilde{\boldsymbol{X}}_{it}$ we include again the same set of covariates as before, capturing age, education, region, citizenship, and marital status. Through the inclusion of the inverse Mills ratio λ^e, we account for the selection into employment.

We then construct the inverse Mills ratio for the selection into working full-time using the estimation results from equation (3.4). This leaves us with two control functions at hand: λ^e for the selection into employment and λ^{ft} for the selection into working full-time. A full representation of both selection steps is provided in Appendix-Figure C.2.

By augmenting specification (3.1) with these two control functions, we are able to estimate a selection-corrected uniform part-time penalty. Our full regression model for hourly wages in our specification with only a uniform part-time indicator therefore becomes:

$$log(w_{it}) = \alpha + \beta \mathrm{PT}_{it} + \boldsymbol{X}_{it}\boldsymbol{\gamma} + \delta_t + \phi^e \lambda_{it}^e + \phi^{ft} \lambda_{it}^{ft} + \epsilon_{it}, \tag{3.5}$$

where λ^e is the inverse Mills ratio obtained after estimating (3.3) and λ^{ft} is the inverse Mills ratio obtained after estimating (3.4). For the regressions with wage growth as the dependent variable, we additionally include $\phi^{e1}\lambda_{it+1}^e$ to account for the selection into working in the following period:

$$\Delta log(w_{it}) = \alpha + \beta \mathrm{PT}_{it} + \boldsymbol{X}_{it}\boldsymbol{\gamma} + \delta_t + \phi^e \lambda_{it}^e + \phi^{e1} \lambda_{it+1}^e + \phi^{ft} \lambda_{it}^{ft} + \epsilon_{it}. \tag{3.6}$$

Selection into different hours brackets. After laying out how we account for endogenous selection in the estimation of a uniform part-time penalty, we now turn to the selection correction in the model with multiple part-time hours choices. To be able to account for the selection into the different working hours choices, we need to introduce one more selection equation layer. Figure 3.2 provides an overview of the three necessary stages, of which we have introduced the first one, i.e., selection into employment, already.

In line with Figure 3.2, we estimate the probability of working more hours than PT$^\circ$ with our sample of working women as the second step. This selection into working *high* hours is similar to the one for full-time in (3.4) and uses the following setup:

$$\Pr\left(\mathrm{PT}^{+}_{it} = 1 \text{ or } \mathrm{FT}_{it} = 1 \mid \cdot\right) = \Phi\left(\alpha^{h} + \boldsymbol{Z}^{h}_{it}\boldsymbol{\beta}^{h} + \tilde{\boldsymbol{X}}_{it}\boldsymbol{\gamma}^{h} + \eta^{h}\lambda^{e}_{it}\right) \quad (3.7)$$
$$\text{if employment}_{it} = 1.$$

The exclusion restriction vector $\boldsymbol{Z}^{h}_{it}$ contains again two components that capture the variation in the tax and transfer system: First, the residualised simulated disposable income from working *low* hours (mean of PT$^-$ and PT$^\circ$) and second, the increment in residualised simulated disposable income between working *low* and *high* hours (mean of PT$^+$ and FT).[26] Furthermore, we include a maternity indicator and the number of children in the age brackets 0 – 6, 7 – 14, 15 – 18, and 19 – 25 in $\boldsymbol{Z}^{h}_{it}$. $\tilde{\boldsymbol{X}}_{it}$ as well as λ^{e} are defined analogous to (3.4). After estimating (3.7), we construct two inverse Mills ratios: i) the inverse Mills ratio of selecting into working *high* hours (λ^{h}), and ii) the inverse Mills ratio of selecting into working *low* hours (λ^{l}).

With these inverse Mills ratios, we then turn to the final selection stage of Figure 3.2. We estimate the probabilities to work PT$^\circ$ (> 16 to ≤ 24h) instead of PT$^-$ (≤ 16h) and to work FT (> 34h) instead of PT$^+$ (> 24 to ≤ 34h) as presented below:

$$\Pr\left(\mathrm{PT}^{\circ}_{it} = 1 \mid \cdot\right) = \Phi\left(\alpha^{hl} + \boldsymbol{Z}^{hl}_{it}\boldsymbol{\beta}^{hl} + \tilde{\boldsymbol{X}}_{it}\boldsymbol{\gamma}^{hl} + \eta^{hl}\lambda^{e}_{it} + \eta^{hl}\lambda^{l}_{it}\right) \quad (3.8)$$
$$\text{if } \mathrm{PT}^{-}_{it} = 1 \text{ or } \mathrm{PT}^{\circ}_{it} = 1.$$
$$\Pr\left(\mathrm{FT}_{it} = 1 \mid \cdot\right) = \Phi\left(\alpha^{hh} + \boldsymbol{Z}^{hh}_{it}\boldsymbol{\beta}^{hh} + \tilde{\boldsymbol{X}}_{it}\boldsymbol{\gamma}^{hh} + \eta^{hh}\lambda^{e}_{it} + \eta^{hh}\lambda^{h}_{it}\right) \quad (3.9)$$
$$\text{if } \mathrm{PT}^{+}_{it} = 1 \text{ or } \mathrm{FT}_{it} = 1.$$

Analogous to (3.7), we use simulated disposable income for the working hours choices in question to construct our main exclusion restrictions: $\boldsymbol{Z}^{hl}_{it}$ contains i) residualised simulated disposable income from working PT$^-$, and ii) the increment in residualised simulated disposable income between working PT$^-$ and PT$^\circ$. $\boldsymbol{Z}^{hh}_{it}$ contains i) residualised simulated disposable income from working PT$^+$, and ii) the increment in residualised simulated disposable income between working PT$^+$ and FT. We additionally include a maternity indicator, the age of the youngest and oldest child, and the number of children

[26] We use the following representative working hours for the simulation of incomes: PT$^-$ = 10 hours, PT$^\circ$ = 20 hours, PT$^+$ = 30 hours, FT = 40 hours.

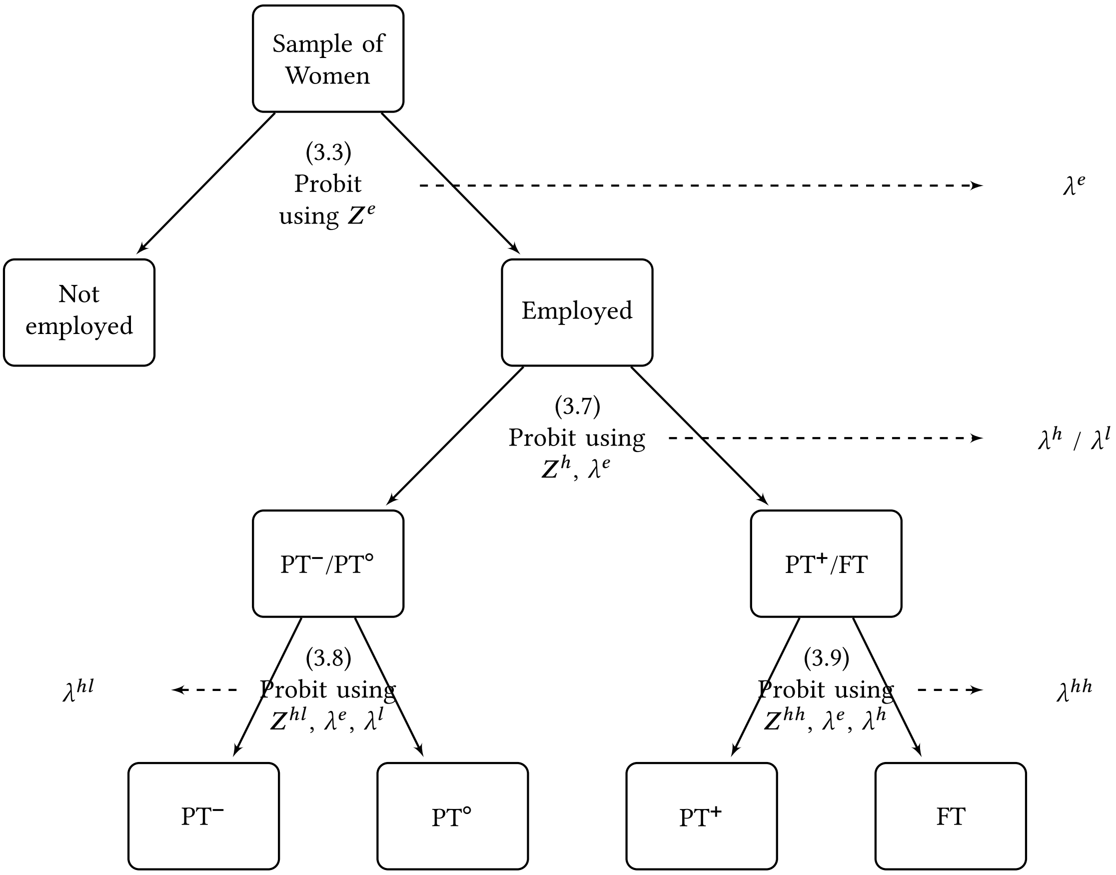


Figure 3.2: Illustration of the selection equations for different working hours choices

Notes: Own illustration of equations (3.3), (3.7), (3.9), and (3.8). λ^e, λ^h, λ^{hh}, and λ^{hl} denote the corresponding inverse Mills ratios obtained after estimating the respective Probit models.

in the age brackets 0 – 6, 7 – 14, 15 – 18, and 19 – 25 in our exclusion restriction vectors $\boldsymbol{Z}_{it}^{hl}$ and $\boldsymbol{Z}_{it}^{hh}$. $\tilde{\boldsymbol{X}}_{it}$ represents once again the set of covariates as used in all selection equations, while λ^e, λ^h, and λ^l are the inverse Mills ratios as described above, included as control functions to capture the respective selection mechanisms.

The estimation of (3.7), (3.8), and (3.9) provides us with three additional control functions that capture the selection into our four working hours choices. We augment specification (3.2) with the inverse Mills ratio for employment and the three just introduced ones to eliminate the selection bias in our coefficients of interest, the part-time penalties for different working hours choices β_1, β_2, and β_3. Therefore, the control function augmented version of (3.2) becomes:

$$\begin{aligned} y_{it} = \alpha + \beta_1 \text{PT}_{it}^{-} + \beta_2 \text{PT}_{it}^{\circ} + \beta_3 \text{PT}_{it}^{+} + \boldsymbol{X}_{it}\boldsymbol{\gamma} + \delta_t + \\ \phi^e \lambda_{it}^e + \phi^l \lambda_{it}^l + \phi^{hl} \lambda_{it}^{hl} + \phi^{hh} \lambda_{it}^{hh} + \epsilon_{it}, \end{aligned} \tag{3.10}$$

where λ^e, λ^l, λ^{hl}, and λ^{hh} are the inverse Mills ratios obtained after estimating equations (3.3), (3.7), (3.8), and (3.9) respectively. For the regressions with wage growth as the dependent variable, we again additionally include $\phi^{e1}\lambda_{it+1}^e$ to account for the selection into working in the following period.

3.4.2 Identification

Our strategy to credibly identify the causal effect of working part-time on wages is based on two components: conditioning on a broad set of observables, and on accounting for endogenous selection into the different working hours choices. First, by conditioning on a large number of individual characteristics as well as occupation and industry fixed effects, we make sure the part-time penalty is not driven by observable heterogeneity. Second, we carefully correct for endogenous selection into different working hours brackets using the control function approach laid out above. The main ingredient of the selection correction procedure is the variation in simulated disposable income induced by reforms of the tax and transfer system, which we discuss in detail below.[27]

We believe that the selection-corrected specifications account for the relevant sources of endogeneity. Nevertheless, we additionally show results for all models augmented with individual fixed effects to show that our findings are not driven by time-invariant unobserved heterogeneity. In these cases, identification relies on within-individual variation in working hours, i.e., on individuals who are observed working in different hours bins across time.

Tax reform induced variation in the incentives to work. To illustrate the source of variation in the tax and transfer system that we use for our main exclusion restrictions, we plot the development of the average tax rate across years and working hours choices for four hourly wage levels in Figure 3.3. The reference household for the underlying simulations is a mother with one 3-year-old child. In line with how we construct our

[27] An early example for this strategy is Blundell, Duncan, and Meghir (1998).

exclusion restrictions, we abstract from spousal earnings.[28] The key point of this illustration is to highlight the non-uniform variation in tax rates across the years and working hours choices.

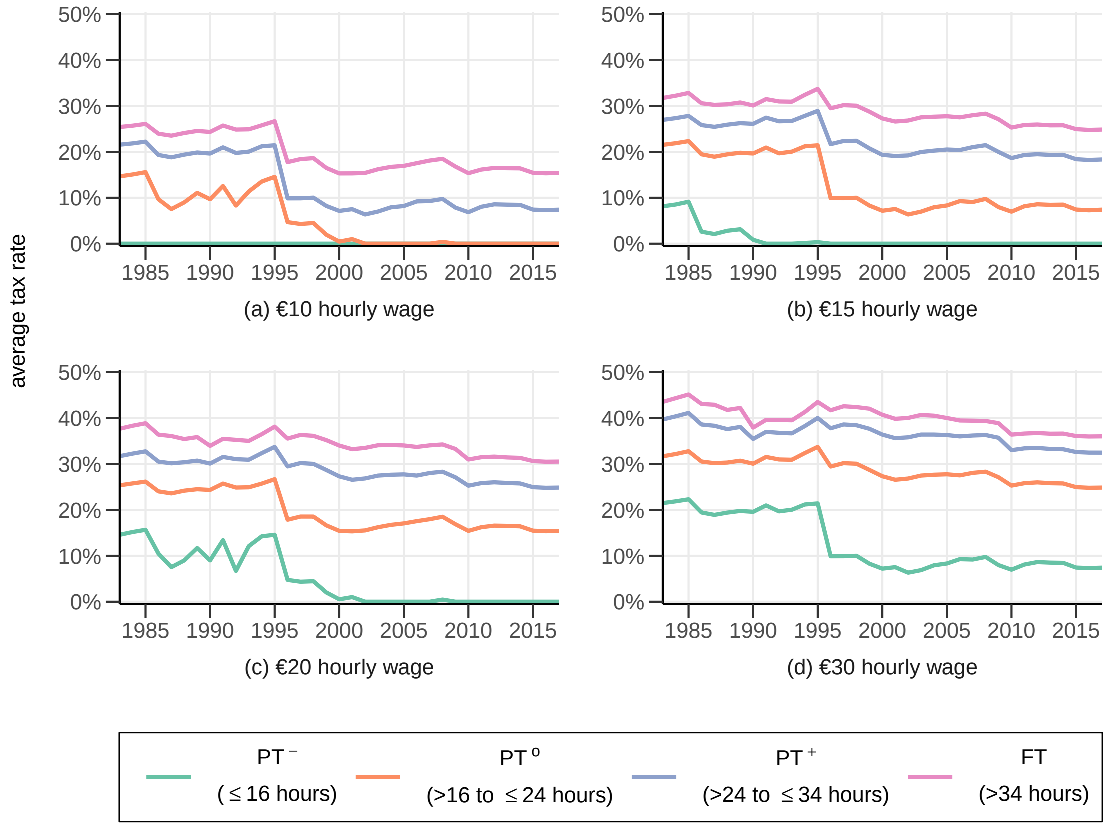


Figure 3.3: Development of average tax rates

Notes: Illustration of the average tax rate of a single earner household with one three year old child for different weekly working hours. All wages are expressed in 2015 prices. Source: modified tax code implementation based on Bick, Brüggemann, et al. (2019).

Starting with Panel (a) in Figure 3.3, we first look at mothers receiving an hourly wage rate of €10. If women only work PT^- at this wage rate, they do not have to pay any taxes. The same holds for PT° from 2002 onward, as several tax reforms changed the average tax rate from 10% to 15% (1980's and 1990's) down to zero. Women working

[28]These graphs abstract for ease of exposition from negative average tax rates, which do occur if individuals earn less than they would be entitled to through welfare payments. The variation in welfare levels, however, also contributes to the variation in the incentive to work different hours (as shown, e.g., in Carrillo-Tudela, Launov, and Robin 2021).

