European University Institute

DEPARTMENT OF ECONOMICS

Working Mothers, Children, and Family Policies

David Koll

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 25 September 2020

European University Institute
Department of Economics

Working Mothers, Children, and Family Policies

David Koll

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Árpád Ábrahám, EUI and University of Bristol, Supervisor
Prof. Dominik Sachs, University of Munich (LMU)
Prof. Peter Haan, Freie Universität Berlin and DIW Berlin
Prof. Johanna Wallenius, Stockholm School of Economics

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, David Koll, certify that I am the author of the work "Working Mothers, Children, and Family Policies" I have presented for examination for the Ph.D. thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

Statement of inclusion of previous work (if applicable):

I confirm that chapter 1 was jointly co-authored with Ms Hélène Turon, Mr Dominik Sachs, and Mr Fabian Stürmer-Heiber and I contributed 40% of the work.

I confirm that chapter 2 was jointly co-authored with Ms Gabriela Galassi and Mr Lukas Mayr and I contributed 33% of the work.

I confirm that chapter 2 draws upon an earlier article we published as IZA Discussion Paper 12595 and Bank of Canada Staff Working Paper 2019-33 under the title "The Intergenerational Correlation of Employment: Is There a Role for Work Culture?".

Signature and Date:

David Koll

January 10, 2020

Abstract

This thesis contains three independent chapters that investigate work decisions and labour market outcomes of mothers and their potential dynamic consequences. Furthermore, it focuses on intended and unintended effects of family policies.

The first chapter, joint work with Dominik Sachs, Fabian Stürmer-Heiber, and Hélène Turon, studies the long-term fiscal implications of childcare subsidies through their impact on maternal labour supply. We explicitly capture life-cycle career aspects in a dynamic structural household model of female labour supply and childcare decisions: higher labour supply of mothers today results in higher expected future earnings. Using German survey data, we provide a structural estimate of the degree to which childcare subsidies are dynamically self-financing through higher labour income tax revenue. Our estimates show that targeting childcare subsidies is a useful tool to increase the ability of these policies to be self-financing.

The second chapter, joint work with Gabriela Galassi and Lukas Mayr, documents a substantial positive correlation of employment status between mothers and their children in the United States. Controlling for ability, education, fertility, and wealth, a one-year increase in maternal employment is associated with six weeks more employment of her child. The intergenerational transmission is stronger to daughters and more pronounced for low-educated and low-income mothers. Investigating potential mechanisms, we provide evidence for a role-model channel, through which labour force participation is transmitted.

The third chapter studies the effect of a divorce law reform on the probability to pay alimony as a divorced father using German administrative data. We show with a difference-in-differences setup that the reform decreased the probability to pay alimony if the youngest common child was aged four to eight compared to sixteen to seventeen. Furthermore, the treatment intensity varies with the age of the youngest child having the largest impact between four and five, thereby decreasing the disposable income of divorced mothers with younger children to a greater extent.

Acknowledgements

I would like to thank my supervisors Árpád Ábrahám and Dominik Sachs for their magnificent encouragement, guidance, and support over the past years. Árpád has always been available to critically discuss my work. I thank him for all the trust he put in me. I have been lucky to have him as my advisor and I learned so much from him. Working with Dominik has been a fantastic experience throughout my PhD which greatly shaped my idea of economic research. He always helped me to get back on track and to never forget the big picture. His positive attitude and his relentless energy are invaluable.

I would like to especially thank Johanna Wallenius and Peter Haan who kindly accepted to become part of my thesis committee for taking the time to read and comment on my dissertation.

My work benefited from numerous discussions with professors at the EUI. Especially, over the last two years, I deeply enjoyed the interaction with Philipp Kircher. Working with Philipp has been an immense source of motivation and a great learning experience. I am extremely grateful for his continuous trust and support.

It is fantastic to work with dedicated people who are able to constantly improve others. I thank my co-authors, Gabriela and Lukas, Fabian, Hélène, and Dominik, for all the discussions, hard work, motivation, and fun that we have had. You always went the extra mile and reminded me that working in teams is so much fun.

I extend my gratitude to the administration of the European University Institute – especially Jessica Spataro, Sarah Simonsen, Rossella Corridori, and Lucia Vigna – who provided immediate solutions to various kinds of problems.

Over the past years in Florence, I had the opportunity to share my life with so many great people, without them it would have been different. My fellow PhD students were a great source of feedback and support with many discussions about research and life. I am especially thankful to Simon and Benedikt, Lukas and João, Vinzenz, Gabriela, Gözde, My, Pascal and Melanie, Tobias, Andresa and Alex, Denise, Giulia and Chiara. A special thanks goes to Sebastian, Ellen, and Anton for their (never-ending) hospitality during my numerous research stays in Munich and to Gabriela und Reinhard for hosting me during my visit in Ottawa. Finally, a big thanks also to my Beachvolleyball crew and my Coppa team.

I gratefully acknowledge financial support by the German Academic Exchange Service (DAAD), the European University Institute, the German Academic Scholarship Foundation (Studienstiftung), and the Einaudi Institute for Economics and Finance in Rome which enabled me to experience this exceptional years at the EUI. I wrote parts of this dissertation during research stays at the CESifo at the LMU in Munich, at the Center for Macroeconomic Research at the University of Cologne, at UC Berkeley, and at the University of Konstanz. I am grateful for their hospitality. Finally, I thank Maurice Brandt and his colleagues from the German Federal Statistical Office for their data support during my work on the third chapter.

I dedicate this thesis to my parents Claudia and Wolfgang and my siblings Miriam and Deborah. I am unbelievably thankful for your unconditional support throughout my entire life. You have always been supportive, understanding, and interested. And finally I want to thank my grandparents, Helga, Elsbeth and Eduard, for your loving support. Without all of you, I would not have had the opportunities which made this thesis possible.

Contents

Abstract **i**

Acknowledgements **iii**

1 The Fiscal Return to Childcare Policies **1**

1.1 Introduction 1

1.2 Background, data, and stylised facts 5

1.2.1 Institutional background in Germany 5

1.2.2 Data 6

1.2.3 Stylised facts 6

1.3 Model 10

1.3.1 Preferences 12

1.3.2 Children 13

1.3.3 Childcare 13

1.3.4 Wages 15

1.3.5 Dynamic problem 16

1.3.6 Dynamic fiscal calculations 17

1.4 Quantification of the Model 19

1.4.1 Childcare need and government policies 20

1.4.2 Wages: heterogeneity and dynamics . . . 22

1.4.3 Fertility types . . . 26

1.4.4 Calibration of homogeneous preference parameters . . . 29

1.4.5 Maximum Likelihood Estimation . . . 30

1.5 Model Fit . . . 37

1.6 Fiscal calculation and Policy Experiments . . . 39

1.7 Conclusion . . . 44

Appendices **47**

1.A Estimation Samples . . . 47

1.B Additional Stylised Facts . . . 47

1.C Quantification of the Model . . . 49

1.C.1 Childcare prices . . . 49

1.C.2 Wage processes . . . 50

1.C.3 Fertility types . . . 55

1.C.4 Maximum Likelihood Estimation . . . 57

1.D Model fit . . . 58

1.D.1 Derivation of the compensated intensive margin labour supply elasticity . 58

1.E Policy Experiments . . . 59

2 The Intergenerational Correlation of Employment: Mothers as Role Models **67**

2.1 Introduction . . . 67

2.2 Data . . . 72

2.3 Empirical strategy . . . 75

2.4 Results . . . 76

2.4.1 Intergenerational correlation of employment . . . 76

2.4.2 Heterogeneous employment correlations 81
2.5 Potential mechanisms 84
2.5.1 Work culture 85
2.5.2 Mechanisms that can be ruled out 89
2.6 Conclusion 91

Appendices **93**
2.A Two-Generations Model 93
2.B Details on the empirical analysis 95
2.B.1 Details on the data 95
2.B.2 Methodological challenges in the measurement of intergenerational persistence of labor market outcomes 97
2.B.3 Details on the robustness exercises 98
2.B.4 Details on the preferences for work in NLSY79 and CNLSY79 99
2.B.5 Evidence favoring role model: Employment during periods of cohabitation versus non-cohabitation 100
2.C Additional Tables 101
2.D Additional Figures 108

3 Less money for divorced mothers? The child-age dependent reform of alimony in Germany **113**
3.1 Introduction 113
3.2 Regulation of alimony and the reform of 2008 117
3.3 Data 119
3.4 Motivational evidence 121
3.4.1 Descriptive evidence 121

3.4.2 Specific court case . . . 121
3.5 Empirical strategy and setup . . . 123
3.5.1 Empirical strategy . . . 123
3.5.2 Empirical setup . . . 124
3.6 Results . . . 127
3.6.1 Baseline results . . . 127
3.6.2 Robustness . . . 131
3.7 Conclusion . . . 138

Appendices **141**

3.A Additional information on the institutional setup . . . 141
3.A.1 German divorce law . . . 141
3.B Additional tables and figures . . . 142
3.B.1 Robustness: Alternative control group . . . 142
3.B.2 Robustness: Treatment intensity . . . 144
3.B.3 Average age of children in age brackets . . . 145

Chapter 1

The Fiscal Return to Childcare Policies

1.1 Introduction

Kleven, Landais, Posch, et al. 2019 document substantial child penalties in earnings for different countries. The penalty is particularly large for Germany, where female earnings after first child birth drop by 80% in the short run and still remain 60% below the pre-birth level ten years later. Between 2005 and 2013 several laws have been adopted to expand subsidised childcare availability. One explicit goal of these policies was to promote female labour supply.[1]

It has been argued that childcare subsidies may be partly self-financing through higher tax revenue: if childcare subsidies promote female labour supply, this will also increase income tax revenue.[2] In this paper we operationalize this idea and propose both a theoretical framework and a structural estimate of this dynamic fiscal effect for Germany.

We show that answers to the following questions are crucial in assessing the fiscal returns to childcare subsidies: (i) which households do change their childcare decision on either the extensive or intensive margin? (ii) To what extent do these households also change their labour supply? (iii) Which households do not change their childcare decision but adjust their labour supply decision due to wealth effects? (iv) How does labour market participation affect the future earnings trajectories of these marginal households?

We provide answers to these questions within a structural dynamic framework in which households choose maternal labour supply and hours of market childcare. Besides the usual time

[1]The idea that such policies can indeed achieve this goal is supported by a sizeable empirical literature, see e.g. Bauernschuster and Schlotter 2015 for the German context, Baker, Gruber, and Milligan 2008 for the Canadian context or Cascio, S. J. Haider, and Nielsen 2015 for a summary of different studies for different countries.

[2]Rainer et al. 2013 provide some back-of-the-envelope calculations for Germany based on static tax revenue effects. Bach et al. 2020 calculate the static fiscal effect of providing full-time daycare for children in primary school.

and budget constraints, households face a 'childcare' constraint, i.e. the age composition of children within each household implies a need for childcare which households have to fulfil. Childcare can be provided through market childcare, home produced childcare or informal, other childcare, (e.g. by grandparents, friends). Households are heterogeneous along observable characteristics, i.e. their current age and both partners' wages, as well as their fertility type and unobserved characteristics. The nature of the latter is threefold: households differ in their preferences for home produced childcare and for maternal leisure, and in their access to free informal caretakers for their children. We show the importance of both observed and unobserved heterogeneity for the answers to the questions above and argue that our structural model is well suited to take these into account.

The quantification of the model is carried out with survey data from the German Socio-Economic Panel (GSOEP) and consists of five steps. First, since tax and childcare price schedules are key to our fiscal calculation, we include a detailed specification of the variations of both of these with family characteristics. Second, we estimate the wage heterogeneity and dynamics assuming a Markovian structure of the wage process including gender- and age-dependency. But most importantly, the transition probabilities – and therefore the expected future wage – depend on the current labour supply decision. Third, we predict for each household a fertility type distribution since the actual type is only observed for households with completed fertility. Each household's fertility type is defined as the age at first child birth and desired completed family size. In the prediction, we use estimated logit models combined with the currently available information on the age at first birth and the current number of children.

The fourth and the fifth components of the quantification are conducted jointly: fourth, we calibrate the homogeneous parameters in the utility function to match reasonable participation and compensated labour supply elasticities. Fifth, we estimate our structural model to obtain the joint distribution of the unobserved preferences for home produced childcare and for female leisure, as well as the access to informal childcare. Our estimation strategy is to find the distributional parameters that maximise the likelihood of matching the observed household choices in terms of female labour supply and used hours of market childcare services with their model counterpart.

The estimated joint distribution of unobserved heterogeneity allows us to simulate the model's response to small perturbations of childcare fees and predict which households and mothers change their behaviour in response to the reform and how they adjust it. This enables us to simulate the long-term marginal fiscal returns of childcare subsidies. We find that an untargeted increase in childcare subsidies is barely creating any return through static and dynamic maternal labour supply responses. The share of additional income taxes generated over the additional costs of childcare subsidies lies at 2.6% over the remaining life cycle of households. In other words, investing 1 Euro in untargeted childcare subsidies, creates additional (discounted) income tax revenues of 2.6 cents.

However targeting childcare subsidies is a useful tool to increase the ability of the policy to be self-financing: first, if the childcare subsidy increase is paid only to working mothers, the degree of self-financing increases to 17.4% over the remaining life cycle. Second, targeting working mothers with children below the age of 3 raises the marginal return to 23.3%. Finally, if the subsidy expansion is contingent on full-time working mothers, it would dynamically refinance itself by 75.7% over the remaining life cycle. Or put differently, an investment of 1 Euro into childcare subsidies of full-time working mothers, creates additional (discounted) income tax revenues of 75.7 cents.

As visible from the previous results, targeting bears a crucial impact on the expected costs and benefits of the policy in the long-term. There are two underlying mechanisms: first, targeting childcare subsidies towards working mothers decreases the share of households that are inframarginal in their labour supply decision and consume a positive amount of market childcare. If subsidies are unconditional, the share of households who receive higher subsidies but do not work more, i.e. do not pay more taxes, amounts to 82.23%. It decreases to 63.05% (12.97%) in the case of work contingent (full-time work contingent) subsidies. Such a decrease in the share of inframarginal households implies a sizeable reduction in the amount of subsidies that do not generate any additional tax revenue. Second, if subsidies are targeted at households with working mothers, the share of households that increase their labour supply in the initial subsidy period rises: if the subsidy is untargeted, 0.17% of the households increase their labour supply, whereas the number rises to 0.6% (0.5%) for work contingent (full-time work contingent) childcare subsidies. This implies that targeting also increases current and future tax revenues of the government compared to an untargeted childcare subsidy.

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

Reduced form studies that show positive effects of childcare subsidies on female labour supply in the German context include Bauernschuster and Schlotter 2015, Gathmann and Sass 2018, Busse and Gathmann 2019, and Müller and Wrohlich 2020.[3] Bauernschuster and Schlotter 2015 use the introduction of a legal entitlement to a place in kindergarten for children above three in 1996 and find a positive impact of (subsidised) childcare availability on maternal

[3]For the effect of childcare on maternal labour supply in other countries, see, for example, 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 labour supply is Haeck, Lefebvre, and Merrigan 2015 who focus on the universal childcare reform in Québec. They find positive effects on maternal labour force participation over a ten year horizon.

employment using an instrumental variable and difference-in-differences approach. Müller and Wrohlich 2020 find a similar effect focusing on children below three by exploiting geographical and time variation of a recent childcare expansion. Focusing on a policy reform which increases the implicit price of childcare in the East German state Thuringia, Gathmann and Sass 2018 show that in response to the reform childcare attendance drops and maternal labour supply declines, especially for vulnerable subgroups such as single, low-income, and low-skilled parents. Busse and Gathmann 2019 use regional variation in childcare prices in Germany combined with birthday cut-offs to evaluate the introduction of universal free childcare which took place in different West German states at different points in time. They find that universal free childcare increases childcare attendance of 2 – 3 year old children which leads to higher labour market attachment of mothers. The responses are stronger for vulnerable subgroups such as poorer families and low-skilled parents.

Combining structural and reduced form methods, Geyer, Haan, and Wrohlich 2015 use quasi-experimental variation in Germany to estimate a structural labour supply model. The authors evaluate the effect of family policies and find a large increase of maternal employment following a joint introduction of parental leave benefits and subsidized childcare.

In contrast to the described studies, our paper focuses on estimating the long-run fiscal effects of childcare subsidies through modelling counter-factual policy scenarios.[4] Our structural approach also allows us to target these policies towards subgroups of the population.

Structural work that focuses on the particular German setup includes Bick 2016 and Wang 2019. The former paper studies the effect of childcare policies and the latter jointly studies the impact of childcare and parental leave policies. Both find that an increase in subsidised childcare raises maternal labour force participation. In addition, Haan and Wrohlich 2011 use variation in taxes and transfers, which includes childcare costs, to pin down the incentives for employment in a structural model for Germany. The authors find that higher childcare subsidies generally increase the labour supply for well educated women as well as those without children. Also related to our paper are the studies by Guner, Kaygusuz, and Ventura 2019 and Laun and Wallenius 2019. Guner, Kaygusuz, and Ventura 2019 study welfare effects of different child-related transfers in the U.S. context, whereas Laun and Wallenius 2019 examine in a structural household model, which is calibrated to Sweden, the effects of family policies 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 home produced childcare, builds on Turon 2019. Lastly, the paper is related to Adda, Dustmann, and Stevens

[4]Eckhoff Andresen and Havnes 2019 refine the typical back-of-the-envelope calculation of the static effect of childcare subsidies on income tax revenues. They estimate the impact of (subsidised) childcare availability on annual earnings – increase of 66,000 NOK (8,000 USD) – as well as the average tax rate on the additionally earned income – 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.

2017 who set up a structural model to determine the career costs of children and Blundell et al. 2016 who structurally estimate the returns to experience for woman in the UK (and thereby also the wage penalties from not working full-time). Our approach is different from all mentioned papers as we add a clear-cut public finance question by assessing the fiscal costs of untargeted and targeted childcare subsidies. For this purpose, we model rich heterogeneity in family structures as well as preferences and estimate their joint distribution which facilitates to identify mothers who are marginal or inframarginal in a given policy scenario.

Many papers in the public finance literature have emphasized that the implied effects on labour supply provide a rationale for subsidizing childcare (e.g. Bastani, Blomquist, and Micheletto 2019; Domeij and Klein 2013; Ho and Pavoni 2019). The goal of this paper is to make this argument more operational from an applied point of view and quantifies the size of the implied fiscal externality of childcare subsidies. The paper is also related to Colas, Findeisen, and Sachs 2019, who study the same question for college education subsidies but with different trade-offs.

This paper is organized as follows. In Section 2 we present some policy background and stylized facts which motivate our analysis. Section 3 contains the model setup and the quantification of the model is described in Section 4. In Section 5 we illustrate the fit of the model with data moments and quasi-experimental studies before we present our policy analysis and fiscal calculations in Section 6. Section 7 concludes.

1.2 Background, data, and stylised facts

1.2.1 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 centres). Approximately around the time when children turn 3 years old, they enter kindergarten and stay there until they start school. Lastly, during school age children may attend after school care centres in the afternoon.[5] The distinction between these institutions matters because childcare prices differ by attended institution and, therefore, by child age. This is foremost due to different institutional cost structures and price regulations.[6] Contrary to e.g. the U.S. 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 percent of childcare institutions are either operated by municipalities or by non-profit organisations.[7]

[5] Note that some children might also attend full-day schooling.

[6] See Appendix 1.C.1.1 for a detailed description of childcare price determinants.

[7] See Authoring Group Educational Reporting 2018.

Recent reforms. While all three types of childcare institutions have been continuously present throughout Germany since the early 1990s, their use and prevalence has changed substantially in the past thirty years. During this time, family policy has increasingly focused on childcare as a tool to allow females to both participate in the labour market and have children. A major first step in this policy effort has been the introduction of a legal entitlement to a place in a kindergarten for children aged 3 until school entry which took place in 1996. While this reform led to a substantial increase in the number of kindergarten slots, childcare supply for children below 3 (i.e. nursery slots) remained scarce until 2005. In that year, the German government committed to creating 230,000 additional nursery slots by October 2010, increasing childcare coverage for children aged 0 to 2 from just 5 to 17 slots per 100 children.

Focusing policy efforts further on nurseries, an extension of the legal entitlement to childcare was passed: in October 2010, it introduced a legal entitlement to a nursery slot for all children below 3 whose parents both work. In August 2013, this entitlement was further extended to cover all children after their first birthday, regardless of the employment status of the parents.[8]

1.2.2 Data

For our analysis, we focus on German females aged between 20 and 65 who are currently not in education and live in a household with a full-time working partner.[9] We track this group over the time span 2000 to 2017 in a representative longitudinal survey data, the German Socio-Economic Panel (GSOEP). The GSOEP is an unbalanced household panel that has been running since 1984 in West Germany and includes East Germany since 1991. 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.[10]

1.2.3 Stylised facts

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

[8] Family policy related law changes: 1996 - Reform of the §SGB VIII, 2005 - Tagesbetreuungsausbaugesetz, 2008 - Kinderfoerderungsgesetz.

[9] In line with our modelling framework described below, we furthermore limit the sample to females who have at most three children and gave birth (if any) only between ages 20 and 41. (87.40% of observations).

[10] A detailed description of the GSOEP can be found in Goebel et al. 2018.

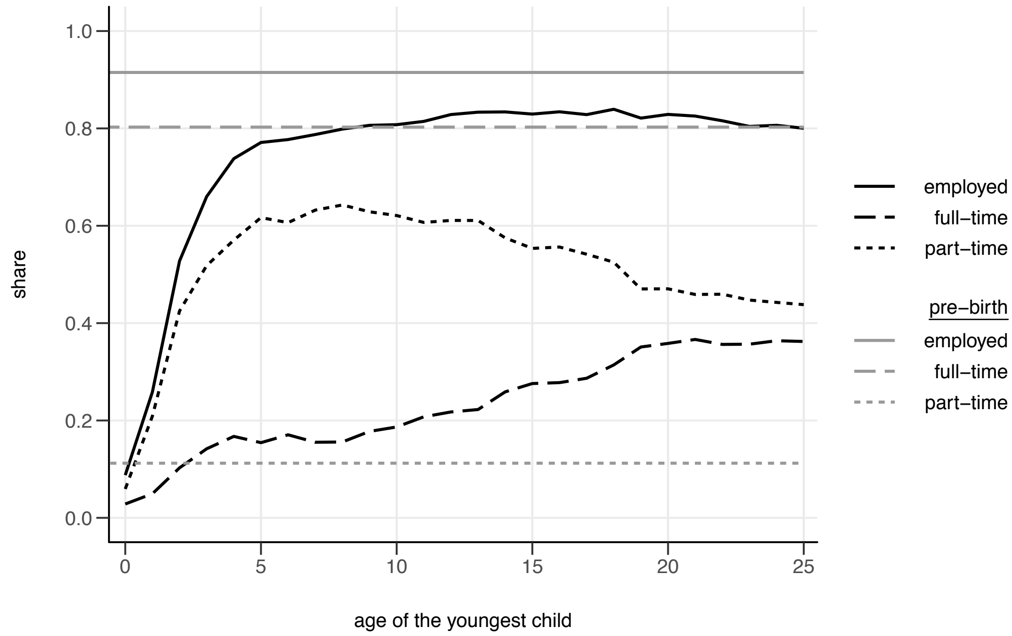


Figure 1.1: Employment status by age of the youngest child

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.

This also holds for Germany where childbirth has a substantial and sustained effect on maternal employment status.[11] Figure 1.1 illustrates this effect in the GSOEP data. It compares maternal pre-birth employment rates to their post-birth developments depending on the age of the youngest child. The average pre-birth employment rate of future mothers is above 90 percent of which 80 percent work full-time – corresponding to 40 hours per week – and the remaining women work part-time. In the first year after childbirth the overall employment rate drops by more than 80 percentage points to below 10 percent and only three percent of mothers with newborns aged 0 to 1 work full-time. It is important to note that until the third birthday of the child mothers may take maternity leave. During such a maternity related temporary 'non-employment'-spell mothers partially receive financial compensation and have the right to return to their pre-birth job.[12]

As children grow older, the share of mothers working full-time rises gradually. While it reaches 15 percent by the end of statutory maternity leave at age 3, from then on it only increases slowly

[11] Kleven, Landais, Posch, et al. 2019 use the event study approach 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.

[12] 78.03% of mothers whose youngest child is aged 0 to 1 state that they are on maternity leave.

to just about 20 percent when the youngest child is 10 years old. Afterwards the full-time share gradually rises up to almost 40 percent by the time the youngest child reaches adulthood. However, this level is still more than 40 percentage points lower than the pre-birth full-time share, suggesting that a substantial fraction of mothers never returns to full-time work.

Part-time contracts, i.e. 20 hours per week, illustrate another side of the employment pattern of mothers: Before the first birth, only 12 percent of future mothers work part-time. This share increases sharply one year after childbirth to reach 62 percent with five year old children. At this point, total employment reaches almost 78 percent. This is in line with many mothers opting for part-time when their children are young, although they worked full-time before parenthood. After the youngest child has turned 10, the part-time share gradually declines to 47 percent by child age 20. Together with the simultaneously increasing full-time share, this trend suggests an intensive margin decision, i.e. mothers returning from part-time to full-time. Summing up, participation remains almost flat around 80 percent from age 5 on.

Taken together, Figure 1.1 shows that childbirth substantially reduces the hours that women work, but that part-time arrangements do play a large role for mothers' labour market attachment. However these part-time jobs yield lower wage growth rates, which we will document in Section 1.4.2, leading to lower earnings over the life cycle.

Reform effects on childcare enrolment and maternal labour supply. One of the explicit goals of the recent family policy reforms in Germany has been to increase maternal labour market participation, both in terms of part-time and full-time work. As laid out in Section 1.2.1, the expansion of subsidised childcare, especially for children below the age of three, has been a major policy instrument to accomplish this goal. Figure 1.2 plots childcare enrolment shares for different child age brackets across recent years.

In 2000, more than 80 percent of 3 to 5 year old children were already attending childcare, while the share was only nine percent for children below 3. Over the years, during which subsidised childcare was substantially expanded,[13] this enrolment share has increased to more than 40 percent in 2017. Overall, childcare attendance of children below 3 has increased by more than 30 percentage points over the last 17 years. Enrolment of 3 to 5 year old children has also increased over that time span by 15 percentage points to 97 percent in 2017.

To relate this pattern of increased childcare attendance to employment rates, Figures 1.3(a) and 1.3(b) plot maternal full-time and part-time participation rates by the age of the youngest child. During the 2000 to 2017 time span, when childcare enrolment of children aged 0 to 2 increased substantially, mothers of children in this age bracket have been increasing their labour force participation both along the extensive and the intensive margin. Part-time rates

[13] As discussed in Section 1.2.1.

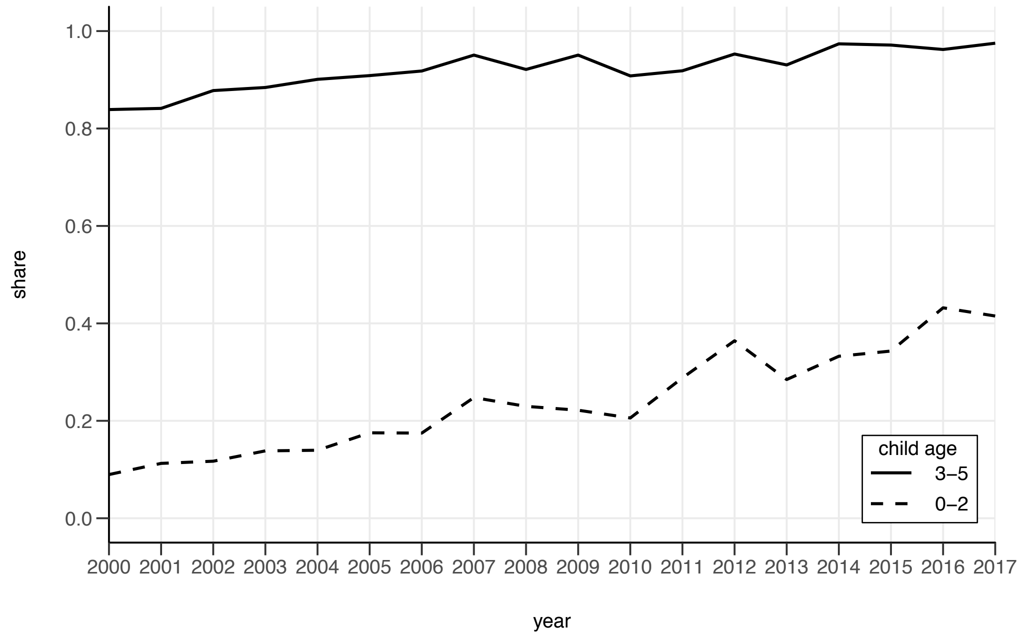


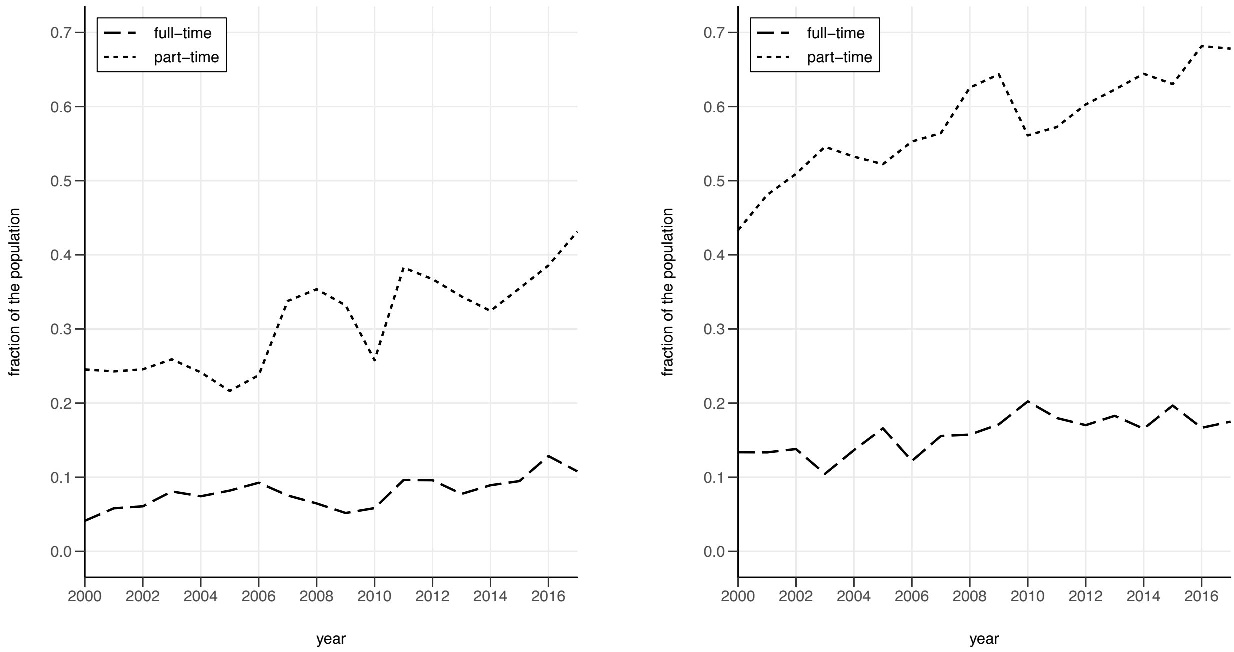
Figure 1.2: Childcare enrollment across time

Notes: Enrolment 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. Source: 2000 to 2017 GSOEP.

among these mothers have increased by more than 15 percentage points, while full-time rates have also increased slightly. Looking at the 3 to 5 year old children in Figure 1.3(b), part-time employment shares have risen substantially from 45 to more than 65 percent between 2000 and 2017. During the same period, the full-time share has increased by around five percentage points.

Childcare price setting. To shed light on one of the key policy instruments, namely childcare prices, Figure 1.4 illustrates how the prices for full-time childcare vary across the income distribution and by the number of children. Two key facts can be observed: first, childcare prices are progressive in income. While one-child households in the first net income quintile pay on average 150 EUR per month for a full-time slot, households in the fifth quintile pay almost twice as much. The same holds for households with two children. Second, the total amount paid for childcare does not increase linearly in the number of children. Comparing one-child and two-children households, the price for the second child is always lower than the price for the first child.

These stylised facts show that childcare subsidies play an important role for increasing the labour supply of mothers and thereby reducing child penalties. Furthermore, we illustrated that the current level of childcare subsidies is targeted by household income and the number of



(a) Children aged 0-2

(b) Children aged 3-5

Figure 1.3: Employment of mothers by age of the child

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 to 2 for panel (a) or 3 to 5 for panel (b). Source: 2000 to 2017 GSOEP.

children. We will now use a structural model, that is consistent with and guided by the stylised facts, to understand whether there is a rationale for targeting childcare subsidies that can be grounded in fiscal externalities.

1.3 Model

We present a dynamic model of households faced with labour 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 and the period discount rate is denoted β. For simplicity, both spouses have the same age t. We assume that fertility is exogenous but allow households to be heterogeneous with respect to their fertility 'type', defined as the age at first child birth and completed family size. Households know their fertility type at all times and thus anticipate the number and timings of future childbirths. Labour supply choices are discrete: the female spouse can work full-time, part-time (half of the full-time hours) or choose not to participate, while the male spouse is assumed to always work full-time. Households can decide between the female spouse caring for the children at home,

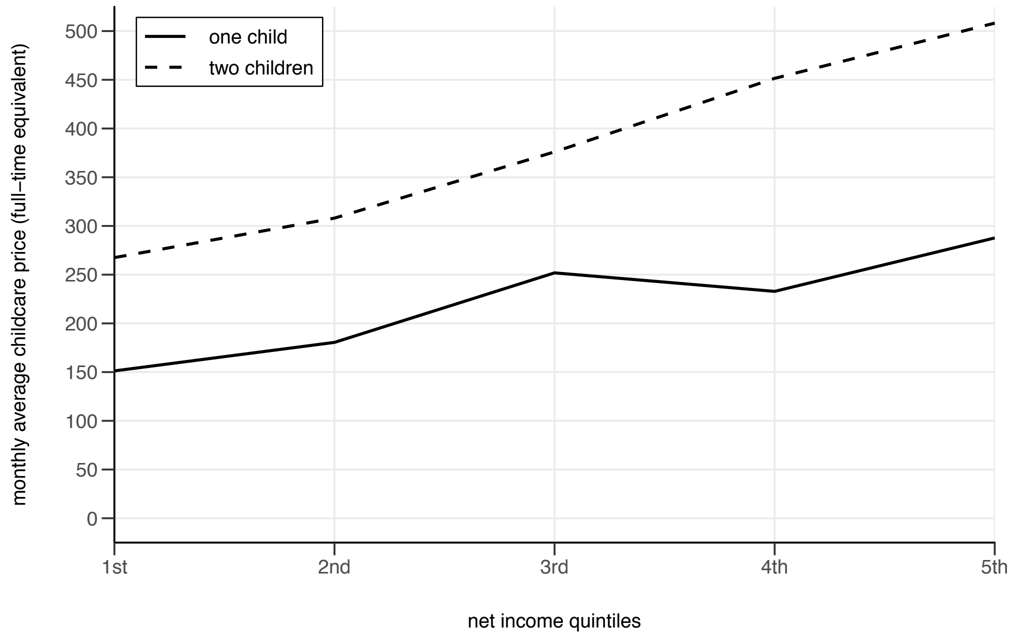


Figure 1.4: Total monthly childcare prices across the income distribution

Notes: Full-time equivalent monthly childcare price vs. household net income quintiles, all adjusted to 2017 prices. Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, conditional on observed positive childcare fees. Source: 2013, 2015, 2017 GSOEP.

which we will call 'home produced childcare', and externally provided childcare. The latter can either be informal childcare by e.g. grandparents (if available to the household), or use of market childcare services. Households are unobservably heterogeneous in three dimensions: their preference for home produced childcare, their taste for the leisure of the female spouse, and their access to free informal childcare. Our aim is to estimate these three dimensions of heterogeneity in order to gain a better understanding of the aggregate labour supply elasticity of mothers to the price of childcare.[14]

The dynamic component of our framework comes from the impact of current labour supply choices on the expected growth rate of potential wages, which is positively affected by hours worked in the current period. Hence, career breaks imply dynamic wage penalties. We assume

[14]Germany reformed its parental leave benefit policy in 2007. We abstract from such benefits because the one-year paid parental leave currently in place has only a small effect on the three-year budget of the household. Further, it would complicate our model substantially as parental leave benefits are state-dependent on the previous labor market choice. Even before that reform, a job guarantee for three years was ensured for mothers. This implies that there will be no wage uncertainty if one returns to work after a three-year leave: a mother gets the same wage as before. Yet, wage penalties can show up directly afterwards because women will not have the same wage increases and career development as they would have without the child-birth related career break. While we do not explicitly model this job protection, we do capture the dynamic effects of these career breaks well by our estimated wage process.

no instantaneous penalty of part-time work on the hourly wage. We abstract from dynamic interactions of home produced childcare, i.e. we assume that current childcare choices do not impact future optimal childcare decisions (marginal utilities of future choices are unaffected). This does allow households to believe in long-run benefits of domestic childcare, but assumes that such considerations are fully captured in the instantaneous utility. This limitation departs from the literature on investments into child quality (e.g. Bernal and Keane 2011 and Del Boca, Flinn, and Wiswall 2014), but keeps the model tractable.

We assume that marriages are formed at the age of 20, both spouses retire at 65 and have a remaining lifespan of 15 years after retirement. We rule out saving and borrowing as well as divorce. We assume that these choices do not interact with our decision of interest, which is the simultaneous choice of labour supply and childcare. We formalise below the various components of our framework and the dynamic optimisation problem that households face.

1.3.1 Preferences

Households value female leisure time L, household consumption c, and the time devoted by the mother to home produced childcare dcc during normal working hours.[15] 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. Preferences are reflected in the following instantaneous utility function:

$$u(c, L, dcc) = \tag{1.1}$$

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

Two elements, α and g, of this utility function capture unobserved heterogeneity across households. α represents the relative taste for female leisure over consumption and g the relative preference for domestic childcare over the consumption-leisure component. $\mathcal{G}$ scales the unobserved heterogeneity g and allows for child-age dependency in the taste for home produced childcare. It captures systematic differences between having younger (i.e. below 3) and older children (i.e. between 3 and 9) without adding an additional dimension to the unobserved het-

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

erogeneities. Based upon the child-age vector K (explained in detail below), the functional form of $\mathcal{G}$ corresponds to

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

$\mathcal{G}$ allows us to capture the sharp difference in the childcare enrollment between $0-2$ year olds and $3-5$ year olds, which Figure 1.2 illustrates. It can be seen as a shortcut to capture social norms regarding the childcare attendance of $3-5$ year olds.[16] The three CRRA coefficients γ_c, γ_L and γ_{dcc} as well as $\overline{L}$ and $\overline{dcc}$ are homogeneous across all households.

1.3.2 Children

Each household's childbearing trajectory is deterministic and characterised by its 'fertility type', denoted f. This type comprises the age at first birth a and the number of children n that the household will eventually have. Age at first birth can be in any of the first seven periods, i.e. between the ages of 20 and 41. In addition, households can have between 0 and 3 children.

$$f = \begin{cases} (\cdot, 0) & \text{if no children} \\ (a, n) & a \in \{1, ..., 7\}, n \in \{1, ..., 3\} \text{ otherwise} \end{cases} \tag{1.2}$$

This setup leaves us with 22 different fertility types. The two underlying assumptions are that each household only has one child per 3-year period and that the children's age difference is equal to the period length (i.e. 3 years).[17] $K(t, f)$ denotes a 3-element vector indicating the presence of a child in the first three age brackets $(0-2)$, $(3-5)$ and $(6-8)$ in any household of type f and age t. For example, a household of type $f = (2, 1)$ is a household that will bear its first (and only) child in the second period, i.e. at age $(23-25)$, and thus will have, in period 4, one child in the third age category and none in the other two categories: $K(4, (2, 1)) = (0, 0, 1)$. By assumption, each element of K can only be 0 or 1 since only one child can be born in each period.

1.3.3 Childcare

In each category $i = 1, 2, 3$ 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, $\bar{t}_i$. For example,

[16] We furthermore assume that every $3-5$ year old has to attend childcare at least half-days, as laid out in Section 1.4.1, for the same reason.

[17] The median birth spacing observed in our sample as defined in Section 1.2.2 is 3 years.

in the first and second categories, the infant needs care all of the time, whereas in the third category, the child needs care in the non-school hours.

