Equity and Efficiency of Childcare Subsidies: A Dynamic Structural Approach*

David Koll† Dominik Sachs‡ Fabian Stürmer-Heiber§ Hélène Turon¶

January 21, 2022

Abstract

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

We then turn to the question how subsidies should vary with income. We compare the dynamic marginal excess burden (MEB) from making the childcare fee schedule marginally more progressive to the dynamic MEB from making the tax schedule marginally more progressive.

JEL Code: H23, H31, J13, J22, J24

Keywords: Female Labour Supply, Childcare, Family Policies, Fiscal Externalities, Dynamic Discrete Choice, Redistribution

---

*We thank Nuan Stahl and Victoria Szabo for excellent research assistance. David Koll gratefully acknowledges funding through the German Research Foundation (through the CRC-TR-224 project A3 and the Gottfried Wilhelm Leibniz-Prize), the German Academic Scholarship Foundation, and the Zukunftskolleg of the University of Konstanz. Fabian Stürmer-Heiber gratefully acknowledges funding through the International Doctoral Program "Evidence-Based Economics" of the Elite Network of Bavaria. We thank seminar participants at CRED Paris, EUI, University of St. Gallen, University of Konstanz, LMU Munich, University of Mannheim, Bank of Canada, SED, NASMES, Barcelona Summer Forum, Conference on Dynamic Structural Estimation, Verein für Socialpolitik, IIPF, and FROGEE Online Workshop on Gender Economics for their valuable feedback.

†University of Mannheim and European University Institute, Florence. david.koll@eui.eu

‡University of St. Gallen. dominik.sachs@unisg.ch

§LMU Munich. fabian.stuermerheiber@econ.lmu.de

¶University of Bristol. helene.turon-lacarrieu@bristol.ac.uk

1 Introduction

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

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

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

We then turn to the redistribution that is entailed in the progressive childcare fee schedule: subsidies decrease with increasing household income to a minimum of around 50%. We show that the high subsidies for high-income families can barely be grounded on efficiency rationales. Specifically, we compare the childcare fee schedule to the income tax schedule in terms of each

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The remainder of this paper is organized as follows: Next, we provide a short review of the literature. In Section 2 we present the institutional background and a number of stylised facts, which motivate our analysis. Section 3 contains the model setup and Section 4 presents the estimation of its components. In Section 5 we illustrate the fit of the estimated model. Section 6 presents our policy analysis and fiscal calculations and Section 7 concludes.

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

Reduced form studies that show short to medium-term positive effects of (subsidized) public childcare on female labour supply in the German context include Bauernschuster and Schlotter (2015), Gathmann and Sass (2018), Busse and Gathmann (2020), and Müller and Wrohlich (2020).[4] Bauernschuster and Schlotter (2015) leverage the 1996 introduction of a legal entitlement to a place in kindergarten for children above 3 for identification and find a positive impact of public childcare on maternal employment. Müller and Wrohlich (2020) find a similar effect by exploiting spatial and temporal variation in a recent public childcare expansion for below 3-year-olds. Focusing on a policy reform which increased the implicit price of childcare in the East German state Thuringia, Gathmann and Sass (2018) show that in response to the reform, public childcare attendance dropped and maternal labour supply declined, especially for vulnerable subgroups such as single, low-income, and low-skilled parents. Busse and Gathmann (2020) use regional variation in childcare fees in Germany combined with birthday cut-offs to evaluate the introduction of free universal public childcare, which took place in some selected West German states at different points in time. They find an increase in public childcare attendance of 2 – 3 year old children, which led to higher labour market attachment of mothers. The responses are stronger for vulnerable subgroups such as poorer families and low-skilled parents.

[4] Positive effects of public childcare on women's labour market outcomes have been documented for a number of countries. See e.g. Baker, Gruber, and Milligan (2008), Bettendorf, Jongen, and Muller (2015), Givord and Marbot (2015), and Nollenberger and Rodríguez-Planas (2015). One of the few studies that evaluates long-term effects on maternal 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. For other countries, however, researchers have also found contrasting results. Havnes and Mogstad (2011) and Kleven, Landais, Posch, et al. (2020) found at most modest effects of public childcare provision on female earnings in Norway and Austria respectively.

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

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

Our approach is different from all mentioned papers as we focus both on the availability of public childcare as well as on the degree of subsidization. Furthermore, as we take a clear public finance perspective, we have to go further than just capturing the aggregate labour supply responses to counterfactual policies. We model rich heterogeneity in family structures as well as preferences to pin down which mothers are marginal or inframarginal for a given policy scenario. This approach allows us to accurately quantify the net fiscal effects of the distribution of behavioural responses to the policies.