PT^{+} or FT did have to pay taxes throughout the entire time span, but the level varies substantially: Hovering around 20% and 25% until 1995, both average tax rates dropped substantially in 1996 and continued with some variation at levels that are around 10 percentage points lower. This large drop in 1996 was due to a court ruling that the basic income tax allowance must be large enough to ensure that a subsistence level income remains untaxed, which forced policymakers to raise the allowance to that level.[29]

Comparing the development across the working hours choices, however, it is clearly visible that the 1996 reform widened the gap in the average tax rate between working PT^{+} and FT, while it narrowed the gap between PT^{+} and PT°. Such changes in the incentives to work different hours are the key source of variation that we exploit for identification.

In Panels (b), (c), and (d), the changing differences between the average tax rates for different working hours are apparent as well. After 1996, the differences between PT°, PT^{+}, and FT increased throughout all three panels compared to the pre-1996 period. Furthermore, there are many instances of differential impacts of tax reforms on the different working hours choices: For example, between 2005 and 2010, the average tax rate steadily decreased across the years for working FT in Panels (c) and (d). Contrary, for working PT^{-}, the average tax rate first increased until 2008 and then decreased again. Another illustrative example is the difference between working PT^{+} and FT in Panel (d). While in the 1980s there was a small gap between the average tax rates, this gap almost completely diminished by 1990, but opened up again post 1996.

Overall, Figure 3.3 illustrates that throughout the years, reforms of the tax and transfer system have induced substantial variation in the average tax rates for different working hours, thereby shaping the incentives differently. As this variation in the tax and transfer rate is plausible to only impact wage levels and growth rates through the working hours choice, it provides us with credibly exogenous variation for our exclusion restrictions.

First stage results. The results for the first stage regressions from Section 3.4.1 are presented in Appendix-Tables C.2 and C.3. They show that the exclusion restrictions based on simulated disposable income have a strong effect on the decision to participate in the labor market and on how many hours to work for almost all choices we investigate. The exception is the decision to work PT^{-} instead of PT°, for which the variation in simulated disposable income alone does not have sufficient explanatory power. One reason for this might be that rather low yearly earnings are not significantly affected by the tax

[29] See Bundesverfassungsgericht (1992).

and transfer system. In this selection equation, the family composition based exclusion restrictions, i.e., the presence, number, and age of children, play a more important role. These more commonly used exclusion restrictions seem also to be valuable additions for the other four first-stage regressions. At the bottom of each table, we present joint tests of the income-based exclusion restrictions and the full set of exclusion restrictions. The full set is jointly significant in each of the five specifications.

3.5 Results on Wage Levels

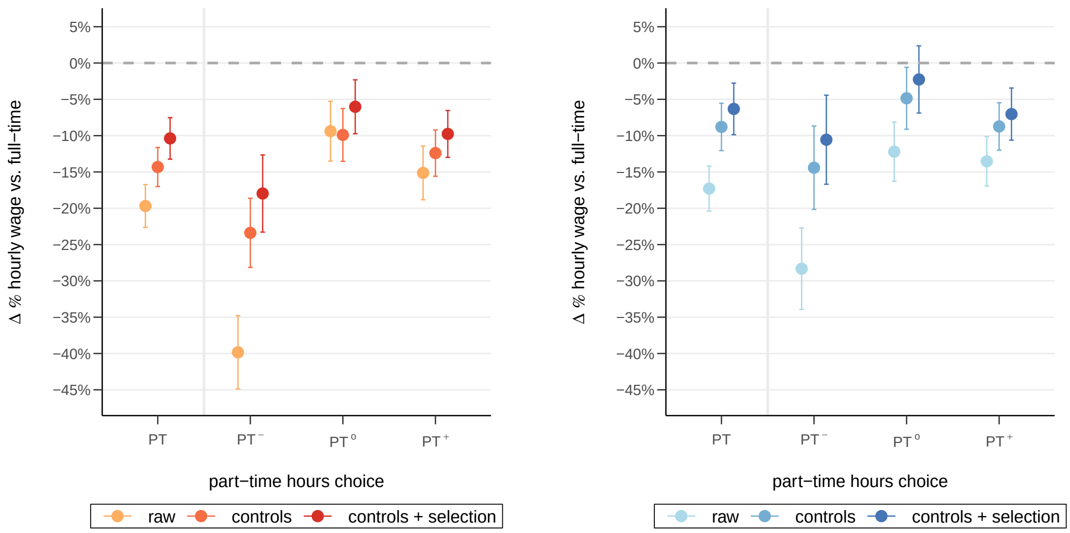
3.5.1 Part-time penalties in wage levels

We now turn to our main set of results on part-time penalties in wage levels. In Figure 3.4a, we present the results from estimating our conditional-on-observables model (3.1) and its selection-corrected version (3.5). Starting with the uniform part-time indicator in the leftmost column of Figure 3.4a, we find significant part-time penalties ranging from -20% without controls to -10% with a broad set of controls and the selection correction included. At an average full-time wage of €11.84, these translate into €−1.11 to €−2.33 lower hourly wages for part-time work.

Splitting the uniform part-time indicator into the three hours bins reveals substantial heterogeneity in the part-time wage penalties. PT^- (≤ 16h) carries a selection-corrected penalty of -18%. This penalty for women who work very few hours is 2 to 3 times larger than the one for women who work more hours (PT° or PT^+). Turning to PT° and PT^+, however, we find a hump shaped pattern: Working medium hours in PT° carries a rather low selection-corrected penalty of just -6%, while working relatively high hours in PT^+ implies a higher penalty of -10%.

Turning to the specifications with individual fixed effects in Figure 3.4b, we find very similar patterns as in Figure 3.4a, but with a level shift upwards (closer to zero). The specification with a uniform part-time indicator now stands at a -6% penalty for part-time work. Working very few hours (PT^-) carries again the largest penalty of -10%, while we find no significant part-time penalty for PT° and again a sizeable penalty of -7% for high part-time hours (PT^+). The hump-shaped pattern is, therefore, also present in the fixed effects specification.

Overall, the results in Figure 3.4 clearly show i) sizeable part-time penalties in hourly wage levels, and ii) that there is substantial heterogeneity in these penalties that is missed by treating part-time as a uniform hours choice. Furthermore, the selection correction plays an important role for the penalties, as is also confirmed by joint tests of the coefficients on the control functions reported below Appendix-Tables C.4 and C.5.



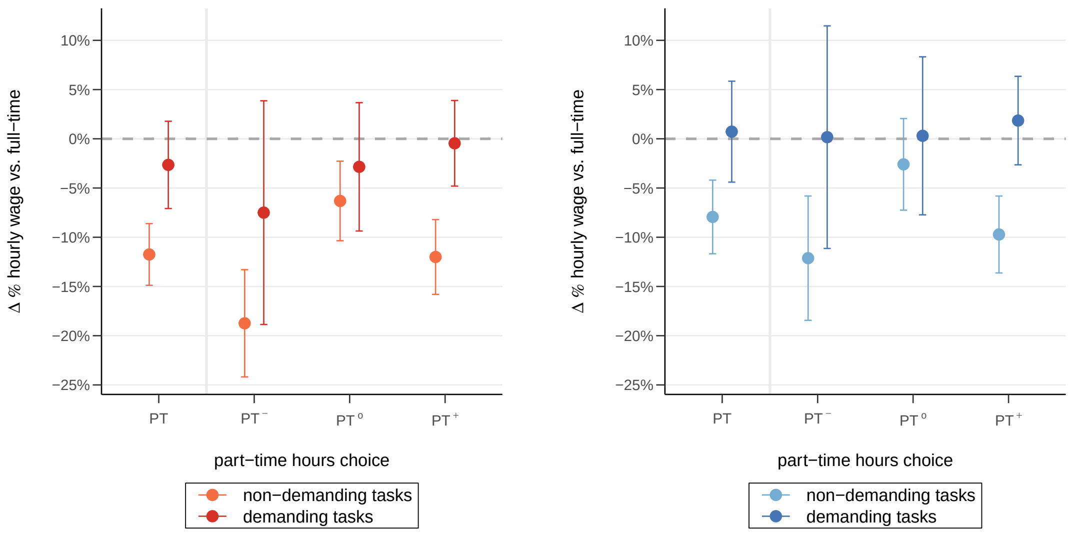
(a) without individual FE (b) with individual FE

Figure 3.4: Part-time penalties in wage levels

Notes: PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. *raw* is the part-time penalty only controlling for year fixed effects, *controls* includes a broad set of individual controls incl. occupation, industry, and year fixed effects and *controls + selection* adds control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Tables C.4 and C.5 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

3.5.2 Heterogeneity in part-time penalties in wage levels

To better understand the observed patterns, we now look at different dimensions of potential heterogeneity. Specifically, we investigate the observable heterogeneity in part-time penalties for different job characteristics by interacting our part-time working hours choices with indicators for each of the job characteristics. The results of the selection-corrected part-time penalties using both identification strategies, i.e., with and without individual fixed effects, along with their respective 95% confidence intervals, are presented in the following figures. Within each figure, the left-out category is always full-time within the specific characteristic. The darker colored estimates in Figure 3.5, for example, are the wage differences between part-time in a demanding occupation and full-time in a demanding occupation.

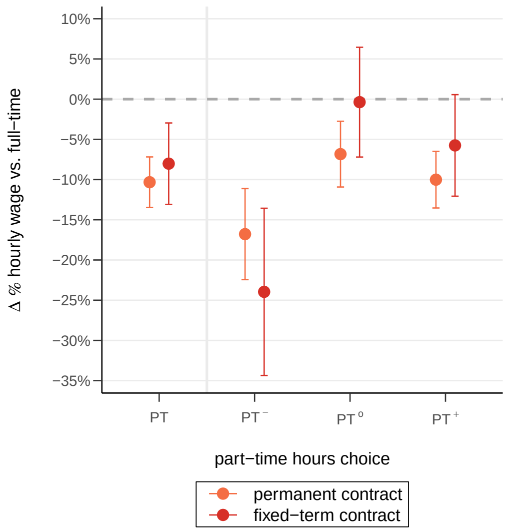


(a) without individual FE (b) with individual FE

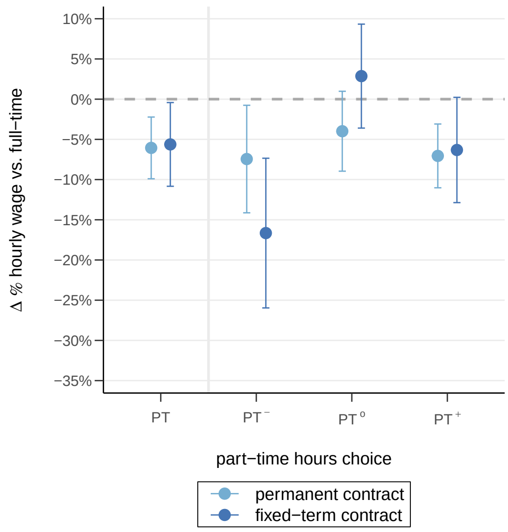
Figure 3.5: Part-time penalties in wage levels by task composition

Notes: The classification ‘demanding tasks’ indicates that at least one-third of the tasks entailed in an individuals occupation can be classified as analytic non-routine. PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.8 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

We start our heterogeneity analysis by looking at the role that the task composition of occupations plays. Figure 3.5 presents the part-time penalty estimates within occupations that consist of a demanding or non-demanding task set (high share of analytic non-routine tasks). The clear takeaway is that the sizeable part-time penalties documented in Figure 3.4 are predominantly driven by occupations with non-demanding tasks. For such occupations, the selection-corrected penalties are more pronounced, ranging from −3% to −19% and translating to up to €−2.05 per hour. Furthermore, the increase in penalties between PT° and PT^+ is also very prominent for non-demanding occupations. Occupations consisting mainly of demanding tasks, on the other hand, display no significant wage penalties, neither for the uniform part-time indicator nor for any of the three hours bins. The same pattern holds for our alternative definition of how demanding an occupation is, i.e., the know-how requirement, as illustrated in Appendix-Figure C.4.



(a) without individual FE



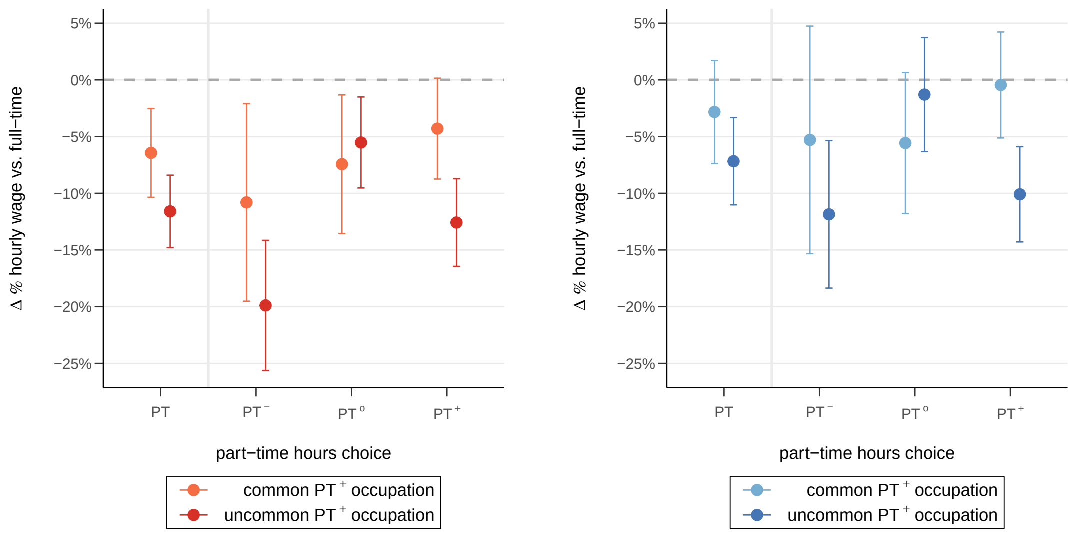
(b) with individual FE

Figure 3.6: Part-time penalties in wage levels by contract type

Notes: PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.10 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Next, we turn to differences between permanent and fixed-term contracts. As displayed in Figure 3.6 for the uniform part-time indicator and the two higher part-time hours bins, PT° and PT^+, the penalties are slightly lower in fixed-term contracts than in permanent contracts. Overall, however, the penalty estimates for both contract types are bunched relatively close together and do not carry very tight confidence intervals. This leads us to conclude that contract types play only a small role for the extent of part-time penalties in hourly wages.

Finally, we use our measure for the prevalence of PT^+ within an occupation to investigate the heterogeneity of our results. Figure 3.7 presents the corresponding penalty estimates by how common or uncommon PT$^+$ is per occupation. The results show clearly that in occupations where PT$^+$ is a common working hours choice, part-time penalties are continuously decreasing in the number of hours worked per week. The opposite is true for occupations where PT$^+$ is uncommon. These occupations are the drivers of the hump shape pattern, i.e., that PT$^+$ carries a higher penalty than PT$^\circ$. Furthermore, we confirm that these results also hold if we use the industry-level perspective to differentiate by the prevalence of PT$^+$ in Appendix-Figure C.5.