There are three ways in which parents can fulfil this childcare need: maternal time, which we denote *dcc* and refer to as 'home produced' childcare, informal childcare denoted *oth*, e.g. by grandparents, or formal childcare services, e.g. by a nursery, denoted *mcc*. Informal childcare is free and only available to some households. This availability, which is captured by *oth*, ranges between 0 and 40 hours a week, is constant over time, and known to the household.[18] Since we cannot directly observe it in the data, we will refer to it as an unobserved heterogeneity. Formal childcare is always available at a price, normalised to full-time use, which depends on the age i of the child, the family structure K and the household gross income y^{gross}:

$$p\left(i, K\left(t, f\right), y_t^{gross}\right)$$

Since households' preferences put value neither on informal childcare nor on market childcare, they will never choose market childcare when informal childcare is available. The amount of market childcare necessary for a child of age i is thus:

$$mcc(i) = \max\left\{0, \bar{t}_i - dcc - oth\right\} \tag{1.3}$$

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

$$Tcc(t, f) = \sum_{i=1}^{3} K_i(t, f) \cdot mcc(i)$$

where $K_i(t, f)$ is the i-th element of the 3-dimensional vector $K(t, f)$ indicating if a child with age i currently lives in the household.

The household expenditure for the care of all its children is given by:

$$Ecc(t, f, y^{gross}) = \sum_{i=1}^{3} K_i(t, f) \cdot p(i, K(t, f), y^{gross}) \cdot mcc(i)$$

Note that we assume that taking care of n children at home requires one hour of home produced or informal childcare but n hours of market childcare.

[18] The introduction of *oth* is motivated by the fact that we observe mothers who work more hours than they buy childcare for. See Appendix 1.B and Figure 1.8 for a discussion of how childcare hours correspond to agreed working hours in the data.

1.3.4 Wages

We denote w^M and w^W the hourly wages of the two spouses. As mentioned above, the male spouse always works full time and w^M refers to his actual wage. The female spouse chooses labour supply discretely among full-time, part-time, or non-participation ($lm_t \in \{0, 0.5, 1\}$). In case of labour market participation (full-time or part-time), w^W refers to her actual wage. If the female chooses not to participate w^W refers to her potential wage. The latter is imputed following the procedure laid out in Section 1.4.2.

In order to keep the state space tractable, we represent wage dynamics as first-order Markov processes. The gender specific wage distributions are characterised by age-dependent wage quintiles. The transition probabilities are governed by age-independent, labour supply specific transition matrices between the wage quintiles. Future wages therefore depend on the current wage quintile and the current labour supply choice lm, i.e. $w_{t+1}(w_t, lm_t)$ as well as age and gender. As we assign potential wages to those who do not work, w^M and w^W evolve continuously over time according to this process.

Despite its simple setup, the wage process is able to capture the key dynamics of more evolved human capital accumulation frameworks (such as e.g. Blundell et al. 2016). First, wage growth differs across ages and the wage distribution. Thereby our use of age-dependent wage quintiles accommodates the fact that the wages of young workers grow at a different rate than those of workers close to retirement. This can be seen as a simplistic approximation of concave wage growth in experience. Furthermore, as the transition probabilities are different for each wage quintile, expected wage growth rates also depend on the current position in the wage distribution.

Second, the wage process also captures the well-documented fact that periods of part-time work or non-participation deteriorate future labour market prospects. With labour supply dependent transition matrices, working part-time instead of working full-time implies an increased probability to move to a lower wage quintile and a decreased probability to move to a higher wage quintile. Furthermore, the extent of these penalties for part-time and non-participation varies across the wage distribution. A female at the top of the wage distribution faces a higher expected drop in her future wages from part-time work than one at the middle of the wage distribution.

As it captures the two aforementioned key facts, the first-order Markov representation is able to deliver the key mechanism of interest: The decision to stay at home to raise children has negative consequences for the wage prospects of a mother, differing in extent with her position in the wage distribution and her age.[19]

[19] We abstract from instantaneous wage penalties for part-time work.

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

1.3.5 Dynamic problem

We now formalise the dynamic optimisation problem faced by households. The observed characteristics that determine the household's behaviour are its current age t, the male wage w_t^M and the female wage w_t^W. Age evolves deterministically, increasing by 1 every period and wages evolve according to the first-order Markov process described above. The unobserved and constant characteristics of the household that will also influence its choice are its fertility type f, the availability of informal childcare oth, and its tastes for home produced childcare and for leisure, g and α respectively. Together with age t, the fertility type f determines the number and age structure of the children in the household. We summarise all characteristics by a seven-dimensional state space vector $\Omega_t = (t, w_t^W, w_t^F, f, oth, g, \alpha)$.

At each age t, the household has to choose female labour supply (lm_t), female leisure (L_t), and the use of home produced (dcc_t) and market (mcc_t) childcare. The labour supply decision is discrete and the other choices are continuous. The three constraints that the household faces are the time constraint for the female spouse in equation (1.4), the need for childcare of current children (represented by (1.3)) and the budget constraint (shown in (1.5)):

$$lm_t + L_t + dcc_t \;=\; 1 \tag{1.4}$$

$$c_t + Ecc(t, f, y_t^{gross}) \;=\; y_t^{net} \tag{1.5}$$

Since we rule out saving and borrowing, households consume all their income net of childcare expenditures. Weekly household net income is calculated as $y_t^{net} = y_t^{gross} - T(y_t^{gross})$, where $T(\cdot)$ is the income tax schedule and the gross household income, y_t^{gross}, is given by:

$$y_t^{gross} = 40 \cdot \left(w_t^M + lm_t \cdot w_t^W\right) \tag{1.6}$$

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

$$V(\Omega_t) = \max_{lm_t, c_t, L_t, dcc_t} u(c_t, L_t, dcc_t|\Omega_t) + \beta\mathbb{E}[V(\Omega_{t+1}|\Omega_t, lm_t)] \tag{1.7}$$

subject to the three constraints above. This problem can be broken down into an intra-temporal choice and and inter-temporal choice.

The intra-temporal decision reflects the optimal time allocation between leisure and home produced childcare for a given female labour supply choice lm_t. The inter-temporal choice is to choose the optimal labour supply, lm_t, given the conditional optimal choices of leisure $L_t(lm_t)$ and home produced childcare $dcc_t(lm_t)$.

The optimal intra-temporal choice consists in solving the following static problem for a given discrete female labour supply decision $lm_t \in \{0, 0.5, 1\}$ and state vector Ω_t:

$$u^*(\Omega_t, lm_t) = \max_{c_t, L_t, dcc_t} u(c_t, L_t, dcc_t | \Omega_t, lm_t) \tag{1.8}$$

subject to the childcare constraint (1.3), the time constraint (1.4) and the budget constraint (1.5).

Finally, the optimal inter-temporal choice consists in maximising lifetime utility $V(\Omega_t)$ by choosing female labour supply lm_t:

$$V(\Omega_t) = \max_{lm_t} u^*(\Omega_t, lm_t) + \beta V(\Omega_{t+1} | \Omega_t, lm_t) \tag{1.9}$$

for given dynamics of the states $\big(\Omega_{t+1} | \Omega_t, lm_t\big)$. The model is solved by backward induction from retirement, where we assume that 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 of in retirement.

1.3.6 Dynamic fiscal calculations

The goal of this paper is to estimate the net fiscal effect of child care subsidies focusing on income tax revenues generated from the response of female labour supply to this policy over the entire life cycle. For this purpose, the key challenge is to estimate which households react to changes in child care policies and how they react. In other words, we will recover households that are marginal or inframarginal at the labour supply and/or the market childcare margin when facing a (small) policy change. In addition, we determine the size of each group to calculate the population-wide net fiscal effect.

In our structural model, for given childcare prices p (see eq. (1.3.3)), the optimal choices of labour supply and use of market child care over the life cycle are determined by observed characteristics, s, and unobserved heterogeneity $h = (g, oth, \alpha)$. We denote the optimal choices for a given childcare price schedule p by $lm^*(\Omega; p)$ and $mcc^*(\Omega; p)$.[20]

Following a permanent increase in the subsidy of market childcare implying a new age-dependent childcare price schedule $p' < p$, we are interested in two different measures that quantify the degree to which childcare subsidies are self-financing: first, we are interested in the fiscal effect

[20]Note that we do not have any general equilibrium or spill-over effects in this model. Therefore, there is no interaction between the choices of different households.

in the period of the policy implementation. Second, we estimate the long-run effect taking not only the current but also future effects of the policy on the government budget into account.

What are the different effects on the government budget? First, let us consider the costs for the government. We denote the exogenous, hourly, and age-dependent cost of market childcare as C.[21] The costs of the reform to the government arise through increased subsidy spending. For a household with a given state vector Ω, the increase in government expenses for subsidies in the current period is given by:

$$\Delta S(\Omega; p, p') = (C - p') \cdot mcc^*(\Omega; p') - (C - p) \cdot mcc^*(\Omega; p) \tag{1.10}$$

We can decompose it into two effects: on the one hand, the *mechanical effect* $\Delta S_p(\Omega) = \big[(C-p')-(C-p)\big] \cdot mcc^*(\Omega; p)$ will capture the budgetary impact of the reform if the consumption of market childcare stays unchanged; on the other hand, the *behavioural effect* $\Delta S_q(\Omega) = (C - p') \cdot \big[mcc^*(\Omega; p') - mcc^*(\Omega; p)\big]$ reflects that households might adjust their use of market childcare. The mechanical effect includes households who can be marginal or inframarginal at the childcare margin, whereas the behavioural effect by definition only captures marginal households, i.e. those that do change their use of market childcare due to the reform.

Second, the reform also affects income tax revenues raised from discrete female labour supply, $lm^* \in \{0, 1/2, 1\}$:

$$\begin{aligned}\Delta T(\Omega; p, p') &= T\big(y^{gross}(\Omega; p')\big) - T\big(y^{gross}(\Omega; p)\big) \\ &= T\big(40 \cdot \big(w^M + lm^*(\Omega; p') \cdot w^W\big)\big) - T\big(40 \cdot \big(w^M + lm^*(\Omega; p) \cdot w^W\big)\big) \end{aligned} \tag{1.11}$$

Note that both $\Delta S(\Omega; p, p')$ and $\Delta T(\Omega; p, p')$ are defined per model period.

Therefore, using these measures we can directly infer the net fiscal effect conditional on the household type Ω when we consider only the impact period of the reform. If we are instead interested in the estimation of the net fiscal effect over the (remaining) life cycle of the households, we need to simulate their (remaining) life cycle to determine the evolution of the states Ω. The households' age t and the male wage w^M evolve exogenously whereas the fertility type f and the dimensions of unobserved heterogeneity h are fixed over time.[22] Therefore, the only endogenous component is the female wage w^W as current female labour supply affects future female wages. By simulating female wage processes using the Markovian structure described in Section 1.3.4, we are able to calculate the expected discounted difference in paid taxes over the life cycle due to the reform for a given current state vector Ω.

[21] We set the costs for a full-time childcare slot to the average public spending per child in full-time care and assume that the costs are linear in hours (see Section 1.4.1).

[22] We rule out that fertility responds to the change in childcare prices which is reasonable given that in our policy experiments in Section 1.6 we increase childcare subsidies by only 50 EUR per month.

As we are interested in the population-wide net fiscal effect – either for the impact period of the reform or over the (remaining) life cycle of households –, we need to weigh the described subsidy and tax effects that depend on Ω with the population density of Ω.

Decomposition of the tax effects. Each household's response in terms of female labour supply depends on their states Ω. We distinguish two effects: first, let us consider households that currently have to cover some childcare need. Due to a decrease in child care prices, some households will adjust their current female labour supply. This in turn affects current and future tax revenues. We call this effect the *direct tax effect* and decompose it further into a *static effect* and a *dynamic effect*. The static tax effect summarizes the instantaneous effect on tax revenues induced by the change in current female labour supply. In addition, the dynamic tax effect describes the effect on future tax revenues: current female labour supply affects future labour supply and earnings through the Markovian wage process that reflects returns to experience and therefore, alters future income tax revenues.

Second, we focus on households who do not have to cover any childcare need in the current period but expect to do so in the future. As the reform is permanent, in anticipation of cheaper market child care in the future, younger households might adjust their decisions today, e.g. current female labour supply. These adjustments have dynamic consequences in terms of accumulated experience. We define this effect as the *anticipation tax effect*. Households who do not have children yet, but expect to have them in the future, will face lower childcare costs. This may increase female labour supply during child-rearing and the expected increase of future labour supply itself affects the incentive to supply labour in the current period.[23]

Higher female labour supply prior to having children affects government tax revenues positively. It also implies a higher accumulation of experience prior to having children. This works as a multiplier because it increases the return to work in every other period before, during and after having children which, everything else equal, will – at the margin – increase female labour supply in those periods.

1.4 Quantification of the Model

We quantify the model in five steps:

1. We calibrate the childcare need and estimate parameters of government policies that govern childcare prices and specify the tax system.

2. We estimate wage quintiles and transition matrices as inputs for our wage process.

[23]Intuitively, a woman who is expecting to have children and to take a long career break during the child-rearing years will be less inclined to invest in her human capital today than if she expects to carry on working during the child-rearing years.

3. We employ a multinomial logit model to estimate the distribution of fertility types.
4. We calibrate various homogeneous preference parameters.
5. We jointly estimate by maximum likelihood the distribution of heterogeneity in the preferences for home produced childcare, female leisure and the availability of informal childcare.

1.4.1 Childcare need and government policies

In this subsection, we first calibrate the childcare need. Second, we quantify the government policies that are required as exogenous inputs for our model: the interest rate, pension benefits, the German tax system, institutional childcare cost structures, and childcare prices.

1.4.1.1 Childcare need

Table 1.1 summarises the assumed childcare need for children of different ages: If a child is younger than 6 years, the childcare need (during working time) 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 1.2), we impose that 20 of the 40 hours required for this age interval have to be covered by market childcare. Half-day market 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 the remaining hours. This yields for each child age i the age-specific weekly hours of childcare needed, $\bar{t}_i$.

Table 1.1: Childcare need within normal working hours

Children's age interval	0 - 2	3 - 5	6 - 8	≥ 9
Hours of childcare needed per week	40	40	15	0
Mandatory market childcare	0	20	0	0

1.4.1.2 Interest rate

We set the interest rate of the government to three percent per three-year model period, which corresponds approximately to one percent per annum.

1.4.1.3 Pensions

We approximate the German pension system by assuming that the households receive 50% of both partners' last period's potential gross full-time earnings throughout retirement, i.e. the pension share, $\mathcal{B}$, is equal to 0.5.

1.4.1.4 Taxes

We use the Matlab implementation of the German tax code provided by Bick et al. 2019 to map gross to net income. It 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 wives differ with their husbands' income and child allowances reduce the taxable income of the household.

1.4.1.5 Childcare cost structures

We approximate the cost structures of public childcare institutions by assuming the cost to be linear in the number of children. This abstracts from any non-linearities driven by capacity constraints, but nevertheless provides a useful benchmark for the public spending per child. We use the figures in Table 1.2 from the German statistical office for 2010, which we adjust to 2017 prices for our fiscal calculations.

Table 1.2: Average annual public spending per child in full-time care in 2010

Children's age interval	0 - 2	3 - 5	6 - 8
Cost in EUR	10,900	7,300	6,200

Notes: See Statistisches Bundesamt 2012, in 2010 prices.

1.4.1.6 Childcare prices

Data. We use data from the 2013, 2015, and 2017 GSOEP waves, which contains information on childcare hours per day and monthly fees.[24] We normalise the monthly fees by the reported daily childcare hours to extract the monthly price of full-time childcare defined by an attendance of eight hours per day. For this purpose, we assume linearity of childcare prices in hours.

Empirical model. We use the following linear model to estimate the childcare price function reflected in the structural model by $p(i, K(t,f), y_t^{gross})$:

$$\begin{aligned} p_{kjs} = \; & \alpha + \beta_1 \cdot y_{js}^{gross} + \beta_2 \cdot y_{js}^{gross} \cdot \mathbb{1}\big\{\text{one sibling with age} < 17 \text{ in HH}\big\}_{js} \\ & + \beta_3 \cdot y_{js}^{gross} \cdot \mathbb{1}\big\{\text{two siblings with age} < 17 \text{ in HH}\big\}_{js} + \epsilon_{kjs} \end{aligned} \tag{1.12}$$

[24] In terms of the sample construction, this estimation is based on the same sample as laid out in Section 1.2.2.

The dependent variable p_{kjs} is the monthly price of full-time childcare paid by household j for child k in survey year s. Our empirical model closely reflects the current childcare price setting in Germany. While the specific price calculation is at the discretion of each municipality, state laws define possible determinants to be household income, the number of children in the household as well as the number of childcare hours per day. In line with these regulations, we allow full-time equivalent monthly childcare prices to differ by household gross income y_{js}^{gross}. The interaction terms of gross household income with dummies for the number of siblings capture discounts granted to families with many children.[25]

We estimate (1.12) as a Tobit regression with censoring at 0 and 725 EUR, the lowest and highest observed monthly childcare payments in our data.[26] As the institutional design of childcare prices differs substantially between the defined age brackets, we estimate equation (1.12) separately for children aged 0 - 2, 3 - 5, and 6 - 8. We abstract from regional variation in order to keep the state space of the structural model parsimonious.

Results. The results of the Tobit regressions are summarised in Table 1.3. Monthly childcare prices increase significantly in gross household income for all age brackets. Average prices 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.

1.4.2 Wages: heterogeneity and dynamics

As described in Section 1.3.4, we use a first-order Markov process to model the dynamics of wages. To calculate the transition matrices between wage quintiles conditional on each labour supply choice requires a complete dataset of hourly wages.

Imputation of potential wages. Using the 2000 to 2017 GSOEP panel, we observe monthly gross labor income as well as agreed hours for males, whom we only include if they work full-time. This allows us to directly compute male hourly wages. Therefore, we can calculate quintiles over the male wage distribution, separately for each model age. For working females we observe the same information as for men which we use to compute gross hourly wages. For females who choose not to participate in the labour market we do not observe any labour income and therefore, we need to impute their *potential* gross hourly wages with the following specification:

[25] As we are mainly interested in predicting childcare prices, we only include covariates that are in line with the institutional setup described in Appendix 1.C.1. The stand-alone sibling dummies are not included as they do not add any explanatory power.

[26] We use a Tobit to account for the fact that 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 1.3: Tobit estimation of childcare prices for different child ages
Dependent variable:
Monthly price of market childcare normalised to 40h/week

	child age		
	0 - 2	3 - 5	6 - 8
gross HH income	0.036	0.021	0.026
	(0.0037)	(0.0015)	(0.0042)
gross HH income × 1 sibling in HH	-0.0099	-0.0060	-0.0039
	(0.0029)	(0.0012)	(0.0034)
gross HH income × 2 siblings in HH	-0.022	-0.011	-0.0040
	(0.0040)	(0.0015)	(0.0040)
constant	84.2	65.3	11.9
	(18.3)	(6.32)	(15.7)
N	368	2002	645

Notes: Sample: Children attending market childcare for whom childcare fees and attended childcare hours are observed. Tobit regressions with censoring at 0 and 725 EUR. All prices are adjusted to 2017 price levels. Source: 2013, 2015, 2017 GSOEP.

Females wages are determined by:

$$log(w_{it}^{f}) = \mathbf{X}_{it}\boldsymbol{\beta} + u_{it} \tag{1.13}$$

where $\mathbf{X}$ contains Mincer-type covariates (age, full-time/part-time experience, education),[27] the number of children below 5, the overall number of children, an urban dummy, a dummy for former East Germany, and year dummies.

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

$$\mathbf{Z}_{it}\boldsymbol{\zeta} + \nu_{it} > 0 \tag{1.14}$$

$\mathbf{Z}$ 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 dummies that indicate the presence of a 0-2, 3-5, 6-8, 9-17, or 18+ year old child in the household. Furthermore, we

[27]See Appendix 1.C.2 for a detailed description.

include the husband's gross wage quintile and the net household income if the female chooses not to work.[28]

In line with the parametric selection correction procedure suggested by Semykina and Wooldridge 2010, we run a probit version of (1.14) for each time period. These also include the individual specific time means across 2000 to 2017 of all covariates included in $\mathbf{Z}$, denoted by $\overline{\mathbf{Z}}$.

$$Pr(\text{participation} = 1)_{it} = \mathbf{Z}_{it}\boldsymbol{\zeta} + \overline{\mathbf{Z}}_i\boldsymbol{\xi} + \upsilon_{it} \tag{1.15}$$

From (1.15) we obtain the inverse Mills ratios $\hat{\boldsymbol{\lambda}}_{it}$, which we use in the selection correction version of the wage equation (1.13):

$$log(w^f_{it}) = \mathbf{X}_{it}\boldsymbol{\rho} + \overline{\mathbf{Z}}_i\boldsymbol{\xi} + \hat{\boldsymbol{\lambda}}_{it}\boldsymbol{\gamma} + u_{it} \tag{1.16}$$

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

Wage quintiles and transitions. After predicting the potential wages of the non-working females, we calculate quintiles over the female wage distribution, separately for each model age. Together with the means for each quintile, this wage quintile assignment characterises the female wage component in the observed state space of the structural model. The full wage quintile matrices for males and females can be found in Tables 1.10 and 1.11 in Appendix 1.C.2. The age-specific calculation allows us to capture the typical hump-shaped pattern of wages over the life cycle, as illustrated in Figure 1.5. As this hump shape is typically driven by human capital accumulation, the age-specific quintiles allow us to approximate the effects of experience on wages in a parsimonious manner within our model framework.

To characterize the transitions between wage quintiles, we calculate the annual shares of individuals that remain in or change their wage quintile in the next period. To ensure that our transition matrices are estimated over a period of constant labour supply, we calculate these at the yearly frequency, over individuals with identical labour supply choices in two consecutive years. The transition matrices that we use in the three-year period model are the estimated yearly matrices cubed. To capture the different effects that full-time, part-time or non-participation have on (expected) future wages, we calculate the transition probabilities separately for each female labour supply choice as well as full-time working men. In other

[28]See Appendix 1.C.2 for a discussion of the identifying assumptions.

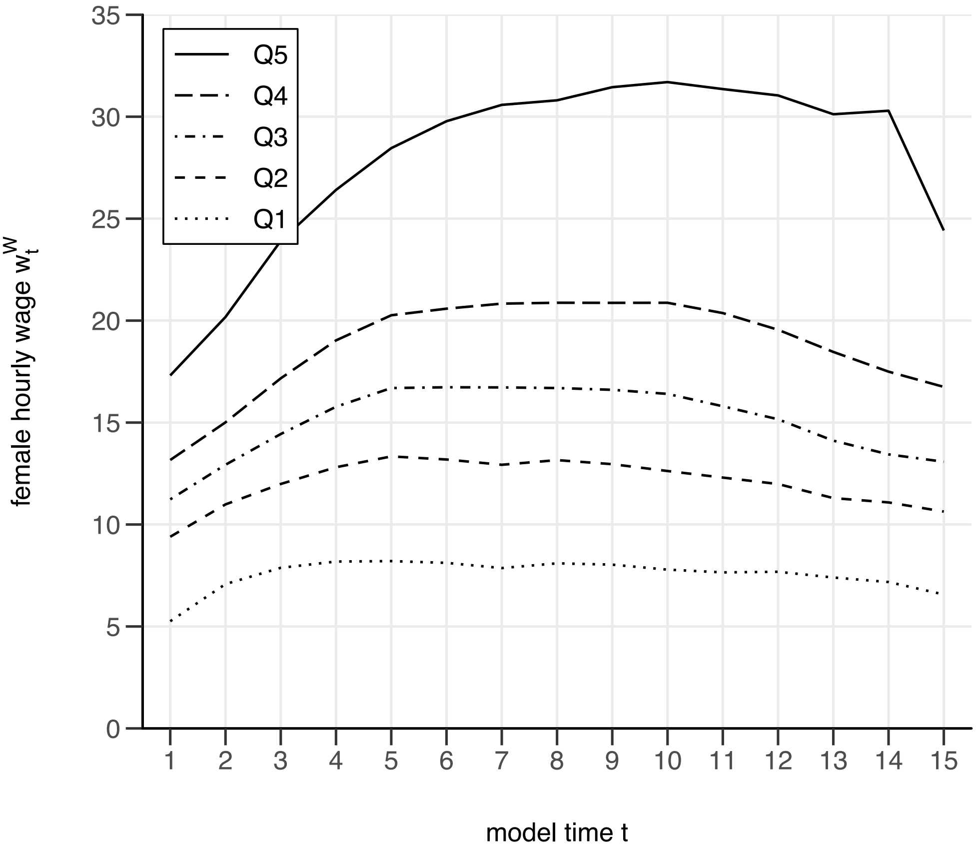


Figure 1.5: Female wage quintile means across the life cycle

Notes: Gross hourly wages in 2017 prices. See Appendix 1.C.2 for the equivalent representation of male wages. Source: 2000 to 2017 GSOEP.

words, we condition the transition probabilities on the current wage quintile as well as current labour supply. This keeps the transition probabilities tightly connected to the *current* labour supply choice and avoids confounding them with direct wage effects of a differing future labour supply choice. As the fully specified transition matrices for women consist of 75 probabilities,[29] we restrict their calculation to be age-independent to avoid small sample issues.

The resulting transition matrices, which are summarised in Tables 1.12 and 1.13 in Appendix 1.C.2, deliver the key properties of the wage process that the human capital accumulation literature has found: working full-time yields a substantially higher (lower) probability to move to a higher (lower) wage quintile than working part-time or not working. This is in line with higher accumulation of human capital through full-time work and positive returns to experience. The same argument holds (to a lesser extent) for part-time vs. non-participation. Between these labour supply choices, the difference in expected future wages is particularly large for the low-wage quintiles and rather small for the high-wage quintiles. It implies that females who earn

[29] $75 = 3$ labour supply choices $\times$ 5 initial wage quintiles $\times$ 5 future wage quintiles.

low wages have a lot to gain from working part-time instead of not working, while females with high wages gain only little from part-time work relatively to not working.

1.4.3 Fertility types

In our theoretical model, we assume that households make choices with full knowledge of their fertility type $f = (a, n)$, i.e. they anticipate accurately the mother's age at the time of the birth of her first child, a, and the number of children the household will have eventually, i.e. completed fertility, n. We assume that women can have children when they are aged between 20 and 40. Since our model time period is 3 years, we group first births (a) into three-year brackets. The decision to set the maximum number of children to three ($n \in \{0, ..., 3\}$) is driven by the data as we observe only few families with more than 3 children.[30] This setup results in 19 possible fertility types $f = (a, n)$, as households who have their first child in the age range $35 - 37$ have either one or two children and households that have their first child between $38 - 40$ can only have that one child.

Fertility type is a constant and partially observed characteristic of households. More specifically, for households where the female spouse is aged 41 and above, (a, n) are perfectly observed by assumption. However, if the spouse is under 41, we may know the age at which she had her first child, but we may not know the size of completed fertility (unless she already has three children). Our strategy will thus be to estimate determinants of the probabilities to belong to each fertility type with the subsample of females aged 41 and above and use these estimates to predict the fertility types of females aged below 41.

More specifically, we base our predictions of the fertility type distribution on the observed (a, n) distribution from the cohorts born between 1955 and 1976.[31] In order to maximize the use of information we have about each household, we estimate the propensities to belong to specific fertility types conditional on the current age, the potentially observed age at first birth, and the current number of children. For females under 41 who have not (yet) given birth, we predict the probabilities to belong to each of the (a, n) combinations, excluding those with an age at first birth a below or equal to the current age of the woman t. For females under 41 for whom we do know their age at first birth a, we only need to predict their completed fertility n, taking into account the number of children the female already has.

Focusing first on females younger than 41 whom we observe without children at model age t, we assume for simplicity that they will not give birth in the current period. Therefore those

[30] In the GSOEP sample used 4.03% of households have more than three children.

[31] This limits the sample to those cohorts who had their first child in 1975 or later, excluding earlier cohorts whose fertility patterns were differently affected by the introduction of contraceptives in the late 1960s. Since 1975, the number of births per woman has been more or less stable in Germany.

observed childless at ages $38 - 40$ are assumed to remain childless and are assigned $f = (\cdot, 0)$. For all younger childless females we use the following algorithm to estimate their probability to give birth to n children in the future:

- For every model age $\tilde{t}$ for which first birth is possible in the next period (until $t = 6$, i.e. ages $35 - 37$):
  - Estimate the multinomial logit (1.17) on the subsample of females with observed $f = (a, n)$, for whom $a > \tilde{t}$ or $a = \cdot$ (first birth *after* $\tilde{t}$ or childless),

$$\mathrm{Prob}(f_i = f|\mathbf{X}_i) = \frac{\exp(\mathbf{X}_i \alpha_f)}{1 + \sum_{k=1}^{19} \exp(\mathbf{X}_i \alpha_k)}, \tag{1.17}$$

    where $\mathbf{X}_i$ includes a constant, a dummy for high education (A-levels and above), and dummy indicating religion at age 20.[32]
  - Predict the propensities for all childless females aged $\tilde{t}$ to belong to each fertility type f.

Focusing second on females for whom we observe age at first birth a and their current number of children at model age t, we again assume that they will not give birth to an additional child in the current period (i.e. taking the number of children as fixed for current period). We then estimate their probability of completing their fertility with n children using the following algorithm:

- For every combination of age at first birth $\tilde{a}$, current number of children $\tilde{n}$, and model age $\tilde{t}$:
  - Estimate a multinomial logit as specified in (1.17) on the subsample of females with observed $f = (\tilde{a}, n)$, for whom $n \geq \tilde{n}$, and with additional births (if any) only occurring at model ages after $\tilde{t}$.[33]
  - Predict the propensities for all females aged $\tilde{t}$ with $\tilde{a}$ and $\tilde{n}$ to complete fertility with n children (i.e. to belong to the corresponding fertility type $f = (\tilde{a}, n)$).[34]

With these estimated fertility type distributions for females below age 41 at hand, we replicate each observation five times and assign a predicted fertility type to each replication according to the estimated probabilities. For a female with three possible fertility types (observed $a = 1$,

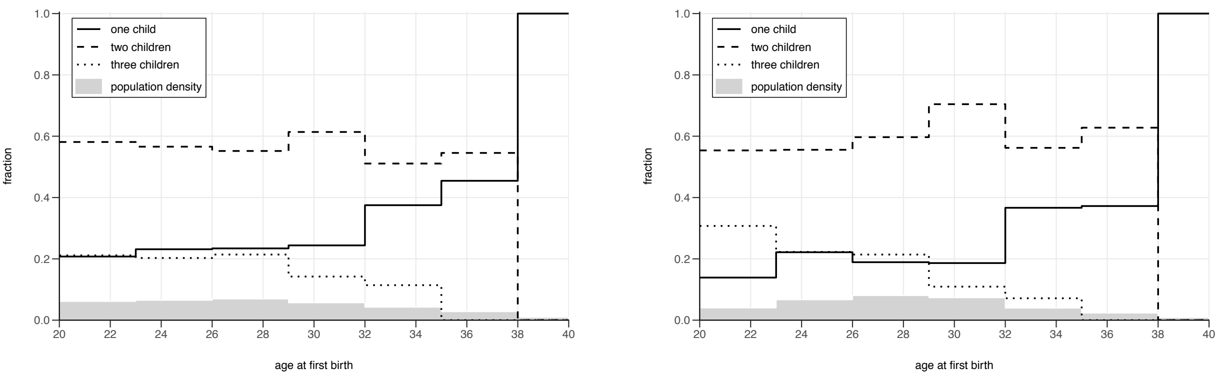
[32]If religion is not observed at age 20, we use the available information closest to age 20.

[33]This effectively limits the number of possible fertility types to at most three ($n \in \{1, \cdots 3\}$) and therefore estimates the probability to complete fertility with n children conditional on a specific a.

[34]See Appendix 1.C.3 for an additional illustrative examples of how the fertility type estimation is conducted.

completed fertility propensities: $n = 1$ with 21%, $n = 2$ with 56%, and $n = 3$ with 23%), we round the propensities to fifths and assign one replication to fertility type $(1, 1)$, three replicates get fertility type $(1, 2)$ and the last replicate gets fertility type $(1, 3)$.[35] This procedure reflects for every woman a fertility distribution over possible types taking uncertainty in the fertility type assignment into account. The expansion is superior to assigning only the most likely fertility type as it creates a considerably more realistic aggregate fertility distribution, while still keeping the number of observations manageable for the counterfactual simulations.

Illustration of results. Figure 1.6 illustrates the outcome of the described fertility type assignment. Figure 1.6(a) plots the observed fertility distribution for females aged 41 and above in the model framework. It can be compared to the fertility distribution directly computed from the data shown in Figure 1.9 in Appendix 1.C.3, which is not adjusted to the time periods used in the model. Figure 1.6(b) on the other hand plots the predicted fertility distributions for females aged below 41. This predicted fertility distribution is shaped (i) by the distribution of the covariates used in (1.17) and (ii) by the partially observed fertility outcomes for the below 41 year olds.



(a) Observed fertility (age 41 and above) (b) Predicted fertility (below age 41)

Figure 1.6: Fertilty distributions

Notes: Omitted childless household shares: panel (a): 11.16%, panel (b): 11.36%. Sample (a): women aged 41 to 65, born between 1955 and 1976, not in education and living with a full-time working partner, first birth (if any) between 20 and 40 and having completed fertility (max. 3 children) by age 41. Sample (b): women aged 20 to 40, born between 1976 and 1995, not in education and living with a full-time working partner, first birth (if any) after 20. Source: 2000 to 2017 GSOEP.

Overall, both fertility distributions are largely similar. However, more females in the younger cohorts have their first child later than the older cohorts, as displayed by the shift in the density

[35]If the rounding to fifths yields a sum of the propensities that is less (greater) then 1, we adjust the rounded values as follows: Starting with the fertility type whose propensity we rounded down (up) the most, we add (substract) one fifth to (from) that fertility type. We continue this procedure with the fertility type with second (and then third/fourth/...) largest rounding adjustement until the sum of the propensities adds up to 1.

(shaded in grey) from the first bracket $20-22$ to later ages. Out of those households aged below 41 who do have their first child early, a higher fraction gets three children and a lower fraction gets just one child. The younger cohorts are slightly more likely to remain childless (11.16 percent observed for the older cohorts vs. 11.36 percent predicted for the younger cohorts).[36] The implied total fertility rate, i.e. number of children per household, is very similar for the predicted younger cohorts (1.70) and the observed older cohorts (1.68).

For the estimation and policy experiments, we will proceed with the fertility types assigned to each household, defined as the actual fertility types for households aged over 41 and the predicted fertility type for younger households.

1.4.4 Calibration of homogeneous preference parameters

In line with Blundell et al. 2016, we set the discount factor β to 0.94 per three-year model period.[37] Those parameters of the utility function (1.1) that do not differ between agents are calibrated to match the participation and compensated intensive-margin labour supply elasticities from the literature (e.g. Chetty, Guren, et al. 2011) as well as data moments of labour supply and childcare take-up. We discuss this further in Section 1.5 where we present the model-data fit. The calibrated values are shown in Table 1.4. We use a log-specification for consumption ($\gamma_c = 1$) and set the CRRA coefficients on leisure γ_L and home produced childcare γ_{dcc} to 2. The leisure endowment $\overline{L}$ is set to 2.5 hours per week and the home produced childcare endowment is set to 10 hours per week, if children below 9 are present. Furthermore, the shifter on the preference for home produced childcare κ is set to 0.25. It implies that households have a much stronger preference to raise toddlers (i.e. children below age 3) at home compared to 3 to 8 year old children which is reflected in the 75% decrease in the preference for home produced childcare. This allows us to capture social norms on the mode of childcare for children below 3.[38]

Table 1.4: Calibrated model parameters

parameter	β	γ_c	γ_L	$\bar{L}$	γ_{dcc}	$\overline{dcc}$	κ
value	0.94	1 (log)	2	2.5	2	10	0.25

[36] See Table 1.14 in Appendix 1.C.3 displays the shares underlying the distributions in Figure 1.6.

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

[38] 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." See GESIS - Leibniz-Institut für Sozialwissenschaften 2017.

1.4.5 Maximum Likelihood Estimation

In our framework, the preference for home produced childcare g, the availability of informal childcare oth, and the taste for female leisure α are all driving forces of female labour supply and childcare decisions during parenthood. Therefore, their distributions determine how female labour supply and women's career paths react to changes in childcare subsidies. In this subsection, we lay out our estimation procedure to identify the underlying joint distribution of permanent, unobserved heterogeneity and provide results.

We first introduce our methodological approach by explaining in detail the components of the likelihood function as well as our assumptions. Then, we outline the identification of the different types of unobserved heterogeneity and summarise the construction of the estimation sample. Finally, we describe our numerical optimisation routine and show the maximum likelihood estimation results of the parameters that govern the joint distribution of unobserved heterogeneity.

1.4.5.1 Methodology

Setup and Likelihood specification. Let us denote by s_n and x_n two vectors of characteristics of individual n that we take from the data. We define $s_n = (t, w_t^M, w_t^W, f)$ as the vector that contains the state variables age, male wage, female wage, and fertility type.[39] [40]

In addition, we denote by x_n the vector of individual characteristics which we allow to influence the joint distribution of the three-dimensional unobserved heterogeneity that is denoted by $l_2(h|x_n)$ where $h = (g, oth, \alpha) \in \boldsymbol{H}$ captures the unobserved heterogeneity and its space $\boldsymbol{H}$. The vectors s_n and x_n might have some overlap. The vector of covariates x_n should only include fixed characteristics as we estimate permanent, unobserved heterogeneity. We discuss the covariates that we use in the estimation in detail in Section 1.4.5.3.

The two observed data outcomes that we intend to fit are the discrete female labour market choice lm_n and the total amount of market childcare used by the household Tcc_n. Both, lm_n and Tcc_n, have a direct counterpart in the structural model: we denote by $\widehat{lm}(s_n, h)$ and $\widehat{Tcc}(s_n, h)$ the respective optimal model-predicted choices.

[39] The subscript n is omitted for the elements of the vector for ease of exposition.

[40] The age of the household t and the male wage w_t^M are taken directly from the data as we restrict our household sample to those with full-time working men. In contrast, the female wage w_t^W and the household's fertility type f might be unobserved. To predict unobserved females wages and fertility types we rely on the Heckman two-step wage estimation (see Section 1.4.2) and the fertility type estimation (see Section 1.4.3).

With these information at hand, we can define the likelihood that household n chooses (lm_n, Tcc_n) given their observed characteristics as:

$$\ell_n = \sum_{h \in \boldsymbol{H}} \Big[l_1(lm_n, Tcc_n | s_n, h) \cdot l_2(h|x_n) \cdot l_3(s_n, x_n) \Big]$$

Note that we discretise the state space of unobserved heterogeneity $\boldsymbol{H}$ (more details on this are in the next paragraph). The *sample likelihood function* that we maximise is then given by:

$$\mathcal{L} = \prod_{n=1}^{N} \ell_n \tag{1.18}$$

The *object of interest* is $l_2(h|x_n)$ which is a discretised version of the joint probability density function of the unobserved heterogeneity conditional on initial characteristics x_n. In the estimation procedure, we find the weights over the grid points $h \in \boldsymbol{H}$, i.e. the parameters that govern the joint distribution function $l_2(h|x_n)$, which maximize $\mathcal{L}$. We describe the underlying functional form and correlation assumptions on $l_2(h|x_n)$ in detail below.

We refer to the first ingredient of the likelihood, $l_1(lm_n, Tcc_n|s_n, h)$, as the model-data match which is defined as an indicator function[41]

$$l_1(lm_n, Tcc_n|s_n, h) = \begin{cases} 1 \textit{ iff} & \widehat{lm}(s_n, h) = lm_n \text{ and } \widehat{Tcc}(s_n, h) = Tcc_n \\ 0 & \text{otherwise.} \end{cases} \tag{1.19}$$

Finally, the third component, $l_3(s_n, x_n)$, is the joint distribution of initial characteristics (s_n, x_n) observed in the data.

Joint distribution of unobserved heterogeneity. We assume that the three marginal distributions are independent conditional on x_n:

$$l_2(\underbrace{g, oth, \alpha}_{=h} | x_n) = l_2^g(g|x_n^g) \cdot l_2^{oth}(oth|x_n^{oth}) \cdot l_2^{\alpha}(\alpha|x_n^{\alpha})$$

where x_n^g, x_n^{oth}, and x_n^{α} are subsets of x_n that might have some overlap. Such overlap creates a correlation between the marginal distributions l_2^g, l_2^{oth}, and l_2^{α} without assuming an explicit correlation structure.

[41]It does not matter if we adjust the weight (currently set to 1), that we assign to each model-data match, multiplicatively because by taking the logarithm of the sample likelihood it only shifts the objective function up or down, but it does not affect the estimated parameter values. This also holds for multiplicative adjustments on the household level.

Marginal distributions of unobserved heterogeneity. We assume that the data-generating process of each type of heterogeneity denoted by $het = \{g, oth, \alpha\}$ is given by:

$$het = \gamma^{het} + x_n^{het}\beta^{het} + u^{het}$$

where $x_n^{het} \subseteq x_n$ is the vector of covariates which we include in the estimation of the heterogeneity type het. We assume that the error term u^{het} follows a normal distribution with zero mean and standard deviation σ^{het}, i.e. $u^{het} \sim \mathcal{N}\left(0, \left(\sigma^{het}\right)^2\right)$. Furthermore, we assume that the three error terms u^g, u^{oth}, and u^α are mutually independent. Any correlation between the dimensions of unobserved heterogeneity will be captured through an overlap of the initial covariates x_n^g, x_n^{oth}, and x_n^α.