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

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

2 Background, data, and stylised facts

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

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, they enter kindergarten and stay there until they start school at age 6. Lastly, during school age, children may attend after school care centres in the afternoon.[6]The distinction between these institutions matters because childcare fees differ by the attended institution and, therefore, by child age.[7] Contrary to e.g. the 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% of childcare institutions are either operated by municipalities or by non-profit organisations.[8]

Recent public childcare expansion. While all three types of childcare institutions have been continuously present throughout Germany since the early 1990s, their use and prevalence have changed substantially in the past thirty years. Public childcare supply for children below 3 (i.e. nursery slots) remained scarce until 2005. In that year, the German government committed

[6]Only a small share of elementary schools operates with mandatory full-day schooling
[7]See Appendix C.1.1 for a detailed description of the determinants of childcare fees.
[8]See Authoring Group Educational Reporting (2018).

to creating 230,000 additional nursery slots by October 2010, increasing the childcare coverage for children aged 0 – 2 from just 5 to 17 slots per 100 children. Focusing policy efforts further on nurseries, a legal entitlement to a public childcare slot for all children aged 1+ was introduced in August 2013.[9] This reform effectively ended the rationing of public childcare.[10]

2.2 Data

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

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 (see e.g. Kleven, Landais, and Søgaard 2019, Angelov, Johansson, and Lindahl 2016, as well as Kleven, Landais, Posch, et al. 2019).

This also holds for Germany where childbirth has a substantial and sustained effect on maternal employment.[13] Figure 1 illustrates this effect in the GSOEP data. It compares maternal pre-birth employment rates to their post-birth development depending on the age of the youngest child. The average pre-birth employment rate of future mothers (illustrated by

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

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

[11]In line with our modelling framework described in Section 3, we furthermore limit the sample to females who have at most three children and gave birth (if any) only between ages 20 and 40 (87.40% of observations). See Appendix A for additional details on the sample.