(a) without individual FE

(b) with individual FE

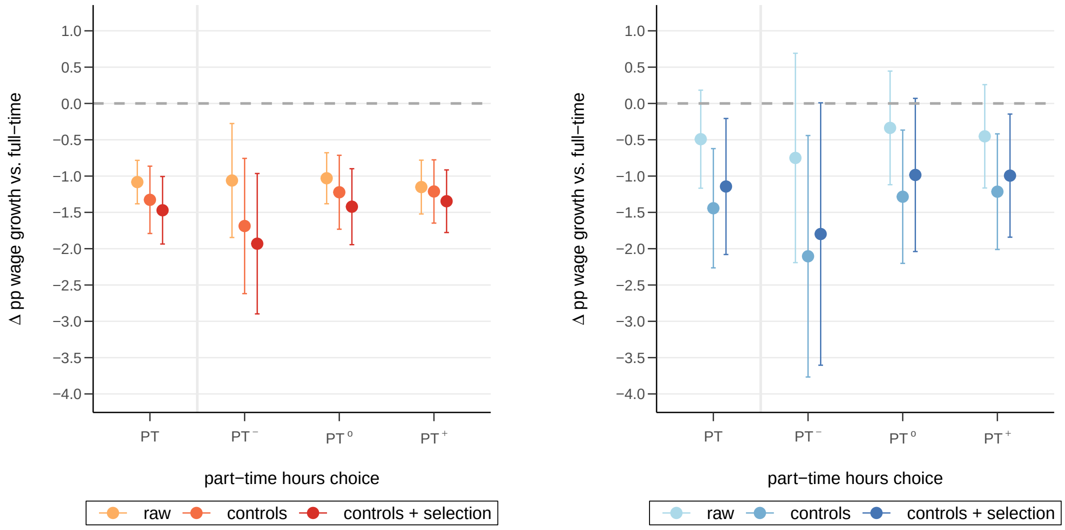
Figure 3.7: Part-time penalties in wage levels by prevalence of PT$^+$

Notes: The classification 'common PT$^+$ occupation' indicates that the prevalence of PT$^+$ relative to PT$^\circ$ is above its median value across all occupations. PT denotes working ≤ 34 hours per week, PT$^-$ denotes working ≤ 16 hours, PT$^\circ$ denotes working > 16 to ≤ 24 hours, PT$^+$ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.11 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

3.6 Results on Wage Growth

3.6.1 Part-time penalties in wage growth

After investigating the effects of part-time work on wage levels, we now turn to our second outcome of interest: annual wage growth rates. Analogous to the previous section, we start in Figure 3.8 with the results for both the conditional-on-observables and the fixed-effects specification.



(a) without individual FE (b) with individual FE

Figure 3.8: Part-time penalties in wage growth

Notes: PT denotes working ≤ 34 hours per week, PT$^-$ denotes working ≤ 16 hours, PT$^\circ$ denotes working > 16 to ≤ 24 hours, PT$^+$ denotes working > 24 to ≤ 34 hours. *raw* is the part-time penalty only controlling for year fixed effects, *controls* includes a broad set of individual controls incl. occupation, industry, and year fixed effects and *controls + selection* adds control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Tables C.6 and C.7 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

For the conditional-on-observables model in Figure 3.8a, we find that part-time work in general (PT) implies a significant −0.9 to −1.4 percentage point decrease in the wage growth rate vs. full-time. This translates into a −30% to −40% penalty compared to the average full-time wage growth rate of 3.30%. Focusing on the selection-corrected results, this wage growth penalty is heterogeneous across our part-time hours bins, but the differences are less pronounced than the ones for wage levels. The largest penalties are again incurred for part-time work with very few hours (−1.9 pp, −60% vs. FT), while working PT$^{\circ}$ or PT^{+} carries similar penalties of around −1.3 pp (−40% vs. FT).

The fixed-effects specification in Figure 3.8b paints a similar picture, again with a (small) level shift upwards and with substantially larger standard errors. Using just the uniform part-time indicator confirms a significant selection-corrected −1 pp wage growth penalty for working less than full-time (−30% vs. FT). The point estimates for the different hours bins are continuously decreasing in working hours, starting at −1.7 pp for PT^{-}, via −1 pp for PT$^{\circ}$ to −0.7 pp for PT^{+}. This translates into penalties of more than −50% vs. FT for PT^{-}, which decrease to −20% for PT^{+}. The larger standard errors render these estimates, however, only significant at the 10% level (see Appendix-Table C.7 for details).

It is also notable that the wage growth penalty is the smallest when not controlling for confounding factors. Including controls increases the penalties, implying that, with respect to wage growth, there is some positive selection on observables into part-time. This pattern is in contrast to the findings for wage levels. A possible explanation would be that part-time workers tend to work in, e.g., occupations that pay lower hourly wages, but grant higher wage growth.

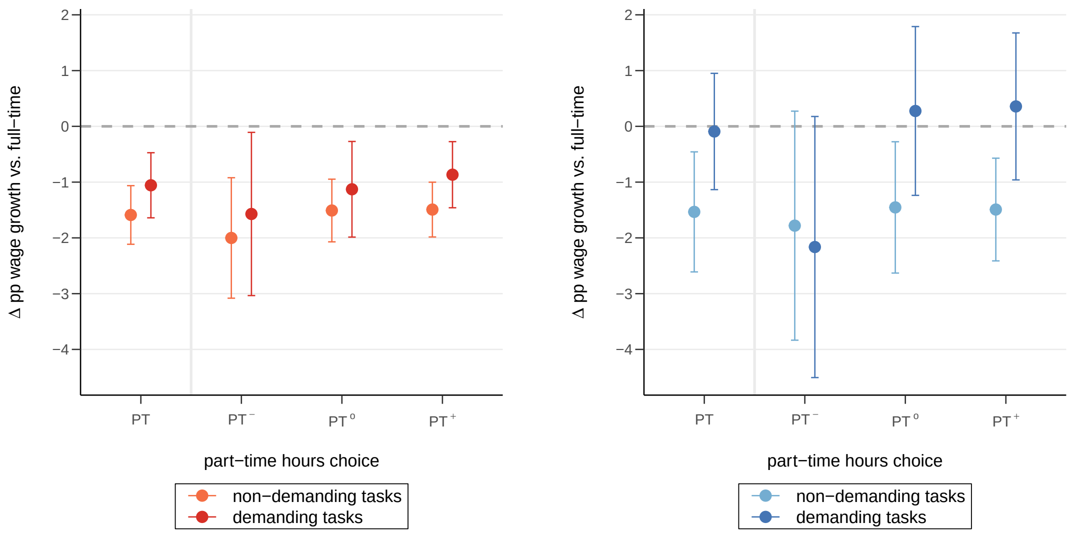
In summary, the results in Figure 3.8 again illustrate that there is considerable heterogeneity within a sizeable uniform part-time penalty in wage growth rates. Throughout these models, we again conduct joint tests of the control functions in Tables C.6 and C.7 which confirm the importance of the selection correction terms.

3.6.2 Heterogeneity in part-time penalties in wage growth

As previously for the hourly wage penalties, we investigate next whether the wage growth penalties are heterogeneous by different job characteristics. In line with Section 3.5.2, the left-out category is always full-time within the specific characteristic.

Starting again with heterogeneity by task composition, we find only slightly lower wage growth penalties for occupations with demanding tasks in Figure 3.9a. In the specifications with individual fixed effects in Figure 3.9b, we do find evidence for lower wage growth penalties for occupations with demanding tasks. The wages of women working part-time on non-demanding tasks grow around 1.5 percentage points slower than in full-time (−45% vs. FT). Contrary, those working part-time on demanding tasks do not incur any wage growth penalty on average (uniform PT indicator) and for PT$^{\circ}$ as well as PT^{+}, but only for PT^{-}. This mixed evidence regarding higher part-time hours in

Figures 3.9a and 3.9b may reflect that permanent unobservables do play a role for the size of the penalties in the small subgroup with demanding tasks.[30]



(a) without individual FE

(b) with individual FE

Figure 3.9: Part-time penalties in wage growth by task composition

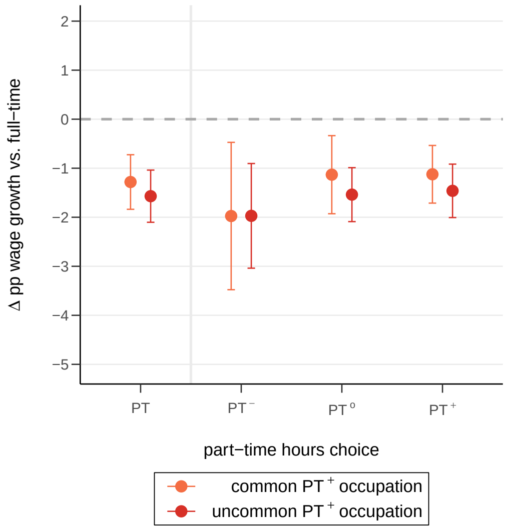
Notes: The classification 'demanding tasks' indicates that at least one-third of the tasks entailed in an individuals occupation can be classified as analytic non-routine. PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.13 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Using the measure of know-how requirements in Appendix-Figure C.6 yields very similar results to the just discussed task-based perspective. In terms of contract types,

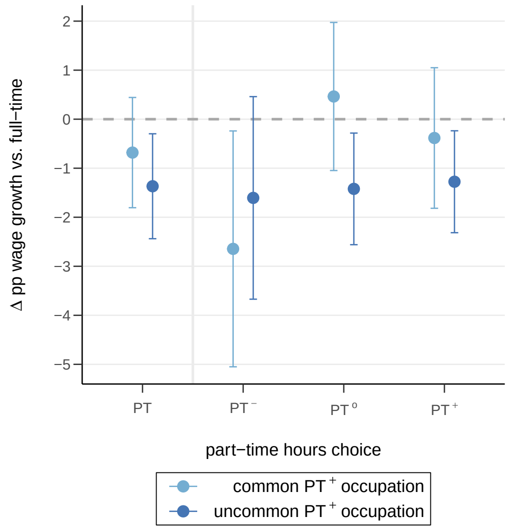
[30]These time-constant unobservables are eliminated by the inclusion of individual fixed effects in Figure 3.9b. Alternatively, the difference between both specifications may be driven by the following: As the fixed effects specification identifies the effects only from individuals who we observe working in different hours at different points in time, it also picks up on a behavioral component, namely the choice to switch working hours. The results may therefore indicate that those who do switch hours and work on demanding tasks differ from those who work on demanding tasks but are not observed switching their working hours.

i.e., differentiating between permanent and fixed-term contracts, we find only little heterogeneity across the hours bins in Appendix-Figure C.7.[31]

Finally, we investigate heterogeneity by the prevalence of PT^+ within an occupation in Figure 3.10.[32] There are no large differences in the wage growth penalties between occupations in which PT^+ is common or uncommon in Figure 3.10a. For the fixed-effects specifications in Figure 3.10b, however, we do find that occupations where PT^+ is common do not carry wage growth penalties for PT° and PT^+, but only for low part-time hours in PT^-. In contrast, occupations where PT^+ is uncommon display very comparable penalties of around 1 to 2 percentage points throughout the different hours choices.



(a) without individual FE



(b) with individual FE

Figure 3.10: Part-time penalties in wage growth by prevalence of PT^+

Notes: The classification 'common PT^+ occupation' indicates that the prevalence of PT^+ relative to PT° is above its median value across all occupations. PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.16 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

[31]The notable exception is part-time with very few hours (PT^-), where permanent contracts appear to carry larger wage growth penalties than fixed-term contracts. However, the large standard errors for these coefficients illustrate high uncertainty in the point estimates.

[32]See Appendix-Figure C.8 for very comparable results using the prevalence of PT^+ within an industry.

3.7 Potential Mechanisms

Given our findings of sizeable part-time penalties and substantial heterogeneity by part-time hours choices, we now discuss a number of plausible drivers of our results. These are differences in firms' cost functions, slower human capital accumulation, differences in task content, and how well workers' skills match job requirements. While our data does not allow us to sharply differentiate between all of these, they illustrate the economic mechanisms that are likely at play.

Differences in firms' cost functions. There are four cost dimensions for firms that can rationalize part-time penalties both in wage levels as well as in wage growth rates: i) recruitment and training cost, ii) capital cost, iii) coordination and communication cost, and iv) set-up cost.

To incorporate workers productively into the value chain of a firm, they need to be recruited and trained first. As both recruitment and training are costly for firms, they can be seen as investments into the workers. These investments then yield a return from the productivity of the workers. Given that recruitment and training costs are incurred per worker, full-time workers, who are productively engaged in the firm 100% of the time, can generate a higher return on the investment than part-time workers. This difference in return on the fixed recruitment and training investments is a plausible rationale for lower wages for part-time workers (Montgomery 1988).[33]

Turning to capital costs, a firm's output typically stems from the productive combination of labor and capital inputs, and most workers require some capital to work with. As part-time workers utilize their assigned capital less than full-time workers, they may generate a lower return on it, unless the capital can be shared among multiple part-time workers. This lower return can rationalize part-time penalties. However, if part-time work and capital use can be targeted to peak hours with high demand, while full-time workers have to be paid during periods with low or no demand as well, this may also give rise to part-time premiums (Manning and Petrongolo 2008).

Given that communication is costly, instructing multiple part-time workers, who each do a share of a given task, is more costly than instructing a single full-time worker, if the task cannot be split into sub-tasks that require separate instructions. Firms may pass

[33] Aside from part-time penalties, this mechanism can also contribute to occupational segregation: If similar hourly wages in part-time and full-time are enforced through, e.g., union contracts, then firms will prefer to hire full-time applicants for jobs that require high recruitment and training investments.

on these costs to workers via part-time penalties. Furthermore, Goldin (2014) underlines the role of costly coordination by showing that firms value workers with regular hours more than workers with irregular hours.

Finally, there may be set-up costs that have to be incurred before productive work can start. If part-time workers spend a larger share of their working time on these unproductive set-up costs, they have a lower hourly productivity and, therefore, may be paid lower wages (Barzel 1973).

All these four cost components are plausible drivers for the penalties on the uniform part-time indicator that we have documented in the previous sections. They rationalize why firms may pay lower wages to part-time workers and especially to workers with very low working hours (PT^-). However, the hump shape pattern of increasing penalties in wage levels from PT° to PT^+ appears counterintuitive, as the impact of the cost components laid out above should decrease with increasing working hours.

A plausible mechanism for these patterns could be that firms' part-time cost functions differ across occupations and industries by the prevalence of PT^+. It might be difficult to incorporate PT^+ workers into the workflows of occupations or industries where only very few individuals are working PT^+. For those employees that nevertheless work in PT^+, the high coordination costs are passed onto them via lower hourly wages. On the contrary, PT^+ workers can be incorporated much easier in jobs that, for example, consist mostly of tasks which are easily scalable by hours worked. Therefore, in these cases, PT^+ workers incur lower part-time penalties than PT° workers instead of higher ones. Through this lens, our measures for the prevalence of PT^+ could be interpreted as proxies for the scalability of tasks between PT° and PT^+.

Slower human capital accumulation. Focusing on wage growth rates, differences in human capital accumulation are an additional rationale for part-time penalties. If human capital accumulation (or retention) is a function of hours worked, then part-time work will imply slower accumulation or even depreciation. As firms directly benefit from the human capital level of their employees, who become more productive with higher human capital, they have a reason to tie wage growth rates to it. The existence and extent of this mechanism is shown, for example, in Blundell, Costa Dias, et al. (2016) and Adda, Dustmann, and Stevens (2017), who find substantially lower human capital accumulation and wage growth rates for part-time work.

Our findings in Section 3.6 are in line with this literature as they show that part-time work is associated with sizeable wage growth penalties. Part-time jobs with low working hours yield especially low wage growth rates, and we find a decreasing pattern of the penalties in hours worked, as we would expect if human capital accumulation is a function of hours worked. However, we also show that sizeable penalties still exist for many part-time jobs with high working hours, only slightly lower than for medium part-time hours PT°. Only for PT° and PT^{+} in occupations with demanding tasks or where PT^{+} is common, we find that the penalties disappear in some specifications.

This highlights that not all part-time jobs with high working hours can be regarded as close substitutes to full-time work. Human capital accumulation may be a nonlinear function of hours worked and there may also be hours-independent contributors to the wage growth penalties. Furthermore, the documented differences across job characteristics illustrate that human capital accumulation is not necessarily a uniform process across occupations.

Differences in task content. Another potential mechanism is linked to within occupation differences in task content. If part-time workers are more likely to carry out less demanding (and therefore less profitable) tasks than full-time workers, such a difference could rationalize part-time wage penalties. Evidence for this hypothesis is provided by Black and Spitz-Oener (2010). They document that a recent shift towards more non-routine analytical tasks and the accompanying wage increases are only observable for full-time workers (within the same occupation).

While our results differentiated by the task composition highlight that the part-time penalties in hourly wages are concentrated in non-demanding occupations, we cannot investigate the task composition at the within-occupation level. This is due to the fact that we do not have task classifications at the individual level, but rely on categorizing the occupational codes.