This setup implies that each dimension of heterogeneity het is normally distributed conditional on x_n^{het} with a covariate-dependent mean $\gamma^{het} + x_n^{het}\beta^{het}$ and a covariate-independent standard deviation σ^{het}, i.e. $het|x_n^{het} \sim \mathcal{N}\left(\gamma^{het} + x_n^{het}\beta^{het}, \left(\sigma^{het}\right)^2\right)$. From the model setup in Section 1.3 it follows that all dimensions of heterogeneity, g, oth, and α are defined on the closed interval $[0, 1]$. Therefore, we truncate the normal distribution of $het|x_n^{het}$ at 0 and 1. Finally, we need to discretise each continuous marginal distribution to reflect the (numerical) discreteness of the joint distribution $l_2(h|x_n)$ in the likelihood function (1.18).[42]

The ultimate goal of the estimation is to find the parameters $(\gamma^g, \beta^g, \sigma^g, \gamma^{oth}, \beta^{oth}, \sigma^{oth}, \gamma^\alpha, \beta^\alpha, \sigma^\alpha)$ which maximize the sample likelihood function given in equation (1.18).[43]

1.4.5.2 Identification

We do not have a formal proof of identification, but here we provide intuition for identification of the parameters that govern the heterogeneity distribution. For simplicity, let us consider households with a single child below 3, given the same states s_n and initial characteristics x_n. We observe four different groups of households with respect to their female labour market choices and market childcare use which allow us to disentangle the different dimensions of heterogeneity:

1. First, the woman works and the amount of market childcare is sufficient to cover her working time.
2. Second, the woman does not work and, nonetheless, the household consumes a positive amount of market childcare.

[42] For all types of heterogeneity we use an equidistant grid between 0 and 1. To obtain well-defined probability mass functions $l_2^g, l_2^{oth}, l_2^\alpha$ we first assign to each grid point the mass of the normal distribution based on equally long intervals around each grid point. Then, we normalize the sum of the weights to 1.

[43] The number of estimated parameters in the likelihood function depends on the number of covariates used in the estimation as β^{het} has the same size as x_n^{het} for $het \in \{g, oth, \alpha\}$.

3. Third, the woman works, but the household does not buy enough market childcare to cover her working hours.

4. Finally, the woman does not work, and does not buy any market childcare, i.e. home produced childcare covers the full childcare need.

Otherwise equivalent households that buy the same amount of market childcare, but choose different female labour supply – comparing group 1 to group 2 – help us to identify the heterogeneity in the taste for leisure α. To identify informal childcare availability oth, the setup allows us to compare households with the same female work decisions, but who choose different amounts of market childcare (comparing group 1 to group 3). Lastly, home produced childcare preferences g can be identified by comparing households with the same female labour supply, but different market childcare choices (group 2 vs. group 4) as well as by contrasting households with different labour supply and market childcare choices (group 1 vs. group 4).

1.4.5.3 Data and initial characteristics x_n

Data and sample definition. We conduct the maximum likelihood estimation with the 2017 GSOEP cross-section under the same sample restrictions as described in Section 1.2.2. Furthermore, we use in the maximum likelihood estimation only females who have at least one child below 9. This leaves us with an estimation sample of 628 households. These households are replicated 5 times to reflect our estimated fertility distribution as detailed in Section 1.4.3.

Initial characteristics x_n. The initial covariates x_n should be constant characteristics as we assume that the unobserved heterogeneity h is permanent. The choice of initial covariates x_n is important in two ways: first, the initial covariates should be chosen to have predictive power in the estimation of the unobserved heterogeneity as well as capturing possible correlation between the types of heterogeneity. Second, the vector x_n might contain elements of s_n that are fixed over the life cycle (e.g. fertility type). Alternatively, x_n might include some fixed, initial characteristics of the household that explain time-invariant components of elements of s_n that do also have a time-varying component (e.g. using education as a fixed characteristic to capture the time-invariant part of wages).

In our current setup we condition each of the marginal distributions l_2^g, l_2^{oth}, and l_2^α on two dummy variables that work as mean shifters of the respective distribution:[44] (i) we allow the distribution of the preference for home produced childcare l_2^g to depend on education of the mother and living in former East Germany. We choose the former as it might explain a permanent component of female wages which are an element of s_n and the latter as we expect

[44]In future versions of this paper, we intend to improve upon the current setup by extending the set of covariates (esp. by including maternal and paternal education as well as approximate measures of the fertility type such as early/late first birth and number of kids when fertility is completed).

the covariate to increase the predictive power of the estimation. (ii) The distribution of other childcare availability l_2^{oth} depends on living in former East Germany and facing completed fertility of 2 or more children. The former is used to increase explanatory power and the latter is a measure of the fertility type which is part of s_n. (ii) The distribution of leisure preferences l_2^{α} depends being catholic at the age of 20 and facing completed fertility of 3 children. Similarly, being catholic has predictive power for the taste of leisure and having three children works as a measure of the fertility type.

1.4.5.4 Optimisation routine

To solve the optimisation problem numerically, we use a basin-hopping algorithm in combination with a Matlab built-in minimisation routine for unconstrained target functions (*fminunc*). 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.[45]

Intuitively, the procedure works as follows: we set an (arbitrary) initial starting point and solve for a (possibly local) minimum using *fminunc*. 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 induces a random perturbation of the parameters of the previously found (potentially local) minimum and restarts the minimization routine *fminunc* at the perturbed parameters.[46] The basin-hopping algorithm then compares the new minimum found to the previously found one and records the point with the lowest target function value as a candidate for the global minimum. The algorithm repeats the procedure described above, 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 optimum 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 50.[47]

1.4.5.5 Results

The estimated coefficients γ, β and σ are shown in Table 1.5. In addition, we plot the resulting conditional truncated normal probability density functions for each heterogeneity in Figure 1.7.

First, considering the preference for home produced childcare, the estimates show that – as one might intuitively expect – higher educated women have on average a lower preference to take care of their children at home, i.e. the mean is shifted by -0.67. Similarly, women that live

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

[46]The precise mechanism that determines the size of the perturbation is described in the Appendix 1.C.4.

[47]We intend to evaluate the robustness of our results to the number of iterations in the future.

Table 1.5: Maximum likelihood estimates

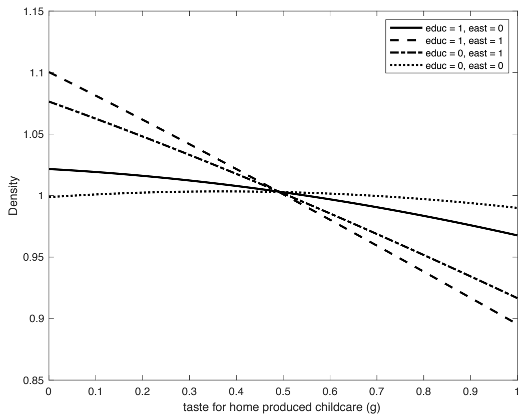
	home produced childcare (l_2^g)	availability of other childcare (l_2^{oth})	leisure (l_2^α)
γ	0.37	-1.84	3.06
β_{educ}	-0.67		
β_{east}	-2.24	-4.36	
$\beta_{\text{compl. fert. 2+}}$		1.40	
β_{catholic}			-1.02
$\beta_{\text{compl. fert. 3}}$			1.37
σ	3.84	0.92	1.05

Notes: The dummy covariates are education of the woman (educ), currently living in former East Germany (east), having a fertility type with two or more children (compl. fert. 2+), woman being catholic at the age of 20 (cath), and having a fertility type with three children (compl. fert. 3). Sample as defined in Section 1.4.5.3.

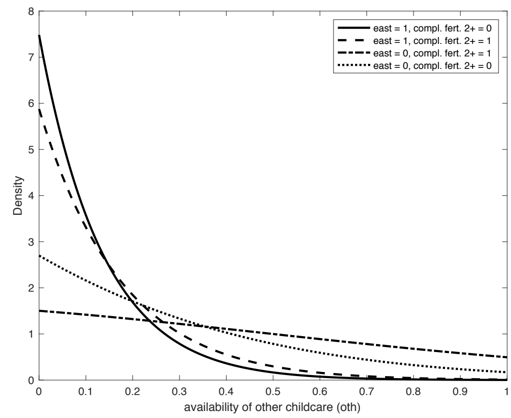
in former East Germany tend to have lower preferences for home produced childcare as those in former West German states which is reflected by a mean shifter of -2.24. Living in former East or West Germany seems to be an important driver: the shift between living in former East instead of former West Germany towards lower preferences for home produced childcare is larger compared to having higher instead of lower education conditional on living in East/West Germany. Therefore, we observe in Figure 1.7(a) that the density at lower values of the taste for home produced childcare is highest for high educated women living in former East Germany.

Second, we estimate the distribution of availability of other childcare conditional on living in former East Germany and having a fertility type f_n that corresponds to completed fertility of two or more children. Living in former East Germany implies that fewer other childcare is available as the mean is shifted by -4.36. This is in line with lower preferences for home produced childcare in former East Germany as both effects might be driven by cultural differences. People providing other modes of childcare such as grandparents, friends, or neighbours seem to share the lower preferences for home produced childcare as well. A potential consequence would be that they restrict their supply of other childcare to the household.[48] This rationalises the considerably higher density for low availability of other childcare in former East Germany as shown in Figure 1.7(b). Furthermore, households with more than two children face higher

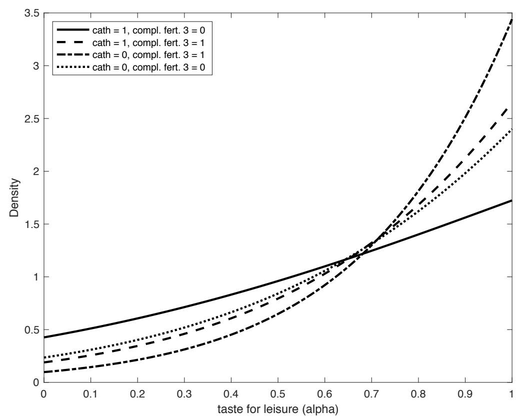
[48] Additionally, young families might face unalike difficulties in finding childcare slots in East or West Germany (i.e. rationing of childcare slots might be more relevant in West Germany). We might partially also capture the differential shortage of childcare slots in this estimation.



(a) Conditional pdfs of taste for home produced childcare (g)



(b) Conditional pdfs of availability of other childcare oth



(c) Conditional pdfs of taste for leisure α

Figure 1.7: Estimated conditional marginal probability density functions (pdfs) of each heterogeneity

Notes: The distributions are conditional on dummy variables (e.g. 'educ = 1, east = 0' indicates that the function is conditional on the woman having high education and not living in former East Germany). The dummy covariates are education of the woman (educ), currently living in former East Germany (east), having a fertility type with two or more children (compl. fert. 2+), woman being catholic at the age of 20 (cath), and having a fertility type with three children (compl. fert. 3). Sample as defined in Section 1.4.5.3.

availability of other childcare as the mean of the distribution is shifted to the right by 1.40. This may reflect that those who live closer to grandparents also tend to have more children.

Third, the marginal distributions for the taste of leisure are estimated conditional on the woman being catholic at the age of 20 and having completed fertility of three children. Catholic women tend to have lower preferences for leisure with a mean shifter of -1.02. Furthermore, women with completed fertility of three children tend to have higher preferences for leisure as their mean is shifted by 1.37. Both effects are visible in Figure 1.7(c).

1.5 Model Fit

As explained in the previous section, we estimate the joint distribution of unobserved heterogeneity and calibrate the homogeneous parameters of the utility function to achieve a good fit of the estimated model with empirical moments of interest. We evaluate the ability of the estimated structural model to match the data moments of the two choices, female labour supply lm and total use of market childcare Tcc. The homogeneous utility parameters are calibrated to match empirical estimates of the participation elasticity as well as the compensated intensive margin labour supply elasticity taken from Chetty, Guren, et al. 2011 (see Section 1.4.4).

First, Table 1.6 shows the model and data moments of female labour supply. We condition on the age of the youngest child in the household. Overall, we are able to fit the labour supply pattern well conditional on child age. The estimation is only a bit off in two directions: first, we overestimate part-time participation of mothers with children between 0 and 2 and therefore, underestimate non-participation as well as full-time work. Second, we overestimate non-participation of mothers with children aged 6 to 8 and underestimate the share of full-time working mothers.

Table 1.6: Labor supply shares conditional on youngest child's age

	Children $0-2$			Children $3-5$			Children $6-8$		
	NP	PT	FT	NP	PT	FT	NP	PT	FT
Model	0.35	0.54	0.11	0.18	0.65	0.17	0.18	0.65	0.17
Data	0.37	0.48	0.15	0.19	0.64	0.17	0.15	0.64	0.21

Notes: Sample as described in Section 1.4.5.3, NP corresponds to non-participation, PT to part-time work, and FT to full-time work.

Second, we evaluate the model-data fit of the total market childcare take-up of households conditional on 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. Therefore, we normalise for each household their total market childcare take-up by the household's total childcare need. This yields the percentage of childcare need that a household covers with market childcare. For example, consider a household with children aged 2 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 20 hours per week, the household will cover 62.5% of their total childcare need by market childcare ($\frac{50}{80} = 0.625$). As their coverage is higher than 50%, we classifying their market childcare take-up as *high*.

Table 1.7 shows the aggregate shares of households who cover less or more than 50% of their total childcare need with market childcare conditional on the age of the youngest child in the household. We achieve an exact fit of the shares if the youngest child is between 0 and 2. For households whose youngest children are 3 to 5, our fit differs only by about 10 percentage points. For children aged 6 to 8, we are not able to achieve a precise fit with the current calibration.[49]

Table 1.7: Total market childcare take-up shares conditional on youngest child's age

	Children $0-2$		Children $3-5$		Children $6-8$	
	low	high	low	high	low	high
Model	0.58	0.42	0.27	0.73	0.42	0.58
Data	0.58	0.42	0.17	0.83	0.62	0.38

Notes: Sample as described in Section 1.4.5.3. *low* if the household covers less or equal than 50% of the required aggregate childcare time by market childcare. *high* if the coverage is larger than 50%.

Third, we match the Hicksian (compensated) intensive margin labour supply elasticity as well as the participation elasticity because getting female labour supply responses right is crucial for our policy experiments in Section 1.6. To obtain the model-counterpart for the empirical estimates, we conduct a policy experiment which increases the gross female wage rate permanently by 1% and calculate the (average) labour supply responses along both margins. The resulting participation elasticity is 0.2 which is perfectly in line with estimates from the quasi-experimental literature as Chetty, Guren, et al. 2011 report 0.26 as the result of their meta-analysis of quasi-experimental studies for the extensive margin elasticity. Our model predicts a Hicksian (wealth constant) intensive margin elasticity of 0.195 which is also in line with the literature (see e.g. Chetty, Guren, et al. 2011).[50] [51]

[49] As this illustration is limited due to the high level of aggregation, we intend to look at the fit on a more continuous basis in the future.

[50] We obtain the model-based Hicksian intensive margin elasticity by following the procedure in Saez 2001 which is explained in detail in the Appendix 1.D.1.

[51] Chetty, Guren, et al. 2011 report a mean of their meta-analysis of quasi-experimental studies of 0.15. They adjust this estimate to achieve a preferred structural Hicksian labour supply elasticity that reflects fixed costs of adjusting working hours. They report this estimate as the main outcome with a value of 0.33 which is still in line with the magnitude of our estimate. Furthermore, Chetty, Guren, et al. 2011 also believe that small differences can occur due to noise or specification choices.

1.6 Fiscal calculation and Policy Experiments

In this section, we use our estimated structural model to conduct counter-factual policy experiments. We are interested in the net fiscal effects of changes in childcare subsidies, as described in Section 1.3.6: to what extent is a subsidy increase self-financing through higher income taxes paid by households? This might differ substantially if the changes in the subsidies are targeted or untargeted. We therefore consider different policy scenarios: first, we consider an untargeted increase in childcare subsidies.[52] Second, we condition the subsidy increase on the mother working at least part-time. Third, we combine the conditioning on employment of the mother with targeting specific child age brackets (below 6 or 3 years). Fourth, we simulate an increase in subsidies that are targeted towards full-time working mothers. In all scenarios, we are deliberately not keeping the budget balanced as our objective is to assess whether households' responses in the short term and in the long term provide extra tax revenues that offset the increase in subsidy spending.

In every policy scenario, we distinguish three ways to answer our postulated question:

(1) We focus on the amount of subsidy spent and income taxes received in the year of the policy introduction and call this setup *Impact period*.

(2) We simulate the remaining life cycle starting from the current age and wage quintiles of the household (as observed in 2017) and derive the net fiscal effect of changes in childcare subsidies. As opposed to (1), this accounts for dynamics and we refer to this setup as *Remaining life cycle*. Note that this setup also includes the effect of the impact period.

(3) We simulate the entire life cycle of each household and calculate the degree to which the policy change is self-financing.[53] As opposed to (2), this setup accounts for anticipation effects and is called *Entire life cycle*. In addition, it includes past years of child-rearing of households who currently have older children. In this setup, the policy change was already in place during those years and might have affected the household's decision.

Distinguishing (1) and (2) allows to determine the relative importance of static and dynamic effects as highlighted in Section 1.3.6. Whether (2) or (3) is the appropriate measure depends

[52] In all policy scenarios, the subsidy is set to 50 EUR for each child who attends childcare full-time (40 hours per week). As the market childcare price is linear in hours, the subsidy is assumed to be linear in hours as well: for example, the subsidy will be 25 EUR for a family whose child attends childcare part-time (20 hours per week).

[53] As we observe households at different current ages, we set their age to the first model period (20 - 22) and simulate their entire life cycle. To provide starting values for male and female wages, we assume that each household starts its life in their wage quintiles observed in 2017.

on whether the reform is anticipated by households who do not have children at the time of the reform.

Untargeted childcare subsidy. Table 1.8(i) reports the degree to which an untargeted increase of childcare subsidies is self-financing. If we consider the impact period, the increased tax revenues only make up for 1.7% of the increased spending in childcare subsidies. Focusing on the remaining life cycle of all households, the number rises to 2.6%, but still only a small share of the expenditures is recovered. Simulating the entire life cycle of households increases the effect further to 3.7%. We provide a decomposition of the static labour supply responses in Table 1.15 in Appendix 1.E. The share of households that are marginal in their labour supply decision is 0.22% of which 0.17 percentage points increase their labour supply and 0.05 decrease it.[54] Some households affect the fiscal return even though they are inframarginal in their labour supply decision but because they use a positive amount of market childcare after the reform. In case of an untargeted increase of childcare subsidies, their share is 82.23%. This includes households that increased their amount of market childcare in response to the reform.

In addition, we calculate the effect separately for each female wage quintile. In the impact period, the effects are homogeneous between the first and the fourth wage quintile. For the highest female wage quintile we however observe a negative degree of self-financing. We need to interpret this result with caution as the income effect of households with women in the highest wage quintile appears to be sizeable, which implies that some mothers reduce their labour supply and consume more leisure as childcare gets cheaper. This reduces tax revenues. Shifting the focus to the dynamics, the largest difference in the degree of self-financing between the impact period and the remaining life cycle arises for the first wage quintile which hints towards a larger dynamic tax effect for lower maternal wages. Once we consider the entire life cycle, the degree of self-financing is almost equal for all wage quintiles.

Finally, we investigate how the degree of self-financing varies with (anticipated) family size in Table 1.25(i) in Appendix 1.E. We generally observe that the degree of self-financing increases in the number of children, ranging from 1.1% for one child to 2.2% for three children for the impact period. The explanation could be the following: having three children over the life cycle implies that the subsidy is (in some periods) higher in absolute terms for these families compared to families with fewer children making market childcare for them cheaper and therefore, lowering their opportunity cost of work to a larger extent.[55] Furthermore, as mothers with many children

[54]Some households decrease their female labour supply because at their childcare demand the income effect outweighs the substitution effect.

[55]The reduction in childcare spending of the household is up to 118.75 EUR per month in case three children attend full-time childcare. Note that the oldest one can attend childcare at maximum for 15 hours per week. For families with two children it is 100 EUR and with one child it is 50 EUR.

tend to stay longer at home, if a policy is able to incentivise some of those mothers to return to work earlier, it generates a lot of tax revenues compared to the baseline scenario.

Table 1.8: Self-financing degree of changes in monthly full-time childcare subsidies

		female wage quintile				
	total	Q1	Q2	Q3	Q4	Q5
(i) *untargeted 50 EUR*						
Impact period	1.7%	2.3%	2.0%	1.9%	2.5%	-0.2%
Remaining life cycle	2.6%	3.8%	3.0%	2.7%	2.8%	0.9%
Entire life cycle	3.7%	3.6%	3.7%	3.8%	3.8%	3.7%
(ii) *work contingent 50 EUR*						
Impact period	12.3%	7.6%	11.0%	12.3%	16.9%	13.4%
Remaining life cycle	17.4%	17.1%	17.2%	17.2%	19.2%	15.9%
Entire life cycle	20.2%	19.1%	20.0%	20.3%	21.0%	20.3%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values. See Tables 1.20 and 1.21 in Appendix 1.E for a detailed decomposition into subsidy expenditures and tax revenues.

Childcare subsidy contingent on maternal work. As we have seen above, an untargeted subsidy does not have a strong effect on the aggregate female labour supply and thereby generates only little additional government revenue. We now consider childcare subsidies that are targeted towards households with a working mother (part-time or full-time).

Table 1.8(ii) shows that for all three setups (1) - (3) work contingent childcare subsidies are considerably more self-financing than untargeted subsidies. In the impact period, 12.3% of overall childcare subsidies are refinanced through income tax revenues generated from increased maternal labour supply. This number increases to 17.4%, when we simulate households over their remaining life cycle, and further to 20.2%, when we focus on the entire life cycle of households.

As is visible from the net fiscal effects, the labour supply responses are considerably stronger in this policy scenario compared to the untargeted subsidy – especially at the extensive margin, i.e. switches from non-participation to part-time work (see Table 1.16 in Appendix 1.E). Compared to the untargeted subsidy, the share of households which are marginal in their labour supply decision triples to 0.62%. Of these, 0.60 percentage points increase their labour supply and

0.02 decrease it. The share of households that are inframarginal in their labour supply decision but buy a positive amount of market childcare after the reform amounts to 63.05%.[56]

Furthermore, in the impact period we observe an increase of the net fiscal effect with females wages. When we consider the remaining life cycle, it is prevalent that the differences across wage quintiles vanish. The difference between both results points to the importance of the dynamic tax effect which is most sizeable at the lowest wage quintile. This is in line with the observation in Section 1.4.2 that switching from non-participation to part-time work increases the expected future wage of women in the lower wage quintiles to a greater extent than for women in the higher wage quinitles.

Finally, we calculate the fiscal return conditional on (anticipated) family size (see Table 1.25 in Appendix 1.E) and observe the same pattern as for untargeted subsidies: families with three children create the largest net fiscal effect which increases from 16.4% in the impact period to 22.4% over the remaining life cycle and 24.9% considering the entire life cycle. The effect for households with two children is approximately 5 to 6 percentage points lower in all setups, whereas it is about 7 to 8 percentage points lower for one-child families.

Overall, making childcare subsidies maternal work contingent improves the fiscal return substantially from about 2.6% to 17.4% considering the remaining life cycle of the households (in the impact period it increases from 1.7% to 12.3%).

Childcare subsidy contingent on maternal work and child age. We further allow the policy to be targeted towards children within a specific age range – in addition to maternal employment. If the subsidies are restricted to cover children below the age of 6, results change only slightly compared to the policy that includes also children aged 6 to 8 (compare Table 1.9(i) with Table 1.8(ii)). Table 1.9(ii) shows that if we condition the eligibility of childcare subsidies on the employment of the mother and the child being between 0 and 2, the net fiscal return increases by about 4 percentage points in all three setups compared to conditioning exclusively on maternal work.

All in all, targeting the childcare subsidies on maternal work *and* towards children below the age of 3 increases the degree of self-financing of childcare subsidies further.

Childcare subsidy contingent on maternal full-time work. Finally, Table 1.9(iii) shows results for the policy scenario in which childcare subsidies are targeted towards full-time working mothers, creating strong full-time work incentives. The degree to which childcare subsidies finance themselves increases to 52.8% when we consider the impact period. Over the remaining life cycle of the households, it increases further to 75.7% and will reach 79.0% if we consider the

[56]These are households where the female does participate in the labour market such that they receive the additional subsidy, but they do not change their labour supply in response to the subsidy.

Table 1.9: Self-financing degree of alternative changes in monthly full-time childcare subsidies

	total	female wage quintile				
		Q1	Q2	Q3	Q4	Q5
(i) *work and child age < 6 contingent 50 EUR*						
Impact period	12.6%	7.6%	11.1%	12.8%	17.3%	13.5%
Remaining life cycle	17.6%	17.2%	17.4%	17.6%	19.5%	15.9%
Entire life cycle	20.5%	19.3%	20.3%	20.7%	21.4%	20.6%
(ii) *work and child age < 3 contingent 50 EUR*						
Impact period	16.8%	9.9%	13.3%	15.5%	21.6%	21.2%
Remaining life cycle	23.3%	21.0%	20.7%	21.7%	26.5%	24.7%
Entire life cycle	25.4%	23.3%	24.9%	25.6%	26.8%	26.7%
(iii) *full-time work contingent 50 EUR*						
Impact period	52.8%	50.5%	53.1%	52.6%	53.9%	52.8%
Remaining life cycle	75.7%	79.4%	75.4%	74.8%	77.3%	73.2%
Entire life cycle	79.0%	77.3%	78.5%	78.9%	80.2%	79.9%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values. See Tables 1.22, 1.23, and 1.23 in Appendix 1.E for a detailed decomposition into subsidy expenditures and tax revenues.

households' entire life cycle. The effects are very homogeneous across female wage quintiles. The sizeable increase compared to the previous policy scenarios can be explained by two effects: First, the share of households that are inframarginal in their labour supply decision and consume a positive amount of market childcare decreases to 12.97% of all households. This share is substantially lower than in the case of work contingent childcare subsidies, which was 63.05%. The aforementioned reduction in the share of inframarginals implies a sizeable decrease in the amount of subsidies that the government spends (compare Table 1.24 and Table 1.21). Second, the share of households that increase their labour supply from part-time to full-time work is 0.50% which implies that the income tax revenues decrease only slightly compared to the work contingent subsidies.

Table 1.26 in Appendix 1.E shows additional results differentiated by the (anticipated) family size. The same patterns as for work contingent policies occur: families with three children show a higher degree of self-financing as families with fewer children and the difference will get amplified if we move from the impact period to the remaining life cycle.

In summary, this section documents results of untargeted and targeted childcare subsidies and emphasises that targeting childcare subsidies is a useful tool to increase the ability of these policies to be self-financing. Specifically, an untargeted policy is self-financing by 2.6% when we consider the remaining life cycle of all households. The value increases to 17.4% if the subsidy is targeted towards working mothers. It surges dramatically to 75.7% for a refined targeting towards full-time working mothers.

1.7 Conclusion

The recent literature has shown that career breaks due to child-rearing have a major and long-lasting impact on maternal earnings (e.g. Kleven, Landais, Posch, et al. 2019). Various family policies attempt to dampen those effects to achieve – among other things – gender equality by increasing female labour force participation and wages. Childcare subsidies are one such policy and aim to enable mothers to stay in work or return to work more easily during the child-rearing years. This yields higher accumulation of human capital and thus a rise in future maternal wages and participation. The increase in (current and future) maternal labor supply as well as future wages increases generated income tax revenues and therefore impacts the net fiscal cost of childcare subsidies. The question at the heart of this paper is: *to what extent are childcare subsidies dynamically self-financing through increased income tax revenues?*

To answer this question, we develop a structural household model in which households choose female labour supply and market childcare use. We estimate the model based on German survey data, the German Socio-Economic Panel (GSOEP). Each household's choices depend on various unobserved and observed household characteristics: first, on observables such as the age and the wage of both partners; second, on the (potentially) unobserved, exogenous fertility type of the household; third, on unobserved heterogeneity, namely the taste for home produced childcare, the taste for female leisure, and the access to other modes of childcare (e.g. by grandparents, relatives, friends). We estimate the joint distribution of unobserved heterogeneity by maximum likelihood. We show that our estimated model fits the empirical moments well and predicts reasonable participation and compensated labour supply elasticities.

To predict the net fiscal return of childcare subsidies, we conduct counter-factual policy scenarios and determine in each scenario which mothers change their behaviour due to the reform and how. Untargeted increases in childcare subsidies generate only a small positive effect on income tax revenues, i.e. 2.6% over the remaining life cycle of all sampled households. We then document that targeting childcare subsidies is a useful tool to increase their ability to be self-financing. First, targeting subsidies towards working mothers increases the degree of self-financing to 17.4% over the remaining life cycle. Second, if the increase in childcare subsidies was contingent on maternal employment and targeted towards children below the age of 3, the result would increase to 23.3%. Finally, if we conditioned the policy to depend on full-time working mothers, it would dynamically refinance itself by 75.7% over the remaining life cycle. In general, we find that the dynamic component of the raised income taxes is higher for lower female wages. Furthermore, the degree of self-financing is the larger the more children the family has.

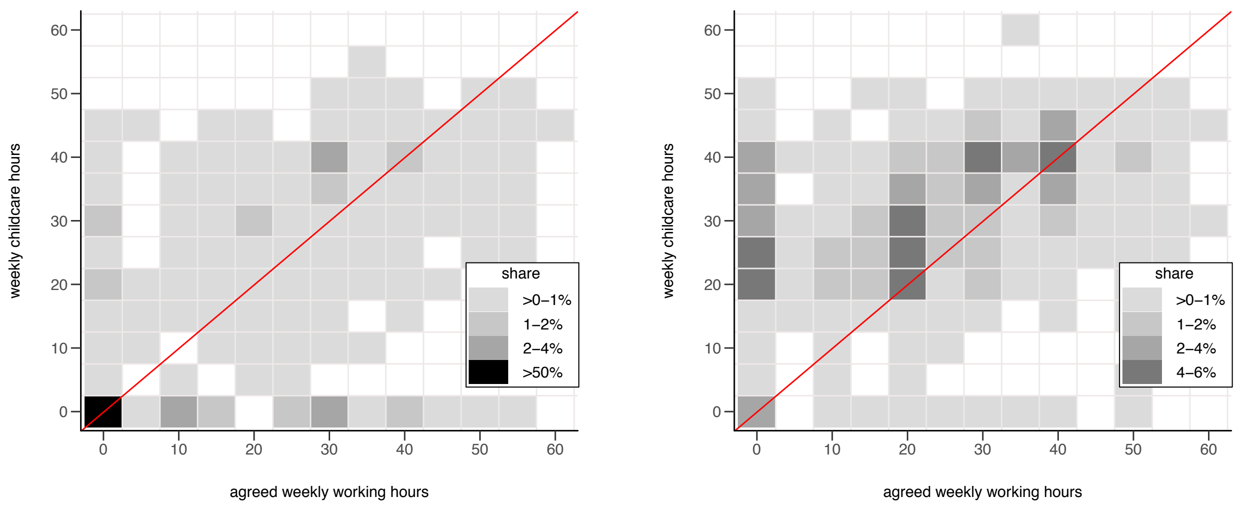
In future work, we consider including rationing of childcare slots on a regional level using administrative German data. Then, we could extend our analysis to policies that increase the provision of childcare slots and determine their dynamic net fiscal costs. Regional variation could also help to refine the estimation of childcare prices. In addition, we plan to incorporate the occupation of the mother to capture that the female wage process – an important ingredient of our model – might be driven by an occupation-specific human capital process.

Appendix

1.A Estimation Samples

In addition to the sample restrictions described in Section 1.2.2, we condition on observing the following covariates for every female: education, religion, state, living in an urban or rural area, hourly wages if working, the number of siblings as well as whether the household lives in one of the partners' home towns. Furthermore, we drop households that are either in the top 1% or bottom 1% of the male or female wage percentiles.

1.B Additional Stylised Facts



(a) Children aged 0-2

(b) Children aged 3-5

Figure 1.8: Mothers working hours vs. childcare hours by age of the child

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 to 2 for panel (a) or 3 to 5 for panel (b). Source: 2009 to 2017 GSOEP.

Childcare hours versus working hours Concerning the joint decision making of mothers concerning the use of market childcare and labour supply, Figures 1.8(a) and 1.8(b) plot agreed weekly working hours of the mother against the number of hours the children spend in childcare. To focus on the trade-off faced by the female, the sample is restricted to married mothers whose partner works full-time.

As Figure 1.8(a) shows, more than half of mothers of below 3 year olds neither work nor send their child to childcare. Of those who do work, not everybody sends their child to childcare at the same time (mass at zero childcare hours and positive working hours). Furthermore, we observe some children in childcare whose mothers are not working at all.

For the 3 to 5 year olds in Figure 1.8(b), the picture looks entirely different. Most children attend childcare and many mothers work to some degree. Similar to the below 3 year olds, a number of mothers does not work while their child is in 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 childcare.

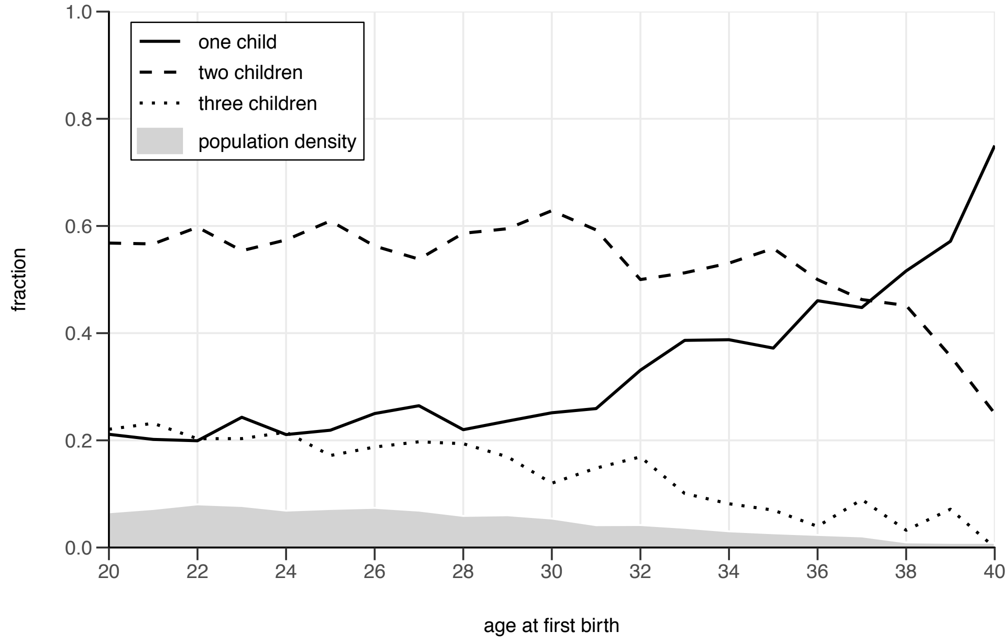


Figure 1.9: Completed fertility across age at first birth

Notes: Populations density illustrates the relative frequencies of age at first birth in the sample. Sample: mothers aged 41 to 65 that are not in education and live with a full-time working partner, conditional having given birth only between ages 20 and 40 to at most three children. Source: 2000 to 2017 GSOEP.

Completed fertility Figure 1.9 plots completed fertility size across the age of the mother at first birth. It helps to understand when in their life cycle mothers start having children and how many. Most mothers give birth for the first time in their early twenties and end up completing fertility with two children. Naturally, those who give birth later have a higher propensity to only have one child. Furthermore, having three children is more common if the first birth occurs early in life.

1.C Quantification of the Model

1.C.1 Childcare prices

1.C.1.1 Determinants of childcare prices

Child age. One of the important determinants of childcare prices 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 1.2.1. Prices are usually higher for younger children since costs of nurseries are higher than those of kindergartens.

Regional variation. Childcare prices in Germany differ further on a regional level because of two reasons: First, the price 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 childcare is part of the education system, different federal states have implemented different childcare pricing regulations. So far, we do not include regional variation in our empirical and structural model, but we intend to do so in the future. Currently, eleven of the sixteen federal states allow for partial or full reduction of parental contributions conditional on the children's age cohort and hourly use of day care.[57]

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

1. *Household income:* In eleven out of sixteen states the household income is used as a determinant and in two out of sixteen states it can be used. [59]

2. *Number of children in the household:* In twelve out of sixteen states, childcare prices 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.

[57]See Authoring Group Educational Reporting 2018: *Education in Germany 2018*, Section C2, p. 70–71 and Tables C2-3a.

[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 (e.g. current year versus previous years).

3. *Agreed hours of day care:* In eight out of sixteen states, the legislation also requires the prices to be determined on the agreed hours of day care between the parents and the day care centre/kindergarten.

1.C.2 Wage processes

Mincer-type covariates in wage equation. In the wage equation (1.13) we include the following Mincer-type covariates: Linear and quadratic terms for age, full-time work experience, and part-time work experience. Furthermore, we include dummies for different education levels, namely a dummy for a lower track school degree and vocational training, a dummy for a high school degree, a dummy for a university degree. The left out category includes individuals with no school degree or a lower track degree without vocational training.

Identifying assumptions. The underlying identifying assumption of the selection model is that the exclusion restrictions must affect female wages w^f only through participation. Controlling for education, age, and the labour market experience of the female, the husbands wage quintile and the net household income in case of non participation, which is partially driven by the tax and benefit system, can both be plausibly excluded from the wage equation. The presence of (young) children however is likely to affect the mothers wage through other channels than participation, for example by inducing the mother to opt for a lower paying job with more child-friendly working hours. Therefore we include the number of (small) children in the wage equation and only include dummies that reflect the detailed child age structure as exclusion restrictions.

Table 1.10: Female wage quintiles

t	$w_t^W = 1$	$w_t^W = 2$	$w_t^W = 3$	$w_t^W = 4$	$w_t^W = 5$
1	5.26	9.40	11.23	13.16	17.31
2	7.08	10.99	12.93	15.01	20.18
3	7.88	11.99	14.44	17.17	23.91
4	8.18	12.81	15.78	19.03	26.41
5	8.21	13.34	16.69	20.26	28.46
6	8.12	13.19	16.73	20.59	29.78
7	7.86	12.93	16.72	20.83	30.58
8	8.10	13.16	16.69	20.88	30.80
9	8.03	12.96	16.61	20.87	31.45
10	7.79	12.62	16.41	20.87	31.70
11	7.65	12.30	15.80	20.37	31.36
12	7.68	11.99	15.16	19.55	31.04
13	7.40	11.30	14.11	18.46	30.12
14	7.18	11.08	13.44	17.50	30.29
15	6.57	10.63	13.08	16.75	24.42

Notes: Gross hourly wages in 2017 prices. Sample as defined in Section 1.2.2. Source: 2000 to 2017 GSOEP.

Table 1.11: Male wage quintiles

t	$w_t^M = 1$	$w_t^M = 2$	$w_t^M = 3$	$w_t^M = 4$	$w_t^M = 5$
1	8.86	12.47	15.25	18.11	26.26
2	9.63	13.50	16.35	19.78	28.96
3	10.66	15.16	18.46	22.36	33.14
4	11.80	16.79	20.76	25.71	37.96
5	12.34	18.03	22.64	28.23	42.66
6	12.46	18.15	23.08	29.10	44.19
7	12.66	18.48	23.35	29.89	44.73
8	12.79	18.75	23.78	30.53	45.94
9	12.50	18.53	23.54	30.43	46.27
10	12.15	18.35	23.33	29.95	45.69
11	12.33	18.23	23.29	30.22	46.16
12	11.93	18.37	23.64	31.21	48.14
13	11.71	18.31	23.96	31.88	47.49
14	11.57	18.34	24.49	33.69	48.50
15	12.03	18.35	23.74	32.43	51.76

Notes: Gross hourly wages in 2017 prices. Sample as defined in Section 1.2.2. Source: 2000 to 2017 GSOEP.