Skill mismatch. Finally, it may be the case that part-time workers are more likely to work in occupations for which they were not trained. If the entry into part-time is related to occupational downgrading (Connolly and Gregory 2010), then the acquired skills of workers may not match those required for the part-time occupation. This job-skill mismatch may drive lower wages for part-time workers. Our findings of large penalties for workers in non-demanding occupations are consistent with this mechanism, as these are likely the occupations that workers downgrade to.

3.8 Conclusion

In this paper, we build a unique long-run panel of hourly wages of German women through the linkage of social security data on earnings and survey data on hours. This high-quality data set allows us to conduct a novel investigation of the scope and heterogeneity of part-time penalties in wage levels and growth rates. By allowing for heterogeneity of these penalties across different part-time working hours choices, we are able to shed light on the effects of these choices on the career paths of women. To credibly account for selection into specific working hours, we follow a novel strategy proposed by Costa Dias, Joyce, and Parodi (2021). Specifically, we leverage variation in the incentive to work different hours induced by reforms to the tax and transfer system.

In terms of wage levels, we find significant selection-corrected part-time penalties and document substantial heterogeneity across different part-time hours choices. Working part-time with low hours (≤ 16h) implies the largest penalties, but we also find that working high part-time hours (> 24 to ≤ 34h) carries penalties that are higher than in medium part-time (> 16 to ≤ 24h). This is especially the case for occupations and industries where such high hours are an uncommon choice and points to heterogeneity in firms' cost functions for part-time work.

Regarding wage growth rates, we also find large and heterogeneous selection-corrected part-time penalties. They are again the largest for working low hours, but are decreasing non-linearly with increasing working hours. The fact that sizeable penalties remain present for high part-time hours suggests that even such part-time jobs are not close substitutes to full-time jobs.

Putting these findings into context, our results show that treating part-time as a uniform working hours choice, such as 20 hours per week, misses out on substantial heterogeneity. Extrapolating that our findings would hold for men as well, the scope of the penalties has important implications for the division of labor among spouses: Take a couple that decides to move from one partner working full-time and one partner working 50% to an equal division of labor with both partners working part-time for 75% of the week instead. This couple would incur a net income loss and move to a lower wage growth trajectory due to the sizeable penalties for working 75%. Part-time penalties are therefore creating substantial disincentives to split working hours equally among spouses, on top of similar disincentives imposed by joint taxation.

Consequently, if policymakers want to further encourage increases in mothers' labor supply to close the remaining gender gaps in labor market outcomes, it will likely not be enough to relax time constraints through, e.g., the provision of public childcare. Instead, part-time work itself also has to become more attractive.

Appendix C1 Additional Descriptives and Summary Statistics

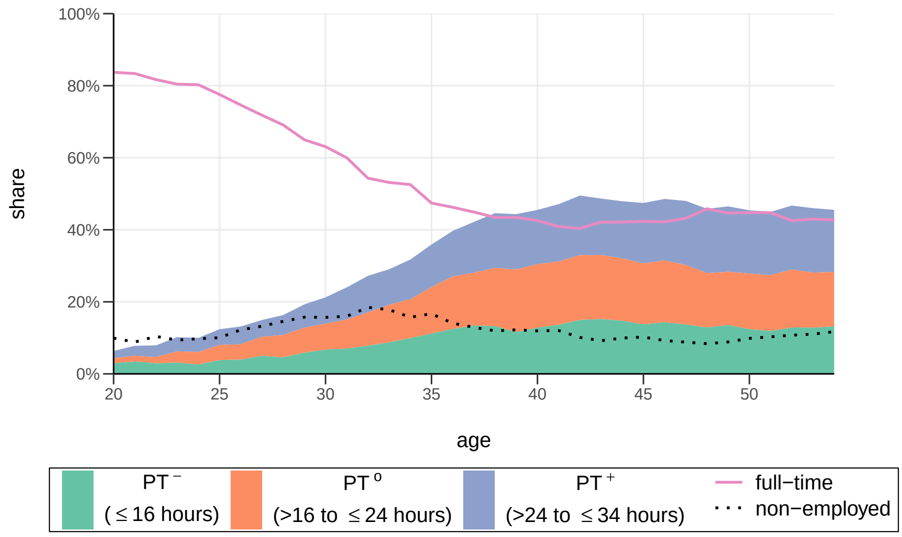


Figure C.1: Women's working hours choices across the lifecycle

Notes: Categorization based on contracted working hours per week, full-time corresponds to > 34 hours. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.1: Summary statistics

	matched indicator sample	unmatched
daily wage	75.29	59.41
contracted hours	32.52	35.77
hourly wage	€11.50	€9.27
wage growth	0.0283†	–
age	36.28	38.94
college	25.81%	24.11%
FT experience	7.28	6.92
PT experience	2.45	3.69
West Germany	84.38%	78.40%
Non-German nationality	6.68%	6.84%
tenure	5.03	5.87
large firm (≥ 200 FT employees)	35.23%	33.39%
family background PC 1	-0.0647	-0.1847
family background PC 2	0.0517	0.0817
married	53.78%	67.46%
mother	53.60%	74.43%
age of youngest child	3.79	5.35
age of oldest child	7.62	10.24
number of children 0 – 3	0.10	0.16
number of children 0 – 6	0.20	0.32
number of children 7 – 14	0.32	0.48
number of children 15 – 18	0.18	0.22
number of children 19 – 25	0.25	0.30
individuals	5,673	2,893
individuals × years	78,914	14,775

Notes: †: smaller subsample of individuals with repeated wage observations (cond. on repeated observations not changing the hours bracket), $N^{\dagger}$ = 64,299. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Appendix C2 Additional Illustrations of the Empirical Strategy

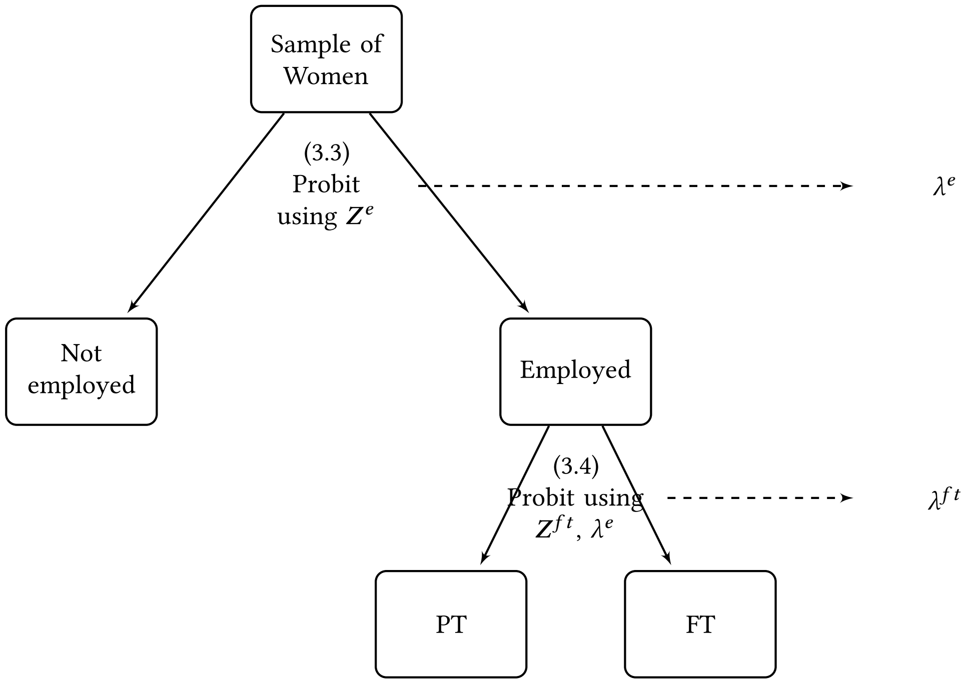


Figure C.2: Illustration of selection equations for uniform part-time vs. full-time

Notes: Own illustration of equations (3.3), and (3.4), λ^e and λ^{ft} denote the corresponding inverse Mills ratios obtained after estimating the respective Probit models.

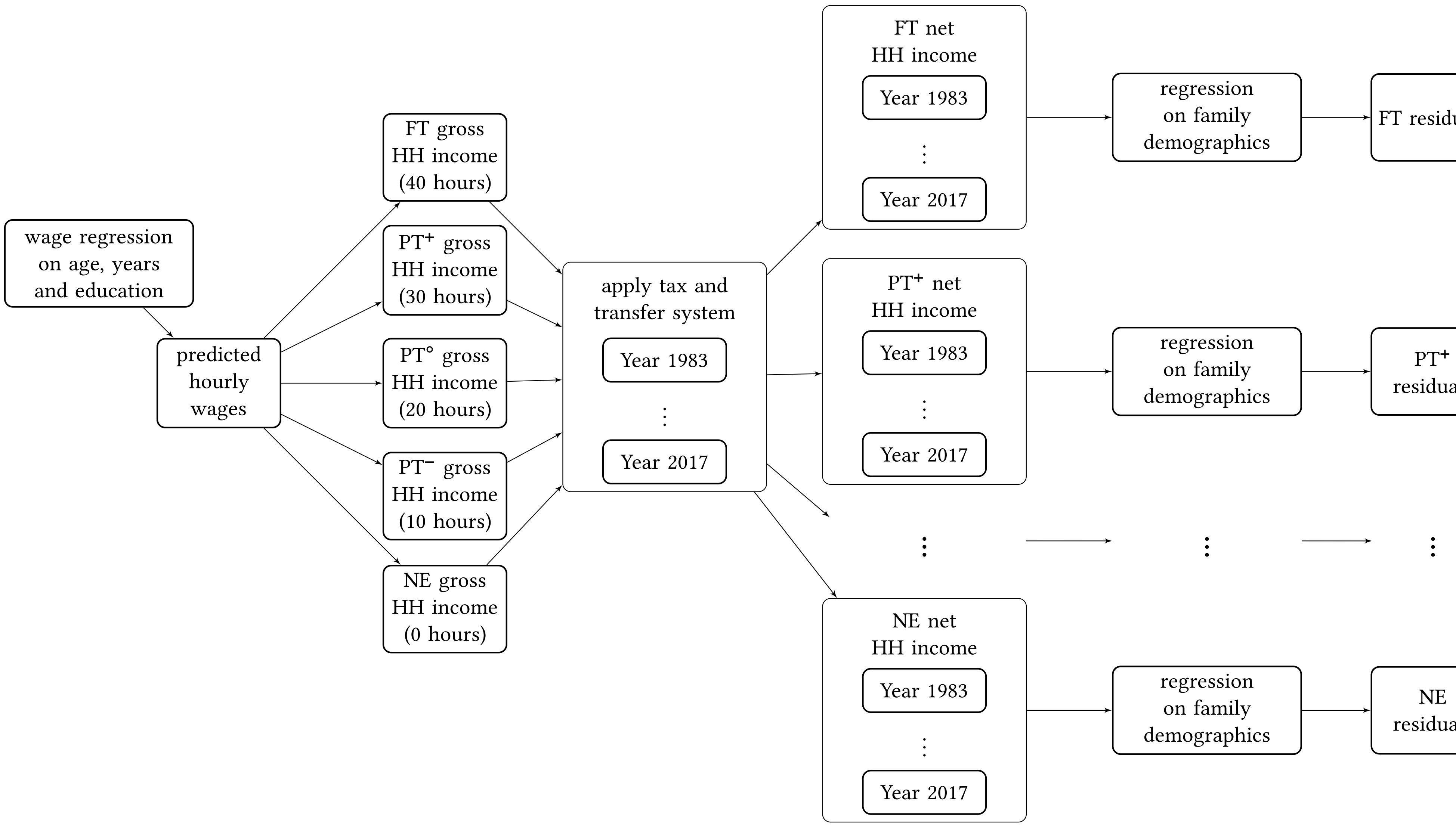


Figure C.3: Construction of the main exclusion restrictions

Notes: Own illustration of the different steps described in Section 3.4.1, constructing residuals that capture variation in the incentives to work different hours. The regressions on family demographics regress net income of the respective working hours choice on a constant, a maternity indicator and the number of children in the age brackets 0 – 6, 7 – 14, 15 – 18, and 19 – 25. This implementation follows Costa Dias, Joyce, and Parodi (2021).

Appendix C3 First Stage Results

Table C.2: First stage selection equations - employment and part-time

	employment	PT
	(1)	(2)
simulated net income not working	-0.0214^{***} (0.0065)	
simulated net income PT		0.1311^{***} (0.0179)
Δ simulated net income FT − PT		0.1793^{***} (0.0243)
mother	-0.2731^{***} (0.0620)	-0.5159^{***} (0.0870)
number of children aged 0 – 3	-0.6341^{***} (0.0326)	
number of children aged 0 – 6		-0.5256^{***} (0.0476)
number of children aged 7 – 14		-0.3775^{***} (0.0364)
number of children aged 15 – 18		-0.1209^{***} (0.0434)
number of children aged 19 – 25		0.0065 (0.0516)
age of the youngest child		-0.0134^{***} (0.0027)
age of the oldest child		−0.0025 (0.0052)
χ^2-test on income excl. restrictions	10.78	74.60
p-value	0.001	0.000
χ^2-test on all excl. restrictions	795.99	665.35
p-value	0.000	0.000
N	89,894	78,896

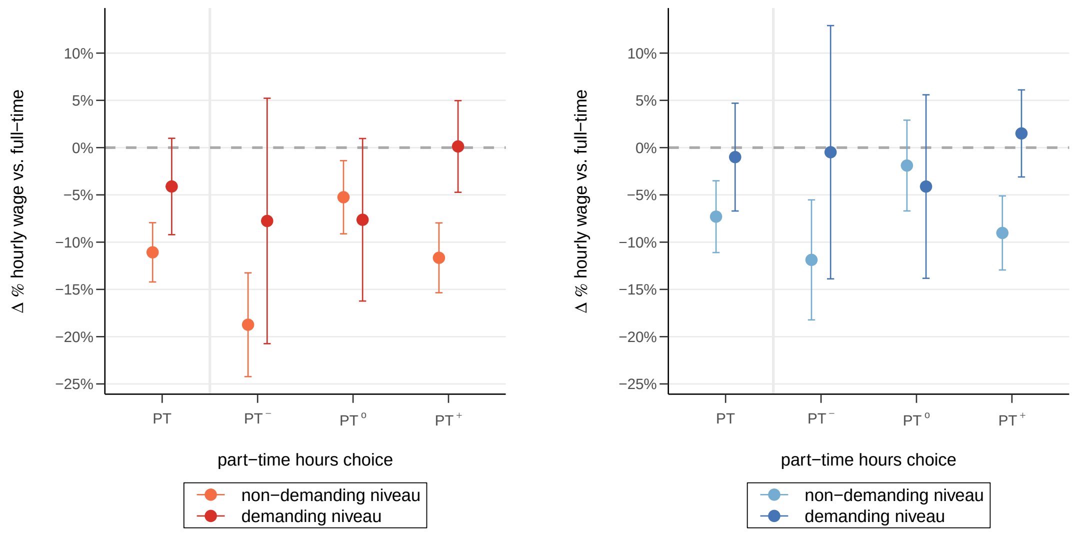
Notes: Probit models with dependent variables as listed in column headers, employment is a dummy variable for working. PT denotes working ≤ 34 hours per week, and FT denotes working > 34 hours. Simulated net incomes (residualised) as described in Section 3.4.1. Robust standard errors clustered at the individual level in parentheses. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.3: First stage selection equations - multiple part-time hours choices

	$\mathbb{1}\{PT^- \textit{ or } PT^\circ\}$	PT^-	PT^+
	(1)	(2)	(3)
simulated net income	0.1082^{***}		
mean(PT^-, PT°)	(0.0209)		
Δ simulated net income	0.1961^{***}		
mean(PT^+, FT) − mean(PT^-, PT°)	(0.0289)		
simulated net income		0.0238	
PT^-		(0.0321)	
Δ simulated net income		0.0165	
$PT^\circ - PT^-$		(0.0403)	
simulated net income			0.1185^{**}
PT^+			(0.0551)
Δ simulated net income			0.1321^{*}
FT − PT^+			(0.0768)
mother	-0.5293^{***}	0.3120^{*}	-0.7633^{***}
	(0.0687)	(0.1692)	(0.2617)
number of children aged 0 – 6	-0.4971^{***}	−0.1382	-0.4725^{***}
	(0.0442)	(0.1108)	(0.1831)
number of children aged 7 – 14	-0.3521^{***}	−0.0851	-0.3656^{***}
	(0.0336)	(0.0974)	(0.1275)
number of children aged 15 – 18	-0.1687^{***}	−0.0561	−0.0991
	(0.0374)	(0.0864)	(0.0724)
number of children aged 19 – 25	0.0032	-0.1545^{*}	−0.0015
	(0.0451)	(0.0878)	(0.0572)
age of the youngest child		−0.0134	-0.0130^{***}
		(0.0027)	(0.0034)
age of the oldest child		−0.0025	0.0024
		(0.0052)	(0.0059)
χ^2-test on income excl. restrictions	51.77	0.63	4.77
p-value	0.000	0.730	0.092
χ^2-test on all excl. restrictions	560.22	18.29	29.43
p-value	0.000	0.032	0.001
N	78,896	18,551	60,363

Notes: Probit models with dependent variables as listed in column headers, where PT^- denotes working ≤ 16 hours per week, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours, and FT denotes working > 34 hours. Simulated net incomes (residualised) as described in Section 3.4.1. Robust standard errors clustered at the individual level in parentheses. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

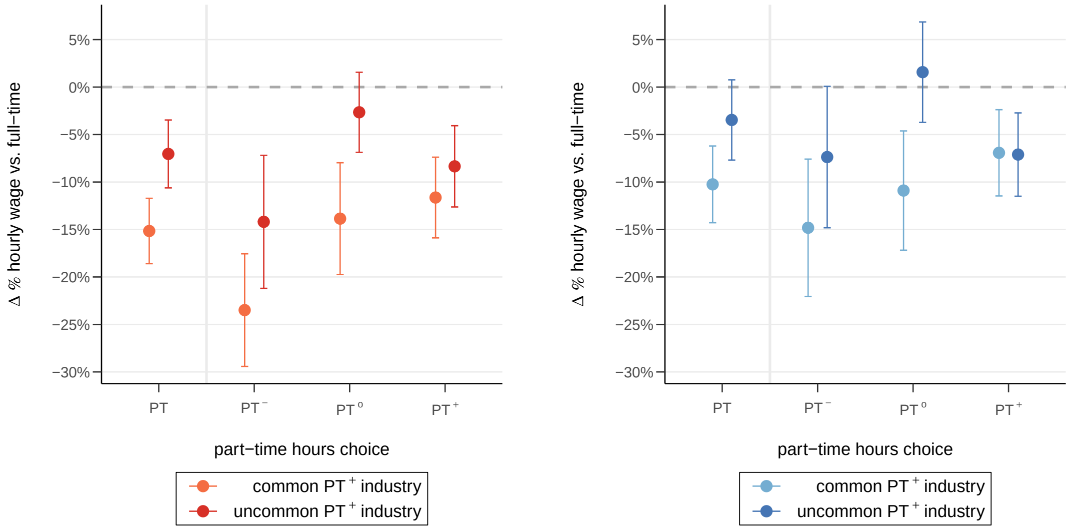
Appendix C4 Additional Results on Effect Heterogeneity



(a) without individual FE (b) with individual FE

Figure C.4: Part-time penalties in wage levels by know-how requirement

Notes: The classification 'demanding know-how' indicates that a specialist or expert skill level is necessary for a job. PT denotes working $\leq$ 34 hours per week, PT^- denotes working $\leq$ 16 hours, PT° denotes working $>$ 16 to $\leq$ 24 hours, PT^+ denotes working $>$ 24 to $\leq$ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.9 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

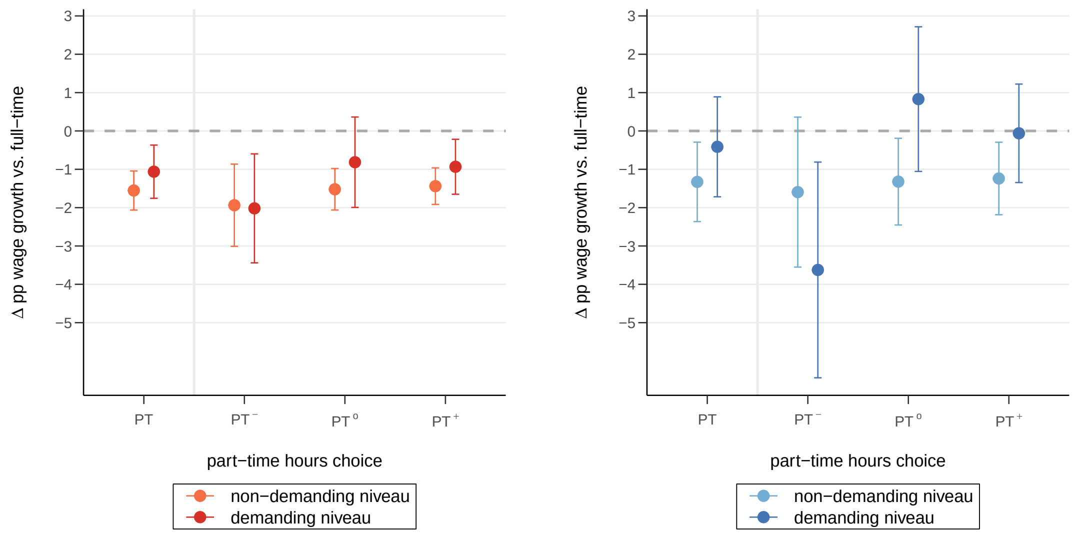


(a) without individual FE

(b) with individual FE

Figure C.5: Part-time penalties in wage levels by prevalence of PT$^+$ (industry based)

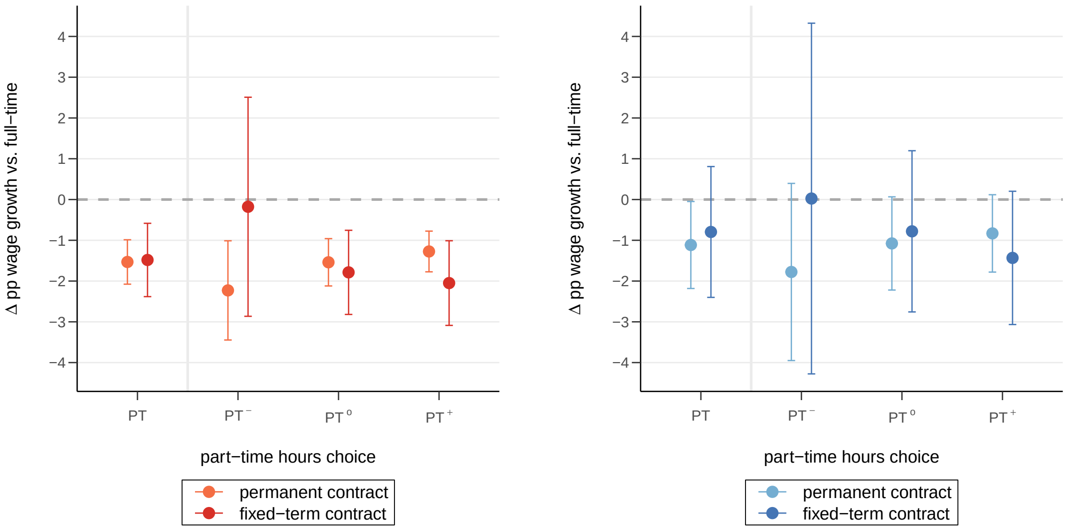
Notes: The classification 'common PT$^+$ industry' indicates that the prevalence of PT$^+$ relative to PT$^\circ$ is above its median value across all industries. PT denotes working ≤ 34 hours per week, PT$^-$ denotes working ≤ 16 hours, PT$^\circ$ denotes working > 16 to ≤ 24 hours, PT$^+$ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.12 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).



(a) without individual FE (b) with individual FE

Figure C.6: Part-time penalties in wage growth by know-how requirement

Notes: The classification 'demanding know-how' indicates that a specialist or expert skill level is necessary for a job. PT denotes working ≤ 34 hours per week, PT$^-$ denotes working ≤ 16 hours, PT$^\circ$ denotes working > 16 to ≤ 24 hours, PT$^+$ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.14 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

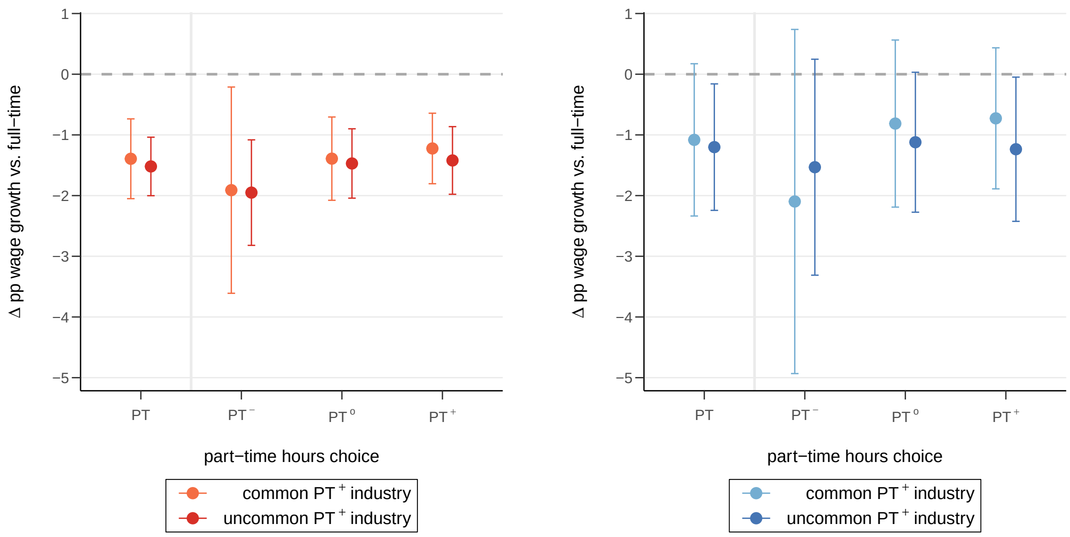


(a) without individual FE

(b) with individual FE

Figure C.7: Part-time penalties in wage growth by contract type

Notes: PT denotes working ≤ 34 hours per week, PT⁻ denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT⁺ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.15 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).



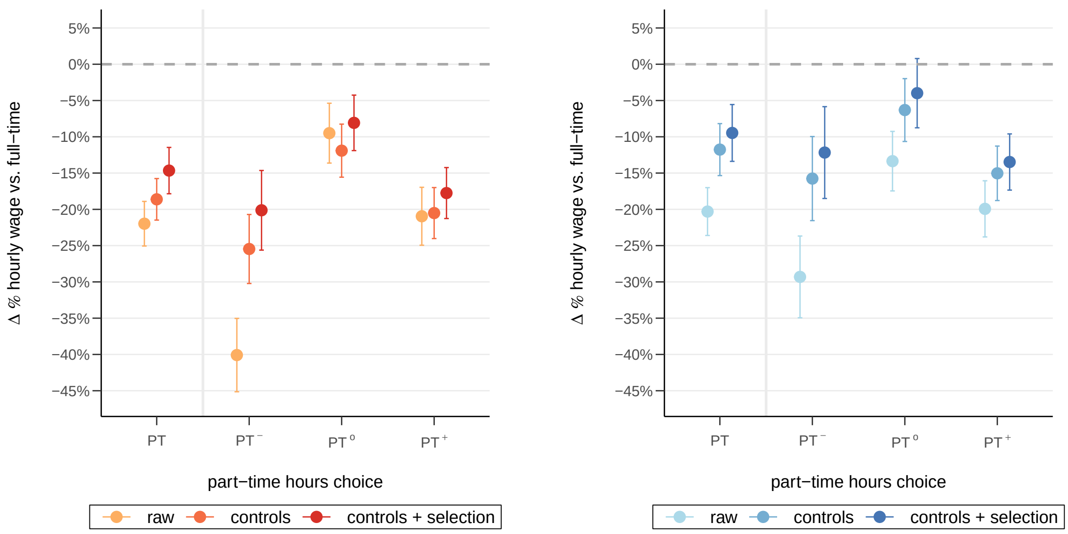
(a) without individual FE

(b) with individual FE

Figure C.8: Part-time penalties in wage growth by prevalence of PT$^+$ (industry-based)

Notes: The classification 'common PT$^+$ industry' indicates that the prevalence of PT$^+$ relative to PT$^\circ$ is above its median value across all industries. PT denotes working ≤ 34 hours per week, PT$^-$ denotes working ≤ 16 hours, PT$^\circ$ denotes working > 16 to ≤ 24 hours, PT$^+$ denotes working > 24 to ≤ 34 hours. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using standard errors clustered at the individual level, based on 1,000 bootstrap replications. See Table C.17 for underlying values. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Appendix C5 Additional Results with Alternative Sample

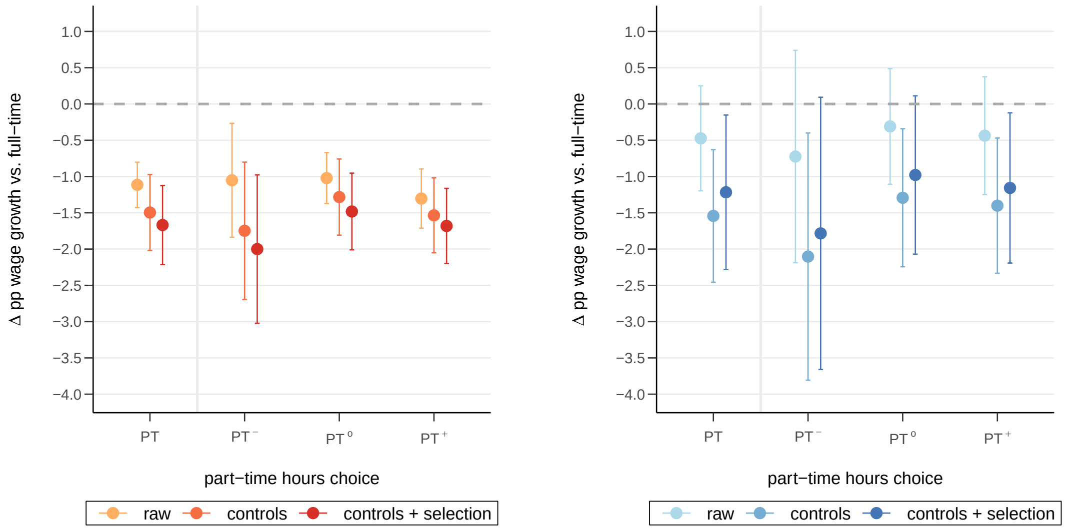


(a) without individual FE

(b) with individual FE

Figure C.9: Part-time penalties in wage levels (alternative sample)

Notes: PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. *raw* is the part-time penalty only controlling for year fixed effects, *controls* includes a broad set of individual controls incl. occupation, industry, and year fixed effects and *controls + selection* adds control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. Sample: women aged 20 to 54, not in education, strict match part-time/full-time status in NEPS and IEB. Years 1975 – 2017. Source: FDZ-IAB (2020).



(a) without individual FE

(b) with individual FE

Figure C.10: Part-time penalties in wage growth (alternative sample)

Notes: PT denotes working ≤ 34 hours per week, PT^- denotes working ≤ 16 hours, PT° denotes working > 16 to ≤ 24 hours, PT^+ denotes working > 24 to ≤ 34 hours. *raw* is the part-time penalty only controlling for year fixed effects, *controls* includes a broad set of individual controls incl. occupation, industry, and year fixed effects and *controls + selection* adds control functions for the selection into working specific hours, as described Section 3.4. Bars show the 95% confidence intervals using robust standard errors clustered at the individual level, based on 1,000 bootstrap replications. Sample: women aged 20 to 54, not in education, strict match part-time/full-time status in NEPS and IEB. Years 1975 – 2017. Source: FDZ-IAB (2020).