Table 1.12: Female wage quintile transition probabilities

lm = NP	$w_{t1}^W = 1$	$w_{t1}^W = 2$	$w_{t1}^W = 3$	$w_{t1}^W = 4$	$w_{t1}^W = 5$
$w_{t0}^W = 1$	0.8230	0.1025	0.0326	0.0268	0.0151
$w_{t0}^W = 2$	0.4243	0.4294	0.0689	0.0461	0.0313
$w_{t0}^W = 3$	0.2927	0.2790	0.2953	0.0731	0.0598
$w_{t0}^W = 4$	0.1846	0.1401	0.2522	0.2909	0.1321
$w_{t0}^W = 5$	0.1449	0.0871	0.1342	0.2801	0.3538
lm = PT	$w_{t1}^W = 1$	$w_{t1}^W = 2$	$w_{t1}^W = 3$	$w_{t1}^W = 4$	$w_{t1}^W = 5$
$w_{t0}^W = 1$	0.4136	0.2265	0.1739	0.1081	0.0779
$w_{t0}^W = 2$	0.2798	0.2512	0.2262	0.1458	0.0970
$w_{t0}^W = 3$	0.1689	0.1942	0.2817	0.2215	0.1337
$w_{t0}^W = 4$	0.1196	0.1344	0.2241	0.2964	0.2256
$w_{t0}^W = 5$	0.0967	0.0991	0.1532	0.2545	0.3965
lm = FT	$w_{t1}^W = 1$	$w_{t1}^W = 2$	$w_{t1}^W = 3$	$w_{t1}^W = 4$	$w_{t1}^W = 5$
$w_{t0}^W = 1$	0.3308	0.2536	0.1719	0.1436	0.1001
$w_{t0}^W = 2$	0.1867	0.2807	0.2310	0.1845	0.1171
$w_{t0}^W = 3$	0.0882	0.1633	0.2915	0.2865	0.1705
$w_{t0}^W = 4$	0.0466	0.0822	0.1902	0.3569	0.3241
$w_{t0}^W = 5$	0.0274	0.0399	0.0951	0.2480	0.5896

Notes: Gross hourly wages in 2017 prices. Sample as defined in Section 1.2.2. Source: 2000 to 2017 GSOEP.

Table 1.13: Male wage quintile transition probabilities

lm = FT	$w_{t1}^M = 1$	$w_{t1}^M = 2$	$w_{t1}^M = 3$	$w_{t1}^M = 4$	$w_{t1}^M = 5$
$w_{t0}^M = 1$	0.5211	0.2636	0.1202	0.0621	0.0329
$w_{t0}^M = 2$	0.2566	0.3454	0.2315	0.1165	0.0500
$w_{t0}^M = 3$	0.1215	0.2357	0.3138	0.2241	0.1050
$w_{t0}^M = 4$	0.0565	0.1205	0.2302	0.3453	0.2475
$w_{t0}^M = 5$	0.0284	0.0524	0.1112	0.2615	0.5464

Notes: Gross hourly wages in 2017 prices. Sample as defined in Section 1.2.2. Source: 2000 to 2017 GSOEP.

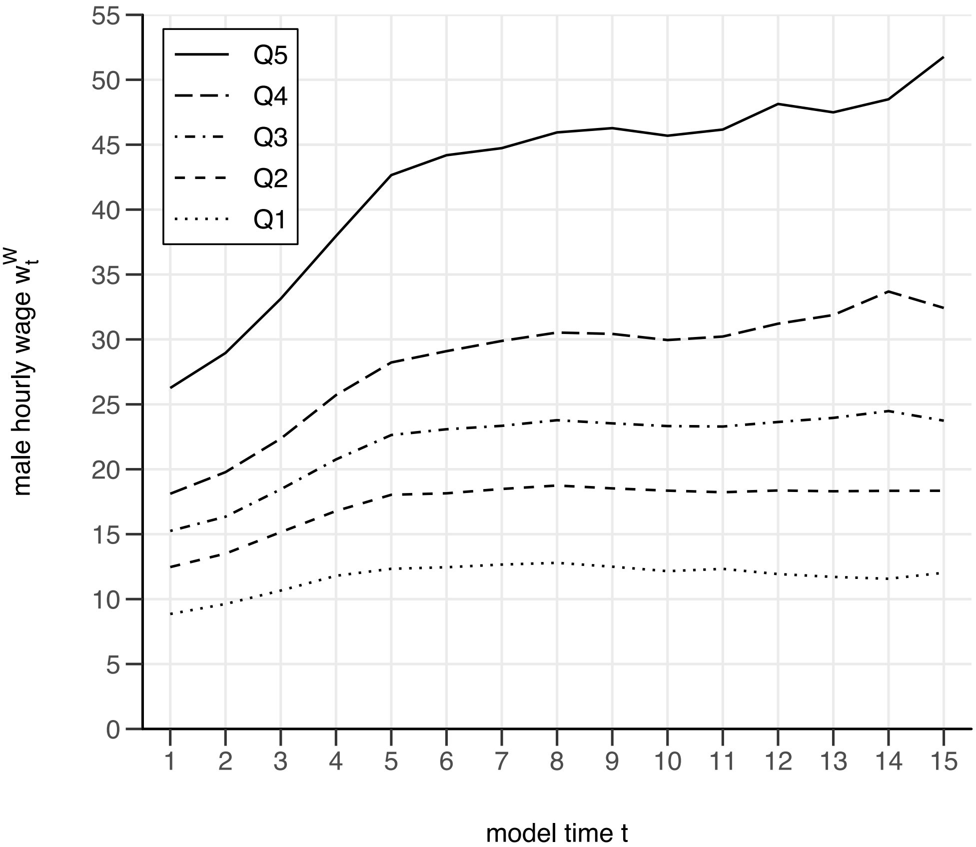


Figure 1.10: Male wage quintile means across the life cycle

Notes: Gross hourly wages in 2017 prices. Source: 2000 to 2017 GSOEP.

1.C.3 Fertility types

Examples for the fertility type prediction. Illustrating the fertility type prediction for the first model age bracket ($t = 1$, i.e. $20 - 22$), the strategy for childless females aged $20 - 22$ looks as follows: We estimate equation (1.17) on the subsample of 41+ year olds who had their first child between 23 and 40 *or* remained childless. Then we use the estimated coefficients in an out-of-sample prediction of the probabilities to belong to each fertility type (a, n) for the $20 - 22$ year old childless females. This excludes fertility types with $a = 1$, which correspond to having the first child between 20 - 22, because we observe the woman *without* children in that age bracket. For each of the subsequent age brackets (up to the $t = 7$, i.e. $38 - 40$), we apply the same strategy.

Illustrating the procedure for women who gave first birth in the model age bracket $t = 1$, i.e. $20-22$: We estimate equation (1.17) on the subsample of 41+ year olds who had their first child between $20-22$ and gave birth to their second child between 23 and 40, if ever. Then we use the estimated coefficients in an out-of-sample prediction of the probabilities for the $20-22$ year old females that first gave birth in that period to belong to each of the three possible fertility types. For females with the same age at first birth ($20 - 22$), but currently in the second age bracket ($23 - 25$), we accordingly use the subsample of second births between age 26 and 40. This strategy takes into account that additional children and therefore higher fertility types become rarer as current age increases. We apply it analogously to females for whom we observe both first and second birth to estimate the probability of a third birth. Individuals observed with three children cannot have more births in our framework and are therefore deterministically assigned into their corresponding fertility type. This is possible as both components of f, age at first birth a and the number of children n, are observed.

Table 1.14: Observed vs. in-sample predicted fertility distribution

fertility type	observed	predicted
1	0.1116	0.1136
2	0.0339	0.0148
3	0.0949	0.0591
4	0.0345	0.0328
5	0.0401	0.0394
6	0.0980	0.0987
7	0.0351	0.0396
8	0.0432	0.0406
9	0.1023	0.1280
10	0.0395	0.0459
11	0.0376	0.0364
12	0.0931	0.1377
13	0.0216	0.0214
14	0.0425	0.0386
15	0.0580	0.0591
16	0.0136	0.0075
17	0.0339	0.0233
18	0.0407	0.0394
20	0.0259	0.0241

Notes: Sample: women aged 41 to 65 from the sample as defined in Section 1.2.2. Source: 2017 GSOEP.

1.C.4 Maximum Likelihood Estimation

Description of the basin-hopping algorithm. As described in Section 1.4.5.4, the basin-hopping algorithm is a stochastic global optimisation algorithm. Starting from an arbitrary pre-specified initial parameter combination, it employs a minimisation routine (in our case Matlabs *fminunc*) to find a (possibly local) first minimum. This first minimum is the first candidate for the global optimum. Using the parameter combination of the minimum found after the first minimisation, the algorithm then applies a random perturbation to these parameters (which we will call "taking a *step*" from now on) and restarts the minimisation routine. If the routine returns a new minimum whose function value is lower than the current global optimum candidate, this parameter combination becomes the current global optimum candidate. The *step*-taking, which is discussed in detail below, is then repeated for either a pre-specified number of times or terminated if the global minimum candidate has not changed for a pre-specified number of *steps*. After termination, the final global optimum candidate is returned as the global solution to the minimisation problem.

The *step*-taking procedure is the key mechanism of the algorithm to search the multidimensional target function for minima. In order to be able to escape the basins of attraction of local minima, the following adaptive procedure is used to conduct the *step*-taking:[60]

Starting with an initial stepsize – 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) stepsize. 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 minimisation 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*.[61]

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 optima candidates, the algorithm retains some flexibility to explore the target function in the proximity of the current global optimum candidate.

The procedure furthermore includes an adaptive adjustment of the search radius (as determined by the size of the perturbations, i.e. the stepsize). After every n-th *step*, the algorithm compares

[60] A *basin of attraction* is the set of possible starting points of a minimisation routine that leads to the same minimum, see Nusse and Yorke 1996.

[61] 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.

the rate at which *steps* are *accepted* to a target rate.[62] 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. Therefore the stepsize is adjusted downwards by 10%, narrowing the search radius. If the *acceptance* rate is below the target rate, this implies that the perturbations are to small, i.e. the algorithm is likely stuck in a basin of attraction. Therefore the stepsize is adjusted upwards by 10%, widening the search radius to be able to escape the current basin of attraction and explore the function outside of it further.

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

1.D Model fit

1.D.1 Derivation of the compensated intensive margin labour supply elasticity

The derivation of the Hicksian (compensated) labour supply elasticity is based on chapter 3.1 and equations 1 to 4 in Saez 2001. For convenience, we follow Saez' notation. Our estimated model-based aggregate hours elasticity is 0.29 which implies together with the participation elasticity of 0.2 a Marshallian (uncompensated) labour supply elasticity of $\xi^u = 0.09$. Following Saez 2001, we define $z = z(1-\tau, R)$ as the Marshallian (uncompensated) earnings supply function. It is based on a utility function $u(c, z)$ with consumption c and earnings z and a linear budget constraint $c = z(1-\tau) + R$ where τ is the marginal tax rate and R is some non-labour income, i.e. lump sum transfer. Income effects are defined as $\eta = (1-\tau)\frac{\partial z}{\partial R}$.

To approximate η we run an additional experiment in which we raise the net income of all households by 1,000 EUR per year. In response to this policy, households reduce their gross labour income by approx. 400 EUR per year via a reduction in aggregate hours by 0.87% and in participation by 0.62%. Decomposing the income effect of 400 EUR further, 225 EUR can be attributed to the extensive margin and 175 EUR from the intensive margin. Therefore, we achieve $\frac{\partial z}{\partial R} = \frac{-175}{1000} = -0.175$. Assuming a marginal tax rate of 40% yields $\eta = (1-0.4)(-0.175) = -0.105$.[63] By the Slutsky equation, we can directly compute the Hicksian (compensated) intensive margin elasticity as $\xi^c = \xi^u - \eta = 0.09 - (-0.105) = 0.195$.

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

[63] 40% is (close to) the top marginal tax rate which yields the most conservative scenario, i.e. the smallest Hicksian (compensated) intensive margin elasticity.

1.E Policy Experiments

Table 1.15: Impact period labour supply reactions to 50 EUR subsidy - untargeted -

	policy scenario		
baseline	NP	PT	FT
NP		0.08%	0.00%
PT	0.03%		0.09%
FT	0.00%	0.02%	

Notes: Share of individuals who change their labour supply. 99.79% do not change their labour supply.

Table 1.16: Impact period labour supply reactions to 50 EUR subsidy - work contingent -

	policy scenario		
baseline	NP	PT	FT
NP		0.51%	0.00%
PT	0.00%		0.09%
FT	0.00%	0.02%	

Notes: Share of individuals who change their labour supply. 99.39% do not change their labour supply.

Table 1.17: Impact period labour supply reactions to 50 EUR subsidy - work and child age < 6 contingent -

	policy scenario		
baseline	NP	PT	FT
NP		0.45%	0.00%
PT	0.00%		0.06%
FT	0.00%	0.02%	

Notes: Share of individuals who change their labour supply. 99.48% do not change their labour supply.

Table 1.18: Impact period labour supply reactions to 50 EUR subsidy - work and child age < 3 contingent -

	policy scenario		
baseline	NP	PT	FT
NP		0.19%	0.00%
PT	0.00%		0.04%
FT	0.00%	0.00%	

Notes: Share of individuals who change their labour supply. 99.76% do not change their labour supply.

Table 1.19: Impact period labour supply reactions to 50 EUR subsidy - full-time work contingent -

	policy scenario		
baseline	NP	PT	FT
NP		0.00%	0.00%
PT	0.00%		0.50%
FT	0.00%	0.00%	

Notes: Share of individuals who change their labour supply. 99.50% do not change their labour supply.

Table 1.20: Decomposition of self-financing degree of changes in full-time childcare subsidies - untargeted subsidy of 50 EUR/month -

		female wage quintile				
	total	Q1	Q2	Q3	Q4	Q5
(i) *Impact period*						
tax revenue	58,913	14,565	14,113	13,368	18,027	-1,160
subsidy spending	-3,426,640	-640,287	-698,956	-697,366	-711,887	-678,143
self-financing	1.7%	2.3%	2.0%	1.9%	2.5%	-0.2%
(ii) *Remaining life cycle*						
tax revenue	168,448	39,673	36,456	37,838	41,948	12,534
subsidy spending	-6,554,948	-1,055,981	-1,235,231	-1,420,547	-1,503,297	-1,339,893
self-financing	2.6%	3.8%	3.0%	2.7%	2.8%	0.9%
(iii) *Entire life cycle*						
tax revenue	468,332	95,897	92,808	91,017	96,509	92,102
subsidy spending	-12,632,132	-2,667,431	-2,530,636	-2,410,591	-2,545,713	-2,477,762
self-financing	3.7%	3.6%	3.7%	3.8%	3.8%	3.7%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values.

Table 1.21: Decomposition of self-financing degree of changes in full-time childcare subsidies - work contingent subsidy of 50 EUR/month -

		female wage quintile				
	total	Q1	Q2	Q3	Q4	Q5
(i) *Impact period*						
tax revenue	350,076	40,481	64,372	71,283	96,955	76,986
subsidy spending	-2,844,838	-529,859	-584,666	-580,426	-575,106	-574,780
self-financing	12.3%	7.6%	11.0%	12.3%	16.9%	13.4%
(ii) *Remaining life cycle*						
tax revenue	936,972	148,301	176,009	201,782	233,660	177,220
subsidy spending	-5,387,610	-867,710	-1,020,548	-1,170,005	-1,216,395	-1,112,952
self-financing	17.4%	17.1%	17.2%	17.2%	19.2%	15.9%
(iii) *Entire life cycle*						
tax revenue	2,022,978	407,169	403,173	389,334	421,855	401,448
subsidy spending	-10,037,358	-2,129,409	-2,013,978	-1,913,920	-2,004,849	-1,975,202
self-financing	20.2%	19.1%	20.0%	20.3%	21.0%	20.3%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values.

Table 1.22: Decomposition of self-financing degree of changes in full-time childcare subsidies - work and child age < 6 contingent subsidy of 50 EUR/month -

		female wage quintile				
	total	Q1	Q2	Q3	Q4	Q5
(i) *Impact period*						
tax revenue	301,760	31,331	52,095	64,160	87,678	66,496
subsidy spending	-2,385,767	-411,830	-470,751	-502,919	-507,719	-492,548
self-financing	12.6%	7.6%	11.1%	12.8%	17.3%	13.5%
(ii) *Remaining life cycle*						
tax revenue	746,537	109,810	134,093	165,233	194,767	142,634
subsidy spending	-4,241,317	-638,112	-771,890	-939,965	-996,471	-894,879
self-financing	17.6%	17.2%	17.4%	17.6%	19.5%	15.9%
(iii) *Entire life cycle*						
tax revenue	1,816,055	364,185	359,855	348,826	380,969	362,220
subsidy spending	-8,878,776	-1,882,459	-1,772,932	-1,686,600	-1,779,282	-1,757,503
self-financing	20.5%	19.3%	20.3%	20.7%	21.4%	20.6%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values.

Table 1.23: Decomposition of self-financing degree of changes in full-time childcare subsidies - work and child age < 3 contingent subsidy of 50 EUR/month -

		female wage quintile				
	total	Q1	Q2	Q3	Q4	Q5
(i) *Impact period*						
tax revenue	149,544	13,274	21,371	31,469	48,888	34,542
subsidy spending	-887,638	-134,288	-161,093	-202,801	-226,724	-162,733
self-financing	16.8%	9.9%	13.3%	15.5%	21.6%	21.2%
(ii) *Remaining life cycle*						
tax revenue	276,056	33,844	42,040	60,899	80,054	59,218
subsidy spending	-1,185,972	-160,980	-203,541	-280,007	-301,708	-239,735
self-financing	23.3%	21.0%	20.7%	21.7%	26.5%	24.7%
(ii) *Entire life cycle*						
tax revenue	985,456	194,070	194,935	189,405	205,918	201,128
subsidy spending	-3,877,211	-834,616	-782,439	-738,765	-769,220	-752,172
self-financing	25.4%	23.3%	24.9%	25.6%	26.8%	26.7%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values.

Table 1.24: Decomposition of self-financing degree of changes in full-time childcare subsidies
- full-time work contingent subsidy of 50 EUR/month -

		female wage quintile				
	total	Q1	Q2	Q3	Q4	Q5
(i) *Impact period*						
tax revenue	340,954	43,216	60,099	68,585	78,349	90,705
subsidy spending	-646,133	-85,570	-113,181	-130,310	-145,303	-171,768
self-financing	52.8%	50.5%	53.1%	52.6%	53.9%	52.8%
(ii) *Remaining life cycle*						
tax revenue	890,701	116,559	147,086	188,472	218,491	220,094
subsidy spending	-1,177,373	-146,822	-195,140	-252,047	-282,752	-300,611
self-financing	75.7%	79.4%	75.4%	74.8%	77.3%	73.2%
(ii) *Entire life cycle*						
tax revenue	1,768,055	350,457	343,151	337,533	361,029	375,884
subsidy spending	-2,239,252	-453,434	-436,863	-427,935	-450,396	-470,625
self-financing	79.0%	77.3%	78.5%	78.9%	80.2%	79.9%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values.

Table 1.25: Self-financing degree of changes in monthly full-time childcare subsidies by the number of children

		number of children		
	total	1	2	3
(i) *untargeted 50 EUR*				
Impact period	1.7%	1.1%	1.8%	2.2%
Remaining life cycle	2.6%	1.5%	2.7%	3.0%
Entire life cycle	3.7%	2.3%	4.0%	3.9%
(i) *work contingent 50 EUR*				
Impact period	12.3%	9.5%	11.7%	16.4%
Remaining life cycle	17.4%	13.8%	16.3%	22.4%
Entire life cycle	20.2%	16.5%	19.0%	24.9%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values. The number of children reflects completed fertility, i.e. number of children over the entire life cycle.

Table 1.26: Self-financing degree of changes in monthly full-time childcare subsidies by the number of children

	total	number of children 1	2	3
(i) *work and child age* < 6 *contingent 50 EUR*				
Impact period	12.6%	9.7%	12.1%	16.7%
Remaining life cycle	17.6%	13.5%	16.4%	22.8%
Entire life cycle	20.5%	16.5%	19.3%	25.4%
(i) *work and child age* < 3 *contingent 50 EUR*				
Impact period	16.8%	14.0%	17.3%	18.4%
Remaining life cycle	23.3%	19.2%	23.2%	25.7%
Entire life cycle	25.4%	22.8%	24.8%	28.3%
(iii) *full-time work contingent 50 EUR*				
Impact period	52.8%	49.0%	53.0%	56.8%
Remaining life cycle	75.7%	70.4%	75.1%	83.0%
Entire life cycle	79.0%	74.5%	77.8%	86.7%

Notes: (i) Impact period: year of the policy introduction (i.e. 2017), (ii) Remaining life cycle: simulating the households in the sample from their observed age in 2017 until age 80, (iii) Entire life cycle: setting the age of the households in the sample to 20 in 2017 and simulating their behaviour until age 80. All underlying tax revenue and subsidy spending is discounted to 2017 values. The number of children reflects completed fertility, i.e. number of children over the entire life cycle.

Chapter 2

The Intergenerational Correlation of Employment: Mothers as Role Models

This paper has been circulated as IZA Discussion Paper 12595 and Bank of Canada Staff Working Paper 2019-33 under the title *The Intergenerational Correlation of Employment: Is There a Role for Work Culture?*.

2.1 Introduction

For several decades, the intergenerational correlation of labor market outcomes has been a subject of interest among both academics and policy-makers. As a key determinant of socio-economic mobility, the correlation of labor earnings between subsequent generations has received particular attention. An extensive literature documents that earnings of individuals are highly correlated with those of their parents (see the comprehensive surveys by Solon 1999, Bowles and Gintis 2002, Black and Devereux 2011, Björklund and Jäntti 2011). The focus of this literature is on the identification and quantification of channels through which the "potential" to earn is transmitted. Such channels include, among others, the genetic inheritance of cognitive skills, higher investments into children's education by parents with higher income, and parents' social networks, which children can take advantage of.

By contrast, in this paper we argue that not only "earnings potential" is transmitted across generations but also the "willingness to work". Specifically, we document a, to the best of our knowledge, novel fact: employment status, or the fraction of individuals' working-age life spent in employment, is highly correlated with their mothers'. Moreover, this correlation remains significant even after controlling for the main determinants of the intergenerational correlation of earnings, according to the literature on the topic. We provide evidence in support of a role-model channel underlying this correlation: children emulate the example of mothers with respect to employment. The positive and strong intergenerational correlation of employment has important implications not only for the analysis of social mobility but, potentially, also for the optimal design of tax-transfer policies.

Why has this fact been overlooked so far? Perhaps the reason is that the empirical literature on intergenerational earnings correlations typically restricts the analyzed sample to individuals and periods for which earnings are observed, thereby neglecting the variation in employment status (i.e. the extensive margin of labor supply) by construction. Such data, of course, still capture some variation in labor supply, namely the variation in hours worked, or the intensive margin of labor supply. However, our analysis shows that while the unconditional intergenerational correlations of both employment and hours are substantial, only the one for employment remains significantly different from zero after controlling for education, ability, fertility, and wealth. This means, on the one hand, that the similarity in the number of hours worked between mothers and children can be mostly explained by similar benefits from work (determined by ability and education) as well as their need to work (determined by wealth).[1] On the other hand, it also means that the same factors are not able to fully explain the high correlation in the decision whether or not to work at all.

We obtain our results by linking data from the National Longitudinal Survey of Youth 1979 (NLSY79) and the Children and Young Adults (CNLSY79) cohort. These data are designed to link mothers from a representative sample born in the US between 1957 and 1964 with their children. Since more mothers than fathers are at the margin between labor force participation and non-participation, we believe the focus on mother-children pairs is reasonable given our goal. Exploiting the longitudinal structure of the data, we first estimate the permanent component of employment status along the life cycle for both, mothers and children. This permanent component measures how much of their active life individuals spend in employment. The information included in this component is different from the permanent component of earnings, which is based only on periods of employment when earnings are observed.

We find a robust, statistically significant and positive correlation of employment status.[2] The unconditional correlation is 0.21, implying that an increase in lifetime employment of mothers by one year is associated with an increase in the employment of her child of around *11 weeks*. After netting out the influence of ability, education, wealth, and some other relevant covariates,

[1] Ability for mothers is measured via the Armed Forces Qualification Test (AFQT), ability for children via the Math score in the Peabody Individual Achievement Test (PIAT). Fertility is included in order to control for potentially correlated fertility attitudes between mothers and their children, which in turn might be correlates of employment.

[2] In the Appendix we provide results for an extensive set of different specifications, all of which confirm our main result.

the correlation remains at 0.12, corresponding to an incremental employment of children of around *six weeks*. This is what we call *residual* correlation of employment.[3]

Furthermore, by splitting the sample into different sub-samples, we find that the residual employment correlation between mothers and their children is heterogeneous across several dimensions. First, it is significantly higher for daughters (0.18) than for sons (0.07). While a one-year increase in lifetime employment of mothers increases the employment of their sons by on average less than *four weeks* (still significant at the 5% level), it increases employment of their daughters by more than *nine weeks*. Second, the intergenerational correlation of employment is decreasing in the degree of maternal education, being significantly positive only for mothers without any college education. Finally, the correlation tends to decrease in maternal family income. For the bottom quintile of the income distribution, a one-year increase in the mother's employment increases employment of her child by around *nine weeks*.

The significant intergenerational correlation of the extensive – rather than the intensive – margin of labor supply is particularly important in light of several existing policies, such as the Earned Income Tax Credit (EITC) in the United States, which aim to encourage labor force participation. This is especially the case since we find the correlation to be higher at the bottom of the income distribution, the target group of the EITC. Our results suggest that there may be a, perhaps unintended, dynamic fiscal benefit of such policies through increased labor market participation of future generations.

However, before such conclusions can be drawn, an understanding of the channels determining this correlation is needed. For example, if the intergenerational transmission of employment was not affected by mothers' behavior but rather the result of a direct transmission of preferences for work,[4] none of the government's costs, of a policy encouraging parental employment, will be recovered through higher participation of their children. In such a situation, children will have the same attitude towards work independent of the existence of such a policy. However, the very opposite is true if children emulate the *behavior* of their parent. Then a policy that increases parental employment, even if it is currently costly, may amortize through increased participation of future generations.

Our data offers three pieces of evidence suggesting that indeed such a role-model effect is in place and that therefore, from a public finances' point of view, policies that move mothers into the labor force may result in increased revenues from future generations. First, as mentioned above, the correlation of employment status is higher for mother-daughter pairs than for mother-son pairs, and role models tend to be more pronounced within the same gender (Bettinger and

[3] To put these numbers into perspective, estimates for the intergenerational earnings elasticity in the US have oscillated around 0.4 (see, for example, Solon 1992, Zimmerman 1992, Chetty, Hendren N., et al. 2014).

[4] By direct preference transmission we refer to a situation in which the mother transmits her preference for work to her children independently of her work behavior.

Long 2005). Second, we construct a measure for maternal work preferences exploiting certain survey questions in the NLSY79. While this measure is significantly correlated with maternal employment, thus confirming its validity, it is uncorrelated with children's employment. This suggests that preferences for work are not directly transmitted across generations. Instead, it seems important that the child actually observes the mother working. This is confirmed by our third and last piece of evidence, which disentangles the direct transmission of preferences from the role-model channel by controlling for periods in which the mother does not cohabit with her child. This measure serves as a proxy for mothers' work preferences. It turns out that the correlation is mainly driven by periods of cohabitation, in which it is arguably easier for the child to emulate the behaviour of the mother.

Finally, we rule out alternative explanations for this residual correlation, such as the effect of networks, occupation-specific human capital, or local labor markets. Particularly, we analyze the heterogeneity in the intergenerational correlation of employment across mother-children pairs that do or do not share industries, occupations, or local labor markets. The lack of difference across groups indicates that these explanations are unlikely to drive the intergenerational correlation of employment status.

Related literature. Our paper contributes to many different branches of the empirical literature studying the transmission of preferences for work across generations. Using tools of the well-established literature on the intergenerational correlations of labor market outcomes,[5] we focus on an unexplored variable, employment, and argue that it bears important information on the transmission of preferences for work.

The gender literature has analyzed the transmission of preferences for work from the perspective of gender roles. An important part of this literature uses the so-called "epidemiological approach". This approach considers the intergenerational transmission of cultural traits when outcomes of second-generation migrants and those of the parents' country of origin are correlated. Fernandez 2007 and Fernandez and Fogli 2009 interpret such correlation in female labor force participation as cultural transmission of women's roles. Another, more structural, strand of the gender literature also looks at cultural transmission. For instance, Fernandez 2013 explains the S-shape in the female labor force participation during the second half of the 20^{th} century with a model that introduces learning across generations about the returns to female work. These studies deal with the transmission of society-wide preferences. We instead analyze preference transmission within the family, from mothers to children. Furthermore, our paper does not limit attention to the transmission of gender roles, as we do not restrict the analysis to mothers and daughters. In this last sense, our paper distances itself from others that have

[5] The literature dealing with methodological issues for measuring intergenerational correlations in reduced form, on which we rely, is vast (see, for example, Solon 1992, Solon 1999, S. Haider and Solon 2006, Grawe 2006, Lee and Solon 2009, Nybom and Stuhler 2016, Nybom and Stuhler 2017, Mazumder 2005).

analyzed the transmission of gender roles (see, for example, Binder 2018, Olivetti, Patacchini, and Zenou 2018).

Another related strand of literature documents that parental welfare benefit reception results in an increased probability of children claiming the benefits themselves. In the context of the Norwegian disability insurance (DI) system, Dahl, Kostøl, and Mogstad 2014 exploit variation in the leniency of appeal judges, who are randomly assigned to decide on cases where individuals were originally denied disability insurance. The authors find that when a parent is allowed DI at the appeal stage, their adult child's DI participation rate increases by 12 percentage points over the following 10 years. This number is surprisingly similar to what we find for employment. In particular, when estimating a linear probability model and looking at a given working-life year of the mother and the child, we find that maternal employment increases the probability of being employed for the child by 12 percentage points. Furthermore, their results are consistent with our suggested mechanism. In particular, in both their paper and ours, differential outcomes of children are not explained by differences in what parents want – all parents in their paper apply for DI – but rather by differences in what parents actually do. Two similar recent contributions are Dahl and Gielen 2018, who use a regression discontinuity design induced by a reform of DI in the Netherlands, which tightened eligibility criteria, and Hartley, Lamarche, and Ziliak 2017, who exploit cross-state variation in the timing of welfare and income support program reforms in the US. We see our contribution complementary to these papers. On the one hand, the quasi-experimental design in these three papers allow them to make causal inferences. On the other hand, the findings of these papers are very specific to the respective institutional setting and restricted to the receipt of a certain kind of welfare benefit. In contrast, we document the transmission of employment between mothers and their children for a representative sample of the US population. The evidence from these papers does not allow for inferences on the transmission of this important labor market outcome.

Closely related to this paper are also studies that infer transmission of work preferences from the intensive margin of labor supply. Estimating an overlapping generations model with data from the Panel Study of Income Dynamics, Toledo 2010 attributes the correlation in hours worked between fathers and sons to the transmission of preferences for work.[6] Using data on mothers and fathers, Altonji and Dunn 1991 confirm both our result as well as Toledo's. In particular, they find that working hours between fathers and sons are positively correlated while those between mothers and their children – both daughters and sons – are not. Neither of these papers analyses the extensive margin of labor supply. As mentioned before, this distinction turns out to be crucial as we find a significant intergenerational correlation at the extensive margin between mothers and children. Taking the evidence in these papers and ours' together,

[6] Altonji and Dunn 2000 reach a similar conclusion using the National Longitudinal Survey of Labor Market Experience and relying on a factor model that allows preferences to influence labor market outcomes.

the transmission in working hours seems to be more substantial between fathers and sons, while transmission in labor market participation is more substantial between mothers and daughters.[7]

Outline. The remainder of the paper is structured as follows. In Section 2.2, we present the data, followed by the empirical strategy in Section 2.3. Section 2.4 documents the main results. In Section 2.5, we discuss potential mechanisms. Section 2.6 concludes.

2.2 Data

We use the National Longitudinal Survey of Youth 1979 cohort (NLSY79) and the Children and Young Adults cohort (CNLSY79). These data are widely used in the analysis of inequality and labor market research. The NLSY79 surveys a representative sample of individuals born in the US between 1957 and 1964. Respondents are 14 to 22 years old in 1979 and are followed since then. Our last observation is 2012, when they are 47 to 56 years old. The frequency is annual between 1979 and 1994, and biannual thereafter. The children of the women in this cohort are surveyed on a biannual basis since 1986, constituting the CNLSY79. They are linked to the original cohort by a unique identifier provided by the US Bureau of Census.[8]

We restrict the analysis to the cross-sectional sub-sample of the NLSY79 that is designed as a representative sample of the US population in 1979. We exclude other sub-samples that oversample particular groups of the population, to avoid weighting the estimates. We restrict to observations during ages 25 to 45 for both cohorts to keep the representativeness of the lifetime employment experience (the oldest individual in the second cohort is 38 years old in 2012). We obtain a final sub-sample of 1,373 mothers paired to 2,339 children.

The data are particularly rich. They provide detailed information on labor market outcomes, education, and further demographic and socio-economic characteristics. Importantly, they contain widely used indicators of ability, which is a key confounder for the estimation of intergenerational transmission of labor market outcomes: the Armed Forces Qualification Test (AFQT) for the mothers and the Peabody Individual Achievement Test (PIAT) for the children; we use the Math score of the latest PIAT assessment for the children cohort, in line with the literature (Abbott et al. 2013). We use information on wealth (net worth), computed as assets (savings, home and vehicle ownership) minus debts (credit cards, students loans, mortgages, vehicle loans, and others).

[7]While the correlation in employment status between mothers and sons is also significantly positive, it is lower in magnitude, as mentioned before.

[8]Although in the NLSY79, only mothers (and not fathers) can be linked to CNLSY79 data, this does not challenge the objective of our paper. As we focus on the extensive margin of the labor supply decision, using maternal employment information is reasonable because female labor force participation is typically lower (through more elastic labor supply) than male labor force participation, particularly during the period of observation of the first cohort.

Table 2.1 provides descriptive statistics of the data (additional descriptives are summarized in Table 2.9 in the Appendix). For most variables, we report the means across individual averages for those observations over the 25 to 45 years old range in our sample. The last two columns refer to the sample of mothers and their children, and the first one shows the characteristics of the total sample of women in the NLSY79 cohort for reference. All monetary values are deflated using the Consumer Price Index (CPI) and expressed in prices of 1980.

Mothers are observed on average for 14 waves, and children for 2.5 waves. The average age is 33 for mothers and 27 for children. The sample of mothers is representative of women with children by design. As compared with the total sample of women in the NLSY79, mothers are slightly less educated and live in poorer households. Women are 22 years old on average when they give birth. The children's cohort is relatively younger than the mothers' by construction, as reflected in the age and other characteristics associated to the life cycle (for example, the proportion married and cohabiting is lower in the children's cohort, and the wealth level as well). Observations of older children correspond to younger mothers at the time of birth. Children are slightly more educated than mothers.

Questions about employment status vary across waves in the survey. Our choice of the particular question used in our analysis balances two objectives: (i) we want to have a measure that is as homogeneous as possible between the samples of mothers and children; (ii) at the same time, the questions should be consistent along the different waves and minimize the number of non-responses. We consider mothers to be employed if they declare that they worked for 10 or more weeks in the year before the interview. We categorize children as employed if their earnings in the year before the interview were equivalent to at least two months of a part-time job at the minimum salary.[9] The employment rate is 73% for mothers and 84% for the children cohort (80% for daughters). Although these figures seem high as compared with official statistics of female employment, they are not at odds, considering that we are taking an annual window for the measurement of employment.

Employed mothers and children work on average 36 and 40 hours a week at an hourly wage rate of $7 and $6 (in 1980 USD), respectively. Earnings amount to $9,750 and $13,316 annually. Net worth is higher for the mothers' than for the children's cohort ($43,064 vs. $9,551), a difference potentially due to the composition of the children's sample explained above, as well as because most children had not inherited yet at the time they were surveyed. No such differences are observed in family income across cohorts, though ($27,226 and $26,029, respectively). The average percentile of maternal cognitive test scores is 40, and it is 49 for children. Mothers take

[9]The lower bound for earnings is arbitrary, although reasonable. It is $450 monthly in 1980 USD, which is equivalent to a job of 17 hours a week (part-time), for 8.5 weeks (2 months) and $3.1 per hour. The main purpose is to exclude casual jobs. We also show that the results are robust to other measures of employment.

Table 2.1: Summary statistics for women and mother-child pairs in NLSY79 and CNLSY79

	Women	**Mothers**	**Children**
Demographics			
Age	32.9	33.2	27.3
	(1.7)	(1.0)	(1.6)
Female	100%	100%	50%
Married/cohabiting	67%	77%	33%
Number of children	1.9	2.6	1.2
	(1.4)	(1.2)	(1.3)
Maternal age at birth			21.7
			(3.4)
Education and Ability			
Years of education	13.7	12.8	13.4
	(2.6)	(2.2)	(2.4)
High school drop-out	7%	10%	12%
High school complete	40%	51%	36%
Some college	25%	26%	28%
Complete college	28%	13%	24%
Percentile in cognitive test	48.8	39.6	49.1
	(28.5)	(26.9)	(27.8)
Age at test	18.0	18.3	11.6
	(4.0)	(4.2)	(4.6)
Labor Market Outcomes			
Employment	79%	73%	84%
Hours/week	37.8	36.4	39.6
	(8.3)	(8.8)	(11.0)
Hourly wage (in USD)	8.1	6.7	6.3
	(8.2)	(9.9)	(4.0)
Annual earnings (in 1,000 USD)	12.9	9.8	13.3
	(9.1)	(6.7)	(9.8)
Wealth and Income			
Net worth (in 1,000 USD)	55.9	43.1	9.6
	(95.9)	(81.6)	(31,2)
Family income (in 1,000 USD)	33.5	27.2	26.0
	(35.6)	(24.5)	(29.1)
Number of interviews	13.2	14.1	2.5
	(3.1)	(1.9)	(1.2)
Individuals	3,040	1,373	2,339

Notes: Averages for quantitative variables (standard deviations in parentheses), percentages for dichotomous variables, for observations in the 25 to 45 years old range in our sample. Cognitive tests are AFQT for parents and PIAT Math for children. Monetary variables are in 1980 USD.

the test when they are 18 years old and children when they are 12. Further details on the data can be found in Appendix 2.B.1.

2.3 Empirical strategy

We follow the literature on intergenerational correlations of labor market outcomes to quantify the persistence in employment status across generations. The unit of observation is the mother-child pair i and our main regression specification relates the permanent component of employment – which can be interpreted as the fraction of lifetime employment – of the mother l_{Mi} to the permanent component of employment of the child l_{Ci}. The reduced-form specification is

$$l_{Ci} = \alpha + \beta l_{Mi} + \phi_M X_{Mi} + \phi_C X_{Ci} + \epsilon_i. \tag{2.1}$$

Our coefficient of interest, β, summarizes the intergenerational persistence of employment. X_{Mi} and X_{Ci} are control variables for mothers and children, respectively. We consider different specifications and control for several confounders, including education, ability, wealth, the number of children of both generations, and the age of the mother at birth.

Computation of permanent components. Equation (2.1) relies on measures of lifetime employment status. The literature on intergenerational correlations is quite rich in terms of how to compute these lifetime or long-run measures. Given the nature of our data, we take an approach that allows for the use of all the periods of information. Following Zimmerman 1992 and Toledo 2010, we obtain these lifetime or permanent components of employment as the fixed effects in a statistical model for the probability of being employed in each period under observation.[10]

We specify a linear probability model,

$$l_{kit} = l_{ki} + \sum_{n=1}^{2} \pi_{nk} A_{kit}^{n} + \lambda_{kt} + \upsilon_{kit}, \tag{2.2}$$

which we run for both generations $k \in \{M, C\}$. Specifically, we assume that the probability of individual i to be employed in year t is a function of a second-order polynomial of the individual's age A_{kit}, a year fixed effect λ_{kt}, and an individual fixed effect l_{ki}. This individual

[10] Using multiple periods has been shown to reduce measurement error (see, for example, Solon 1992, Mazumder 2005, S. Haider and Solon 2006). This strategy is simpler than a factor model that explicitly models such error (see, for example, Lochner, Park, and Shin 2018), but we consider it effective, particularly for employment, the main focus of this paper. Lee and Solon 2009 recommend an efficient approach by using all the children's observations in a version of the intergenerational equation (2.1). Our approach also uses all the information of the children, but in a two-step procedure that we deem accurate according to the Frisch-Waugh-Lovell theorem.

fixed effect represents the permanent component of employment status, abstracting from life-cycle fluctuations (absorbed by age effects), and from business-cycle fluctuations (absorbed by year effects). We can interpret the permanent component of employment as the proportion of lifetime each individual is in employment.

Regression versus correlation coefficient. An alternative to the regression coefficient β for measuring persistence in labor market outcomes across generations is the correlation coefficient,[11]

$$\rho = \beta \frac{\sigma_M}{\sigma_C}, \tag{2.3}$$

where σ_M (σ_C) denotes the standard deviation of mothers' (children's) employment. Because the variability of mothers' and children's employment is very similar, there is not a big difference between the reported regression coefficients and the correlation coefficients.[12] We hence present only the regression coefficients throughout the main text and refer to the coefficient of interest, β, as the correlation of intergenerational employment status. More details about methodological issues in measuring the intergenerational persistence of labor market outcomes can be found in Appendix 2.B.2.

2.4 Results

2.4.1 Intergenerational correlation of employment

In this section, we document the intergenerational correlation of employment status for the United States. Table 2.2 shows the regression coefficients for maternal employment and covariates estimated using equation (2.1). Standard errors are clustered at the mother level to account for possible auto-correlation in siblings' error terms.

The first column (without controls) shows an unconditional correlation of employment of 0.21. Children are on average employed for an additional 11 weeks when their mother is employed one more year ($0.21 \times 52 \approx 11$).[13] This finding of a substantial association in employment across generations is, to the best of our knowledge, a novel fact.

[11] Note that the correlation coefficient is conditional on covariates X_{Mi} and X_{Ci} if included in the regression.

[12] The standard deviations of the permanent components l_{Mi} and l_{Ci} are $\sigma_M = 0.30$ and $\sigma_C = 0.32$.

[13] As a comparison, estimates for the intergenerational elasticity of income for the US have oscillated around 0.4 in early work based on survey data (Solon 1992, Zimmerman 1992) to above 0.5 in recent work using administrative data (Chetty, Hendren N., et al. 2014). Smaller figures correspond to other outcomes related to employment; for example, Toledo 2010 estimates 0.2 intergenerational correlation in hours, and Macmillan 2011, 0.1 for non-employment.

Table 2.2: Baseline regression

Dependent variable: Employment - child (l_{Ci})

Specification	(1)	(2)	(3)	(4) Baseline
Employment - mother l_{Mi}	0.21***	0.14***	0.12***	0.12***
	(0.029)	(0.028)	(0.027)	(0.027)
Ability - mother		0.05*	0.04	0.01
		(0.029)	(0.027)	(0.027)
Ability - child		0.12***	0.07***	0.07**
		(0.029)	(0.026)	(0.026)
Yrs. schooling - mother		0.00	0.00	-0.00
		(0.004)	(0.004)	(0.004)
Yrs. schooling - child		0.02***	0.01***	0.01***
		(0.003)	(0.003)	(0.003)
Net worth - mother			0.01	0.00
			(0.005)	(0.005)
Net worth - child			-0.01	-0.01*
			(0.006)	(0.006)
Number of children - mother				0.00
				(0.006)
Number of children - child				-0.04***
				(0.006)
Control age at birth - mother	NO	NO	NO	YES
Observations	2,339	2,237	1,969	1,969
Adjusted R^2	0.04	0.08	0.05	0.09

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

In the remaining specifications, we further include covariates that typically influence the outcome variable, i.e. employment. In specification (2) we control for ability and education, of both mother and child. We observe that the main predictors, significant at the 1% level, are ability and education of the child. The mother's ability is significant at the 10% level, while her education is insignificant. Importantly, the coefficient on the mother's employment declines by one-third compared to specification (1), being 0.14, still statistically significant at the 1% level.

In specification (3), we include net worth to control for potential wealth effects on labor supply. Although we observe that the sign on children's net worth is negative as expected, both coefficients are small and insignificant. The coefficient on employment of the mother declines only slightly, to 0.12.

Finally, in specification (4) we additionally control for the number of children of both generations and the age of the mother at birth using dummies. This is the specification we will use in everything that follows, unless stated otherwise. The number of children is intended to control for correlated fertility attitudes, which in turn would affect labor supply. An additional grandchild reduces lifetime employment of the child by 4%, statistically significant at the 1% level.

In all specifications from (2) to (4), the coefficient on the maternal employment is significantly positive. The value in the baseline specification with all the controls is 0.12. Hence, an increase of the mother's employment by one year increases employment of the child by around *six weeks* ($0.12 \times 52 \approx 6$). Human capital variables (education and ability) seem to play an important role in the intergenerational correlation of employment, as most of the difference between the coefficient of 0.12 in the regression with all the controls and the coefficient of 0.21 in the regression without controls occurs when these variables are included. However, there is a big part of the intergenerational correlation of employment that cannot be explained by either human capital or the other controls.

Extensive versus intensive margin of labor supply. In the baseline results in Table 2.2, we focus on the extensive margin of labor supply, the main interest of our investigation. To put these results into perspective, we include now a measure of intensive margin of labor supply: weekly working hours. Table 2.3 repeats the estimates of β for employment status in the first two columns (specification (1) and (4) in Table 2.2). In the last two columns, we show the analogous coefficients of a regression using hours worked per week instead of employment (we include the periods of non-employment with zero hours worked).

The unconditional regression coefficients in columns one and three are both significantly positive. However, once we introduce the relevant controls in columns two and four, the coefficient in weekly hours is not significantly different from zero anymore.

This result is in line with existing research on the intergenerational correlation of hours. Using a similar methodology but different data, Altonji and Dunn 1991 find no significant correlation between mothers' working hours and those of their children – both daughters and sons – after appropriately controlling.[14] By contrast, they, as well as Toledo 2010, find a substantial intergenerational correlation in hours between fathers and their sons. Taking the evidence together one can conclude that there is a significant transmission of work behavior along the intensive

[14]To be more precise, they use the natural logarithm of weekly hours. Using the natural logarithm of weekly hours instead of the the level, we obtain similar results to Table 2.3 (the coefficient without controls is 0.17 and drops to 0.06 when controls are added).

Table 2.3: Margins of labor supply

Dependent variables: Employment - child and weekly hours - child

	Employment - child		Weekly hours - child	
Employment - mother	0.21***	0.12***		
	(0.029)	(0.027)		
Weekly hours - mother			0.13***	0.04
			(0.031)	(0.031)
Controls	NO	YES	NO	YES
Observations	2,339	1,969	2,433	2,034
Adjusted R^2	0.04	0.09	0.01	0.06

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns two and four, we use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth.

margin of labor supply only from fathers to sons, but not from mothers to their children. However, as we show, there is a substantial transmission of employment from mothers to their children, especially their daughters. Work behavior hence is transmitted also from mothers. Previous studies simply did not find this because they restricted data to periods where mothers were employed, ruling out such transmission by design.

Spousal employment. So far, we have focused exclusively on maternal labor supply variables. It is important to determine whether a father's labor supply choices also influence the employment status of the children. It may be that the unexplained association between employment of mothers and children is due to the influence of the father. Unfortunately, the NLSY79 is not designed to match fathers to their children. However, the data provide information on the employment status of spouses as reported by mothers, which we use as a proxy for father's employment.

The first column of Table 2.4 repeats the baseline result for the sub-sample in which we also observe the spousal employment status (specification (4) in Table 2.2). Column two shows the regression output, where, instead of the maternal employment, we regress child employment on the spouse's employment status and covariates. We observe no significant effect on the children's lifetime employment. In the third column, we include both the maternal and spousal employment status and observe that the coefficient on maternal lifetime employment is not different from the one in the baseline specification (4) in column one, whereas the coefficient on spousal employment is insignificant. Finally, when we also introduce an interaction term be-

Table 2.4: Spousal employment status

Dependent variable: Employment - child (l_{Ci})

Employment - mother	0.13***		0.13***	0.13***
	(0.032)		(0.032)	(0.032)
Employment - spouse		0.05	0.04	0.06
		(0.074)	(0.077)	(0.089)
Emp. - mother x Emp. - spouse				0.22
				(0.215)
Controls	YES	YES	YES	YES
Observations	2,086	2,086	2,086	2,086
Adjusted R^2	0.09	0.08	0.09	0.09

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In all columns, we use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth. The regressions correspond to the triplets spouse-mother-child for which a spouse is reported. Note that not all mothers report having a spouse in all the waves, nor are their spouses the same across waves.

tween mothers' and spouses' employment status (fourth column), this coefficient is positive and large but not statistically significant because of considerably large standard errors. Nonetheless, it suggests that when the employment status of the mother and her spouse are similar, this has a positive effect on the child's employment.

Robustness. The main result of a positive and significant correlation between maternal and children's lifetime employment is robust to several changes in the specification. Variants in the specification are presented in more detail in the Appendix (Section 2.B.3 explains additional details of some exercises, and the tables with results are shown in Section 2.C). First, as is usual for the estimation of earnings correlations, we estimate equation (2.1) with logs of the permanent components (Table 2.10). Second, following Chetty, Hendren N., et al. 2014, we estimate rank-rank regressions for average employment status of mothers and children (Table 2.11). Third, we adopt two alternatives in computing the permanent components: (i) simple averages of the employment status as the permanent component (without controlling for life-cycle or business-cycle fluctuations) as in the early literature (for example, Solon 1992); and (ii) including controls for demographic events into the calculation of the permanent components (Table 2.12). Finally, we also show that our results are robust to the use of other questions in the survey that allow for the inference of employment status but are less comparable across cohorts or less complete across years (Table 2.13).[15]

2.4.2 Heterogeneous employment correlations

In this section, we analyze whether the established fact of a significant and positive intergenerational correlation of lifetime employment differs across relevant dimensions, such as gender (daughters in comparison to sons) and socio-economic background (maternal education and income). We hence partition the sample in three different ways:

(i) according to the child's gender: $\mathcal{G}_1 = \{\text{sons, daughters}\}$

(ii) according to the highest formal maternal education: $\mathcal{G}_2 = \{$incomplete high school, complete high school, incomplete college, complete college$\}$

(iii) according to the mother's family income quintile: $\mathcal{G}_3 = \{$quintile 1,..., quintile 5$\}$

For all $k \in \{1, 2, 3\}$ the estimated models follow the specification,

$$l_{Ci} = \alpha + \sum_{G \in \tilde{\mathcal{G}}_k} \zeta_G \mathbb{I}_{i \in G} + \beta l_{Mi} + \sum_{G \in \tilde{\mathcal{G}}_k} \beta_G \mathbb{I}_{i \in G} l_{Mi} + \phi_M X_{Mi} + \phi_C X_{Ci} + \epsilon_i, \tag{2.4}$$

where the first group of each partition is our reference group (for example, sons in partition $\mathcal{G}_1$) and $\tilde{\mathcal{G}}_k$ denotes the partition without this first group (for example, $\tilde{\mathcal{G}}_1 = \{\text{daughters}\}$). The indicator variable $\mathbb{I}_{i \in G}$ takes the value one when child i belongs to group G and zero otherwise. In the following we discuss the coefficient β_G and/or the marginal effect $\beta + \beta_G$ of mother's employment on the employment of children in the corresponding group G.

Gender. The first column of Table 2.5 shows the results of estimating equation (2.4) with $\mathcal{G}_1 = \{\text{daughters, sons}\}$. The coefficient on the interaction between employment of mothers and the daughter dummy is positive and statistically significant. The intergenerational correlation of employment is 0.18 for girls and 0.07 for boys.[16] Put differently, an increase in mothers' lifetime employment by one year increases employment of their daughters' by *more than nine weeks* on average, but their sons' employment by *less than four weeks*. The stronger link between mothers and daughters in terms of employment is interesting in light of the findings

[15] Further robustness exercises, such as using education-level dummies or including interactions of covariates, also confirm the findings of the baseline estimation. They are not included in the paper but are available upon request.

[16] Note that the coefficient for boys coincides with the marginal effect, as boys are the reference group in the regression. The numbers are the regression coefficients. The corresponding correlation coefficients (see equation (2.3)) are 0.21 and 0.06, respectively. The difference across genders increases as a consequence of disparities in standard deviations of lifetime employment.

in the literature on intergenerational correlations of earnings that report lower estimates for daughters than for sons (see, for example, Chadwick and Solon 2002, Olivetti and Paserman 2015). It is also suggestive of a role-model effect, as role models are intuitively more likely to be gender specific. Nevertheless, the correlation between mothers' and sons' employment is still significantly positive, suggesting that the role-model effect exceeds a pure transmission of gender roles.

Table 2.5: Gender differences

Dependent variable: Employment - child (l_{Ci})		
	Equation (2.4)	Marginal effect
Employment - mother	0.07**	0.07**
	(0.033)	(0.033)
Employment - mother × Daughter	0.11**	0.18***
	(0.051)	(0.041)
Controls	YES	
Observations	1,969	
Adjusted R^2	0.11	

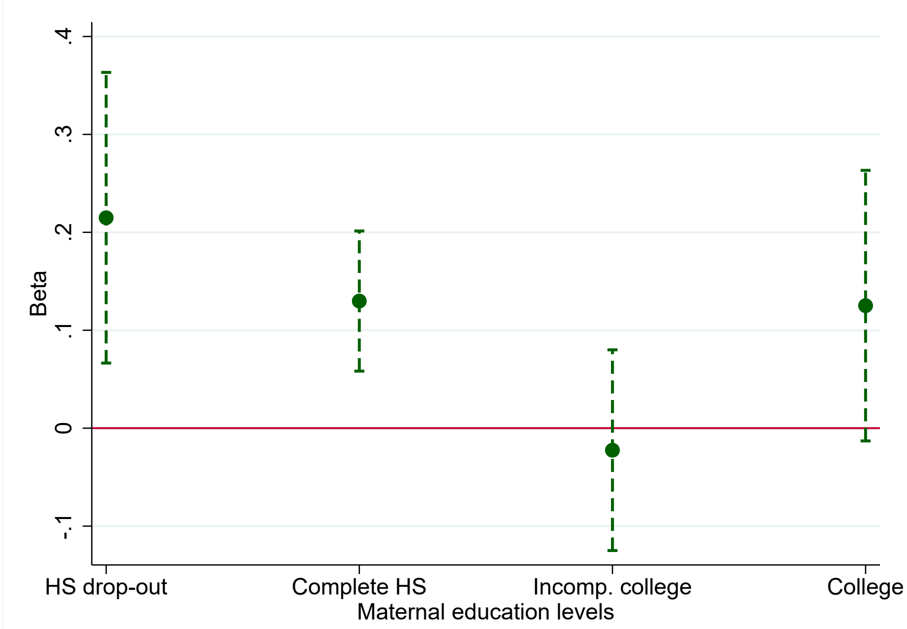
Notes: Standard errors clustered at the mother level in parentheses; standard errors calculated using the delta method for the marginal effects. *** p<0.01, ** p<0.05, * p<0.1. In all columns, we use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth.

Maternal education. The intergenerational correlation of employment status is stronger the more disadvantaged the educational background of the mother. Figure 2.1 depicts the marginal effects of mothers' employment for each education level in $\mathcal{G}_2$. It is the highest and significantly positive for mothers with no degree (0.21) or a high-school degree (0.13). It is close to zero for mothers who attended college but did not complete it. Interestingly, if they obtained a college degree, the coefficient is again positive, suggesting a non-linearity in the transmission of employment status.[17]

Maternal family income. Since education is a crucial determinant of income, it should not be surprising that similar conclusions hold true when we consider maternal family income. Figure 2.2 shows the marginal effects of mothers' employment on children for each income quintile. The estimated coefficient is highest for children from mothers in the lowest income quintile (0.18). It then monotonically decreases in income, reaching zero at the fourth quintile.

[17]The corresponding regression results are reported in Table 2.14 in the Appendix. It can be seen that the interaction of mothers' employment with incomplete college is statistically significant.

Figure 2.1: Intergenerational correlation of employment status by maternal education



Notes: Standard errors clustered at mother level, determined using the delta method. 95% confidence level intervals. The dependent variable is the permanent component of the employment status of the children. The maternal education is the maximum attained and observed education level. We use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth.

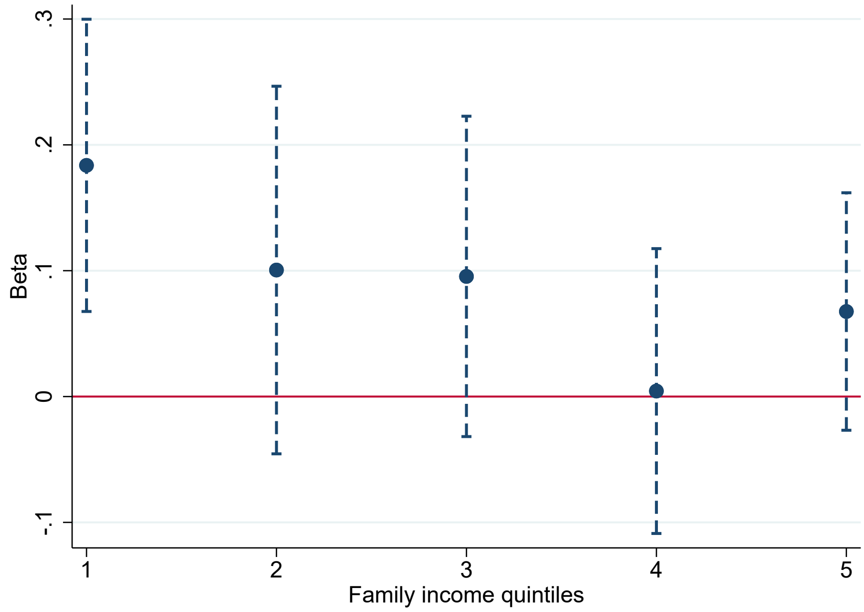
Interestingly, the point estimate for the fifth quintile is again positive. However, only for the first income quintile is the coefficient significant at the 5% level.

This pattern – a higher transmission of employment status at the bottom of the income distribution – is the same for daughters and sons, as Figure 2.9 in the Appendix shows.[18] In particular, mothers from low-income families tend to transmit their employment status to their daughters much more than mothers with higher family income. By contrast, Olivetti, Patacchini, and Zenou 2018 find that gender roles are transmitted more at the top of the income distribution. This discrepancy supports our claim that the residual employment correlation we document is not entirely the result of a transmission of gender roles.

The fact that the transmission of employment status is strongest for low-income earners is particularly interesting in light of existing income tax credits for low-income families with children, such as the EITC in the United States. Such programs directly encourage labor force participation of eligible recipients. If participation of these recipients is transmitted to their children (and hence their children's children, etc.), it may indirectly generate higher labor

[18]Figure 2.9 further shows that education also affects the transmission of employment to girls and boys similarly.

Figure 2.2: Intergenerational correlation of employment status by family income quintiles



Notes: Standard errors clustered at mother level, determined using the delta method. 95% confidence level intervals. The dependent variable is the permanent component of the employment status of the children. Quintiles of family income correspond to the quintile observed in the majority of the survey years. We use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth.

income tax revenues in the following generations. Hence, there may be a dynamic fiscal benefit of such programs. However, before drawing normative conclusions from our – so far positive – analysis, it is necessary to get a better understanding of the precise mechanism through which employment status is transmitted. This is the focus of the remainder of this paper.

2.5 Potential mechanisms

In this section we evaluate potential mechanisms that could explain the significantly positive intergenerational correlation of employment status between mothers and their children. In the first part we discuss how far the transmission of attitudes toward work – or "work culture" – could explain the observed results. Particularly, we provide some evidence suggesting that there may be a role-model effect.

In the second part, we rule out several other mechanisms that could in theory explain the facts. Neither networks, occupation-specific human capital nor local labor markets seem to be a driving force behind the main result in Section 2.4.

2.5.1 Work culture

One way to interpret the results is that parental preferences for work or employment of parents affect the attitude that children have towards work.[19] Therefore, when children inherit work attitudes from their parents, it is important to distinguish two potential channels, through which these attitudes may be transmitted. They are schematically represented in Figure 2.3. First, it could be that preferences are transmitted directly: a mother who dislikes working tends to have children who dislike working independent of her working behavior. Second, it could be a role-model effect: observing the mother participating in the labor market influences the child to develop a more positive attitude towards work.

This differentiation is important for policy analysis or dynamic scoring. For example, when evaluating the desirability of in-work benefits, only in the presence of a role-model channel will such benefits lead to higher income tax revenue raised from future generations. By contrast, if preference transmission does not operate through a role model, for example if children learn from what parents express or if genes play a role, such policies may increase the employment of mothers, but this increase will not spill over to their children and hence will have no effect on future income tax revenue.

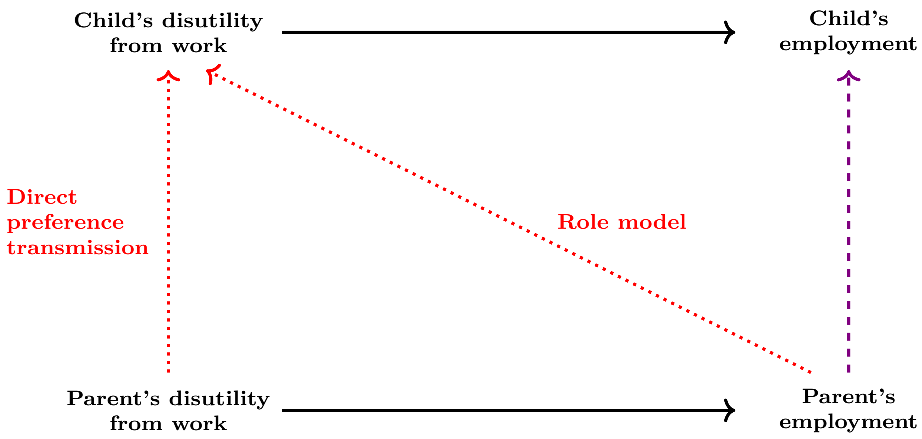
Figure 2.3 illustrates these ideas. We observe a link between children's and parents' employment choices (purple line), and we infer that, after controlling for relevant observed factors (mainly ability, education, and wealth), there is a relation with preferences for work generating this link (red lines). The relation may arise either through direct preference transmission (relating parents' preferences and children's preferences directly) or through a role model (parents' employment choices influence children's preferences) or through a combination of both.

Disentangling the two potential channels is a difficult task because preferences are not directly observable. However, our data provides three pieces of evidence favouring the existence of a role-model channel.

Role models are more pronounced within the same gender. The first piece of evidence was already presented above in the context of our heterogeneity analysis (see Section 2.4.2). Specifically, we showed that the intergenerational correlation in employment between mothers and daughters is significantly higher than the one between mothers and sons (Table 2.5). Role models are more pronounced within the same gender. For example, Bettinger and Long 2005 document that having a female instructor in an initial course at university makes female stu-

[19] In Appendix 2.A we formalize this idea within a simple two-generations model based on Solon 1999. In this model, we allow children's preferences to be affected by parental employment. The child's optimality condition is an intergenerational equation comparable to the one estimated above, with the coefficient on parent's (lifetime) employment precisely capturing this effect. A theory of work culture is hence consistent with the observed significant correlation between mothers' and children's employment.

Figure 2.3: Direct preference channel versus role-model channel



dents more likely to select courses or major in the same subject later on. If preferences were transmitted only directly, we should not observe such an effect.

Measures of work preferences. To obtain the second piece of evidence, we create a measure of work preferences for mothers and directly control for this measure in our regression analysis. While, as mentioned above, preferences cannot be directly observed, two questions in the NLSY79 are related to work preferences and we will make use of them in the following analysis:[20]

(i) Women's place is in the home, not in the office or shop.

(ii) Women are much happier if they stay at home and take care of the children.

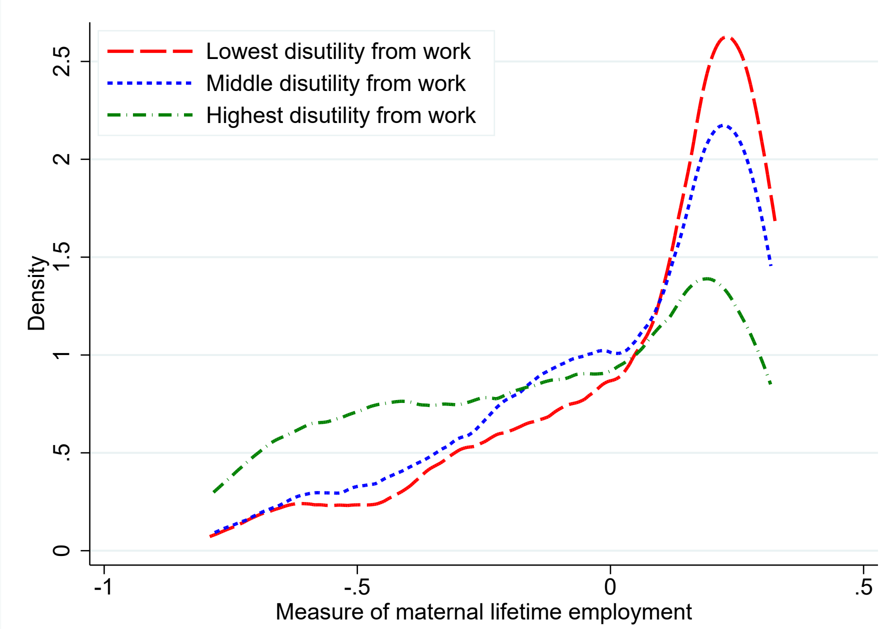
While these survey questions relate foremost to gender roles, they also contain information on mothers' preferences for work. The answers in the survey are given qualitatively. We hence construct a quantitative variable, for which we code the answers of each question such that a higher value represents a higher disutility from work.[21]

Figure 2.4 shows that there is indeed a sensible relationship between our constructed measure of disutility from work and the employment behavior of the mothers' cohort. The three lines correspond to the three terciles of our constructed measure. We observe the expected relationship: the distribution of the permanent employment component of mothers with low disutility from work has more mass at the right and less mass at the left than the distribution of the middle tercile; by contrast, the distribution conditional on high disutility of work has more mass at the left and less at the right.

[20]See Appendix 2.B.4 for details. We use information only on the mothers because the analogous data for the children's cohort do not seem accurate, besides not being informative for male respondents.

[21]The resulting variable is directly comparable to the disutility parameter θ in the model of Appendix 2.A.

Figure 2.4: Distribution of measure of maternal lifetime employment by levels of disutility from work



Notes: Disutility from work computed from questions on women's roles: (i) Women's place is in the home, not in the office or shop, and (ii) Women are much happier if they stay at home and take care of the children. Included in survey years 1979, 1982, 1987, and 2004. We assign the values: (a) strongly agree 1.5, (b) agree 0.5, (c) disagree -0.5, and (d) strongly disagree -1.5. We average across questions and across years.

Using our constructed measure for disutility of work θ_{Mi}, we then run the regression,

$$l_{Ci} = \alpha + \beta l_{Mi} + \omega\theta_{Mi} + \phi_M X_{Mi} + \phi_C X_{Ci} + \epsilon_i. \tag{2.5}$$

Table 2.6 shows the results. The first column repeats the baseline estimation for comparison. The second column introduces our measure of disutility from work of the mother and excludes employment of the mother. The third column shows the results of including the preferences for work of the mother in our baseline specification, i.e. the estimation results of equation (2.5). The coefficient on employment of the mother does not change, and the coefficient on the disutility from work is close to zero. Finally, column four shows the same estimation but restricting the sample to daughters. The results are qualitatively the same, and the previous finding of a higher coefficient of maternal employment for daughter's employment behavior is confirmed. Again, the coefficient on mother's work preferences is close to zero and insignificant.

Importantly, while our measure of disutility from work is significantly negatively correlated with the employment behavior of mothers (the correlation coefficient is -0.27, statistically significant at the 99% confidence level), it does not affect the employment behavior of children.

Furthermore, including this measure in the baseline specification does not affect the coefficient on the mother's employment. These results suggest that the role-model channel is an important driver of the intergenerational correlation of employment, while there seems little or no direct transmission of work preferences.

Table 2.6: Direct preference transmission vs. role model: Measures of work preferences

Dependent variable: Employment - child (l_{Ci})

Specification	Baseline	Maternal preferences	Full	Only daughters
Employment - mother	0.12***		0.12***	0.17***
	(0.027)		(0.027)	(0.043)
Disutility from work - mother		0.02	0.03*	0.01
		(0.014)	(0.014)	(0.022)
Controls	YES	YES	YES	YES
Observations	1,969	1,969	1,969	984
Adjusted R^2	0.09	0.08	0.09	0.14

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth. Disutility from work computed from questions on women's roles: (i) Women's place is in the home, not in the office or shop, and (ii) Women are much happier if they stay at home and take care of the children. Included in survey years 1979, 1982, 1987, and 2004. We assign the values: (a) strongly agree 1.5, (b) agree 0.5, (c) disagree -0.5, and (d) strongly disagree -1.5. We average across questions and across years.

Cohabitation. The third and last piece of evidence, supporting the existence of a role-model channel, results from controlling for mothers' permanent component of employment based on periods when they do not live together with the child. This measure serves as another proxy for mothers' work preferences that would be transmitted directly. The idea is that a role-model channel is at work only when children actually observe the behavior of their mothers, which is facilitated during cohabitation.

For each child, we split the observations of the child's mother into those when they are both cohabiting and those when they are not. Non-cohabitation includes periods before the child's birth and after the child leaves home, independent of whether other children are living in the household. We estimate the permanent component for mothers using only the non-cohabitation period and re-estimate the intergenerational equation introducing this variable to control for mothers' preferences for work. We only use those mother-child pairs for which we have periods of both cohabitation and non-cohabitation.[22]

The results are presented in Table 2.7: also when controlling for maternal preferences for work in the described way, the role of maternal lifetime employment remains relevant and predominant. Furthermore, these periods of non-cohabitation do not seem to add information once lifetime employment is taken into account. This supports the preponderance of the role-model channel.

Table 2.7: Direct preference transmission vs. role model: Periods of non-cohabitation

Dependent variable: Employment - child (l_{Ci})

Specification	Baseline	Maternal preferences	Full
Employment - mother	0.16***		0.14***
	(0.037)		(0.041)
Employment - mother when...		0.07***	0.02
...not cohabiting with child		(0.027)	(0.030)
Controls	YES	YES	YES
Observations	1,123	1,123	1,123
Adjusted R^2	0.11	0.10	0.11

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We use the same covariates as we do in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth. Periods of non-cohabitation are specific for each child-mother pair. Only pairs with both periods of cohabitation and non-cohabitation are included. As this affects the composition of mother-child pairs included in the regression, the baseline results change slightly compared to Table 2.2.

2.5.2 Mechanisms that can be ruled out

While the presented evidence suggests that work culture, or, more specifically, a role-model channel, is responsible for the observed intergenerational correlation in employment status, there are other factors that may well explain this correlation. In this section we briefly discuss three other candidate mechanisms and provide evidence that neither of them is likely to be the driving force behind the results.

[22] In an alternative specification, we use the periods of cohabitation and non-cohabitation to compute two distinct permanent components (see Appendix 2.B.5). The results, shown in Table 2.17, are perfectly in line with the findings in Table 2.7: the effect of maternal employment during periods of cohabitation has a positive and significant effect (0.12) on children's lifetime employment, while employment during non-cohabitation periods is not significantly different from zero.

Networks or occupation-specific human capital. Parents might help children find a job through their connections, or even transmit occupation-specific human capital or preferences leading to correlations in job-finding probabilities across generations.[23] In order to test whether those mechanisms are plausible explanations for the residual intergenerational correlation of employment, we do the following. We split the sample between mother-child pairs who are employed in the same type of business (proxied by industry and sector) or have the same occupation, and those who have different industry/occupations.[24] Industries, sector and occupation are assigned to the individuals according to the category observed most of the survey years. In particular, we estimate equation (2.4) using the partitions $\mathcal{G}_4 = \{$different industry-sector, same industry-sector$\}$ and $\mathcal{G}_5 = \{$different industry-occupation, same industry-occupation$\}$.

The first two columns of Table 2.8 show the results. They suggest that the correlation of employment is not different for mother-child pairs who share the same type of business or occupation. The estimation is imprecise because there are few observations for which child and mother share these traits (18% for industry and sector, and 5% for industry and occupation). However, the point estimates suggest a pattern opposite to what a network channel or transmission of specific human capital would suggest: the correlation of employment is smaller when child and mother are in the same type of business or occupation. This evidence does not support a story of employment correlations driven by networks or specific human capital transmission.

Local labor markets. As a last exercise, we evaluate whether local conditions of the labor market could explain our correlation. So far, our argumentation has revolved around labor supply decisions. However, the estimated correlation could also be driven by market conditions that are determined by labor demand: if mothers and children live in the same region, both generations face similar labor market conditions, i.e. similar separation and job-finding probabilities.

The general version of the NLSY79 contains three different geographic variables but not a precise regional identifier. We hence undertake the following strategy. First, we condition our analysis on the mother-child pair living in the same broadly defined region.[25] Second, we define a variable that indicates if both the mother and the child live in the same region as well

[23] The role of nepotism and preferences for occupations in the intergenerational correlation of earnings has been documented in the literature. See, for example, Corak and Piraino 2011 and Lo Bello and Morchio 2018.

[24] Industries according to the three-digit Census classifications are grouped in 14 aggregate categories, and a similar aggregation is done for occupations to 18 categories. The sectors considered are private, public, self-employment, and family businesses.

[25] The variable region indicates whether the individual lives in one of four areas, Northeast, North Central, South or West. 93% of the mother-child pairs share the region of residence.

Table 2.8: Intergenerational correlation of employment status by (i) same industry-sector, (ii) same industry-occupation, (iii) same region, and (iv) same region-SMSA-urban/rural

Dependent variable: Employment - child (l_{Ci})

	Industry-sector	Industry-occupation	Region	Region-SMSA-urban/rural
Employment - mother	0.13***	0.12***	0.26**	0.12***
	(0.030)	(0.028)	(0.105)	(0.030)
Employment - mother × Same	-0.07	-0.11	-0.15	-0.01
	(0.059)	(0.080)	(0.109)	(0.061)
Controls	YES	YES	YES	YES
Observations	1,969	1,969	1,969	1,969
Adjusted R^2	0.09	0.09	0.09	0.09

Notes: Standard errors clustered at the mother level in parentheses; standard errors calculated using the delta method for the marginal effects. *** p<0.01, ** p<0.05, * p<0.1. Industry, sector, occupation, region, SMSA and urban/rural are assigned as the category that is observed in the majority of the survey years. In all columns, we use the same covariates as in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth.

as in an urban or rural area and in a Standard Metropolitan Statistical Area (SMSA).[26] We assign residence according to the category observed in the majority of survey years, and we compute the intergenerational correlation of employment distinguishing mother-child pairs for which their categories coincide or not.

The last two columns of Table 2.8 present the estimates. Residence in the same location does not significantly affect the employment correlation. Again, the marginal effects for pairs that share geographical variables are smaller than the effects for pairs whose variables differ. Therefore, we do not find evidence that local labor markets can explain the significant intergenerational correlation of employment.

2.6 Conclusion

This paper contributes to the literature on the intergenerational correlation of labor market outcomes. Differently from the existing literature, we focus on the extensive margin of labor supply. Using the NLSY79 and the CNLSY79 we document a robust, statistically significant,

[26] The measure is still imperfect because it could be that both live in an urban area within the same broad region and in an SMSA that could be a different metropolitan city. But only 24% of the observations correspond to pairs living in the same combination of geographical variables.

and positive intergenerational correlation of employment status between mothers and their children. After controlling for the channels that drive the transmission of earnings potential, we find that an increase in lifetime employment of mothers by one year is associated with an increase in the employment of her child by *six weeks*. The correlation is higher for mother-daughter pairs than for mother-son pairs. Furthermore, it is lower when the maternal education and the family income are higher.

While the analysis of this paper is a purely positive one, it has potentially important normative implications. For example, in-work benefits, such as the EITC in the United States, paid to the currently working generation may indirectly increase the employment – and thus income tax revenue – of future generations. This is especially the case if these programs are targeted to low-income families with children. More generally, dynamic scoring of any redistributive policy that affects incentives to work should take this transmission channel into account.

However, a comprehensive policy analysis requires a clear understanding of the mechanism, through which employment status is transmitted across generations. We show that the results are consistent with a theory of work culture and provide suggestive evidence that in their employment decisions, mothers act as a role model for their children, especially for their daughters. We are able to rule out network effects, occupation-specific human capital, and local labor markets as driving forces behind the result.

Appendix

2.A Two-Generations Model

The model is a simple two-generations framework based on Solon 1999. The main addition to it is that children's preferences towards work are (potentially) affected by parental labor force participation.

There is a continuum of families, each consisting of one parent and one child.[27] Generations are indexed by $k \in \{M, C\}$ for parents and children, respectively. Parents are altruistic but discount their child's expected utility by a factor $\alpha \in [0, 1)$. They decide on consumption c_M, labor supply l_M, and human capital investment for their child H. Children decide on consumption c_C and labor supply l_C, but they do not have any offspring and hence do not invest in human capital. Agents are heterogeneous in ability e_k and disutility from labor θ_k.[28] Abilities are correlated across generations, accounting for genetic inheritance.

The parents' optimization problem is given by

$$\begin{aligned}
V_M(\theta_M, e_M, v_M) = &\max_{c_M, l_M, H} \frac{c_M^{1-\sigma}}{1-\sigma} - \theta_M \frac{l_M^{1+\chi}}{1+\chi} + \alpha \mathbb{E}\big[V_C(\theta_C, w_C)\big] \\
\text{s.t.} \quad & c_M + pH = w_M \, l_M \\
& \log(w_M) = \log(e_M) + v_M \\
& \log(\theta_C) = \kappa_0 - \kappa_1 \log(l_M) + \eta_C. \qquad (2.6)
\end{aligned}$$

We assume that utility is additively separable in consumption and labor. The parameter $\sigma > 0$ is the coefficient of relative risk aversion and $\chi > 0$ is the inverse of the Frisch elasticity of labor supply. Parents finance consumption c_M and investment in their child's human capital

[27] The exposition of the model uses the word parent for the sake of generality, even if we use mothers in the empirical analysis. For consistency with the notation in the empirical setup, we denote the parents with the indicator M.

[28] Whereas differences in productivity among children are captured explicitly by both e_C (ability) and H (education), e_M represents for parents a combination of abilities and education, the latter not being modeled.

H, a unit of which costs p, with labor earnings $w_M l_M$. The wage of the parent is determined through ability e_M and a random term v_M, which captures labor market luck.

The last equation (2.6) is the process of intergenerational transmission of preferences for work. Children's disutility from labor, θ_C, (potentially) depends on the parental labor supply decision l_M, through a parameter κ_1. A value of κ_1 different from zero means that parents' labor supply has an effect on children's preferences for work. We do not impose any prior on the direction of the effect. If $\kappa_1 > 0$, then the more parents work, the less children dislike working, and the opposite is the case for $\kappa_1 < 0$. If $\kappa_1 = 0$, then parental employment does not have any influence on children's preferences for work. The parameter η_C is an idiosyncratic preference shock.

Similarly, the child's optimization problem is given by

$$V_C(\theta_C, w_C) = \max_{c_C, l_C} \frac{c_C^{1-\sigma}}{1-\sigma} - \theta_C \frac{l_C^{1+\chi}}{1+\chi} \tag{2.7}$$

$$\text{s.t.} \quad c_C = w_C\, l_C \tag{2.8}$$

$$\log(w_C) = \log(e_C) + \psi \log(H) + v_C \tag{2.9}$$

$$\log(e_C) = \lambda \log(e_M) + u_C. \tag{2.10}$$

Children finance their consumption with labor earnings. Wages w_C of children depend on their ability, e_C, on the acquired human capital H (which has a return ψ), and v_C, which captures labor-market luck. The last equation states that ability is partially inherited. To be specific, the parent's and child's ability are linked via an AR(1) process with persistence $\lambda \in (0, 1)$.

Note that in the model, l_M and l_C are continuous variables, although we focus on the extensive margin of labor supply. In the model, we think of l_M and l_C as the time share in employment over the whole lifetime. This maps well into our empirical analysis, in which we employ the permanent component of employment status.

The Solution. We focus on the solution of the children's problem because it enables us to summarize the relevant model predictions. To be specific, we take parental decisions and realizations of shocks as given. Then, the first-order condition for labor supply l_C can be written as

$$\log(l_C) = -\frac{1}{\sigma+\chi} \log(\theta_C) + \frac{1-\sigma}{\sigma+\chi} \log(w_C). \tag{2.11}$$

We can substitute for $\log(\theta_C)$ with (2.6) and $\log(w_C)$ with (2.9) and obtain

$$\log(l_C) = \alpha + \beta \log(l_M) + \gamma \log(e_M) + \delta \log(H) + \epsilon, \tag{2.12}$$

where the coefficients α, β, γ and δ are functions of structural model parameters. Specifically,

$$\beta = \frac{\kappa_1}{\sigma + \chi}. \tag{2.13}$$

This resulting intergenerational equation of employment status (2.12) is similar in many respects to the models we estimated in Section 2.4.1. It relates children's and parents' employment decisions once human capital decisions and ability transmission have been taken into account. Importantly, employment decisions conditional on human capital and ability are related across generations through the coefficient β. β is proportional to, and has the same sign as, κ_1, which determines how parents' labor supply translates into children's attitude towards work. Equation (2.12) thus provides an empirical test for the presence of the transmission of preferences for work. Because in our estimation $\beta > 0$, according to our theory the child's disutility from work decreases with parental labor supply.

Although the essence of the solution (2.12) coincides with the type of estimated models in Section 2.4.1 (see Table 2.2), there are some differences. Apart from some factors not present in the model, for simplicity (for example, wealth, fertility), the specification in the model is in logs, whereas the empirical specification is linear. This choice responds to simplicity both in the model and in the empirical estimation.[29] As we showed already, the empirical results are robust to a vast set of changes in the specification.

2.B Details on the empirical analysis

2.B.1 Details on the data

NLSY79 and CNLSY79. The data is collected and provided freely by the Bureau of Labor Statistics (BLS) in the US. The NLSY79 consists of three sub-samples: (i) the cross-sectional sample (6,111 individuals) is a representative sample of the US population in 1979, (ii) the supplemental sample (5,295 individuals) over-samples disadvantaged groups (Hispanic or Latino, black and poor people), and (iii) the military sample (1,280 individuals) over-samples the population participating in the army. As explained in the main text, we use only the cross-sectional sample and restrict ages to 25 to 45 years old. Figure 2.5 provides an example for a mother-child pair in the data.