Appendix C6 Results Tables

Table C.4: Part-time penalties in wage levels

	dep. variable: log hourly wage $log(w_{it})$ – mean in FT = 11.83 –					
	(1)	(2)	(3)	(4)	(5)	(6)
PT (≤ 34h)	-0.1969^{***}		-0.1432^{***}		-0.1038^{***}	
	(0.0150)		(0.0137)		(0.0146)	
PT^- (≤ 16h)		-0.3984^{***}		-0.2338^{***}		-0.1797^{***}
		(0.0258)		(0.0243)		(0.0271)
PT° (> 16 to ≤ 24h)		-0.0939^{***}		-0.0990^{***}		-0.0603^{***}
		(0.0210)		(0.0185)		(0.0189)
PT^+ (> 24 to ≤ 34h)		-0.1513^{***}		-0.1240^{***}		-0.0976^{***}
		(0.0188)		(0.0163)		(0.0165)
controls			✓	✓	✓	✓
selection correction					✓	✓
indiv. FE						
F-Test control fun.					55.70	43.71
p-value					0.000	0.000
N	78,896	78,896	78,896	78,896	78,896	78,896

Notes: Categorization based on contracted working hours per week. All models include year fixed effects, controls includes a broad set of individual controls incl. occupation, and industry fixed effects. Selection correction via the inclusion of control functions for the selection into working and the hours choice, see Section 3.4 for details on the regression models. F-Test of control functions is a joint test of the inverse Mills ratios included to control for selection. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.5: Part-time penalties in wage levels (fixed effects specifications)

	dep. variable: log hourly wage $log(w_{it})$ – mean in FT = 11.83 –					
	(1)	(2)	(3)	(4)	(5)	(6)
PT ($\leq$ 34h)	-0.1729^{***}		-0.0880^{***}		-0.0632^{***}	
	(0.0158)		(0.0167)		(0.0181)	
PT^- ($\leq$ 16h)		-0.2832^{***}		-0.1442^{***}		-0.1056^{***}
		(0.0287)		(0.0293)		(0.0313)
PT° (> 16 to $\leq$ 24h)		-0.1221^{***}		-0.0485^{**}		-0.0226
		(0.0208)		(0.0218)		(0.0235)
PT^+ (> 24 to $\leq$ 34h)		-0.1354^{***}		-0.0873^{***}		-0.0703^{***}
		(0.0174)		(0.0166)		(0.0183)
controls			✓	✓	✓	✓
selection correction					✓	✓
indiv. FE	✓	✓	✓	✓	✓	✓
F-Test control fun.					39.59	37.58
p-value					0.000	0.000
N	78,896	78,896	78,896	78,896	78,896	78,896

Notes: Categorization based on contracted working hours per week. All models include year fixed effects, controls includes a broad set of individual controls incl. occupation, and industry fixed effects. Selection correction via the inclusion of control functions for the selection into working and the hours choice, see Section 3.4 for details on the regression models. *F*-Test of control functions is a joint test of the inverse Mills ratios included to control for selection. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.6: Part-time penalties in wage growth

	dep. variable: wage growth $\Delta log(w_{it})$ – mean in FT = 0.0330 –					
	(1)	(2)	(3)	(4)	(5)	(6)
PT ($\leq$ 34h)	-0.0092^{***}		-0.0127^{***}		-0.0144^{***}	
	(0.0016)		(0.0025)		(0.0024)	
PT^- ($\leq$ 16h)		-0.0093^{**}		-0.0162^{***}		-0.0191^{***}
		(0.0044)		(0.0051)		(0.0051)
PT° (> 16 to $\leq$ 24h)		-0.0085^{***}		-0.0120^{***}		-0.0140^{***}
		(0.0018)		(0.0027)		(0.0028)
PT^+ (> 24 to $\leq$ 34h)		-0.0100^{***}		-0.0115^{***}		-0.0131^{***}
		(0.0018)		(0.0022)		(0.0022)
controls			✓	✓	✓	✓
selection correction					✓	✓
indiv. FE						
F-Test control fun.					9.21	10.29
p-value					0.027	0.067
N	64,299	64,299	64,299	64,299	64,299	64,299

Notes: Categorization based on contracted working hours per week. All models include year fixed effects, controls includes a broad set of individual controls incl. occupation, and industry fixed effects. Selection correction via the inclusion of control functions for the selection into working and the hours choice, see Section 3.4 for details on the regression models. F-Test of control functions is a joint test of the inverse Mills ratios included to control for selection. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.7: Part-time penalties in wage growth (fixed effects specifications)

	dep. variable: wage growth $\Delta log(w_{it})$ – mean in FT = 0.0330 –					
	(1)	(2)	(3)	(4)	(5)	(6)
PT ($\leq$ 34h)	-0.0031		-0.0107^{**}		-0.0097^{*}	
	(0.0036)		(0.0045)		(0.0050)	
PT^- ($\leq$ 16h)		-0.0064		-0.0173^{*}		-0.0170^{*}
		(0.0084)		(0.0097)		(0.0101)
PT° ($>$ 16 to $\leq$ 24h)		-0.0027		-0.0105^{**}		-0.0094^{*}
		(0.0040)		(0.0047)		(0.0056)
PT^+ ($>$ 24 to $\leq$ 34h)		-0.0014		-0.0074^{*}		-0.0067
		(0.0036)		(0.0042)		(0.0043)
controls			✓	✓	✓	✓
selection correction					✓	✓
indiv. FE	✓	✓	✓	✓	✓	✓
F-Test control fun.					6.22	12.02
p-value					0.102	0.035
N	64,299	64,299	64,299	64,299	64,299	64,299

Notes: Categorization based on contracted working hours per week. All models include year fixed effects, controls includes a broad set of individual controls incl. occupation, and industry fixed effects. Selection correction via the inclusion of control functions for the selection into working and the hours choice, see Section 3.4 for details on the regression models. F-Test of control functions is a joint test of the inverse Mills ratios included to control for selection. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.8: Part-time penalties in wage levels by task composition

	dep. variable: log hourly wage $log(w_{it})$			
	(1)	(2)	(3)	(4)
non-demanding tasks ×	– mean in FT = 10.94 –			
PT (≤ 34h)	-0.1175^{***}		-0.0794^{***}	
	(0.0160)		(0.0191)	
PT^{-} (≤ 16h)		-0.1874^{***}		-0.1212^{***}
		(0.0278)		(0.0322)
PT° (> 16 to ≤ 24h)		-0.0631^{***}		-0.0260
		(0.0206)		(0.0237)
PT^{+} (> 24 to ≤ 34h)		-0.1201^{***}		-0.0972^{***}
		(0.0194)		(0.0199)
demanding tasks ×	– mean in FT = 14.51 –			
PT (≤ 34h)	-0.0265		0.0073	
	(0.0226)		(0.0261)	
PT^{-} (≤ 16h)		-0.0750		0.0016
		(0.0580)		(0.0577)
PT° (> 16 to ≤ 24h)		-0.0285		0.0030
		(0.0333)		(0.0409)
PT^{+} (> 24 to ≤ 34h)		-0.0046		0.0185
		(0.0222)		(0.0229)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	78,896	78,896	78,896	78,896

Notes: The classification 'demanding tasks' indicates that at least one-third of the tasks entailed in an individuals occupation can be classified as analytic non-routine. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.9: Part-time penalties in wage levels by know-how requirement

	dep. variable: log hourly wage $log(w_{it})$			
	(1)	(2)	(3)	(4)
non-demanding niveau ×	– mean in FT = 11.16 –			
PT (≤ 34h)	-0.1108^{***}		-0.0730^{***}	
	(0.0160)		(0.0194)	
PT$^-$ (≤ 16h)		-0.1874^{***}		-0.1188^{***}
		(0.0280)		(0.0324)
PT$^\circ$ (> 16 to ≤ 24h)		-0.0525^{***}		−0.0190
		(0.0197)		(0.0245)
PT$^+$ (> 24 to ≤ 34h)		-0.1166^{***}		-0.0903^{***}
		(0.0188)		(0.0200)
demanding niveau ×	– mean in FT = 14.97 –			
PT (≤ 34h)	−0.0411		−0.0100	
	(0.0260)		(0.0291)	
PT$^-$ (≤ 16h)		−0.0776		−0.0049
		(0.0662)		(0.0683)
PT$^\circ$ (> 16 to ≤ 24h)		-0.0763^{*}		−0.0412
		(0.0439)		(0.0495)
PT$^+$ (> 24 to ≤ 34h)		0.0012		0.0150
		(0.0247)		(0.0235)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	78,896	78,896	78,896	78,896

Notes: The classification 'demanding know-how' indicates that a specialist or expert skill level is necessary for a job. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.10: Part-time penalties in wage levels by contract type

	dep. variable: log hourly wage $log(w_{it})$			
	(1)	(2)	(3)	(4)
permanent contract ×	– mean in FT = 11.90 –			
PT (≤ 34h)	−0.1033***		−0.0606***	
	(0.0161)		(0.0196)	
PT⁻ (≤ 16h)		−0.1678***		−0.0745**
		(0.0289)		(0.0341)
PT° (> 16 to ≤ 24h)		−0.0683***		−0.0399
		(0.0209)		(0.0253)
PT⁺ (> 24 to ≤ 34h)		−0.1001***		−0.0706***
		(0.0180)		(0.0202)
fixed-term contract ×	– mean in FT = 11.06 –			
PT (≤ 34h)	−0.0802***		−0.0563**	
	(0.0258)		(0.0265)	
PT⁻ (≤ 16h)		−0.2396***		−0.1665***
		(0.0531)		(0.0475)
PT° (> 16 to ≤ 24h)		−0.0037		0.0287
		(0.0348)		(0.0330)
PT⁺ (> 24 to ≤ 34h)		−0.0575*		−0.0632*
		(0.0322)		(0.0334)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	78,896	78,896	78,896	78,896

Notes: Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.11: Part-time penalties in wage levels by prevalence of PT^+

	dep. variable: log hourly wage $log(w_{it})$			
	(1)	(2)	(3)	(4)
common PT^+ occupation ×	– mean in FT = 12.21 –			
PT (≤ 34h)	-0.0643^{***}		-0.0283	
	(0.0200)		(0.0232)	
PT^- (≤ 16h)		-0.1080^{**}		-0.0529
		(0.0444)		(0.0512)
PT° (> 16 to ≤ 24h)		-0.0743^{**}		-0.0557^{*}
		(0.0312)		(0.0317)
PT^+ (> 24 to ≤ 34h)		-0.0430^{*}		-0.0045
		(0.0227)		(0.0238)
uncommon PT^+ occupation ×	– mean in FT = 11.68 –			
PT (≤ 34h)	-0.1159^{***}		-0.0717^{***}	
	(0.0163)		(0.0196)	
PT^- (≤ 16h)		-0.1989^{***}		-0.1186^{***}
		(0.0293)		(0.0332)
PT° (> 16 to ≤ 24h)		-0.0552^{***}		-0.0129
		(0.0205)		(0.0256)
PT^+ (> 24 to ≤ 34h)		-0.1258^{***}		-0.1009^{***}
		(0.0197)		(0.0214)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	78,896	78,896	78,896	78,896

Notes: The classification 'common PT^+ occupation' indicates that the prevalence of PT^+ relative to PT° is above its median value across all occupations. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.12: Part-time penalties in wage levels by prevalence of PT$^+$ (industry-based)

	dep. variable: log hourly wage $log(w_{it})$			
	(1)	(2)	(3)	(4)
common PT$^+$ industry ×	– mean in FT = 11.55 –			
PT (≤ 34h)	-0.1516^{***}		-0.1025^{***}	
	(0.0176)		(0.0206)	
PT$^-$ (≤ 16h)		-0.2349^{***}		-0.1482^{***}
		(0.0302)		(0.0369)
PT$^\circ$ (> 16 to ≤ 24h)		-0.1386^{***}		-0.1090^{***}
		(0.0300)		(0.0320)
PT$^+$ (> 24 to ≤ 34h)		-0.1163^{***}		-0.0693^{***}
		(0.0217)		(0.0232)
uncommon PT$^+$ industry ×	– mean in FT = 12.00 –			
PT (≤ 34h)	-0.0704^{***}		-0.0346	
	(0.0182)		(0.0216)	
PT$^-$ (≤ 16h)		-0.1419^{***}		-0.0737^{*}
		(0.0357)		(0.0380)
PT$^\circ$ (> 16 to ≤ 24h)		-0.0266		0.0157
		(0.0215)		(0.0270)
PT$^+$ (> 24 to ≤ 34h)		-0.0835^{***}		-0.0711^{***}
		(0.0218)		(0.0224)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	78,896	78,896	78,896	78,896

Notes: The classification 'common PT$^+$ industry' indicates that the prevalence of PT$^+$ relative to PT$^\circ$ is above its median value across all industries. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.13: Part-time penalties in wage growth by task composition

	dep. variable: wage growth $\Delta log(w_{it})$			
	(1)	(2)	(3)	(4)
non-demanding tasks ×	– mean in FT = 0.0335 –			
PT (≤ 34h)	-0.0159^{***}		-0.0153^{***}	
	(0.0027)		(0.0055)	
PT^{-} (≤ 16h)		-0.0200^{***}		-0.0178^{*}
		(0.0055)		(0.0105)
PT° (> 16 to ≤ 24h)		-0.0151^{***}		-0.0145^{**}
		(0.0029)		(0.0060)
PT^{+} (> 24 to ≤ 34h)		-0.0149^{***}		-0.0149^{***}
		(0.0025)		(0.0047)
demanding tasks ×	– mean in FT = 0.0316 –			
PT (≤ 34h)	-0.0106^{***}		-0.0009	
	(0.0030)		(0.0053)	
PT^{-} (≤ 16h)		-0.0157^{**}		-0.0216^{*}
		(0.0075)		(0.0119)
PT° (> 16 to ≤ 24h)		-0.0113^{**}		0.0028
		(0.0044)		(0.0077)
PT^{+} (> 24 to ≤ 34h)		-0.0087^{***}		0.0036
		(0.0030)		(0.0067)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	64,299	64,299	64,299	64,299

Notes: The classification 'demanding tasks' indicates that at least one-third of the tasks entailed in an individuals occupation can be classified as analytic non-routine. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.14: Part-time penalties in wage growth by know-how requirement

	dep. variable: wage growth $\Delta log(w_{it})$			
	(1)	(2)	(3)	(4)
non-demanding niveau ×	– mean in FT = 0.0336 –			
PT (≤ 34h)	−0.0155***		−0.0133**	
	(0.0026)		(0.0053)	
PT$^-$ (≤ 16h)		−0.0194***		−0.0159
		(0.0055)		(0.0100)
PT$^\circ$ (> 16 to ≤ 24h)		−0.0152***		−0.0132**
		(0.0028)		(0.0058)
PT$^+$ (> 24 to ≤ 34h)		−0.0144***		−0.0124**
		(0.0024)		(0.0048)
demanding niveau ×	– mean in FT = 0.0302 –			
PT (≤ 34h)	−0.0106***		−0.0041	
	(0.0035)		(0.0066)	
PT$^-$ (≤ 16h)		−0.0202***		−0.0362**
		(0.0072)		(0.0144)
PT$^\circ$ (> 16 to ≤ 24h)		−0.0082		0.0083
		(0.0060)		(0.0096)
PT$^+$ (> 24 to ≤ 34h)		−0.0093**		−0.0006
		(0.0037)		(0.0066)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	64,299	64,299	64,299	64,299

Notes: The classification 'demanding know-how' indicates that a specialist or expert skill level is necessary for a job. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.15: Part-time penalties in wage growth by contract type

	dep. variable: wage growth $\Delta log(w_{it})$			
	(1)	(2)	(3)	(4)
permanent contract ×	– mean in FT = 0.0329 –			
PT (≤ 34h)	−0.0153***		−0.0112**	
	(0.0028)		(0.0054)	
PT⁻ (≤ 16h)		−0.0223***		−0.0178
		(0.0062)		(0.0111)
PT° (> 16 to ≤ 24h)		−0.0154***		−0.0108*
		(0.0030)		(0.0058)
PT⁺ (> 24 to ≤ 34h)		−0.0127***		−0.0083*
		(0.0025)		(0.0048)
fixed-term contract ×	– mean in FT = 0.0363 –			
PT (≤ 34h)	−0.0148***		−0.0080	
	(0.0046)		(0.0082)	
PT⁻ (≤ 16h)		−0.0018		0.0002
		(0.0137)		(0.0219)
PT° (> 16 to ≤ 24h)		−0.0179***		−0.0078
		(0.0053)		(0.0101)
PT⁺ (> 24 to ≤ 34h)		−0.0205***		−0.0143*
		(0.0053)		(0.0083)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	64,299	64,299	64,299	64,299