[29] Using the linear relationship has the advantage of avoiding arbitrary transformations of the data. Not all permanent components are above 0. Hence, to be able to use the log-specification, we need to shift all permanent components to ensure that they are above 0. But these shifts complicate the interpretation of the coefficients because they are not invariant to the size of the shift. Furthermore, the interpretation of results is very intuitive in the linear setup.

It is worth noting some features of the sample we use for the analysis. Figure 2.6 shows the distribution of the number of interviews. The mode for mothers is 15, with around 75% of the mass concentrated between 14 and 16 interviews. For children, the mode is 3, and only 50% have 3 or more interviews. The left panel of Figure 2.7 shows the distribution of the age of mothers at birth. Of the observations, 80% come from mothers who gave birth between 20 and 23 years old. The right panel of Figure 2.7 shows the same distribution, broken down by number of interviews of the children. Mothers of children with more interviews were younger when their children were born. Figure 2.8 shows the employment-age profiles of mothers and children. The composition of the children's sample, biased towards younger children, as explained in the main text, is also behind the atypical employment-age profile for the cohort. Employment rates decline and become more volatile with age because older children are fewer and belong to mothers who were younger at birth, something the empirical strategy accounts for when computing the permanent components. Furthermore, the dip in the employment rate at the age of 35 to 36 for children reflects the 2008 crisis, which particularly affected younger cohorts.

Ability is measured in the 1979 cohort by the Armed Services Vocational Aptitude Battery (ASVAB), which was collected around 1980 when mothers were between 15 and 23 years old. The scores correspond to the AFQT, which is a composite of test results in arithmetic reasoning, word knowledge, paragraph comprehension, and numerical operations. We use the version of the AFQT revised in 2006 to control for differences in cohorts within the NLSY79. Similar measures of cognitive abilities have been collected for the children cohort since 1986. In particular, we use the latest measurement for each child of the Peabody Individual Achievement Test (PIAT) for Math, considered the most appropriate measure of ability among the test scores available in the data for the younger cohort (Abbott et al. 2013). These measures may capture not only genetic ability, but also some components of scholastic skills. This is not a problem for our analysis, as we are interested in accounting for productivity jointly with education.

Another relevant variable in the analysis, wealth, is introduced as net worth, i.e. assets minus debts. The variable is provided by the BLS for the NLSY79 cohort, and we follow the definition in the CNLSY79, where such a computed variable is not provided. In terms of assets, we include savings in liquid accounts and in financial assets, the market value of the main house and other properties, and the market value of own vehicles. The debts comprise credit card balances, outstanding mortgage value and other property debts, debts for vehicles, and other debts. The net worth variable constructed by the BLS uses imputed assets and debts when there is no response, and values are top-coded. No such procedures are followed in the children's cohort, and also there are some slight changes in the definitions of assets and debts over time.

Earnings is also a variable used throughout the analysis. We use an annual measure, the most comparable variable across cohorts: wages and salaries received during the last calendar year.

Earnings are top-coded for both the parents' and children's cohorts. We construct weekly hours of work, dividing total annual hours by total number of weeks worked during the last calendar year for the mothers' cohort. For the children's cohort, we use weekly hours worked in all jobs, as reported in the survey.

Industries are available according to different versions of the three-digit US Census classification. For the comparison of industries across generations, they are grouped into 14 categories: agriculture, forestry, fisheries; mining; construction; manufacturing of non-durables; manufacturing of durables; transportation, communications, and other public utilities; wholesale trade; retail trade; finance, insurance, and real estate; business and repair services; personal services; entertainment and recreation services; professional and related services; public administration. Similarly, the classification of occupations also corresponds to three-digit US Census classification. They are collapsed into 18 categories: management, business, and financial operations; computer and mathematical; architecture and engineering; life, physical, and social services; community and social services; legal; education, training, and library; arts, design, entertainment, sports, and media; health-care practitioners and technical and support; protective service; food preparation and serving related; building and grounds cleaning and maintenance; personal care and service; sales and related; office and administrative support; farming, forestry, and fishing; construction and extraction, installation, repair and maintenance, and production; transportation and material moving. The variable accounting for sectors refers to private, public, self-employment, and family businesses.

The geographical information on the publicly available version of the NLSY79 is not very detailed. The variables are limited to region (Northeast, North Central, South, or West), urban or rural, and an indicator of residence in an SMSA, which are highly populated areas. Whenever we need to construct a measure of “location,” we use a combination of these three variables.

2.B.2 Methodological challenges in the measurement of intergenerational persistence of labor market outcomes

The data we use feature desirable characteristics for coping with some estimation issues identified in the literature on the intergenerational correlation of earnings. First, Zimmerman 1992 and Solon 1992 show that early estimations based on single-year measures of parents' and children's outcomes are subject to substantial measurement error. This is because single-year measures are subject to transitory deviations from the long-run means. This means that single-year measures are not good proxies for lifetime or permanent components, which yields attenuation bias as a consequence. This problem is particularly relevant for parental outcomes, the explanatory variables in the intergenerational equations. Mazumder 2005 estimates the potential reduction in the bias by increasing the number of observations. The longitudinal nature of the NLSY79 allows for the use of several observations for both generations, particularly in the case of mothers, who are observed on average in 14 waves in our sample (only 4% of the sample has fewer than 10 interviews).

Second, the lack of heterogeneity in the samples aggravates the measurement error (Solon 1992, Solon 1999).[30] We use a representative sample of the US population in 1979, namely the cross-sectional sub-sample of the NLSY79, which is several times bigger than cohorts formed from the Survey Research Center (SRC) component, the analogous of the PSID typically employed in empirical studies of intergenerational earnings' correlations (see, for example, Solon 1992).

Finally, the literature emphasizes a life-cycle bias that arises when parents' and children's observations are not representative of their lifetime outcomes due to non-stable trajectories along the life (S. Haider and Solon 2006, Grawe 2006, Nybom and Stuhler 2016, Nybom and Stuhler 2017). Measurement error is not homogenous along the life cycle, with higher noise for early and late years (Mazumder 2005). To mitigate this problem, the literature recommends using observations for ages between 30 and 50 (Black and Devereux 2011). Our sample restriction to individuals between 25 and 45 years old and the netting out of age effects from the permanent components are intended to mitigate this bias.

2.B.3 Details on the robustness exercises

In order to provide scale-invariant estimates of the persistence in employment, we follow the literature by providing a log-log and a rank-rank specification. It is worth noting that for the log-log specification, we take the logarithm of the permanent components, which are the fixed effects backed out in the estimation of (2.2). As these permanent components include negative values, to take the natural logarithm we add a constant such that the minimum value for each generation is 0.001. For the rank-rank specification, we sort individuals within each generation in ascending order in terms of proportion of periods employed during the 25 to 45 years old window. We assign each individual their position, divided by the total number of individuals (when an employment value is repeated, we average across positions corresponding to that value).

For the robustness exercise, in which we control for demographic events when computing the permanent components, we estimate the following slightly modified model,

$$l_{kit} = l_{ki} + \sum_{n=1}^{2} \pi_{nk} A^{n}_{kit} + \lambda_{kt} + Demo'_{kit}\varsigma + \upsilon_{kit},$$

[30]The interaction between, on the one hand, transitory fluctuations and measurement error, and, on the other hand, the homogeneity in the sample, is discussed in Solon 1989.

where $k \in \{M, C\}$ and $Demo_{kit}$ are controls for demographic events, including births, couple formation and dissolution, job loss and finding by partner, presence of children 0 to 3 years old in the household with/without child care, and presence of older children in the household. We also include controls for education level, region, urban area, living in own dwelling, conjugal status, and whether the partner works.

The alternative variables used to measure employment status are (i) the preferred employment questions without including the requirement of a minimum time or earnings as in the main estimation; (ii) answers to the Current Population Survey (CPS)-type employment status question in the mothers' cohort, and response to whether they have any employer at the time of the survey, for the children's cohort; (iii) questions about hours and earnings (employment corresponds to a positive number of hours and earnings, in the last year for the mother's cohort, and in the year of the survey for the children); and (iv) questions about hours only (last year for mothers, current year for children). As discussed, these questions are less comparable across generations than our preferred measure, and are only available for fewer periods. We also include labor force status for mothers, for whom unemployment questions are also available (this is not the case for the children's cohort).

2.B.4 Details on the preferences for work in NLSY79 and CNLSY79

As referred to in the main text, the questions about women's roles that provide information on preferences for work are (i) Women's place is in the home, not in the office or shop, and (ii) Women are much happier if they stay at home and take care of the children. The questions are included in survey years 1979, 1982, 1987, and 2004 for mothers and 1994, 1996, 1998, 2002, 2006, and 2010 for children. These are qualitative questions, which we quantify with a range centered at zero. We assign the following values: (a) strongly agree 1.5, (b) agree 0.5, (c) disagree -0.5, and (d) strongly disagree -1.5. We average across the two questions for each year and across the years.[31]

Figure 2.10 depicts the distribution of the resulting variable of *maternal disutility from work*. It is slightly skewed to the right, which means that there is an over-representation of mothers with low disutility from work, which is in agreement with a considerably high employment rate (73%).

[31]If information on a variable is missing in a year, we use only the available information for the other variables for that year. This way, we put equal weight on all years. We checked that alternatives (either averaging only the information on the first or the second question) do not change the results. Our preferred measure of disutility from work is negatively correlated with the permanent component of employment (the correlation is -0.27 unconditional, and -0.15 once we controlled for all the covariates in the previous regressions, both statistically significant at the 99% confidence level).

Furthermore, we take terciles of the variable, which gives us three classes that we describe as low, medium, and high maternal preferences for work. Summary statistics for the disutility from work by terciles are shown in Table 2.15.

Table 2.16 shows that whereas the proportion of children employed clearly varies with the employment of the mother, it does not show the same gradient with respect to the maternal disutility from work.

2.B.5 Evidence favoring role model: Employment during periods of cohabitation versus non-cohabitation

As mentioned in the main text, we perform an additional exercise whose results support the existence of role models to drive the intergenerational correlation of employment. Differently from the exercise in the last part of Section 2.5.1, we include the permanent components of mothers' employment both when cohabiting and when not cohabiting with each respective child.

The idea behind this exercise is that the role model will only be transmitted when mother and child cohabit, but the direct transmission of preferences for work is independent of the status of cohabitation. Then, the permanent component of the mother's employment during non-cohabitation with the child will control for maternal preferences for work. Consistent with the results documented in the main text (see Table 2.7), in Table 2.17 we show that the coefficient of employment during cohabitation is significantly different from zero and of similar size as the baseline correlation in Table 2.2. In contrast, employment during periods of non-cohabitation does not play a crucial role. These results are additional evidence for the empirical relevance of the role-model channel.

2.C Additional Tables

Table 2.9: Additional summary statistics for women and mother-child pairs in NLSY79 and CNLSY79

	Women	**Mothers**	**Children**
White	80%	74%	73%
Black	13%	18%	19%
Migrant	5%	5%	0%
Public sector employees	10%	9%	4%
Private sector employees	87%	86%	92%
Self-employed	3%	4%	2%
Father at home			56%
Living in own dwelling	92%	94%	74%
Partner works	64%	68%	51%
Children 0 to 3 y.o. not in child care	17%	20%	23%
Children 0 to 3 y.o. in child care	9%	9%	9%
Children 4 to 5 y.o.	17%	23%	19%
Children 6 to 12 y.o.	40%	61%	29%
Children 13 to 15 y.o.	15%	26%	3%
Children 16 to 18 y.o.	11%	20%	1%
Births	14%	14%	18%
Couple dissolution	4%	5%	7%
Couple formation	6%	4%	18%
Partner job loss	5%	5%	6%
Partner job finding	6%	5%	16%
Individuals	3,040	1,373	2,339

Notes: Percentages for observations in the 25 to 45 years old range in our sample. For the sector of employment, the category most often observed is assigned to the individual. Similar criterium applies for the variable regarding the father living at home. The variables living in own dwelling, partner works, children of different ages, births, couple dissolution and formation, and partner job loss and job finding capture the number of observations for which they take the value 1 (the event occurs); they help understanding the nature of our sample.

Table 2.10: Robustness: Log-log regressions

Dependent variable: Log-employment - child ($\log(l_{Ci})$)

Specification	(1)	(2)	(3)	(4)
Log-employment - mother	0.21***	0.15***	0.15***	0.14***
	(0.034)	(0.036)	(0.037)	(0.034)
Ability - mother		0.08	0.05	0.01
		(0.061)	(0.055)	(0.056)
Ability - child		0.27***	0.14***	0.14**
		(0.061)	(0.054)	(0.055)
Yrs. schooling - mother		0.01	-0.00	-0.00
		(0.008)	(0.007)	(0.008)
Yrs. schooling - child		0.04***	0.03***	0.02***
		(0.007)	(0.006)	(0.006)
Net worth - mother			0.01	0.00
			(0.011)	(0.011)
Net worth - child			-0.01	-0.01
			(0.010)	(0.010)
Number of children - mother				-0.00
				(0.013)
Number of children - child				-0.06***
				(0.013)
Control age at birth - mother	NO	NO	NO	YES
Observations	2,339	2,237	1,969	1,969
Adjusted R^2	0.04	0.08	0.05	0.07

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In all columns, we use the same covariates as we do for the main results in Table 2.2, except that we also take the logarithm of maternal employment.

Table 2.11: Robustness: Rank-rank regressions

Dependent variable: Employment rank - child

Specification	(1)	(2)	(3)	(4)
Employment rank - mother	0.12***	0.08***	0.06***	0.06***
	(0.018)	(0.018)	(0.017)	(0.017)
Ability - mother		0.04*	0.03	0.01
		(0.020)	(0.019)	(0.018)
Ability - child		0.08***	0.05***	0.04**
		(0.019)	(0.018)	(0.018)
Yrs. schooling - mother		0.00	0.00	0.00
		(0.002)	(0.002)	(0.002)
Yrs. schooling - child		0.01***	0.01***	0.01***
		(0.002)	(0.002)	(0.002)
Net worth - mother			0.01	0.00
			(0.004)	(0.004)
Net worth - child			-0.01*	-0.01**
			(0.005)	(0.004)
Number of children - mother				0.00
				(0.004)
Number of children - child				-0.03***
				(0.004)
Control age at birth - mother	NO	NO	NO	YES
Observations	2,339	2,237	1,969	1,969
Adjusted R^2	0.03	0.08	0.05	0.10

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In all columns, we use the same covariates as we do for the main results in Table 2.2.

Table 2.12: Robustness: Alternative measures of the permanent components

Dependent variable: Alternative permanent component employment - child ($\overline{l_{Ci}}$)

	Simple averages		Demographics	
Specification	(1)	(4)	(1)	(4)
Employment - mother (averages)	0.21***	0.12***		
	(0.028)	(0.027)		
Employment - mother (demographics)			0.21***	0.12***
			(0.029)	(0.028)
Controls	NO	YES	NO	YES
Observations	2,339	1,969	2,245	1,877
Adjusted R^2	0.04	0.08	0.03	0.04

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns one and two, we use simple averages for l_{Ci} and l_{Mi}. In columns three and four, we add to the standard estimation of the permanent components demographic events as additional controls.

Table 2.13: Robustness: Alternative survey questions for employment status of children and mothers

Dependent variable: Alternative data measure of employment - child ($\widehat{l_{Ci}}$)

	Alternative measure of child employment				
	1	2	3	4	LFP
Employment - mother (different measure)	0.14***	0.05***	0.07**	0.06*	0.14***
	(0.028)	(0.019)	(0.029)	(0.031)	(0.032)
Controls	YES	YES	YES	YES	YES
Observations	1,969	2,017	1,996	2,034	1,969
Adjusted R^2	0.09	0.14	0.07	0.07	0.09

Notes: Robust standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In all columns, we use the same covariates as we do for the baseline specification (4) in Table 2.2. The employment variables in each column are the following: (1) mothers with a positive number of weeks employed in the last year and children with positive earnings in the last year (no minimum time or earnings); (2) Current Population Survey (CPS)-type employment status question in the mothers' cohort (available for a few selected years), and declaring any employer at the time of the survey for children's cohort; (3) employed if a positive number of hours and earnings is declared in the last year for the mothers' cohort and in the year of the survey for the children's; and (4) positive hours declared in last year for mothers and in current year for children.

Table 2.14: Heterogeneity: Intergenerational correlation of employment status by (i) family income (quintiles) and (ii) mother's education level

Dependent variable: Employment - child (l_{Ci})

	Baseline	Family income	Maternal education
Employment - mother	0.12***	0.18***	0.21***
	(0.027)	(0.059)	(0.076)
Employment - mother × Quintile 2		-0.08	
		(0.095)	
Employment - mother × Quintile 3		-0.09	
		(0.086)	
Employment - mother × Quintile 4		-0.18**	
		(0.081)	
Employment - mother × Quintile 5		-0.12	
		(0.074)	
Employment - mother × Complete high-school			-0.09
			(0.083)
Employment - mother × Incomplete college			-0.24***
			(0.091)
Employment Mother × Complete college			-0.09
			(0.101)
Controls	YES	YES	YES
Observations	1,969	1,969	1,969
Adjusted R^2	0.09	0.09	0.10

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Quintiles of family income correspond to the quintile of family income observed most often. The maternal education is the maximum attained education level. In all columns, we use the same covariates as we use in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth.

Table 2.15: Descriptive statistics for mothers' disutility from work by terciles

	min	max	mean	sd	Observations
Low disutility from work	-1.500	-0.750	-0.960	0.189	831
Medium disutility from work	-0.667	-0.375	-0.518	0.099	804
High disutility from work	-0.333	1.250	0.037	0.308	798
All observations	-1.500	1.250	-0.487	0.462	2433

Notes: Disutility from work computed from questions on women's roles: (i) Women's place is in the home, not in the office or shop, and (ii) Women are much happier if they stay at home and take care of the children. These questions are included only in survey years 1979, 1982, 1987, and 2004. We assign the values (a) strongly agree 1.5, (b) agree 0.5, (c) disagree -0.5, and (d) strongly disagree -1.5. We average across questions and across years.

Table 2.16: Employment – proportion of periods employed in the lifetime – of children and mothers, by terciles of mothers' employment and disutility from work

Employment	Children	Mothers
Mother's employment		
1^{st} Tercile	0.77	0.34
2^{nd} Tercile	0.86	0.80
3^{rd} Tercile	0.89	0.99
Mother's disutility from work		
1^{st} Tercile	0.84	0.78
2^{nd} Tercile	0.85	0.74
3^{rd} Tercile	0.83	0.60

Notes: Employment of mother and child correspond to the averages across years and individuals. Disutility from work: (i) Women's place is in the home, not in the office or shop, and (ii) Women are much happier if they stay at home and take care of the children. We assign the values (a) strongly agree 1.5, (b) agree 0.5, (c) disagree -0.5, and (d) strongly disagree -1.5. We average across questions and across years.

Table 2.17: Direct preference transmission vs. role model: Periods of cohabitation versus periods of non-cohabitation

Dependent variable: Employment - child (l_{Ci})

Specification	Baseline	Non-Cohabitation	Cohabitation	Both
Employment - mother	0.16***			
	(0.037)			
Employment - mother when...			0.13***	0.12***
...**cohabiting** with child			(0.033)	(0.035)
Employment - mother when...		0.07***		0.04
...**not cohabiting** with child		(0.027)		(0.027)
Controls	YES	YES	YES	YES
Observations	1,123	1,123	1,123	1,123
Adjusted R^2	0.11	0.10	0.12	0.13

Notes: Standard errors clustered at the mother level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We use the same covariates as we do in the baseline specification (4) in Table 2.2: ability, years of schooling, net worth, and number of children for mothers and children, as well as mother's age at birth. Periods of non-cohabitation are specific for each child-mother pair. Only pairs with both periods of cohabitation and non-cohabitation are included. As this affects the composition of mother-child pairs included in the regression, the baseline results change slightly compared to Table 2.2.

2.D Additional Figures

Figure 2.5: Visual example of a mother-child pair

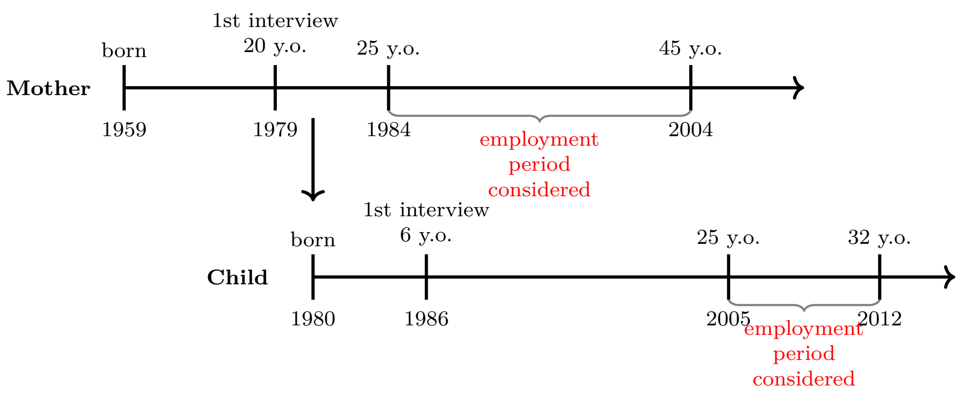


Figure 2.6: Number of interviews of mothers (left) and children (right)

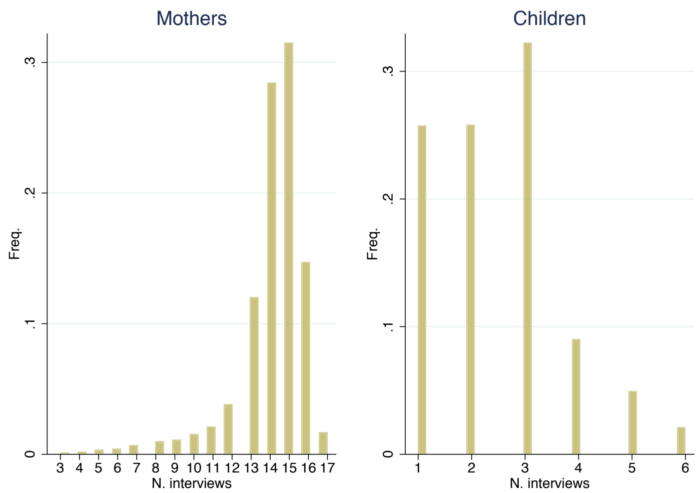


Figure 2.7: Age of mothers at birth of child

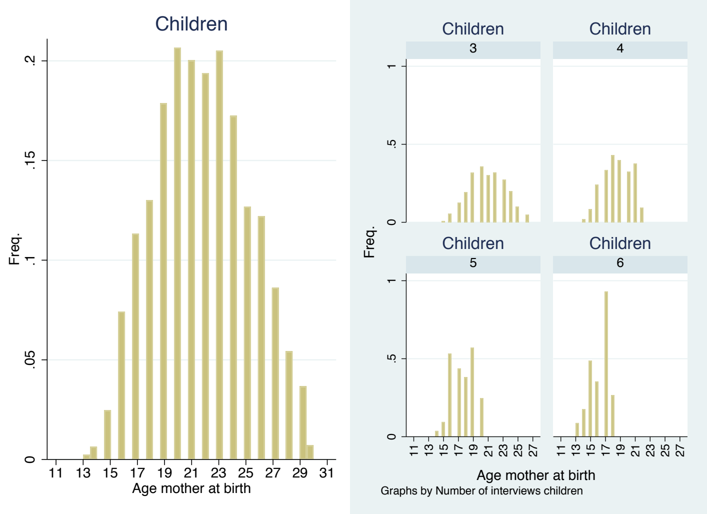


Figure 2.8: Employment-age profiles of mothers (left) and children (right)

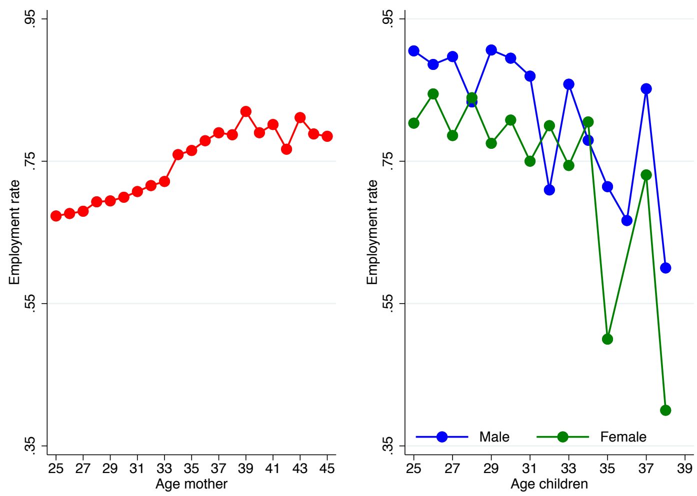
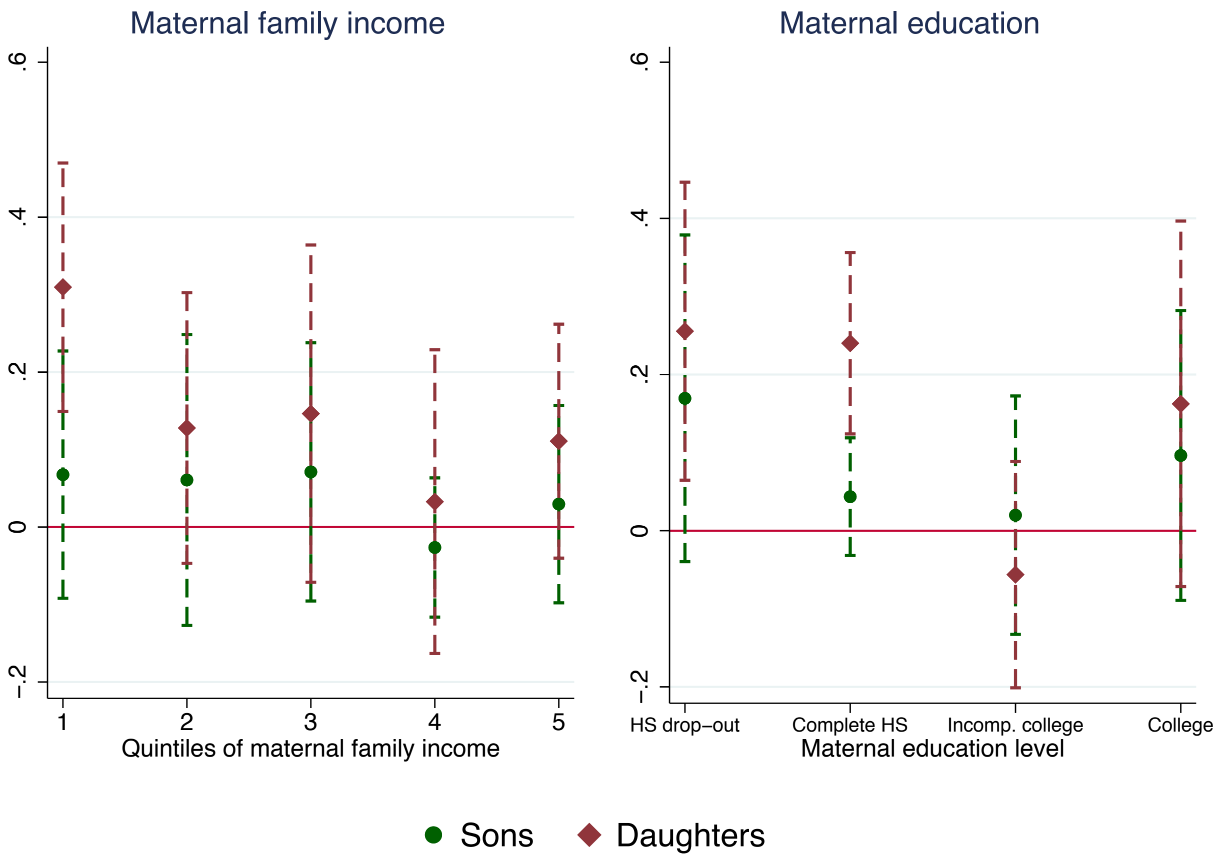
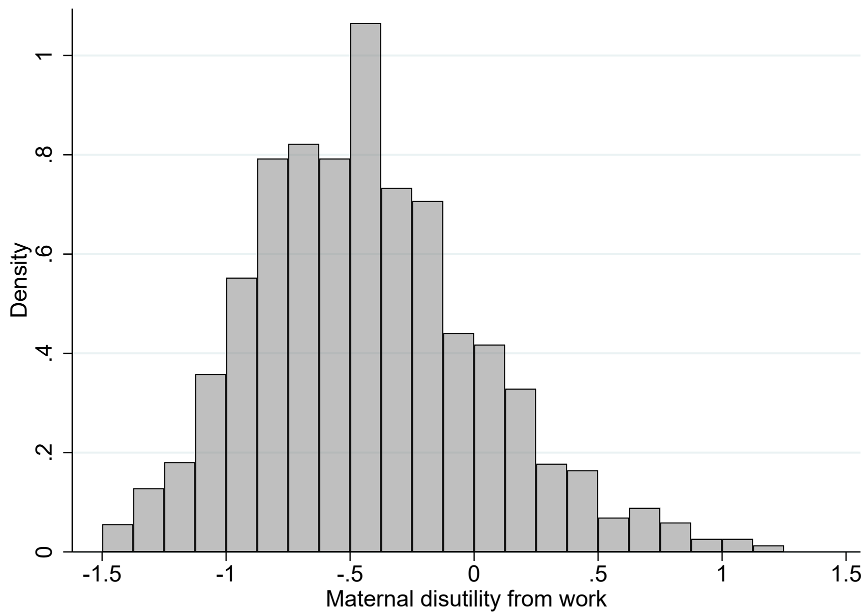


Figure 2.9: Intergenerational correlation of employment status by mother's income (left) and education (right) for sons and daughters



Note: The dependent variable is the permanent component of the natural logarithm of employment status of the children. Standard errors are clustered by mother ID. Mother's position in the income distribution is attributed according to the quintile observed during the maximum number of waves. The education level of mothers is the maximum attained.

Figure 2.10: Distribution of maternal disutility from work



Notes: Disutility from work computed from questions on women's roles: (i) Women's place is in the home, not in the office or shop, and (ii) Women are much happier if they stay at home and take care of the children. These questions are included only in survey years 1979, 1982, 1987, and 2004. We assign the values (a) strongly agree 1.5, (b) agree 0.5, (c) disagree -0.5, and (d) strongly disagree -1.5. We average across questions and across years.

Chapter 3

Less money for divorced mothers? The child-age dependent reform of alimony in Germany

3.1 Introduction

About 40% of all marriages eventually end with a divorce in Germany.[1] Only in 2008, 150,000 children under the age of eighteen experienced their parents getting divorced.[2] In the same year, 16% of all children below eighteen lived with a single parent.[3] Those single parents also face considerably higher poverty risk compared to married or cohabiting parents: in 2011, the poverty rate of households with a single parent was 37.1% compared to 10.5% for married or cohabiting parents in Germany.[4] Combining those facts provides ample reason to investigate the immediate and dynamic effects of the divorce law and, in particular, of regulations that govern (post-marital) maintenance payments on household income, female labour supply, and child development. Dynamic interactions will become even more relevant if we consider effects on behaviour such as human capital accumulation or fertility decisions.

Maintenance payments in Germany are divided into child support and alimony. The eligibility of alimony is based on two concepts: caretaker alimony and post-marital alimony.[5] On January 1, 2008, a *divorce law reform* was enacted in Germany which generally targeted alimony paid

[1] In 2008, 4 out of 10 married couples got divorced within their first 25 years of marriage (see Statistisches Bundesamt 2018b, page 35, or this link for an english website).

[2] See Statistisches Bundesamt 2018b, page 28, or this link for an english website.

[3] 15% of all children lived with their mother and 1% with their father (see Bundesministerium für Familie Senioren Frauen und Jugend 2010, page 22).

[4] See Statistisches Bundesamt 2018a, page 40.

[5] The institutional setup is described in more detail in Section 3.2.

after divorce. As the reform left child support payments unchanged, this paper focuses solely on alimony payments. The reform affected already divorced couples as well as those who got or will get divorced after the implementation of the reform. The purpose of the reform was twofold: first, to strengthen the well-being of children and second, to encourage individual self-responsibility of both spouses after divorce by reducing financial dependencies between them. It will appear questionable if both goals can be achieved at the same time. Policy makers intentionally decided to reduce the entitlements to alimony and thereby, to increase the labour supply of the recipients of alimony payments.[6]

Generally, alimony payments work as a continuation of risk-sharing and as a compensation for foregone human capital accumulation during marriage.[7] But the amount and duration of alimony after divorce also affects labour supply and human capital accumulation of both spouses during and after marriage – and possibly even before marriage – as well as the allocation of housework tasks.

This paper takes a first step by evaluating the effect of the divorce law reform on the probability to pay alimony as a divorced father. We focus on the extensive margin of alimony because the reform was intended to reduce the eligibility to receive alimony by tightening the criteria for post-marital alimony but especially by restricting caretaker alimony depending on the age of the youngest common child. The differential change in the eligibility of divorced mothers to receive caretaker alimony depending on the age of the youngest common child is the key variation used in this paper. We verify the existence of this variation to use it in later studies focusing on different outcomes. Therefore, this paper provides the key first step in the identification strategy of future research projects. As the age of the child is exogenous, the results of this paper suggest that we can use – in future research – child age as an instrumental variable for the treatment intensity of the reform to study the effect of a reduction in alimony payments on other outcomes such as (maternal) labour supply, household income, child development and well-being, and divorce probabilities.

We use a large, administrative panel dataset from Germany – the Income Taxpayer Panel (TPP) which covers a period from 2001 to 2014 – to investigate the impact of the divorce law reform on alimony along its extensive margin. This dataset is suitable to achieve the goal of the paper for three reasons. First, the dataset contains information on alimony payments as they are tax deductible. Second, the TPP covers 5% of all income tax payers between 2001 and 2014, who are observed for at least five years. Therefore, the sample size is large enough to condition the analysis on the age of the youngest common child which is the identifying variation that we

[6]The instrument to increase the well-being of children was to prioritise their entitlement to receive support payments over those of former spouses and other claimants.

[7]Chiappori and Mazzocco 2017 provide an excellent summary of different household models and also emphasise the inter-temporal role of divorce laws.

use in this paper. Third, the panel structure of the dataset allows to follow fathers over time and to observe when they got divorced. It also enables us to use a specification with individual fixed effects. There is no other dataset that fulfils these criteria which makes the TPP the only source to study the impact of the reform.

To the best of our knowledge, this paper is the first paper to show empirically two effects of the reform: first, the general decline in probability to receive alimony for divorced mothers with minor children and, second, those with a youngest common child between four and eight were significantly more affected compared to those with older children. Regarding the first effect, the paper provides descriptive evidence for a substantial decrease after 2008 in the share of divorced men who pay alimony to their former wife and have a common minor child: between 2002 and 2007, the share moved between 24% and 22%, but decreased from 22% in 2008 to just about 15% in 2014 (see Figure 3.1). These numbers suggest that the reform had a direct negative impact on the disposable income of divorced, single mothers with children below the age of eighteen. Regarding the second effect, we show that the reform affected divorced couples differently depending on the age of their youngest common child. Using a difference-in-differences setup, we find in our main specification that the reform decreased the probability to pay alimony for divorced fathers whose youngest common child is between four and eight years old compared to those whose child is between sixteen and seventeen.[8] The estimated decrease in the probability to pay alimony is between 1.6 and 5.9 percentage points depending on the specification of the empirical model. The effect appears sizeable given that the (previously reported) share of divorced fathers with a common child below eighteen who pay alimony in the years prior to the reform is between 22% and 24%. Hence, everything else equal, the reform decreased the disposable income of divorced, single mothers with younger children to a greater extent. Both results are in stark contrast to a declared goal of the reform by policy makers: the improvement of children's well-being.

The results are also robust to different adjustments: first, as mentioned earlier various econometric specifications with and without control variables confirm the decrease in the probability to pay alimony. Second, adjusting the control group to include divorced fathers who have a youngest common child between fourteen and seventeen yields very similar results. Third, if we only condition on those couples who got divorced before the reform, the size of the estimates gets smaller but the long-run effects persist. Finally, we show that the treatment intensity of the reform varies continuously with the age of the youngest child.

Related Literature. Our proposed approach is similar to Low et al. 2018 who study the effects of an introduction of time limits on the welfare receipt of households. They use variation across states and demographic groups of a welfare reform in 1996 in the United States. In

[8]We verify common pre-trends in the outcome of interest of the control and the treatment group.

their quasi-experimental setup, they use that mothers with younger children experienced more severe cuts in their welfare support compared to those with older children. Our approach focuses exclusively on the variation in demographics, specifically the age of the youngest child. Differently from their paper, we use a reform that changes regulations on alimony, i.e. on private transfers between divorced spouses.

Two papers that study potential effects of the German divorce law reform are Fahn, Rees, and Wuppermann 2016 and Bredtmann and Vonnahme 2019. In comparison to our work, both papers compare cohabiting and married couples conditional on the age of the youngest child focusing on different outcomes such as labour supply, divorce probabilities or fertility decisions. Their identification relies on the fact that the reform supposedly left cohabiting couples unaffected, but changed the regulations for married couples, especially if their youngest child is between four and eight. But different from this study, none of the two papers provides direct evidence that the reform had a differential effect on the alimony payments of their control and treatment group. In addition, to the best of our knowledge, no study has shown so far that the intensity of the treatment of divorced couples through the reform varies with the age of the youngest common child.

Bredtmann and Vonnahme 2019 undertake a difference-in-differences estimation using couples that cohabited within three years prior to the reform as a control and those who got married in the same period as a treatment group. They find that married couples have a higher probability to separate due to the reform than cohabiting couples but no significant difference in their labour market outcomes and leisure using the German Socio-Economic Panel (GSOEP). However, their analysis is limited because it estimates a lower bound of the true treatment effect on married couples. The reason is that the current behaviour of cohabiting couples could be equally affected by the reform since they might get married in the future. In a robustness exercise, they redo their estimations focusing their sample on married and cohabiting couples with the youngest child between four and eight. But, in general, the sample size in the GSOEP is small leading to large standard errors. We provide an additional empirical approach by showing that the age of the youngest child of married couples reflects variation in the treatment intensity.

Fahn, Rees, and Wuppermann 2016 set up a theoretical model in which couples make fertility and subsequent separation choices to analyse the effect of alimony payments and separation costs on fertility. They use the variation between married and cohabiting couples with children between four and eight to confirm their model prediction that the reform has decreased in-wedlock fertility compared to out-of-wedlock fertility. However, – as pointed out before – this approach is very different from ours as the authors compare married versus cohabiting couples and do not consider the age of the youngest child as a treatment intensity.

The remainder of the paper is organized as follows. We continue in Section 3.2 with a detailed description of the institutional setup and the reform in 2008 in Germany. In Section 3.3, we describe the Taxpayer Panel and the estimation sample. Afterwards, we focus on descriptive evidence of the reform in Section 3.4. We continue in Section 3.5 with the description of the empirical strategy and the model setups. Section 3.6 provides the results of the baseline setup and the robustness analyses. Section 3.7 concludes.

3.2 Regulation of alimony and the reform of 2008

This section summarises the legal regulation of maintenance payments in Germany and details the changes in the eligibility to receive alimony payments caused by the reform in 2008 on which we build our identification strategy.

Regulation of maintenance payments. The German law specifies that a divorced mother might be eligible to receive alimony payments on the basis of two different legal specifications, post-marital alimony and caretaker alimony. First, post-marital alimony occurs between formerly married partners due to specific reasons such as the length of marriage, health, unemployment, etc. Second, caretaker alimony payments can be claimed if one partner is the caretaker of a common child.[9] In addition, child support payments exist to ensure the well-being of the child and are independent of the former family status of the parents. As the reform did not affect child support payments and we do not observe them in the income tax data, we do not consider them throughout this paper.

Divorced couples can agree upon the amount and duration of alimony payments in a private contract but most of the time they are determined by a family court jointly with the divorce decision. Courts follow common guidelines to set the amount of alimony at 3/7 of the difference in the total net incomes of both former spouses which was unaffected by the reform.[10] The duration depends on the described eligibility.

Reform of 2008. In this paragraph we describe the reform and emphasise the key reform change which we investigate in this paper. The reform took place on January 1, 2008 and had two major goals: first, to decrease the dependence of former partners onto each other by promoting employment of former spouses who had been secondary earners during marriage and, second, to increase the well-being of children.

[9]The law regulating caretaker alimony also applies to unmarried parents, although the rules have been different for married and unmarried parents before the reform.

[10]The total net income does not include regular payments such as interest for mortgages.