Notes: Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.16: Part-time penalties in wage growth by prevalence of PT$^+$

	dep. variable: wage growth $\Delta log(w_{it})$			
	(1)	(2)	(3)	(4)
common PT$^+$ occupation ×	– mean in FT = 0.0301 –			
PT (≤ 34h)	−0.0128***		−0.0068	
	(0.0028)		(0.0057)	
PT$^-$ (≤ 16h)		−0.0198**		−0.0264**
		(0.0077)		(0.0123)
PT$^\circ$ (> 16 to ≤ 24h)		−0.0113***		0.0046
		(0.0041)		(0.0077)
PT$^+$ (> 24 to ≤ 34h)		−0.0112***		−0.0038
		(0.0030)		(0.0073)
uncommon PT$^+$ occupation ×	– mean in FT = 0.0342 –			
PT (≤ 34h)	−0.0157***		−0.0137**	
	(0.0027)		(0.0055)	
PT$^-$ (≤ 16h)		−0.0197***		−0.0161
		(0.0054)		(0.0105)
PT$^\circ$ (> 16 to ≤ 24h)		−0.0154***		−0.0142**
		(0.0028)		(0.0058)
PT$^+$ (> 24 to ≤ 34h)		−0.0146***		−0.0128**
		(0.0028)		(0.0053)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	64,299	64,299	64,299	64,299

Notes: The classification 'common PT$^+$ occupation' indicates that the prevalence of PT$^+$ relative to PT$^\circ$ is above its median value across all occupations. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Table C.17: Part-time penalties in wage growth by prevalence of PT^+ (industry-based)

	dep. variable: wage growth $\Delta log(w_{it})$			
	(1)	(2)	(3)	(4)
common PT^+ industry ×	– mean in FT = 0.0311 –			
PT (≤ 34h)	−0.0139***		−0.0108*	
	(0.0034)		(0.0064)	
PT^- (≤ 16h)		−0.0191**		−0.0210
		(0.0087)		(0.0145)
PT° (> 16 to ≤ 24h)		−0.0139***		−0.0081
		(0.0035)		(0.0070)
PT^+ (> 24 to ≤ 34h)		−0.0122***		−0.0073
		(0.0030)		(0.0059)
uncommon PT^+ industry ×	– mean in FT = 0.0341 –			
PT (≤ 34h)	−0.0152***		−0.0120**	
	(0.0025)		(0.0053)	
PT^- (≤ 16h)		−0.0195***		−0.0153*
		(0.0044)		(0.0091)
PT° (> 16 to ≤ 24h)		−0.0147***		−0.0112*
		(0.0029)		(0.0059)
PT^+ (> 24 to ≤ 34h)		−0.0142***		−0.0124**
		(0.0028)		(0.0061)
controls	✓	✓	✓	✓
selection correction	✓	✓	✓	✓
indiv. FE			✓	✓
N	64,299	64,299	64,299	64,299

Notes: The classification 'common PT^+ industry' indicates that the prevalence of PT^+ relative to PT° is above its median value across all industries. Categorization based on contracted working hours per week. All specifications include a broad set of individual controls incl. occupation, industry, and year fixed effects, as well as control functions for the selection into working specific hours, as described Section 3.4. Robust standard errors clustered at the individual level in parentheses, based on 1,000 bootstrap replications. Significance at the 10%, 5%, and 1% level denoted by *, ** and ***. Sample: women aged 20 to 54, not in education. Years 1975 – 2017. Source: FDZ-IAB (2020).

Bibliography

Aaronson, D. and E. French (2004). “The Effect of Part-Time Work on Wages: Evidence from the Social Security Rules”. *Journal of Labor Economics* 22.2, 329–252.

Adda, J., C. Dustmann, and K. Stevens (2017). “The Career Costs of Children”. *Journal of Political Economy* 125.2, 293–337.

Akerman, A., I. Gaarder, and M. Mogstad (2015). “The Skill Complementarity of Broadband Internet”. *The Quarterly Journal of Economics* 130.4, 1781–1824.

Altonji, J. G. and R. M. Blank (1999). “Chapter 48: Race and gender in the labor market”. *Handbook of Labor Economics.* Vol. 3. C. Elsevier, 3143–3259.

Angelov, N., P. Johansson, and E. Lindahl (2016). “Parenthood and the Gender Gap in Pay”. *Journal of Labor Economics* 34.3, 545–579.

Antoni, M., A. Schmucker, S. Seth, and P. vom Berge (2019). “Sample of Integrated Labour Market Biographies (SIAB) 1975 - 2017”. FDZ-Datenreport 02/2019. Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Authoring Group Educational Reporting (2018). *Education in Germany 2018.*

Bachbauer, N. and C. Wolf (2020). “NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB 7518)”. FDZ-Datenreport 04/2020. Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Bagger, J., R. V. Lesner, and R. Vejlin (2019). “The Gender Wage Gap: The Roles of Job Search, Productivity, Parental Leave, and Experience Accumulation”.

Baker, M., J. Gruber, and K. Milligan (2008). “Universal Child Care, Maternal Labor Supply, and Family Well-Being”. *Journal of Political Economy* 116.4, 709–745.

Baker, M., J. Gruber, and K. Milligan (2019). “The Long-Run Impacts of a Universal Child Care Program”. *American Economic Journal: Economic Policy* 11.3, 1–26.

Bardasi, E. and J. C. Gornick (2008). "Working for less? Women's part-time wage penalties across countries". *Feminist Economics* 14.1, 37–72.

Bargain, O., K. Orsini, and A. Peichl (2014). "Comparing Labor Supply Elasticities in Europe and the United States New Results". *Journal of Human Resources* 49.3, 723–838.

Barzel, Y. (1973). "The Determination of Daily Hours and Wages". *The Quarterly Journal of Economics* 87.2, 220–238.

Bastani, S., S. Blomquist, and L. Micheletto (2020). "Child Care Subsidies, Quality, and Optimal Income Taxation". *American Economic Journal: Economic Policy* 12.4, 1–37.

Bauernschuster, S., T. Hener, and H. Rainer (2016). "Children of a (policy) Revolution: The Introduction of Universal Child Care and Its Effect on Fertility". *Journal of the European Economic Association* 14.4, 975–1005.

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

Bayer, C. and M. Kuhn (2020). "Which Ladder to Climb? Decomposing Life Cycle Wage Dynamics". IZA Discussion Paper 12473. IZA Institute of Labor Economics.

BBSR - Bundesinstitut für Bau-, Stadt- und Raumforschung (2021). *INKAR - Indikatoren und Karten zur Raum- und Stadtentwicklung*.

Bettendorf, L. J., E. L. Jongen, and P. Muller (2015). "Childcare subsidies and labour supply? Evidence from a large Dutch reform". *Labour Economics* 36, 112–123.

Bick, A. (2016). "The Quantitative Role of Child Care for Female Labor Force Participation and Fertility". *Journal of the European Economic Association* 14.3, 639–668.

Bick, A., B. Brüggemann, N. Fuchs-Schündeln, and H. Paule-Paludkiewicz (2019). "Long-term changes in married Couples' labor supply and taxes: Evidence from the US and Europe since the 1980s". *Journal of International Economics* 118, 44–62.

Bick, A. and N. Fuchs-Schündeln (2018). "Taxation and labour supply of married couples across countries: A macroeconomic analysis". *The Review of Economic Studies* 85.3, 1543–1576.

Bien, W., T. Rauschenbach, and B. Riedel (2006). "Wer betreut Deutschlands Kinder? - DJI-Kinderbetreuungsstudie". Wissenschaftliche Texte. Deutsches Jugendinstitut.

Biewen, M. and S. Seifert (2018). "Potential Parenthood and Career Progression of Men and Women – A Simultaneous Hazards Approach". *The B.E. Journal of Economic Analysis & Policy* 18.2, 1–22.

Black, S. E. and A. Spitz-Oener (2010). "Explaining women's success: technological change and the skill content of women's work." *The Review of Economics and Statistics* 92.1, 187–194.

Blau, F. D. and L. M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". *Journal of Economic Literature* 55.3, 789–865.

Blossfeld, H.-P., H.-G. Roßbach and J. von Maurice, eds. (2011). "Education as a Lifelong Process – The German National Educational Panel Study (NEPS)". *Zeitschrift für Erziehungswissenschaft* 14.2.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). "Female Labor Supply, Human Capital, and Welfare Reform". *Econometrica* 84.5, 1705–1753.

Blundell, R., A. Duncan, and C. Meghir (1998). "Estimating Labor Supply Responses Using Tax Reforms". *Econometrica* 66.4, 827–861.

BMFSFJ - Bundesministerium für Familie, Senioren, Frauen und Jugend (2016). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2015.*

BMFSFJ - Bundesministerium für Familie, Senioren, Frauen und Jugend (2020). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2019.*

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019). "Inaccurate Statistical Discrimination". NBER Working Paper 25935. National Bureau of Economic Research.

Booth, A. L. and M. Wood (2008). "Back-to-Front Down Under? Part-Time/Full-Time Wage Differentials in Australia". *Industrial Relations: A Journal of Economy and Society* 47.1, 114–135.

Bourguignon, F. and A. Spadaro (2012). "Tax–benefit revealed social preferences". *The Journal of Economic Inequality* 10.1, 75–108.

Bronson, M. A. and P. S. Thoursie (2020). "The Wage Growth and Within-Firm Mobility of Men and Women: New Evidence and Theory". Working Paper 1706. Department of Economics, Georgetown University.

Bundesministerium für Familie, Senioren, Frauen und Jugend (2004). "Entwurf eines Gesetzes zum qualitätsorientierten und bedarfsgerechten Ausbauder Tagesbetreuung und zur Weiterentwicklung der Kinder- und Jugendhilfe (Tagesbetreuungsausbaugesetz – TAG)". Drucksache 15/3676. Deutscher Bundestag.

Bundesverfassungsgericht (1992). "Beschluss des Zweiten Senats vom 25. September 1992". 2 BvL 5/91, Rn. 1-97.

Busse, A. and C. Gathmann (2020). "Free daycare policies, family choices and child development". *Journal of Economic Behavior & Organization* 179, 240–260.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2021). "Difference-in-Differences with a Continuous Treatment".

Callaway, B. and P. H. C. Sant'Anna (2020). "Difference-in-Differences with multiple time periods". *Journal of Econometrics* (forthcoming).

Card, D., A. R. Cardoso, and P. Kline (2016). "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women". *The Quarterly Journal of Economics* 131.2, 633–686.

Carrillo-Tudela, C., A. Launov, and J.-M. Robin (2021). "The fall in german unemployment: A flow analysis". *European Economic Review* 132, 103658.

Carta, F. and L. Rizzica (2018). "Early kindergarten, maternal labor supply and children's outcomes: Evidence from Italy". *Journal of Public Economics* 158, 79–102.

Cascio, E. U., S. J. Haider, and H. S. Nielsen (2015). "The effectiveness of policies that promote labor force participation of women with children: A collection of national studies". *Labour Economics* 36.C, 64–71.

Chaisemartin, C. de and X. D'Haultfœuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". *American Economic Review* 110.9, 2964–2996.

Chaisemartin, C. de and X. D'Haultfœuille (2021). "Difference-in-Differences Estimators of Intertemporal Treatment Effects".

Chetty, R. (2012). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80.3, 969–1018.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins". *American Economic Review* 101.3, 471–475.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2013). "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities". *NBER Macroeconomics Annual* 27, 1–56.

Chhaochharia, V., S. Ghosh, A. Niessen-Ruenzi, and C. Schneider (2021). "Public Child Care Provision and the Motherhood Penalty".

Colas, M., S. Findeisen, and D. Sachs (2021). "Optimal Need-Based Financial Aid". *Journal of Political Economy* 129.2, 492–533.

Connolly, S. and M. Gregory (2008). "The part-time pay penalty: earnings trajectories of British Women". *Oxford Economic Papers* 61.1, i76–i97.

Connolly, S. and M. Gregory (2010). "Dual tracks: part-time work in life-cycle employment for British women". *Journal of Population Economics* 23.3, 907–931.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). "Who benefits from universal child care? Estimating marginal returns to early child care attendance". *Journal of Political Economy* 126.6, 2356–2409.

Cortes, P., J. Pan, L. Pilossoph, and B. Zafar (2021). "Gender Differences in Job Search and the Earnings Gap: Evidence from Business and Majors". NBER Working Paper 28820. National Bureau of Economic Research.

Costa Dias, M., R. Joyce, and F. Parodi (2021). "The gender pay gap in the UK: children and experience in work". *Oxford Review of Economic Policy* 36.4, 855–881.

Dengler, K., B. Matthes, and W. Paulus (2014). "Occupational Tasks in the German Labour Market - An alternative measurement on the basis of an expert database". FDZ-Methodenreport 12/2014. Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Domeij, D. and P. Klein (2013). "Should Day Care be Subsidized?" *The Review of Economic Studies* 80.2, 568–595.

Doran, E. L., A. P. Bartel, and J. Waldfogel (2018). "Gender in the Labor Market: The Role of Equal Opportunity and Family-Friendly Policies". NBER Working Paper 25378. National Bureau of Economic Research.

Dustmann, C., A. Lindner, U. Schoenberg, M. Umkehrer, and P. vom Berge (2020). "Reallocation Effects of the Minimum Wage". CReAM Discussion Paper Series 2007. Centre for Research and Analysis of Migration, Department of Economics, University College London.

Eckhoff Andresen, M. and T. Havnes (2019). "Child care, parental labor supply and tax revenue". *Labour Economics* 61, 101762.

Eisenhauer, P., P. Haan, B. Ilieva, A. Schrenker, and G. Weizsäcker (2020). "Biased wage expectations and female labor supply: a structural approach".

Ermisch, J. F. and R. E. Wright (1993). "Wage Offers and Full-Time and Part-Time Employment by British Women". *The Journal of Human Resources* 28.1, 111–133.

Exley, C. L. and J. B. Kessler (2019). "The Gender Gap in Self-Promotion". NBER Working Paper 26345. National Bureau of Economic Research.

FDZ-DJI - Forschungsdatenzentrum Deutsches Jugendinstitut (2017). *KiBS - DJI-Kinderbetreuungsstudie*. DOI: 10.17621/kibs2017.

FDZ-IAB - Forschungsdatenzentrum der Bundesagentur für Arbeit am Institut für Arbeitsmarkt und Berufsforschung (2020). *National Educational Panel Study (NEPS), Starting*

Cohort 6 (SC6) survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB) – Version 7518 v1. DOI: 10.5164/IAB.NEPS-SC6-ADIAB7518.de.en.v1.

FDZ-IAB - Forschungsdatenzentrum der Bundesagentur für Arbeit am Institut für Arbeitsmarkt und Berufsforschung (2021a). *Weakly anonymous Version of the Establishment History Panel (BHP) – Version 7517 v1.* DOI: 10.5164/IAB.BHP7517.de.en.v1.

FDZ-IAB - Forschungsdatenzentrum der Bundesagentur für Arbeit am Institut für Arbeitsmarkt und Berufsforschung (2021b). *Weakly anonymous Version of the Sample of Integrated Labour Market Biographies (SIAB) – Version 7517 v1.* DOI: 10.5164/IAB.SIAB7517.de.en.v1.

FDZ-LIfBi - Forschungsdatenzentrum des Leibniz-Institut für Bildungsverläufe (2019). *National Educational Panel Study (NEPS), Starting Cohort 6 (SC6) – Version SC6:10.0.0.* DOI: 10.5157/NEPS:SC6:10.0.0.

FDZ-SOEP - Forschungsdatenzentrum Sozioökonomisches Panel am Deutschen Institut für Wirtschaftsforschung (2019). *Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2017.* DOI: 10.5684/soep.v34.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2020a). *Mikrozensus 2014, On-Site-Zugang.* DOI: 10.21242/12211.2014.00.00.1.1.1.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2020b). *Mikrozensus 2018, On-Site-Zugang.* DOI: 10.21242/12211.2018.00.00.1.1.3.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2021). *Regionaldatenbank Deutschland.*

Felfe, C. (2012). "The motherhood wage gap: What about job amenities?" *Labour Economics* 19.1, 59–67.