What had been the major changes in the regulation of alimony to achieve these goals?[11] First, the law specifies that the age of the youngest common child determines the work obligation of the divorced caretaker as well as the eligibility to receive caretaker alimony payments. Both became considerably stricter through the reform which creates the key variation in this study. Prior to the reform, divorced recipients of caretaker alimony payments were not expected to take up employment until the eighth birthday of the youngest child. Between age eight and eleven the work obligation of the caretaker was determined by courts based on individual characteristics such as the number of children, well-being of the children (i.e. disabilities, etc.) and availability of daycare. Once the youngest child turned eleven, the caretaker was expected to pick up a part-time job. With older children, caretakers were eligible for alimony payments until the youngest child turned fifteen because, from then on, they were expected to work full-time.[12] After the reform, a divorced ex-spouse who takes care of common children has the right to receive caretaker alimony until the youngest child's third birthday. After that, the caretaker is expected to pick up a job and sustain her living without receiving alimony payments.[13]

The described change creates potential variation in the eligibility for caretaker alimony conditional on child age. In this paper, we verify this variation by showing that the probability to receive alimony indeed changed conditional on the age of the youngest common child.

A reason for the adjustment had been that prior to the reform the regulations for caretaker alimony differed if the parents had been married or not. Even before the reform, caretaker alimony for previously unmarried parents was only granted until the third birthday of the youngest child. The reform intended to harmonize the law for all separated parents, which Fahn, Rees, and Wuppermann 2016 and Bredtmann and Vonnahme 2019 use in their respective identification strategies.

The second major change of the reform was the introduction of the principle of self-responsibility: each spouse became responsible to sustain his or her living after a divorce. Before the reform, a spouse would be eligible to receive post-marital alimony payments from her former partner if she was unable to sustain her living and ensure a similar way of life by herself after divorce. The reform restricted the entitlement to post-marital alimony substantially.

[11] Brudermüller and Diederichsen 2008 describe the legal changes and implications in great detail (in German). The change in the caretaker alimony is described on pp. 3 - 6.

[12] These thresholds were not explicitly defined in the law, but had been established by decisions of higher and supreme courts led by their interpretation of the law. The aforementioned individual characteristics might also affect the age thresholds marginally as judges decide about caretaker alimony case by case. Brudermüller and Diederichsen 2008 provide an excellent description of the case law.

[13] In some cases, the court's decision about the work obligation might still depend on the above mentioned individual characteristics. But following the large expansion of public childcare in the early 2000s, it was generally available for children above the age of three. In 2008 more than 90% of children aged three to five went to kindergarten at least part-time (see Section 1.2.3 for own calculations based on the German Socio-Economic Panel).

Third and finally, the ranking of the claimants changed: after the reform, under-aged children are ranked first. Former married or unmarried partners will be ranked second if they are eligible for alimony payments due to child-rearing or if the marriage was of a long duration; if they do not fulfil the former conditions, ex-spouses will be ranked third. All claims are met as long as the liable spouse's income is above the deductible income and, thereby, the ranking only matters if the liable spouse is constrained.

The intention of the final adjustment was to fulfil the second goal of the reform – the increase in the well-being of children – while the first and second adjustments targeted the promotion of employment of former spouses. However, limiting the eligibility for caretaker alimony might have unintended side effects on the well-being of children.

The reform applied to all divorced couples, i.e. to those that got divorced before and after 2008. The reform did not have an automatic effect on existing alimony payments since they are determined by courts and retrials can only be requested if considerable changes (approx. $\geq$10%) are expected.

3.3 Data

The German Federal Statistical Office provides restricted access to the *Income Taxpayer Panel (TPP)*, an administrative panel dataset, which consists of a 5% representative sample of tax payers who pay income taxes in Germany for at least five years between 2001 and 2014. The dataset contains 1,579,443 tax payers. As the dataset is a stratified representative sample, we use population weights in all regressions.[14] Until 2011, the sample is based on tax payers who file a tax declaration for the respective year. From 2012 to 2014, the sample additionally includes tax payers that do not file a tax declaration, but pay income taxes.

Germany uses a system of joint taxation during marriage. We can directly infer from the data if a couple is married and additionally if they file jointly or separately. [15] The dataset contains for all married couples most of the variables twice (for example, gross labour income for person A and gross labour income for person B). Once a married couple gets divorced, from then onwards only one of the spouses is kept in the dataset as this person keeps the tax identification number. In about 3/4 of the cases, the dataset follows former husbands and only in the remaining quarter of divorces, it is possible to identify divorced women.

[14]The strata are based on the state of residence (Bundesland), joint or separate filing, the predominant source of income, the median of total revenues over the observed years and the relative variation of total revenues between the years. Additional information on the Taxpayer Panel is available here (only in German).

[15]Unmarried, cohabiting couples are not allowed to file jointly and need to file separately. Unfortunately, we cannot link two cohabiting partners in the income tax data.

The dataset has some major advantages. First of all, the size of the TPP and its panel structure, which allows us to follow (married and divorced) individuals over time, are crucial. Furthermore, it also contains some demographics such as gender, state of residence, religion, birth year, and children (including their birth year). Especially, the details on children are essential for this study. Finally, the data contain a variable that specifies the *deduction on taxable income due to paid (caretaker or post-marital) alimony*.

Generally, received alimony payments are not classified as taxable income in Germany. Nonetheless, they can be deducted from the taxable income of the payer of alimony if the former partner and recipient of the alimony payments agrees to declare the same amount as taxable income in her tax declaration. This procedure is called 'Realsplitting' and implemented as a continuation after divorce of the joint taxation during marriage (called 'Ehegattensplitting').[16] If the 'Realsplitting' is requested by the payer of alimony, the recipient will be obliged by law to include the alimony payments in her tax declaration. But the law also specifies that in this case the recipient needs to get compensated by her former spouse for the additionally paid taxes. Because of this legal setup, we would not expect a selection effect into using alimony payments as tax deductible. The 'Realsplitting' will be beneficial if the marginal tax rates of both ex-partners are different from each other. But this should not imply any selection because divorced couples with similar marginal tax rates usually have similar net incomes and if net incomes of divorcees are similar, alimony payments will presumably be close to zero as they are calculated as 3/7 of the net income differences.

As mentioned earlier, this paper focuses on the alimony payments of divorced men conditional on the age of the youngest child that they have in common with their former spouse. Besides conditioning on the age of the child, we cannot distinguish in the data if the alimony payments paid by the divorced father are due to post-marital and caretaker alimony. But the distinction is irrelevant for the results of this paper because we will show that the reform had a differential effect on the probability to pay alimony depending on the age of the youngest child.

All in all, the size of the dataset enables us to condition our analysis on the age of the youngest common child of divorced couples. In combination with the ability to follow married and divorced individuals over time and to observe paid alimony in the data, it makes this dataset unique for the analysis undertaken in this paper.

Sample definition. As noted above, we identify divorced individuals as those who had been observed to be married at some point in the data and at some point started to file as an individual indicating a separation. We exclude those from the set of divorced individuals

[16]The tax deduction is limited to 13,805 Euro per year which implies censoring of the data. But the censoring is not relevant in this version of the paper as we focus on the extensive margin. It will become relevant though once the focus is extended to the intensive margin of alimony payments.

who directly filed as married in the year after the divorce occurs, i.e. a tax payer that files individually only for one year and as married before and after that year. Furthermore, we exclude the states of Lower Saxony and North Rhine-Westphalia as the variable on alimony is not consistently recorded over the years in those two states. We focus on divorced men as payers of alimony payments because in the data the overwhelming majority of alimony payers are men.[17]

3.4 Motivational evidence

3.4.1 Descriptive evidence

Before we causally identify the effects of the reform using a difference-in-differences setup, this section provides descriptive evidence on the evolution of the extensive margin of alimony payments over time.

Figure 3.1 shows the evolution of the share of divorced men who pay alimony to their former wife over time across Germany excluding the states Lower Saxony and North Rhine-Westphalia. The data contains only those men whose youngest child, that they have in common with their former spouse, is below the age of 18.

We observe a slight decrease in the share in the years prior to the reform from about 24% in 2002 to 22% in 2007 and 2008.[18] The decrease accelerates substantially after the reform in 2008 resulting in a share of 15% in 2014. The average annual growth rate between the years 2002 to 2007 is -1.5% compared to -5.9% between 2009 and 2014. To visualize and emphasize this result, Figure 3.2 depicts two-year growth rates.[19] We observe a strong decrease in the two-year growth rates after 2008. The reform substantially decreased the probability to pay alimony to the former spouse and mother of the common child on an aggregate level. In the following sections, we will derive differential effects of the reform in a causal fashion.

3.4.2 Specific court case

To give some further anecdotal evidence that the reform had a non-negligible effect on divorced caretakers before turning to the causal econometric analysis and results, let us consider the

[17]In the GSOEP, between 2005 and 2010 about 95% of the recipients of alimony payments were women (Bredtmann and Vonnahme 2019).

[18]Note that this share only refers to alimony payments and does not include child support. In combination with restrictive eligibility criteria and possibly small income differences between ex-spouses, it might explain the share of 22 to 24%. Nonetheless, it could also reflect low compliance of those responsible to pay alimony.

[19]The growth rate in year t is derived as $\frac{s_t - s_{t-2}}{s_{t-2}}$ where s_t denotes the share of divorced men paying alimony in year t. We use two-year annual growth rates to obtain smoother results.

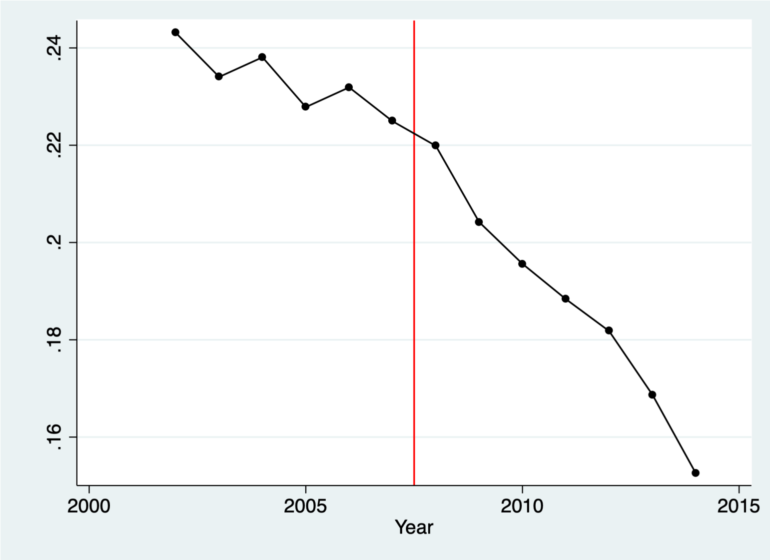


Figure 3.1: Share of divorced men paying alimony to former spouse over time

Notes: Average of the binary variable that specifies if a divorced man pays alimony to his former spouse according to his tax declaration. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child under the age of 18**. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

following court case which took place after the reform in 2008 and 2009:[20] The mother worked for eighteen hours per week as a secondary school teacher and was the sole caretaker of a six-year old son. The son used to go to school in the morning followed by daycare where he usually stayed until 4pm. As the former husband and father of the child did not agree to pay caretaker alimony, the mother went to court to claim it. A lower court granted the mother caretaker alimony of 800 Euro per month, but the ex-husband appealed against the decision. The case went to the Federal Supreme Court for matters of private law (Bundesgerichtshof) which remands the case back to the lower court emphasising that after the reform the parental care taking of children older than three years does not have priority any more. This led to an updated verdict. Since the child was six years old, the mother did not receive caretaker alimony payments any longer. By contrast, before the reform she would have received 800 Euro per month until the child turns fifteen years old.[21] Without any discounting, the net income difference for the mother amounts to a sizeable loss of 86.400 Euro over the course of nine years.

[20]The court case went to the German supreme court for all matters of criminal and private law (Bundesgerichtshof) (BGH 18/3/2009, XII ZR 74/08).

[21]The mother already fulfilled the part-time work requirement which would need to be met once the child turns eleven at the latest.

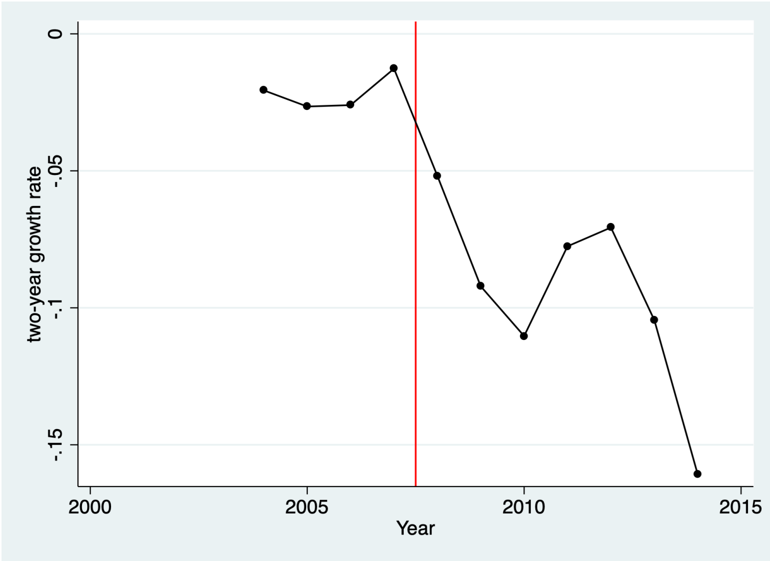


Figure 3.2: Two-year growth rate in share of divorced men paying alimony to former spouse

Notes: Two-year growth rates in the average of the binary variable that specifies if a divorced man pays alimony to his former spouse according to his tax declaration. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child under the age of 18**. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

3.5 Empirical strategy and setup

3.5.1 Empirical strategy

Our empirical strategy intends to investigate whether the reform implied a differential treatment of divorced couples depending on the age of their youngest common child (for details see Section 3.2). If the child was between four and eight, prior to the reform the care-taking mother would not be required to work and could expect to receive caretaker alimony. The reform abolished this right. By contrast, if the child was between sixteen and seventeen, care-taking mothers would not be legally entitled to caretaker alimony before and after the reform.[22] All other reform changes are expected to affect couples with the youngest common child between four and eight or sixteen and seventeen in a similar way. To determine the effect of the reform, we apply a difference-in-differences approach. In the baseline approach, the control group consists of divorced fathers with a youngest common child aged sixteen to seventeen, whereas the treatment group contains divorced fathers with a child between four and eight. The identifying assumption is that the probability to receive alimony payments would have had the same trend for both the treatment and the control group in the absence of the treatment. One threat to this assumption is that the reform might have changed the composition of divorcees, which

[22] Nevertheless, they could still receive alimony payments as former spouses can agree upon these payments individually or as mothers could be eligible due to post-marital alimony (see Section 3.2).

may imply differences in the propensity to pay alimony between treatment and control. A robustness exercise deals specifically with this threat by restricting the sample to those fathers who got divorced before the reform took place in 2008.

3.5.2 Empirical setup

Our empirical analysis relies on three different model setups: first, a linear probability model (LPM) with group fixed effects; second, a linear probability model with individual fixed effects (i.e. a two-way linear fixed effects model); third, a probit model with group fixed effects. The results of all three models are discussed in Section 3.6.1 for the baseline setup.

In general, we include in our estimation sample those divorced fathers on a year-by-year basis who are either in the control or the treatment group in the respective year. Therefore, in the baseline setup the respective data is unbalanced because fathers will drop out of the sample if their children grow out of the age brackets (4 - 8 or 16 - 17).[23] This also implies that some fathers might switch from the treatment to the control groups when their children grow older.[24] We discuss the implications for each setup below.

Linear probability model with group fixed effects. In the simplest setup, we use a difference-in-differences (DiD) specification to estimate the *post-reform treatment effect* α:

$$y_{it} = \alpha \cdot \mathbb{1}\big\{t \geq t_{reform}\big\}_t \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + \delta \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + f_t + X_{it}\beta + \epsilon_{it} \tag{3.1}$$

where $t = 2002, ..., 2014$ denotes the current year, $c_{it} = 1, 2$ the age bracket of individual i's youngest common child (4 - 8 or 16 - 17) in year t, $\mathbb{1}\big\{t \geq t_{reform}\big\}_t$ is a post-reform dummy, and $\mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it}$ a dummy for belonging to the treatment group c_{treat} in year t, i.e. having a youngest common child between four and eight in year t. y_{it} is the binary variable which indicates if a divorced father i pays alimony to his former partner in year t. ϵ_{it} is the error term. As y_{it} is a binary outcome variable, the error terms will be heteroscedastic. Hence, we use robust standard errors within the OLS regression. α is the coefficient of interest on the interaction between post-reform and treatment. We include time-invariant group fixed effects, denoted by $\delta \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it}$, and year fixed effects, f_t, as common in difference-in-differences estimations. X_{it} is a set of additional controls: first, we add state fixed effects. Second, we calculate the difference in total net income between former spouses in the year prior to divorce. We derive quartiles over the entire time period and assign to each individual a time-invariant quartile which we use as a (categorical) control variable in the regression, i.e. we add dummy variables for belonging to the first, second, or third quartile. Similarly, we derive each divorced man's

[23]This will not be the case in Section 3.6.2.3 unless the youngest child gets older than seventeen.

[24]For example, consider a divorced father with the youngest common child being eight in 2004 and 16 in 2012. The father will be in the treatment group in 2004 and in the control group in 2012.

current total net income and calculate deciles over the distribution in the current year which are included as time-variant dummies in the regression. Finally, we add indicator variables for an additional common child under eighteen and for two or more additional common children under eighteen. We estimate the model specified in (3.1) with and without the set of controls X_{it}. The identification of α relies on all observed divorced fathers irrespective if they are observed in the data before and/or after the reform.

To study the dynamics of the propensity to pay alimony payments, in an extension to setup (3.1), we interact the treatment indicator $\mathbb{1}\{c_{it} = c_{treat}\}_{it}$ with year dummies $\mathbb{1}\{t = \tau\}_t$ from 2002 to 2014 (excluding 2007). This approach yields *year-specific treatment effects* α_t which we will use to verify the common pre-trend assumption across control and treatment groups:

$$y_{it} = \sum_{\tau=2002}^{2014} \alpha_\tau \cdot \mathbb{1}\{t = \tau\}_t \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + \delta \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + f_t + X_{it}\beta + \epsilon_{it} \qquad (3.2)$$

We estimate model (3.2) also with and without the set of additional controls X_{it} and still use robust standard errors to address heteroscedasticity. We set α_t to zero in 2007, thereby identifying the year-specific treatment effects relative to the year 2007. Similar to the post-reform specification (model (3.1)), the identification of α_t relies on all divorced fathers who are observed in at least two years including year t. Obviously, some fathers need to be observed in the baseline year (2007) as well.

Linear probability model with individual and group fixed effects. The possibly most demanding setup is to add individual fixed effects to the described linear probability models (3.1) and (3.2). This yields a two-way linear fixed effects regression model. Using the same notation as above, we still interpret α as the *post-reform treatment effect*:

$$y_{it} = \alpha \cdot \mathbb{1}\{t \geq t_{reform}\}_t \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + \delta \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + f_i + f_t + X_{it}\beta + \epsilon_{it} \qquad (3.3)$$

where f_i denote the time-invariant individual fixed effects which control for time-invariant unobserved differences between individuals. They capture all omitted variables that are time-invariant.

α is still the coefficient of interest on the interaction between post-reform and treatment. But in comparison to the LPM (model (3.1)), the identifying variation changes. To get intuition why this is the case, consider the implementation of the individual fixed effects as demeaning all variables by their time averages. As our dataset is unbalanced, divorced fathers who enter the dataset only after the reform have no variation in their interaction term of post-reform and treatment, i.e. $\mathbb{1}\{t = \tau\}_t \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it}$. Demeaning this interaction term implies that the demeaned independent variable which captures the post-reform treatment effect α is always zero for all divorced fathers entering the data after the reform. For example, divorced fathers with children aged four to eight who enter after the reform, do not add to the identification of α while they did in the LPM (model (3.1)). Therefore, the identifying variation in this specification comes exclusively from those divorced fathers that entered the dataset prior to the reform and are still observed after the reform. Hence, this setup is slightly restrictive especially in combination with fathers dropping out of the sample once their children grow out of the specified age range (4 - 8 or 16 - 17).

To appropriately study the dynamics of the propensity to pay alimony payments, we extend the analysis in a similar fashion as before which yields *year-specific treatment effects* α_t:

$$y_{it} = \sum_{\tau=2002}^{2014} \alpha_\tau \cdot \mathbb{1}\{t = \tau\}_t \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + \delta \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + f_i + f_t + X_{it}\beta + \epsilon_{it} \quad (3.4)$$

We also estimate both models (3.3) and (3.4) with and without the set of controls X_{it}. We cluster standard errors at the individual level.[25]

As in model (3.2), we identify α_t relative to 2007. The identification of the post-reform and the year-specific treatment effects differ significantly from each other once we include individual fixed effects. As explained above, demeaning all variables by their time averages implies that the identification of α in model (3.3) relies only on divorced fathers who appear in the data pre- and post-reform. In contrast, in model (3.4) divorced fathers help to identify the year-specific treatment effects in all years in which they are observed as the interaction of the year t dummy and the treatment dummy varies over time. Thereby, demeaning each interaction variable by its time average does not have the same implication as it had in model (3.3). Hence, the identification of α_t relies on divorced fathers who have been observed in year t and at least one additional year. The difference in the identifying variation between models (3.3) and (3.4) has implications for the results (see Section 3.6.1).

Probit model with group fixed effects. Finally, we use the probit model as a non-linear, binary response model to estimate the probability of paying alimony for individual i at time t with a youngest common child in age bracket c_{it}

$$\begin{aligned} \text{Prob}\big(y_{it} = 1 | \mathbf{Z}_{it}\big) &= \Phi\big(\mathbf{Z}_{it} \cdot \xi\big) \\ &= \Phi\Big(\alpha \cdot \mathbb{1}\{t \geq t_{reform}\}_t \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + \delta \cdot \mathbb{1}\{c_{it} = c_{treat}\}_{it} + f_t + X_{it}\beta\Big) \end{aligned} \quad (3.5)$$

[25]In the future, we intend to cluster the standard errors on a lower level such as birth cohort of the child or year of divorce.

where Φ is the cumulative distribution function of a Gaussian standard normal distribution and we define for notational purposes

$$\mathbf{Z}_{it} \cdot \xi = \alpha \cdot \mathbb{1}\big\{t \geq t_{reform}\big\}_t \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + \delta \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + f_t + X_{it}\beta.$$

The parameter α does not have a linear interpretation as in the linear probability model, but its significance would still imply the existence of a post-reform treatment effect.

We can also study the dynamics implied by the probit model. Therefore, we estimate the following model which features a comparable extension as the linear models above:

$$\begin{aligned}\text{Prob}\big(y_{it} = 1 | \tilde{\mathbf{Z}}_{it}\big) &= \Phi\big(\tilde{\mathbf{Z}}_{it} \cdot \xi\big) \\ &= \Phi\Big(\sum_{\tau=2002}^{2014} \alpha_\tau \cdot \mathbb{1}\big\{t = \tau\big\}_t \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + \delta \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + f_t + X_{it}\beta\Big)\end{aligned} \tag{3.6}$$

assuming for notational purposes

$$\tilde{\mathbf{Z}}_{it} \cdot \xi = \sum_{\tau=2002}^{2014} \alpha_\tau \cdot \mathbb{1}\big\{t = \tau\big\}_t \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + \delta \cdot \mathbb{1}\big\{c_{it} = c_{treat}\big\}_{it} + f_t + X_{it}\beta.$$

Again, the coefficients α_t cannot be interpreted directly, but the estimation results are used to plot average adjusted predictions over time.

Both Probit models (3.5) and (3.6) are estimated with and without the set of controls X_{it}. We use non-robust standard errors for inference in the estimation of the probit model as the errors in the probit model are expected to be homoscedastic.[26] The identification in the probit model is equivalent to the LPM with group fixed effects (models (3.1) and (3.2)).

3.6 Results

3.6.1 Baseline results

Table 3.1 shows the estimation results of the post-reform treatment effect for the three different model specifications. Column 1 contains the results for the linear probability model with group fixed effects and no additional controls (besides year fixed effects). The probability to pay alimony as a divorced father with a common youngest child between four and eight decreases due to the reform by 5.9 percentage points compared to a divorced father who has a common youngest child between sixteen and seventeen. The effect remains significant and sizeable (3.7

[26]We do not extend the probit model to include individual fixed effects as these models are known to produce biased estimates.

pp) after controlling for the additional number of minor children, the income difference of the spouses in the year prior to the divorce, and the current income decile of the divorced father (column 2 of Table 3.1).

Table 3.1: Baseline estimation of the probability to pay alimony using models (3.1), (3.3), and (3.5)

Dependent variable: y_{it}

	Linear Probability Model				Probit	
α (coefficient of interest)	-0.0586***	-0.037***	-0.0155	-0.018	-0.2092***	-0.145**
	(0.0100)	(0.0095)	(0.0178)	(0.0178)	(0.0552)	(0.0645)
controls		✓		✓		✓
year FE	✓	✓	✓	✓	✓	✓
group FE	✓	✓	✓	✓	✓	✓
individual FE			✓	✓		
Observations	41418	41418	41418	41418	41418	41418

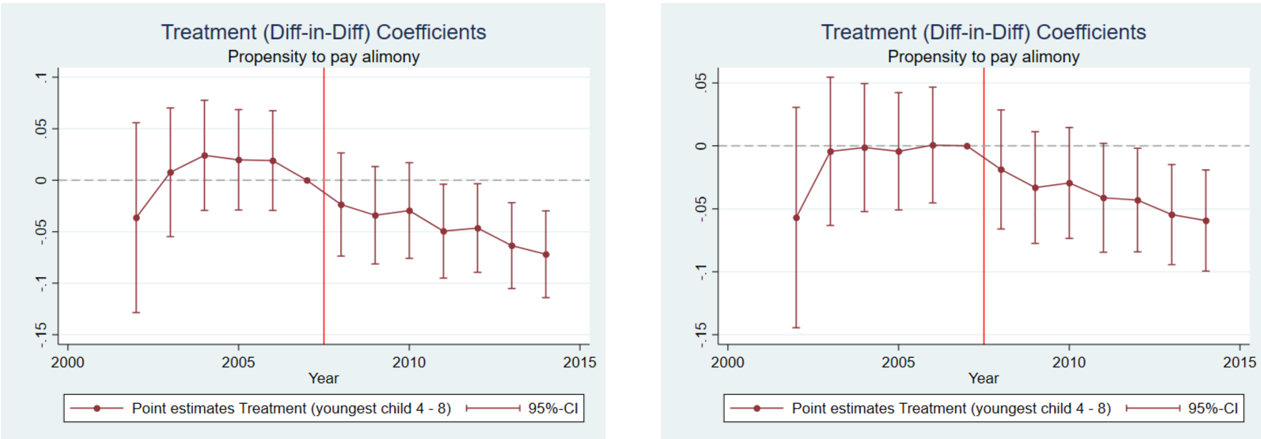
Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Notes: Columns 1 and 2 show results for model (3.1) with and without controls. We use robust standard errors. Columns 3 and 4 show results for model (3.3) with and without controls. Standard errors are clustered at the individual level. Columns 5 and 6 show results for model (3.5) with and without controls. We use regular standard errors. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 16 and 17**. Controls: state fixed effects, the quartile of the difference in total net income between former spouses in the year prior to divorce (last year of marriage) [quartiles are taken over all years], the decile of the total net income in the current year [deciles are taken over the current year], indicator variables for an additional common child under 18 and for two or more additional common children under 18. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

Once we include individual fixed effects (column 3 and 4), the post-reform treatment coefficient shrinks to 1.8 percentage points and its significance disappears. As explained above, demeaning all variables by their time averages, i.e. adding individual fixed effects, implies that the identification of α in model (3.3) relies only on divorced fathers who appear in the data pre- and post-reform. In addition, as also detailed above, we restrict the sample on a year-by-year basis to those who have a child in the age range 4 - 8 or 16 - 17. Therefore, the effect α is identified only via fathers who are observed in a small window around the year of the reform which explains the considerably smaller estimates in columns 3 and 4 compared to columns 1 and 2. As further explained below when we focus on the dynamics, the LPM with and without individual fixed effects yield similar dynamic treatment effects namely smaller short-run but larger long-run effects (compare Figures 3.3(a), 3.3(b), 3.4(a), and 3.4(b)). Therefore, we should rather rely on the estimates in column 1 and 2 instead of 3 and 4.[27]

The estimates resulting from the Probit model confirm the results of the LPM with group fixed effects (column 5 and 6). The post-reform treatment coefficient reflects a significantly stronger decrease in the probability to pay alimony conditional on having a younger child. We cannot compare the size of the regression coefficients to those of the LPM because the Probit model is a non-linear specification. But in Figure 3.5(a) and 3.5(b), we plot the average adjusted predictions over time.

The dynamics of the treatment effect depicted in Figures 3.3(a) - 3.5(b) constitute the *main result of this paper*. As described in Section 3.5.2, we normalise the treatment effect to zero in 2007. Therefore, the graphs show the point estimates of the treatment effect over time relative to year 2007 including the 95% confidence intervals. The confidence intervals will allow us to infer if in a given year the treatment effect is significantly different from the baseline effect in 2007.



(a) Controls: Year and *group* fixed effects

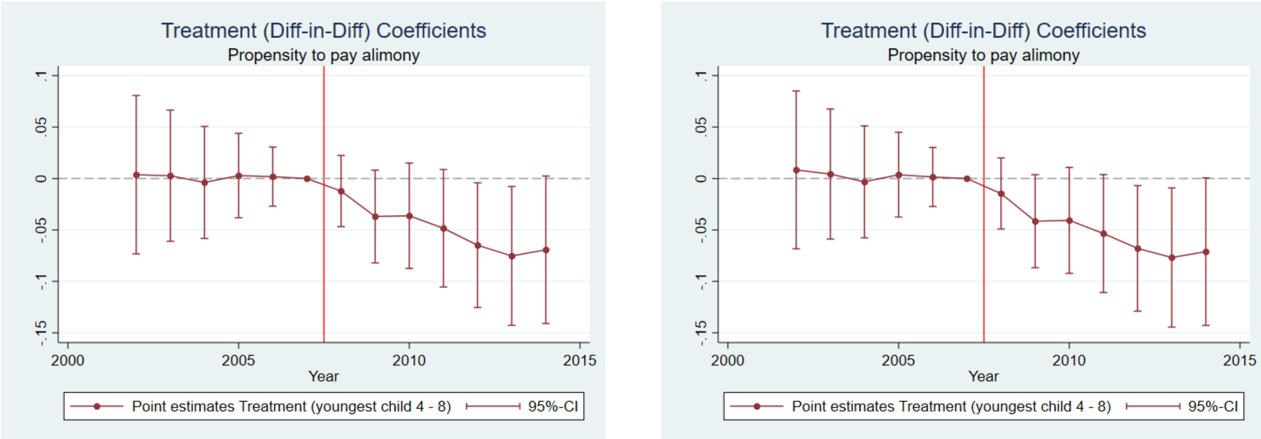
(b) Plus Controls

Figure 3.3: Linear probability model (OLS) with *group* fixed effect

Notes: Both figures plot the interaction effect of treatment indicator and year indicator on the propensity to pay alimony payments of divorced men over time based on a linear probability model with *group* and year fixed effects and robust standard errors. The effect is normalized to 0 in 2007. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 16 and 17**. Controls in Figure 3.3(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

We draw two main conclusions: first, all Figures show common pre-trends of the control and the treatment group which justifies the assumption of parallel trends in the absence of the treatment which is necessary to interpret the effects causally. Second, the negative effect of the

[27] We observe only little difference between including or not including covariates because those are mostly time-invariant and therefore, captured by the individual fixed effects.

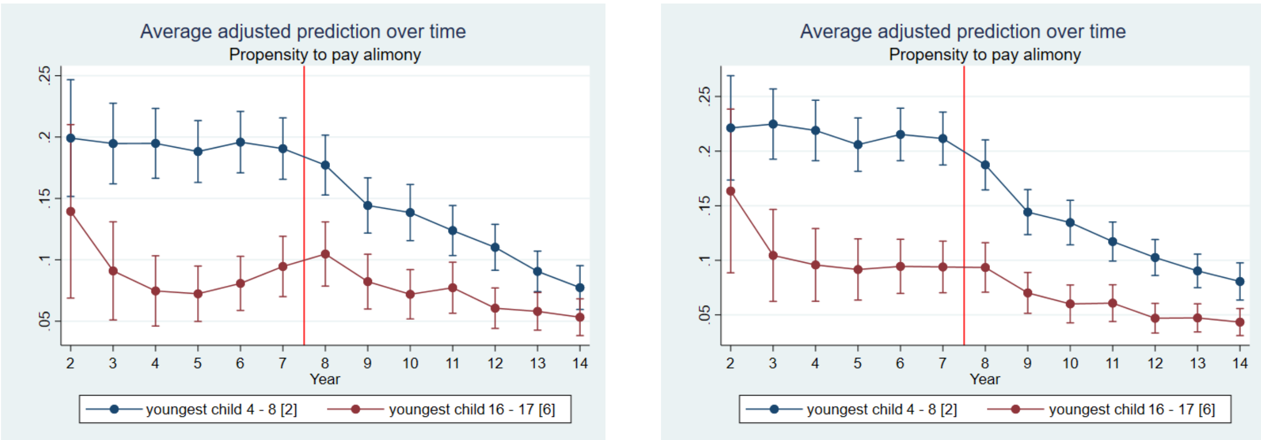


(a) Controls: Year and *individual* fixed effects

(b) Plus Controls

Figure 3.4: Linear probability model with *individual* fixed effect (Two-way fixed effects model)

Notes: Both figures plot the interaction effect of treatment indicator and year indicator on the propensity to pay alimony payments of divorced men over time based on a linear probability model with *individual* and year fixed effects with standard errors clustered at the individual level. The effect is normalized to 0 in 2007. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 16 and 17**. Controls in Figure 3.4(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.



(a) Controls: Year and *group* fixed effects

(b) Plus Controls

Figure 3.5: Probit model with *group* fixed effect

Notes: Both figures plot the average adjusted predictions of the propensity to pay alimony payments of divorced men over time based on a probit model with *group* and year fixed effects with non-robust standard errors. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 16 and 17**. Controls in Figure 3.5(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

reform on the probability to pay alimony magnifies continuously after the reform and becomes significant after 2011/2012 in all specifications.

Figure 3.4(a) (without controls) and 3.4(b) (with controls) show the results of the two-way linear fixed effects regression model including interactions of treatment and year. We observe an approximately linear decline after the implementation of the reform resulting in a treatment effect of 6.5 - 7.7 percentage points in the years 2012 to 2014. The interaction term was precisely zero before the reform confirming common pre-trends. Similar results are obtained using group fixed effects instead of individual fixed effects (Figures 3.3(a) and 3.3(b)). The decrease after the reform is more immediate in the model with group fixed effects, but the long run effects are a bit smaller (4.3 - 7.2 pp), although more precisely estimated. Notably, Figures 3.5(a) and 3.5(b) plot the average adjusted predictions over time based on the Probit model. We observe a stable gap between the predictions prior to the reform, but it starts to close right after. In 2014, the gap between the average adjusted predictions of the treatment and the control group has been the closest over the time horizon. The average adjusted prediction of the propensity to pay alimony for the control group decreases by 5 percentage points from 10% in 2007 to 5% in 2014. But in comparison, the decline for the treatment group is considerably larger and amounts to approximately 14 percentage points between 2007 and 2014 (from 22% to 8%). These results emphasise that the variation in the age of the youngest common child allows us to capture the treatment intensity of the reform.[28]

3.6.2 Robustness

In this section, we extend our analysis and address some concerns that might arise from our estimated baseline setup. First, we use an alternative control group by including divorced fathers with a youngest common child between fourteen and seventeen. Second, we restrict the sample to fathers who got divorced before the reform to deal with sample selection as a threat to identification because the reform might have changed the composition of divorcees. Finally, we evaluate the effects of the reform conditional on children aged between zero and fifteen using two-year and four-year child age brackets. This yields a notion of treatment intensity.

3.6.2.1 Alternative control group

We consider an extension of the control group to include all divorced fathers with a youngest common child between fourteen and seventeen. The models and identification are equivalent to the baseline setup. The results in Table 3.5 in Appendix 3.B.1 are very close to those of

[28]Figure 3.13 verifies that the average child age within the brackets is constant over time in the sample. We condition on the child age brackets 4 - 8, 9 - 13, and 14 - 17 due to data confidentiality. But it seems reasonable to infer from this that the average child age in the group of 16 - 17 year old children does not change over time either.

the baseline setup (Table 3.1). For example, the post-reform treatment effects in the LPM with group fixed effects implies in both setups a decrease in the probability to pay alimony by 3.7 percentage points when we use controls in the regressions. Without controls, the decrease amounts to 5.3 percentage points with the alternative control group and 5.9 percentage points in the baseline setup. Figures 3.9(a) - 3.11(b) in the Appendix confirm the main results derived in Section 3.6.1: first, we find common pre-trends and second, the effect of the reform occurs gradually, not being significant upon impact but later on.

3.6.2.2 Divorced pre-reform

To deal with sample selection as a challenge to identification, we restrict the sample in the baseline setup to those fathers who got divorced before the reform.[29] The post-reform treatment effect α is listed for all model specifications in Table 3.2. The magnitude of the post-reform treatment effect decreases and it is only significant for the LPM with group fixed effects and no controls. Nonetheless, the estimation results are reassuring because of two points: first, Figures 3.6(a) to 3.8(b) confirm common pre-trends for all specifications. Second, we observe strong dynamic effects of the reform which start after 2010 in all model specifications.[30] In the LPM with group fixed effects the treatment coefficient is significant after 2013 (see Figure 3.6(a) and 3.6(b)). We observe in Figures 3.8(a) and 3.8(b) that the Probit model predicts the same propensity to pay alimony in 2013. Before the reform, there was a constant difference of more than 10 percentage points. Finally, the estimates of the LPM with individual fixed effects show the same pattern, but are estimated less precisely (see Figure 3.7(a) and 3.7(b)).

The delayed, but strong dynamic effects might be due to some 'stickiness' of alimony payments. For those couples that got divorced before the reform, an adjustment in alimony payments is not automatic but alimony needs to get re-evaluated and determined by a court. This might be costly and time-intensive such that payers of alimony might wait to observe the interpretation of the new law by the supreme court before actually taking their case to a lower court. By contrast, couples who get a divorce after the reform receive an immediate, joint decision of divorce and alimony and thereby, we observe a more gradual decrease in the overall sample (see Section 3.6.1). This makes it likely that the effect of the reform on already divorced couples lags behind the effect on those that got divorced after the reform.

Finally, focusing on the sample of fathers who got divorced prior to the reform bears some challenges. First, Figure 3.14 in the Appendix shows that the average age of the children in the age bracket 4 - 8 increases over time if the father got divorced before the reform. Therefore, the

[29] The identifying variation is similar to the baseline setup with the additional restriction of only relying on father who got divorced prior to the reform.

[30] We can confirm the results when we use the larger control group (14 - 17) and condition on fathers getting divorced before the reform. Results are available upon demand.

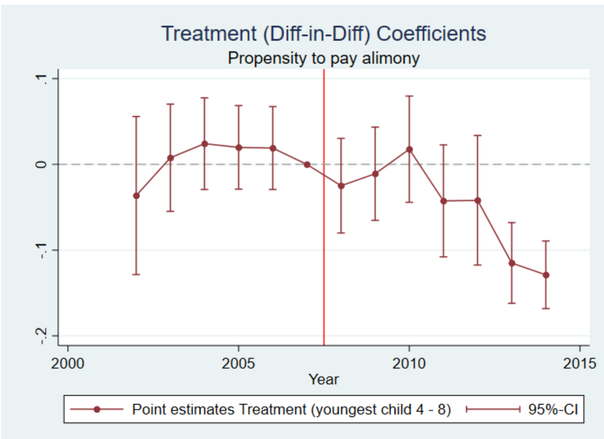
Table 3.2: Robustness – estimation of the probability to pay alimony using models (3.1), (3.3), and (3.5) restricted to the **sample of pre-reform divorced fathers**

Dependent variable: y_{it}

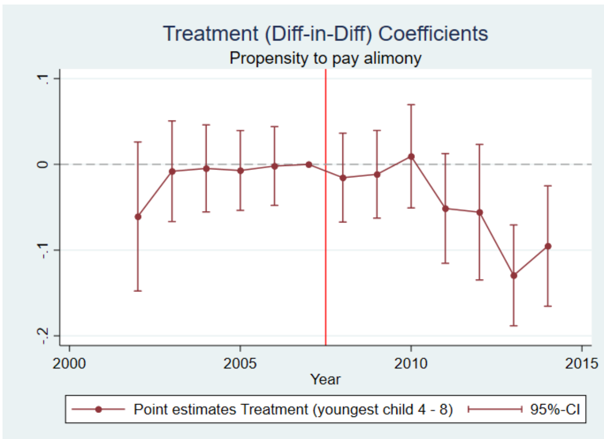
	Linear Probability Model				Probit	
α (coefficient of interest)	-0.032**	-0.0148	-0.0097	-0.0107	-0.0932	0.0023
	(0.0139)	(0.0133)	(0.0178)	(0.0179)	(0.0733)	(0.0828)
controls		✓		✓		✓
year FE	✓	✓	✓	✓	✓	✓
group FE	✓	✓	✓	✓	✓	✓
individual FE			✓	✓		
Observations	24791	24791	24791	24791	24791	24791

Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Notes: The only difference to the baseline setup and results in Table 3.1 is a restriction to the **sample of fathers who got divorced prior to the reform** with a **youngest common child between 4 and 8 or 16 and 17** and residence in any German state besides Lower Saxony and North Rhine-Westphalia. Otherwise, the same notes as in Table 3.1 apply. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.