Felfe, C. and R. Lalive (2018). "Does early child care affect children's development?" *Journal of Public Economics* 159, 33–53.

Fernández-Kranz, D. and N. Rodríguez-Planas (2011). "The part-time pay penalty in a segmented labor market". *Labour Economics* 18.5, 591–606.

Fitzenberger, B. and A. Seidlitz (2020). "The 2011 break in the part-time indicator and the evolution of wage inequality in Germany". *Journal for Labour Market Research* 54.1, 1–14.

Folke, O. and J. Rickne (2020a). "All the Single Ladies: Job Promotions and the Durability of Marriage". *American Economic Journal: Applied Economics* 12.1, 260–287.

Folke, O. and J. Rickne (2020b). "Sexual Harassment and Gender Inequality in the Labor Market". CEPR Discussion Paper 14737. Centre for Economic Policy Research.

Folke, O., J. Rickne, S. Tanaka, and Y. Tateishi (2020). "Sexual Harassment of Women Leaders". *Daedalus* 149.1, 180–197.

Fouarge, D. and R. Muffels (2009). "Working Part-Time in The British, German and Dutch Labour Market: Scarring for the Wage Career?" *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften* 129.2, 217–226.

Francesconi, M. and M. Parey (2018). "Early gender gaps among university graduates". *European Economic Review* 109, 63–82.

Gallego Granados, P. (2019). "The part-time wage gap across the wage distribution". DIW Discussion Paper 1791. Deutsches Institut für Wirtschaftsforschung (DIW).

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program". *Journal of Political Economy* 128.7, 2502–2541.

Gathmann, C. and B. Sass (2018). "Taxing Childcare: Effects on Childcare Choices, Family Labor Supply, and Children". *Journal of Labor Economics* 36.3, 665–709.

Geiler, P. and L. Renneboog (2015). "Are female top managers really paid less?" *Journal of Corporate Finance* 35, 345–369.

Geis-Thöne, W. (2020). "Kinderbetreuung: Über 340.000 Plätze für unter Dreijährige fehlen". IW-Kurzbericht 96/2020. Institut der deutschen Wirtschaft.

Geis, W. (2018). "Kinderbetreuung: Es fehlen immer noch fast 300.000 U3-Plätze". IW-Kurzbericht 11/2018. Institut der deutschen Wirtschaft.

Gelblum, M. (2020). "Preferences for Job Tasks and Gender Gaps in the Labor Market".

GESIS - Leibniz-Institut für Sozialwissenschaften (2017). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2016*. DOI: 10.4232/1.12796.

Geyer, J., P. Haan, and K. Wrohlich (2015). "The effects of family policy on maternal labor supply: Combining evidence from a structural model and a quasi-experimental approach". *Labour Economics* 36, 84–98.

Giannetti, M. and T. Y. Wang (2020). "Public Attention to Gender Equality and the Demand for Female Directors". CEPR Discussion Paper 14503. Centre for Economic Policy Research.

Gibbons, R. and M. Waldman (1999). "A Theory of Wage and Promotion Dynamics Inside Firms". *The Quarterly Journal of Economics* 114.4, 1321–1358.

Ginja, R., A. Karimi, and P. Xiao (2020). "Employer Responses to Family Leave Programs". IZA Discussion Paper 13833. IZA Institute of Labor Economics.

Givord, P. and C. Marbot (2015). "Does the cost of child care affect female labor market participation? An evaluation of a French reform of childcare subsidies". *Labour Economics* 36, 99–111.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). "The German Socio-Economic Panel (SOEP)". *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics* 239.2, 345–360.

Goldin, C. (2014). "A Grand Gender Convergence: Its Last Chapter". *American Economic Review* 104.4, 1091–1119.

Goodman-Bacon, A. (2018). "Difference-in-Differences with Variation in Treatment Timing". NBER Working Paper 25018. National Bureau of Economic Research.

Guner, N., R. Kaygusuz, and G. Ventura (2020). "Child-Related Transfers, Household Labour Supply, and Welfare". *The Review of Economic Studies* 87.5, 2290–2321.

Haan, P. and K. Wrohlich (2011). "Can child care policy encourage employment and fertility? Evidence from a structural model". *Labour Economics* 18.4, 498–512.

Haeck, C., P. Lefebvre, and P. Merrigan (2015). "Canadian evidence on ten years of universal preschool policies: The good and the bad". *Labour Economics* 36 (Supplement C), 137–157.

Hank, K., K. Tillmann, and G. G. Wagner (2001). "Institutional child care in eastern Germany before and after Unification. A comparison with western Germany in the years 1990-1999". *Zeitschrift für Bevölkerungswissenschaft* 26.1, 55–65.

Hardoy, I. and P. Schøne (2006). "The Part-Time Wage Gap in Norway: How Large is It Really?" *British Journal of Industrial Relations* 44.2, 263–282.

Havnes, T. and M. Mogstad (2011). "Money for nothing? Universal child care and maternal employment". *Journal of Public Economics* 95.11, 1455–1465.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error". *Econometrica* 47.1, 153–161.

Heckman, J. J. and R. Robb (1985). "Alternative methods for evaluating the impact of interventions". *Longitudinal Analysis of Labor Market Data*. Ed. by J. J. Heckman and B. S. Singer. Econometric Society Monographs. Cambridge University Press, 156–246.

Heckman, J. J. and R. Robb (1986). "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes". *Drawing Inferences* *from Self-Selected Samples*. Ed. by H. Wainer. New York, NY: Springer New York, 63–107.

Hendren, N. and B. Sprung-Keyser (2020). “A Unified Welfare Analysis of Government Policies”. *The Quarterly Journal of Economics* 135.3, 1209–1318.

Hirsch, B. T. (2005). “Why Do Part-Time Workers Earn Less? The Role of Worker and Job Skills”. *ILR Review* 58.4, 525–551.

Ho, C. and N. Pavoni (2020). “Efficient Child Care Subsidies”. *American Economic Review* 110.1, 162–199.

Hotz, V. J., P. Johansson, and A. Karimi (2017). “Parenthood, Family Friendly Workplaces, and the Gender Gaps in Early Work Careers”. NBER Working Paper 24173. National Bureau of Economic Research.

Hüsken, K. (2011). “Kita vor Ort - Betreuungsatlas auf Ebene der Jugendamtsbezirke 2010”. Wissenschaftliche Texte. Deutsches Jugendinstitut.

Hüsken, K. (2010). “Kindertagesbetreuung 2008. Betreuungsatlas auf Ebene der Jugendamtsbezirke”. Wissenschaftliche Texte. Deutsches Jugendinstitut.

Imai, K. and I. S. Kim (2021). “On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data”. *Political Analysis* 29.3, 405–415.

Jacobs, B., E. L. Jongen, and F. T. Zoutman (2017). “Revealed social preferences of Dutch political parties”. *Journal of Public Economics* 156, 81–100.

Jepsen, M., S. P. O’Dorchai, R. Plasman, and F. Rycx (2005). “The wage penalty induced by part-time work: the case of Belgium”. *Brussels Economic Review* 48.1-2, 73–94.

Jessen, J., R. Jessen, and J. Kluve (2019). “Punishing potential mothers? Evidence for statistical employer discrimination from a natural experiment”. *Labour Economics* 59, 164–172.

Jones, E. B. and J. E. Long (1979). “Part-Week Work and Human Capital Investment by Married Women”. *The Journal of Human Resources* 14.4, 563–578.

Kiessling, L., P. Pinger, P. K. Seegers, and J. Bergerhoff (2019). “Gender Differences in Wage Expectations: Sorting, Children, and Negotiation Styles”. IZA Discussion Paper 12522. IZA Institute of Labor Economics.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). “Field of Study, Earnings, and Self-Selection”. *The Quarterly Journal of Economics* 131.3, 1057–1111.

Klaauw, B. van der and A. Dias da Silva (2011). “Wage dynamics and promotions inside and between firms”. *Journal of Population Economics* 24.4, 1513–1548.

Klemm, K. (2014). "Ganztagsschulen in Deutschland: Die Ausbaudynamik ist erlahmt". Auftragsstudie. Bertelsmann Stiftung.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019). "Child Penalties across Countries: Evidence and Explanations". *AEA Papers and Proceedings* 109, 122–126.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2020). "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation". NBER Working Paper 28082. National Bureau of Economic Research.

Kleven, H., C. Landais, and J. E. Søgaard (2019). "Children and Gender Inequality: Evidence from Denmark". *American Economic Journal: Applied Economics* 11.4, 181–209.

Kosteas, V. D. (2011). "Job Satisfaction and Promotions". *Industrial Relations: A Journal of Economy and Society* 50.1, 174–194.

Kulms, J. and A. Nehls (2018). *Warum in Deutschland 300.000 Kitaplätze fehlen.* Deutschlandfunk. URL: https://www.deutschlandfunk.de/kindertagesstaetten-warum-in-deutschland-300-000.724.de.html?dram:article_id=424386 (visited on 08/31/2021).

Kuziemko, I., J. Pan, J. Shen, and E. Washington (2018). "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?" NBER Working Paper 24740. National Bureau of Economic Research.

Laun, T. and J. Wallenius (2021). "Having It All? Employment, Earnings, and Children". *The Scandinavian Journal of Economics* 123.1, 353–381.

Liu, K. (2016). "Explaining the gender wage gap: Estimates from a dynamic model of job changes and hours changes". *Quantitative Economics* 7.2, 411–447.

Lockwood, B. B. and M. Weinzierl (2016). "Positive and normative judgments implicit in U.S. tax policy, and the costs of unequal growth and recessions". *Journal of Monetary Economics* 77, 30–47.

Lorenz, N. and D. Sachs (2016). "Identifying Laffer Bounds: A Sufficient-Statistics Approach with an Application to Germany". *The Scandinavian Journal of Economics* 118.4, 646–665.

Lucifora, C., D. Meurs, and E. Villar (2017). "Children, Earnings and Careers in an Internal Labor Market".

Lundborg, P., E. Plug, and A. W. Rasmussen (2017). "Can Women Have Children and a Career? IV Evidence from IVF Treatments". *American Economic Review* 107.6, 1611–1637.

Manning, A. and B. Petrongolo (2008). "The Part-Time Pay Penalty for Women in Britain". *The Economic Journal* 118.526, F28–F51.

Manning, A. and J. Swaffield (2008). "The gender gap in early-career wage growth". *The Economic Journal* 118.530, 983–1024.

Matteazzi, E., A. Pailhé, and A. Solaz (2014). "Part-Time Wage Penalties for Women in Prime Age: A Matter of Selection or Segregation? Evidence from Four European Countries". *ILR Review* 67.3, 955–985.

Matthes, B., H. Meinken, and P. Neuhauser (2015). "Occupational sectors and occupational segments on the basis of KldB 2010". Methodenbericht. Statistik der Bundesagentur für Arbeit.

McCue, K. (1996). "Promotions and Wage Growth". *Journal of Labor Economics* 14.2, 175–209.

Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation". *The Review of Economic Studies* 38.2, 175–208.

Montgomery, M. (1988). "On the Determinants of Employer Demand for Part-Time Workers". *The Review of Economics and Statistics* 70.1, 112–117.

Müller, D. and K. Strauch (2017). "Identifying mothers in administrative data". FDZ-Methodenreport 13/2017. Research Data Centre (FDZ) of the German FederalEmployment Agency (BA) at the Institute for Employment Research (IAB).

Müller, K.-U. and K. Wrohlich (2020). "Does subsidized care for toddlers increase maternal labor supply? Evidence from a large-scale expansion of early childcare". *Labour Economics* 62, 101776.

Mulligan, C. B. and Y. Rubinstein (2008). "Selection, Investment, and Women's Relative Wages over Time". *The Quarterly Journal of Economics* 123.3, 1061–1110.

Mumford, K. and P. N. Smith (2009). "What Determines the Part-Time and Gender Earnings Gaps in Britain: Evidence from the Workplace". *Oxford Economic Papers* 61, i56–i75.

Neumark, D. and G. Vaccaro (2020). "The Career Evolution of the Sex Gap in Wages: Discrimination vs. Human Capital Investment". NBER Working Paper 28191. National Bureau of Economic Research.

Nollenberger, N. and N. Rodríguez-Planas (2015). "Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from Spain". *Labour Economics* 36, 124–136.

Nusse, H. E. and J. A. Yorke (1996). "Basins of Attraction". *Science* 271.5254, 1376–1380.

OECD (2017). *Pensions at a Glance 2017.*

OECD (2021a). *Gender wage gap (indicator).* DOI: 10.1787/7cee77aa-en.

OECD (2021b). *OECD Family Database - LMF1.2. Maternal employment rates.*

Olivetti, C. and B. Petrongolo (2017). "The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries". *Journal of Economic Perspectives* 31.1, 205–230.

Paul, M. (2016). "Is There a Causal Effect of Working Part-Time on Current and Future Wages?" *The Scandinavian Journal of Economics* 118.3, 494–523.

Peto, R. and B. Reizer (2021). "Gender differences in the skill content of jobs". *Journal of Population Economics* 34.3, 825–864.

Preston, A. and S. Yu (2015). "Is there a part-time/ full-time pay differential in Australia?" *Journal of Industrial Relations* 57.1, 24–47.

Roussille, N. (2021). "The central role of the ask gap in gender pay inequality".

Saez, E. and S. Stantcheva (2016). "Generalized Social Marginal Welfare Weights for Optimal Tax Theory". *American Economic Review* 106.1, 24–45.

Schmucker, A., A. Ganzer, J. Stegmaier, and S. Wolter (2018). "Establishment History Panel 1975-2017". FDZ-Datenreport 09/2018. Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Schneider, U. C. (2017). "Life Cycle Cost of Overconfidence: Evidence from Maternity Leave Reforms".

Schrenker, A. (2020). "Do Women Expect Wage Cuts for Part-Time Work?"

Semykina, A. and J. M. Wooldridge (2010). "Estimating panel data models in the presence of endogeneity and selection". *Journal of Econometrics* 157.2, 375–380.

Statistisches Bundesamt (2012). *Finanzen der Kindertageseinrichtungen in freier Trägerschaft 2010.*

Statistisches Bundesamt (2014). *Statistiken der Kinder- und Jugendhilfe - Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege am 01.03.2014.*

Stern, S., A. Schultheiss, J. Fliedner, R. Iten, and C. Felfe (2015). "Analyse der Vollkosten und der Finanzierung von Krippenplätzen in Deutschland, Frankreich und Österreich im Vergleich zur Schweiz". Beiträge zur sozialen Sicherheit 3/15. Bundesamt für Sozialversicherungen BSV.

Thomas, M. (2020). "The Impact of Mandated Maternity Benefits on the Gender Differential in Promotions: Examining the Role of Adverse Selection".

Tinios, P., F. Bettio, and G. Betti (2015). "Men, women, and pensions". European Commission Report. European Commission - Directorate-General for Justice.

Turon, H. (2019). "Home production of childcare and labour supply decisions in a collective household model". IZA Discussion Paper 12148. IZA Institute of Labor Economics.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors (2019). "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python". *arXiv e-prints* 1907.10121.

Wales, D. J. and J. P. K. Doye (1997). "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms". *The Journal of Physical Chemistry A* 101.28, 5111–5116.

Wang, H. (2019). "Fertility and Family Leave Policies in Germany: Optimal Policy Design in a Dynamic Framework".

Wolf, E. (2002). "Lower wage rates for fewer hours? A simultaneous wage-hours model for Germany". *Labour Economics* 9.5, 643–663.

Wrohlich, K. (2008). "The excess demand for subsidized child care in Germany". *Applied Economics* 40.10, 1217–1228.

Xiao, P. (2020). "Wage and Employment Discrimination by Gender in Labor Market Equilibrium".

Zucco, A. (2019). "Occupational Characteristics and the Gender Pay Gap". DIW Discussion Paper 1794. Deutsches Institut für Wirtschaftsforschung (DIW).

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

München, der 16.09.2021

Fabian Martin Stürmer-Heiber