(a) Controls: Year and *group* fixed effects

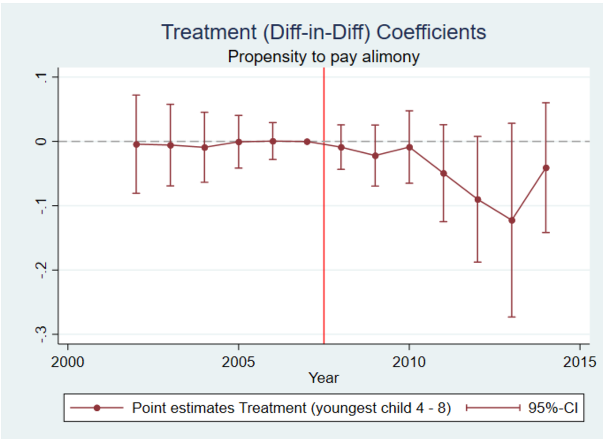


(b) Plus additional controls

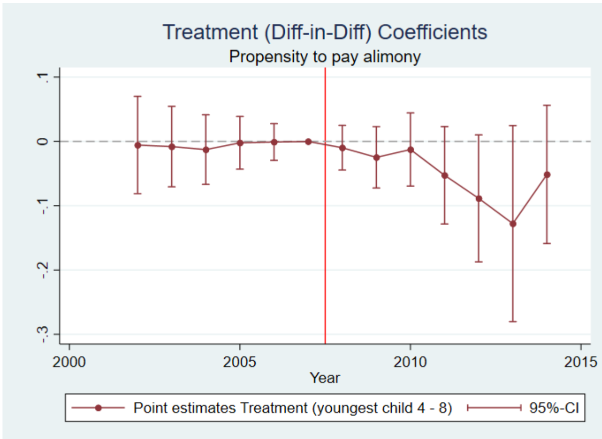
Figure 3.6: Linear probability model (OLS) with *group* fixed effect

Notes: Both figures plot the interaction effect of treatment indicator and year indicator on the propensity to pay alimony payments of divorced men over time based on a linear probability model with *group* and year fixed effects and robust standard errors. The effect is normalized to 0 in 2007. We use population weights in the regressions. **Sample definition: fathers (as in Section 3.3) who got divorced prior to the reform with a youngest common child between 4 and 8 or 16 and 17**. Controls in Figure 3.6(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

treatment effect in the later years might be biased towards zero because the effect of the reform will be higher if the youngest common child is four to five compared to six to seven or eight to nine (see Table 3.4 in the following Section 3.6.2.3). Second, the sample of divorced fathers on which we identify the year-specific treatment effect changes over time. Let us consider, for example, the year-specific treatment effects in 2008 and 2014. A father, who got divorced prior to the reform and has a child in the age group four to eight in 2008, got divorced when the child was between zero and seven. By contrast, a pre-reform divorced father with a child in the age group four to eight in 2014, got divorced when the child was between zero and one. A similar argument can be established for the control group (child between fifteen and sixteen). This implies that the sample composition differs between years which might have an effect on the dynamics of the treatment effect.[31]



(a) Controls: Year and *individual* fixed effects



(b) Plus additional controls

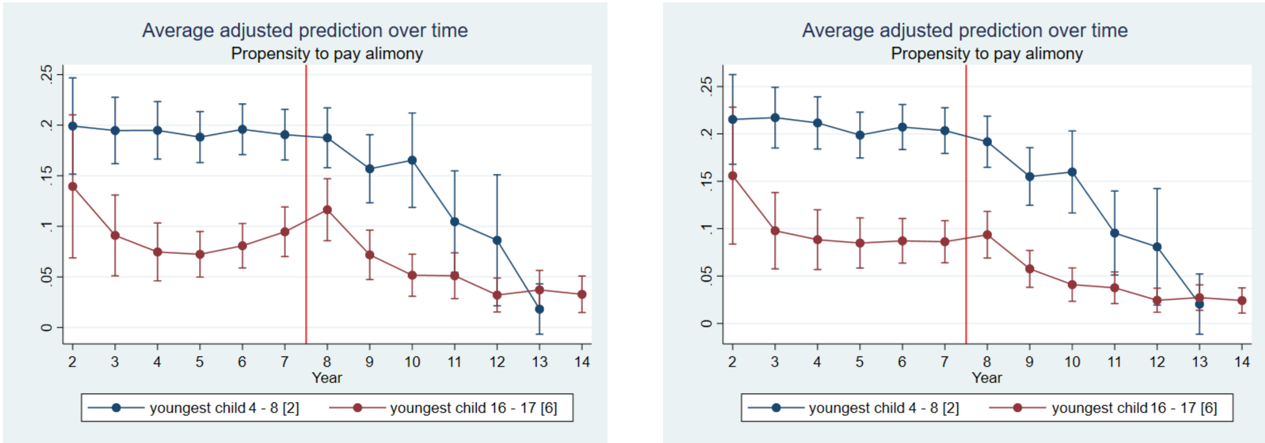
Figure 3.7: Linear probability model with *individual* fixed effect (Two-way fixed effects model)

Notes: Both figures plot the interaction effect of treatment indicator and year indicator on the propensity to pay alimony payments of divorced men over time based on a linear probability model with *individual* and year fixed effects with standard errors clustered at the individual level. The effect is normalized to 0 in 2007. We use population weights in the regressions. **Sample definition: fathers (as in Section 3.3) who got divorced prior to the reform** with a **youngest common child between 4 and 8 or 16 and 17**. Controls in Figure 3.7(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

3.6.2.3 Treatment intensity

Finally, we investigate how the reform influenced the probability to pay alimony depending on differently aged children that are younger than seventeen. This leads to a notion of treatment intensity conditional on the child's age. From the reform changes described in Section 3.2, we

[31]We plan to investigate this further in the future.



(a) Controls: Year and *group* fixed effects

(b) Plus Controls

Figure 3.8: Probit model with *group* fixed effect

Notes: Both figures plot the average adjusted predictions of the propensity to pay alimony payments of divorced men over time based on a probit model with *group* and year fixed effects with non-robust standard errors. We use population weights in the regressions. **Sample definition: fathers (as in Section 3.3) who got divorced prior to the reform** with a **youngest common child between 4 and 8 or 16 and 17**. Controls in Figure 3.8(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

would expect the largest negative effect for fathers with children at the age of four followed by a rather continuous increase with child age.

First, we consider four-year brackets of child age: if the child is between zero and three, four and eight, or nine and thirteen, we will consider their fathers being treated and therefore divide them in three different groups. Then, we compare each group to the control group of divorced fathers with a child between fourteen to seventeen. The resulting estimates are shown in Table 3.3. We find a significant and sizeable effect for all three treatment groups. Once we use controls, the results are largest for the child age brackets 0 - 3 and 4 - 8: the LPM with group fixed effects predicts a significant decrease of about 3.7 percentage points for both groups, which is – for the bracket 4 - 8 – also in line with the results in the baseline estimation (Table 3.1). We visualize the average adjusted predictions from the Probit model in Figure 3.12(a) and 3.12(b) in the Appendix which emphasises common pre-trends of the treatment and control groups except for the child age bracket 0 - 3. This might arise from having too few observations in this treatment group and therefore, we should be careful in interpreting the estimated effect on this group causally. By contrast, the other treatment groups (4 - 8 and 9 - 13) show common pre-trends with the control group (14 - 17) and after the reform we observe the gap between their average adjusted predictions closing over time resulting in almost the same prediction in 2014 for the three groups.

Table 3.3: Robustness – Treatment intensity I: Extension of models (3.1), (3.3), and (3.5) to include **all divorced fathers with a youngest common child below 18**

	Dependent variable: y_{it}			
	Linear Probability Model		Probit	
α if child 0 - 3	-0.0671***	-0.0375**	-0.2534***	-0.1437
	(0.0173)	(0.0165)	(0.0784)	(0.091)
α if child 4 - 8	-0.0519***	-0.0364***	-0.1734***	-0.1426***
	(0.0088)	(0.0084)	(0.0436)	(0.0505)
α if child 9 - 13	-0.0376***	-0.0202***	-0.1498***	-0.072
	(0.0078)	(0.0074)	(0.0421)	(0.0488)
controls		✓		✓
year FE	✓	✓	✓	✓
group FE	✓	✓	✓	✓
Observations	100495	100495	100495	100495

Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Notes: α is the coefficient of interest. Columns 1 and 2 show results for the extension of model (3.1). We use robust standard errors. Columns 3 and 4 show results for the extension of model (3.5). We use regular standard errors. We use population weights in the regressions. **Sample definition: divorced fathers (as in Section 3.3) with a youngest common child below 18.** The **treatment** is defined as having a youngest common child in either of the three **child brackets 0 - 3, 4 - 8, or 9 - 13** with the **control** group defined by the **bracket 14 - 17**. Controls are the same as in Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

Second, to gain an understanding of a more continuous dependency of child age and treatment intensity, we disaggregate the effects further by splitting up the sample of divorced fathers into groups conditional on their youngest common child being in either of the two-year brackets 0 - 1, 2 - 3, ..., 16 - 17. We estimate a LPM and Probit model, both with group fixed effects (see Table 3.4). First and as expected from the institutional setup, the reform has no different effect if the child is between zero and one compared to sixteen to seventeen. The magnitude of the treatment effect is large for the child age bracket 2 - 3 (-4.6 pp in the LPM with group fixed effects and controls). In line with the reform, we observe that fathers with children between four and five are the most treated with a significant post-treatment effect of -5.0 percentage points (LPM with group fixed effects and controls), which might be driven by measurement error in child age as we do not have information on children's birth months. The effect converges slowly towards zero with increasing child age, but remains significant at the 5% level until child age bracket 10 - 11 in the LMP with group fixed effect and controls.[32] The results of the

[32]The insignificance of α for the child age bracket 14 - 15 justifies our specification of the alternative control group in Section 3.6.2.1.

Probit model confirm the dependency of child age and treatment intensity, but using controls decreases the significance of the estimates leaving only the bracket of four to five year old children significant at the 5%-level. This might be due to too few observations per child age bracket driving up the standard errors.

Table 3.4: Robustness – Treatment intensity II: Extension of models (3.1), (3.3), and (3.5) to include **all divorced fathers with a youngest common child below 18**

	Dependent variable: y_{it}			
	Linear Probability Model		Probit	
α if child 0 - 1	-0.0135	-0.0014	-0.048	-0.0143
	(0.0377)	(0.0354)	(0.203)	(0.2413)
α if child 2 - 3	-0.0847***	-0.0465**	-0.3235***	-0.1613
	(0.0195)	(0.0187)	(0.0899)	(0.1044)
α if child 4 - 5	-0.0738***	-0.0504***	-0.2693***	-0.2074**
	(0.0148)	(0.014)	(0.0707)	(0.0817)
α if child 6 - 7	-0.0541***	-0.034***	-0.1872***	-0.1135
	(0.0128)	(0.0121)	(0.0641)	(0.0741)
α if child 8 - 9	-0.0466***	-0.0278**	-0.1722***	-0.0975
	(0.0118)	(0.0112)	(0.0623)	(0.0719)
α if child 10 - 11	-0.0475***	-0.0241**	-0.1914***	-0.0721
	(0.0117)	(0.011)	(0.0629)	(0.0725)
α if child 12 - 13	-0.0342***	-0.0131	-0.1533**	-0.0373
	(0.0113)	(0.0107)	(0.064)	(0.075)
α if child 14 - 15	-0.0128	-0.0035	-0.0647	-0.0013
	(0.0104)	(0.01)	(0.0649)	(0.0764)
controls		✓		✓
year FE	✓	✓	✓	✓
group FE	✓	✓	✓	✓
Observations	100495	100495	100495	100495

Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Notes: α is the coefficient of interest. Columns 1 and 2 show results for the extension of model (3.1) with and without controls. We use robust standard errors. Columns 3 and 4 show results for the extension of model (3.5) with and without controls. We use regular standard errors. We use population weights in the regressions. **Sample definition: divorced fathers (as in Section 3.3) with a youngest common child below 18.** The **treatment** is defined as having a youngest common child in either of the nine **child brackets 0 - 1, 2 - 3, 4 - 5, 6 - 7, 8 - 9, 10 - 11, 12 - 13, 14 - 15** with the **control** group defined by the bracket **16 - 17**. Controls are the same as in Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

3.7 Conclusion

We use a large, administrative dataset and apply a difference-in-differences setup to establish the robust result that the German divorce law reform in 2008 affected the probability to receive alimony as a divorced mother conditional on the current age of her youngest child born within marriage. The results are robust to various econometric specifications and the long-term impact still persists once we condition on couples getting divorced prior to the reform.

In future research we plan to extend the current paper along different dimensions. We intend to focus on the intensive margin and verify that the effect of the reform on the amount of alimony was limited. Further, we want to conduct two additional robustness exercises to potentially control for selection problems. The German law specifies that the recipient of alimony loses her legal entitlement to alimony once she remarries. Unfortunately, we cannot observe if the former spouse gets remarried. But younger women might have a higher chance to remarry and therefore, lose their eligibility implying that their husbands do not need to pay alimony any longer. Hence, we plan to include the age of the woman at the time of divorce to control for the probability to get married again. Similarly, to control for each woman's value on the marriage market, we might include the years since divorce.

The key value of this paper is that its results imply that the margin of the age of the youngest common child can be used as an instrumental variable for the treatment intensity of the reform in future research, i.e. as an instrument for the decrease in probability to receive or pay alimony after the reform.

The variation created by the German reform provides a good setup to investigate the impact of a decrease in the probability to receive alimony – which is, in expectation, a negative shock to the household income – onto labour supply, household income, divorce probabilities, as well as child development.[33] It enables us to provide, for example, estimates of the labour supply elasticities of mothers which are of great interest to the economic research profession and policy makers.

We can estimate the effect on the husband's earnings as well as on divorce probabilities using the Taxpayer Panel. The TPP follows in 3/4 of the cases the former husband and the remaining quarter covers former wives. This allows us to estimate the effect of the reform on male and

[33] We intend to complement the literature that studies the effects of alimony payments on the mentioned outcomes by exploiting the variation created by the German reform. First, Rangel 2006 and Chiappori, Iyigun, et al. 2017 focus on reforms of alimony payments in Brazil and Canada and estimate the effects on female labour supply. In addition, Foerster 2019 set up a dynamic limited commitment model to study the effect of post-marital maintenance payments – i.e. alimony and child support payments – on inter-temporal decisions and welfare of couples using Danish data. Reynoso 2019 combines an equilibrium model of the marriage market with a household model allowing for divorce in a limited commitment framework including non-cooperative behaviour of ex-spouses as well as alimony payments. Finally, Flinn 2000 finds that higher enforcement of child support payments have only weak effects on the welfare of kids of divorced couples.

female gross earnings, possibly focusing on different income categories such as labour income, capital income, etc. In addition, the Taxpayer Panel is very suitable to estimate the effect of the entitlement to alimony payments on divorce probabilities because of its size and panel structure.

To obtain reliable estimates of the effect on labour supply as well as on child development or children's school achievements, we will need to rely on survey data because these variables are not available in the income tax data. There exist two promising data sources: first, the Microcensus which is an annual, cross-sectional survey dataset that contains about 380,000 households and 820,000 individuals per year. It is also administered by the German Federal Statistical Office. Second, the GSOEP which is an unbalanced panel survey that has fewer observations (11,000 households per year), but it covers in its longitudinal form the period from 1984 to 2016. Both datasets contain information on family status, employment and labour force participation, education, children and their school attendance and achievements. Nonetheless, using the Microcensus as a repeated cross-sectional dataset might be useful because its large sample size enables us to condition on the age of the youngest common child of divorced parents. Building on this paper, the estimation of the effect of a decrease in alimony payments on each of these other outcomes is left for future research.

Appendix

3.A Additional information on the institutional setup

3.A.1 German divorce law

A marriage gets divorced in Germany following the request of one spouse or both spouses by a judgement of a family court when it can be concluded that the marriage had failed. Reasons for the failure and each spouse's responsibility for it are not relevant for the court decision. The main basis for the decision is the previous period of separation. A marriage would be presumed to have failed, if both spouses lived separately for one year and the divorce is consensual. If the divorce is unilateral, the spouses need to have lived separately for three years or lived separately for one year requiring an additional hearing of evidence by the court. But according to the divorce statistics of the German Federal Statistical Office, a bit more than 93% of all divorces were consensual in 2008.[34]. Hence, it should be quantitatively sufficient to consider a separation period of one year.[35]

[34]Own calculation based on numbers in Statistisches Bundesamt 2018b, page 12)

[35]A document describing the German divorce law system (in German) can be accessed here.

3.B Additional tables and figures

3.B.1 Robustness: Alternative control group

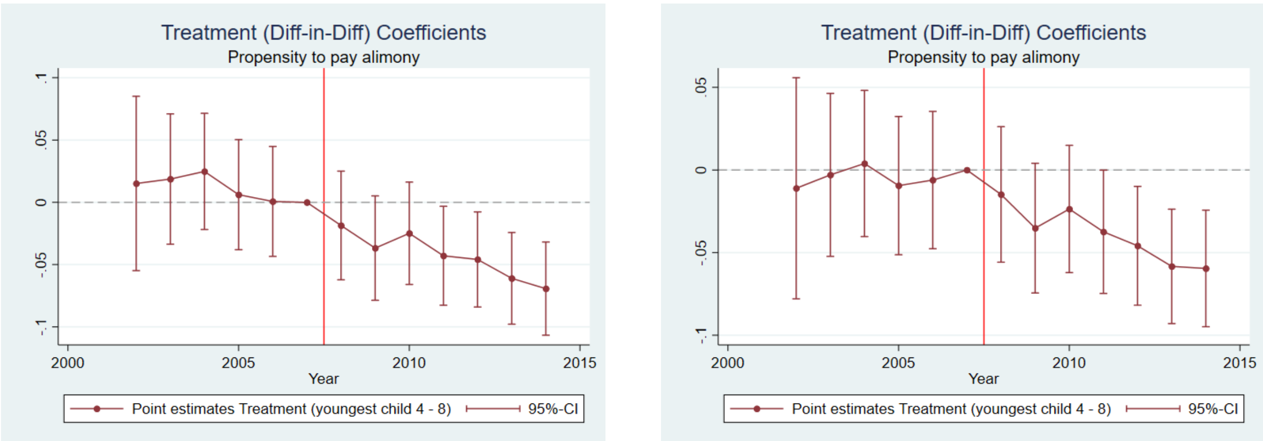
Table 3.5: Robustness 1 – Control group conditional on youngest common child aged 14 - 17

Dependent variable: y_{it}

	Linear Probability Model				Probit	
α (coefficient of interest)	-0.0526***	-0.0371***	-0.0139	-0.0147	-0.1779***	-0.1453***
	(0.0088)	(0.0084)	(0.0131)	(0.0130)	(0.0437)	(0.0505)
controls		✓		✓		✓
year FE	✓	✓	✓	✓	✓	✓
group FE	✓	✓	✓	✓	✓	✓
individual FE			✓	✓		
Observations	57609	57609	57609	57609	57609	57609

Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Notes: The only difference to the baseline setup and results in Table 3.1 is an extended control group definition which covers divorced fathers with a **youngest common child between 14 and 17** (baseline setup: between 16 and 17). Otherwise, the same notes as in Table 3.1 apply. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

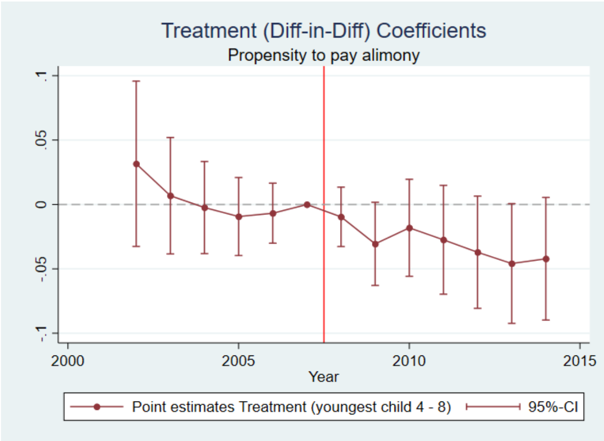


(a) Controls: Year and *group* fixed effects

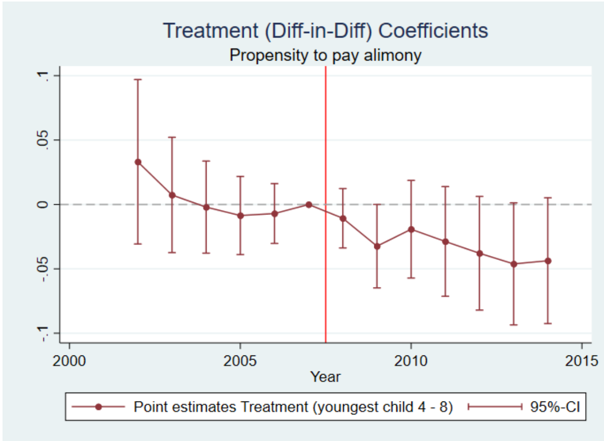
(b) Plus Controls

Figure 3.9: Linear probability model (OLS) with *group* fixed effect

Notes: Both figures plot the interaction effect of treatment indicator and year indicator on the propensity to pay alimony payments of divorced men over time based on a linear probability model with *group* and year fixed effects and robust standard errors. The effect is normalized to 0 in 2007. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 14 and 17**. Controls in Figure 3.9(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.



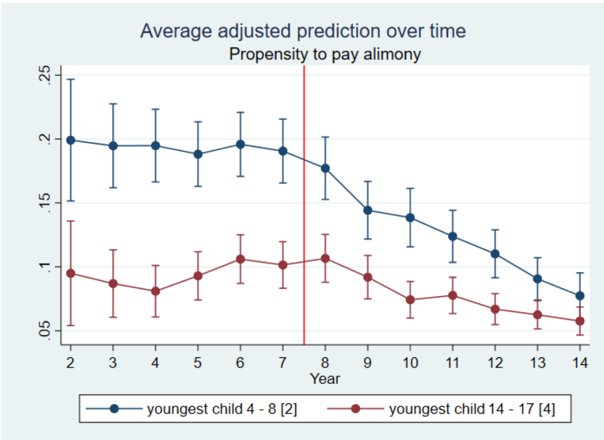
(a) Controls: Year and *individual* fixed effects



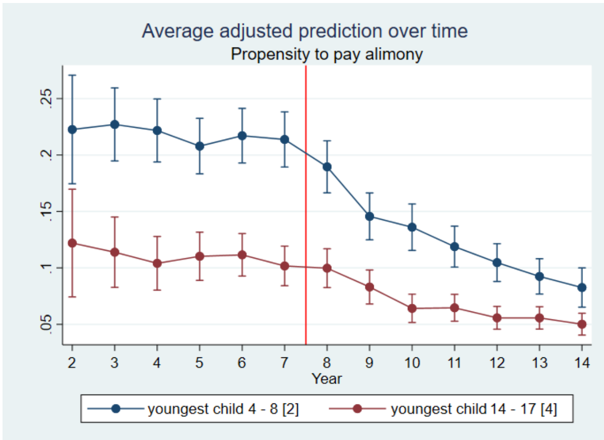
(b) Plus Controls

Figure 3.10: Linear probability model with *individual* fixed effect (Two-way fixed effects model)

Notes: Both figures plot the interaction effect of treatment indicator and year indicator on the propensity to pay alimony payments of divorced men over time based on a linear probability model with *individual* and year fixed effects with standard errors clustered at the individual level. The effect is normalized to 0 in 2007. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 14 and 17**. Controls in Figure 3.10(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.



(a) Controls: Year and *group* fixed effects

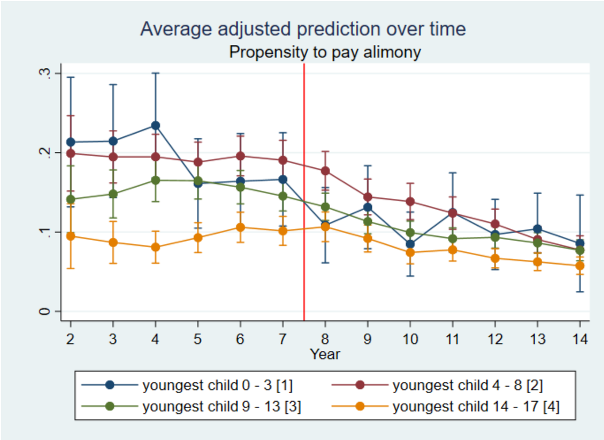


(b) Plus Controls

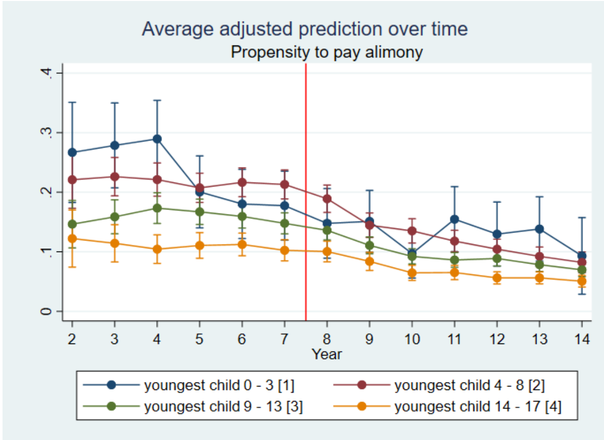
Figure 3.11: Probit model with *group* fixed effect

Notes: Both figures plot the average adjusted predictions of the propensity to pay alimony payments of divorced men over time based on a probit model with *group* and year fixed effects with non-robust standard errors. We use population weights in the regressions. Sample definition: divorced fathers (as in Section 3.3) with a **youngest common child between 4 and 8 or 14 and 17**. Controls in Figure 3.11(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

3.B.2 Robustness: Treatment intensity



(a) Controls: Year and *group* fixed effects



(b) Plus Controls

Figure 3.12: Probit model with *group* fixed effect

Notes: Both figures plot the average adjusted predictions of the propensity to pay alimony payments of divorced men over time based on a probit model with *group* and year fixed effects with non-robust standard errors. We use population weights in the regressions. **Sample definition: divorced fathers (as in Section 3.3) with a youngest common child below 18.** The **treatment** is defined as having a youngest common child in either of the three **child brackets 0 - 3, 4 - 8, or 9 - 13** with the **control** group defined by the **bracket 14 - 17**. Controls in Figure 3.12(b) as defined in footnote of Table 3.1. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

3.B.3 Average age of children in age brackets

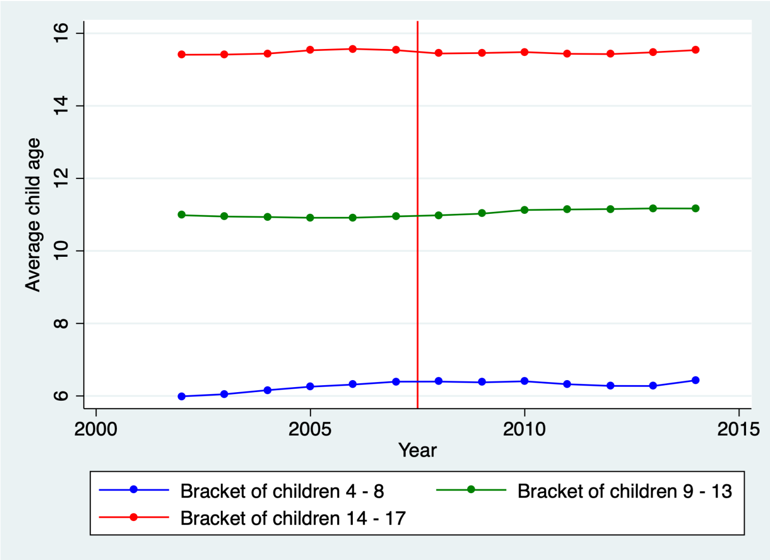


Figure 3.13: Average age of children in age brackets: *entire sample*

Notes: Average child age of the youngest common child of divorced fathers per bracket. Sample definition: *all divorced fathers* (as in Section 3.3) with a **youngest common child under the age of 18** that is in one of the relevant age brackets. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

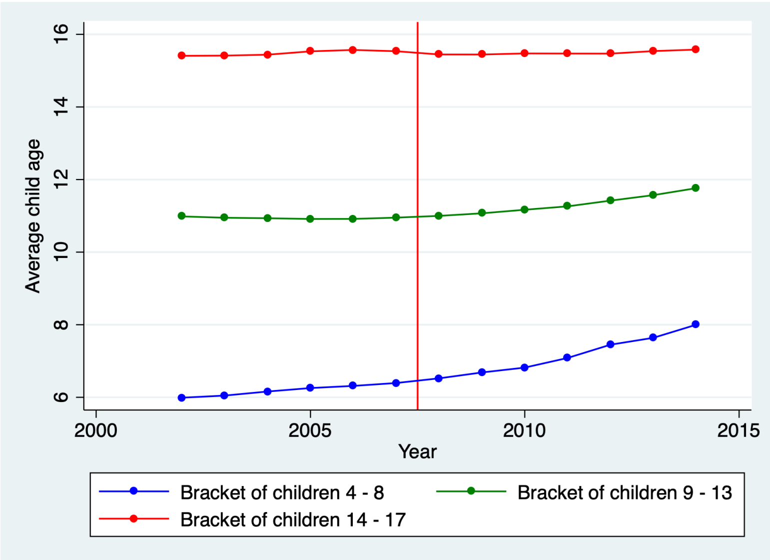


Figure 3.14: Average age of children in age brackets: *sample of men divorced pre-reform*

Notes: Average child age of the youngest common child of divorced fathers per bracket. Sample definition: *fathers who got divorced before the reform, i.e. before (excluding) 2008* (as in Section 3.3) with a **youngest common child under the age of 18** that is in one of the relevant age brackets. Source: RDC of the Federal Statistical Office and Statistical Offices of the Lander, Taxpayer Panel, 2001 - 2014, own calculations.

Bibliography

Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2013). "Education policy and intergenerational transfers in equilibrium". Tech. rep. NBER Working Paper 18782.

Adda, J., C. Dustmann, and K. Stevens (2017). "The Career Costs of Children". *Journal of Political Economy* 125.2, 293–337.

Altonji, J. and T. Dunn (1991). "Relationship among the family incomes and labor market outcomes of relatives". Tech. rep. NBER Working Paper 3724.

Altonji, J. and T. Dunn (2000). "An intergenerational model of wages, hours, and earnings". *Journal of Human Resources* 35.2, 221–258.

Angelov, N., P. Johansson, and E. Lindahl (2016). "Parenthood and the Gender Gap in Pay". *Journal of Labor Economics* 34.3, 545–579.

Authoring Group Educational Reporting (2018). *Education in Germany 2018.*

Bach, S., J. Jessen, P. Haan, F. Peter, C. K. SpieSS, and K. Wrohlich (2020). "Fiskalische Wirkungen eines weiteren Ausbaus ganztägiger Betreuungsangebote für Kinder im Grundschulalter".

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.

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.

Bastani, S., S. Blomquist, and L. Micheletto (2019). "Child Care Subsidies, Quality, and Optimal Income Taxation".

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.

Bernal, R. and M. P. Keane (2011). "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers". *Journal of Labor Economics* 29.3, 459–512.

Bettendorf, L. J. H., E. L. W. Jongen, and P. Muller (2015). "Childcare subsidies and labour supply? Evidence from a large Dutch reform". *Labour Economics* 36, 112–123.

Bettinger, E. P. and B. T. Long (2005). "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students". *American Economic Review: Papers and Proceedings* 95.5, 152–157.

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.

Binder, A. (2018). “Inter-generational transmission and the 21st century rise in skilled mothers’ labor supply”. Tech. rep. University of Michigan.

Björklund, A. and M. Jäntti (2011). “Intergenerational income mobility and the role of family background”. *Oxford Handbook of Economic Inequality.* Ed. by B. Nolan, W. Salverda, and T. Smeeding. Elsevier, 1–36.

Black, S. and P. Devereux (2011). “Recent developments in intergenerational mobility”. *Handbook of Labor Economics, Volume 4, Part B.* Ed. by O. Ashenfelter and D. Card. Elsevier, 1487–1541.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). “Female Labor Supply, Human Capital, and Welfare Reform”. *Econometrica* 84.5, 1705–1753.

Bowles, S. and H. Gintis (2002). “The inheritance of inequality”. *Journal of Economic Perspectives* 16.3, 3–30.

Bredtmann, J. and C. Vonnahme (2019). “Less money after divorce – how the 2008 alimony reform in Germany affected spouses’ labor supply, leisure and marital stability”. *Review of Economics of the Household* 17.4, 1191–1223.

Brudermüller, G. and U. Diederichsen (2008). “Gesetz zur Änderung des Unterhaltsrechts”. *Palandt. Bürgerliches Gesetzbuch (BGB).* 67th ed. Chap. Addendum, 1–48.

Bundesministerium für Familie Senioren Frauen und Jugend (2010). *Familien Report 2010.*

Busse, A. and C. Gathmann (2019). “Free Daycare Policies, Family Choices and Child Development”.

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, 64–71.

Chadwick, L. and G. Solon (2002). “Intergenerational income mobility among daughters”. *American Economic Review* 92.1, 335–344.

Chetty, R., Hendren N., P. Kline, and E. Saez (2014). “Where is the land of opportunity? The geography of intergenerational mobility in the United States”. *Quarterly Journal of Economics* 129.4, 1553–1623.

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.

Chiappori, P. A., M. Iyigun, J. Lafortune, and Y. Weiss (2017). “Changing the Rules Midway: The Impact of Granting Alimony Rights on Existing and Newly Formed Partnerships”. *The Economic Journal* 127.604, 1874–1905.

Chiappori, P. A. and M. Mazzocco (2017). “Static and Intertemporal Household Decisions”. *Journal of Economic Literature* 55.3, 985–1045.

Colas, M., S. Findeisen, and D. Sachs (2019). “Optimal Need-Based Financial Aid”.

Corak, M. and P. Piraino (2011). “The intergenerational transmission of employers”. *Journal of Labor Economics* 29.1, 37–68.

Dahl, G. and A. Gielen (2018). “Intergenerational spillovers in disability insurance”. Tech. rep. IZA Discussion Paper N. 11334.

Dahl, G., A. Kostøl, and M. Mogstad (2014). “Family welfare cultures”. *Quarterly Journal of Economics* 129.4, 1711–1752.

Del Boca, D., C. Flinn, and M. Wiswall (2014). “Household Choices and Child Development”. *The Review of Economic Studies* 81.1, 137–185.

Domeij, D. and P. Klein (2013). “Should Day Care be Subsidized?” *Review of Economic Studies* 80.2, 568–595.

Eckhoff Andresen, M. and T. Havnes (2019). “Child care, parental labor supply and tax revenue”. *Labour Economics* 61, 101762.

Fahn, M., R. Rees, and A. Wuppermann (2016). “Relational contracts for household formation, fertility choice and separation”. *Journal of Population Economics* 29.2, 421–455.

Fernandez, R. (2007). “Women, work, and culture”. *Journal of the European Economic Association* 5.2-3, 305–332.

Fernandez, R. (2013). “Cultural change as learning: The evolution of female labor force participation over a century”. *American Economic Review* 103.1, 472–500.

Fernandez, R. and A. Fogli (2009). “Culture: An empirical investigation of beliefs, work, and fertility”. *American Economic Journal: Macroeconomics* 1.1, 146–177.

Flinn, C. (2000). “Modes of Interaction between Divorced Parents”. *International Economic Review* 41.3, 545–578.

Foerster, H. (2019). “The Impact of Post-Marital Maintenance on Dynamic Decisions and Welfare of Couples”.

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.

GESIS - Leibniz-Institut für Sozialwissenschaften (2017). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2016.*

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.

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 (2018). “The German Socio-Economic Panel (SOEP)”. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics* aop.

Grawe, N. (2006). “Lifecycle bias in estimates of intergenerational earnings persistence”. *Labour Economics* 13.5, 551–570.

Guner, N., R. Kaygusuz, and G. Ventura (2019). “Child-Related Transfers, Household Labor Supply and Welfare”. *Review of Economic Studies* forthcoming.

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, 137–157.

Haider, S. and G. Solon (2006). “Life-cycle variation in the association between current and lifetime earnings”. *American Economic Review* 96.4, 1308–1320.

Hartley, R., C. Lamarche, and J. Ziliak (2017). “Welfare reform and the intergenerational transmission of dependence”. Tech. rep. IZA Discussion Paper N. 10942.

Ho, C. and N. Pavoni (2019). “Efficient Child Care Subsidies”. *American Economic Review* forthcoming.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimueller (2019). “Child Penalties Across Countries: Evidence and Explanations”. *American Economic Review: Papers and Proceedings* 109, 122–126.

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.

Laun, T. and J. Wallenius (2019). “Having It All? Employment, Earnings, and Children”. *The Scandinavian Journal of Economics* forthcoming.

Lee, C. and G. Solon (2009). “Trends in intergenerational income mobility”. *Review of Economics and Statistics* 91.4, 766–772.

Lo Bello, S. and I. Morchio (2018). “Like father, like son: Occupational choice, intergenerational persistence and misallocation”. Tech. rep. Mimeo.

Lochner, L., Y. Park, and Y. Shin (2018). “Wage dynamics and returns to unobserved skill”. Tech. rep. NBER Working Papers 24220.

Low, H., C. Meghir, L. Pistaferri, and A. Voena (2018). “Marriage, Labor Supply and the Dynamics of the Social Safety Net”. Working Paper 24356. National Bureau of Economic Research.

Macmillan, L. (2011). “Measuring the intergenerational correlation of worklessness”. Tech. rep. CMPO Working Paper N. 11/278.

Mazumder, B. (2005). “Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data”. *Review of Economic and Statistics* 87.2, 235–255.

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.

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.

Nybom, M. and J. Stuhler (2016). “Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation”. *Journal of Human Resources* 51.1, 239–268.

Nybom, M. and J. Stuhler (2017). “Biases in standard measures of intergenerational income dependence”. *Journal of Human Resources* 52.3, 800–825.

Olivetti, C. and D. Paserman (2015). “In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850-1940”. *American Economic Review* 105.8, 2695–2724.

Olivetti, C., E. Patacchini, and Y. Zenou (2018). “Mothers, peers and gender-role identity”. *Journal of the European Economics Association* forthcoming.

Rainer, H., W. Auer, S. Bauernschuster, N. Danzer, A. Fichtl, T. Hener, C. Holzner, J. Reinkowski, and M. Werding (2013). “Öffentlich geförderte Kinderbetreuung in Deutschland: Evaluierung der Auswirkungen auf die Arbeitsmarktbeteiligung von Müttern”. *ifo Schnelldienst* 66.07, 31–40.

Rangel, M. A. (2006). “Alimony Rights and Intrahousehold Allocation of Resources: Evidence from Brazil”. *The Economic Journal* 116.513, 627–658.

Reynoso, A. (2019). “The impact of divorce laws on the equilibrium in the marriage market”.

Saez, E. (2001). “Using Elasticities to Derive Optimal Income Tax Rates”. *The Review of Economic Studies* 68.1, 205–229.

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.

Solon, G. (1989). “Biases in the estimation of intergenerational earnings correlations”. *Review of Economics and Statistics* 71.1, 172–174.

Solon, G. (1992). “Intergenerational income mobility in the United States”. *American Economic Review* 82.3, 393–408.

Solon, G. (1999). “Intergenerational mobility in the labor market”. *Handbook of Labor Economics, Volume 3*. Ed. by O. Ashenfelter and D. Card. Elsevier, 1761–1800.

Statistisches Bundesamt (2012). *Finanzen der Kindertageseinrichtungen in freier Trägerschaft 2010*.

Statistisches Bundesamt (2018a). *Alleinerziehende in Deutschland 2017*.

Statistisches Bundesamt (2018b). *Bevölkerung und Erwerbstätigkeit – Statistik der rechtskräftigen Beschlüsse in Eheauflösungssachen (Scheidungsstatistik) und Statistik der Aufhebung von Lebenspartnerschaften 2016*.

Toledo, M. (2010). “On the intergenerational persistence of work hours”. Tech. rep. Mimeo.

Turon, H. (2019). “Home production of childcare and labour supply decisions in a collective household model”.

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. J. Carey, I. 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”.

Zimmerman, D. (1992). “Regression towards mediocrity in economic stature”. *American Economic Review* 82.3, 409–429.