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

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

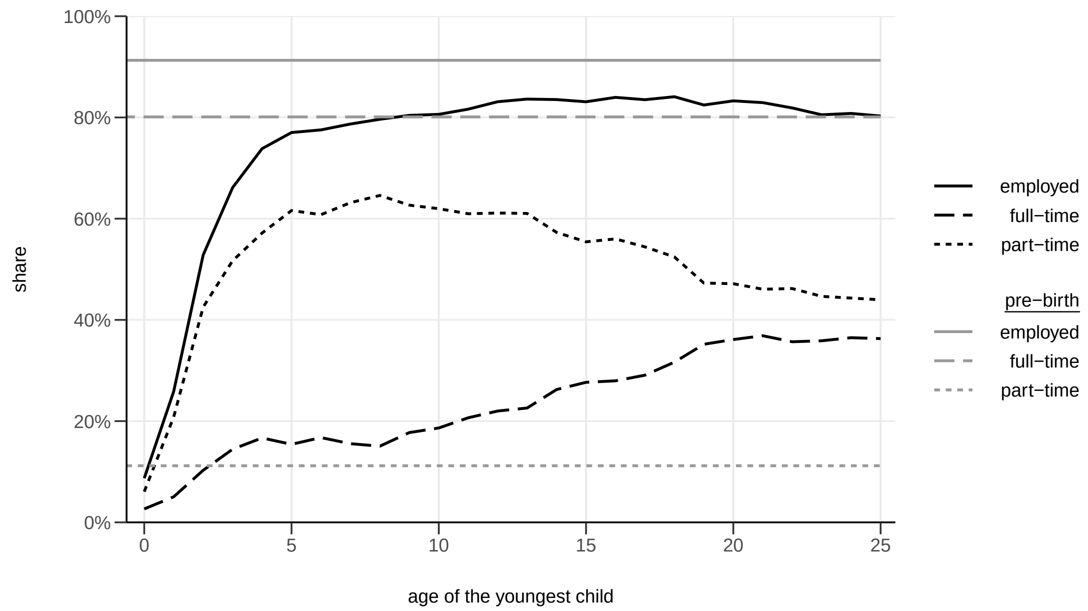


Figure 1: Maternal employment by the age of the youngest child

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

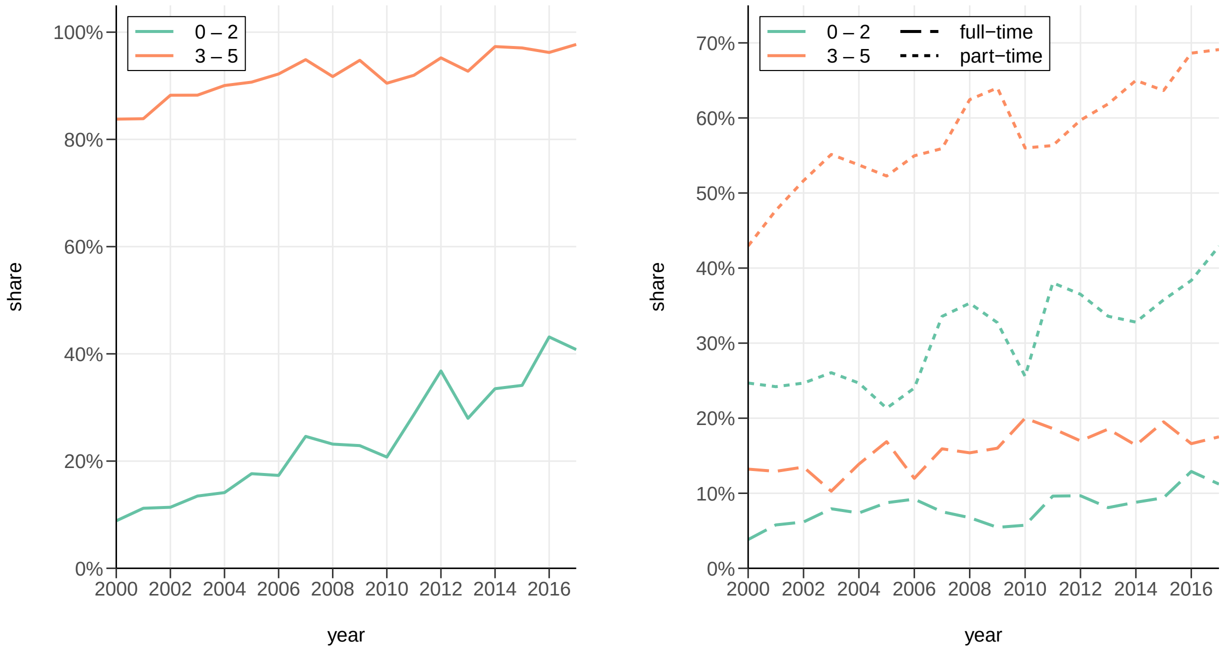
the horizontal lines) is above 90%, of which 80% work full-time (≈ 40 hours per week). In the first year after childbirth, the employment rate drops by more than 80 percentage points.[14]

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

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

Reform effects on childcare enrolment and maternal labour supply. Figure 2a plots childcare enrolment shares for different child age brackets across recent years. In 2000, more

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



(a) Public childcare enrolment

(b) Employment of mothers

Figure 2: Development of public childcare enrolment and maternal employment

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, conditional on having a child aged 0 – 2 or 3 – 5. Source: 2000 to 2017 GSOEP, FDZ-SOEP (2019).

than 80% of 3 – 5 year old children were already attending childcare, while the share was only 9% for 0 – 2 year olds. Over the years during which public childcare for 0 – 2 year olds was substantially expanded, this enrolment share has increased to more than 40% in 2017. Overall, childcare attendance of 0 – 2 year olds has increased by more than 30 percentage points over the last 17 years. Enrolment of 3 – 5 year old children has also increased, especially during the early 2000s, reaching 97% in 2017.

To relate this pattern of increased childcare attendance to employment rates, Figure 2b plots maternal full-time and part-time shares by the age of the youngest child. During the 2000 to 2017 time span, when childcare enrolment of children aged 0 – 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 among these mothers have increased by more than 15 percentage points, while full-time rates have also increased slightly. Looking at mothers with 3 – 5 year olds, part-time employment shares have risen substantially from 45 to more than 65% between 2000 and 2017. During the same period, the full-time share has increased by around 5 percentage points.

3 Model

We now present our 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. Marriages are formed at the age of 20 and are stable. Both spouses are of the same age, retire at 65, and have a remaining lifespan of 15 years after retirement. Fertility follows an exogenous stochastic process, which captures the substantial empirical heterogeneity in family composition and the age of parents at first birth.

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

A distinct feature of our model is the large amount of heterogeneity. Households differ in education, which is an important component in the stochastic wage and fertility processes. Besides education, female wages, male wages, and family composition, households are heterogeneous in three further dimensions: their preference for home-produced childcare, their taste for the leisure of the female spouse, and their access to free informal childcare. The latter three are unobservable to the econometrician. As we argue below, accounting for this unobserved heterogeneity is key to capture the large heterogeneity in childcare and labour supply choices of households.

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

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

3.1 Children

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

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

3.2 Preferences

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

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

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

[17]Only 4.03% of households have more than three children in our data.

Preference heterogeneity. Households' preferences are heterogeneous in two dimensions: α and g. α represents the relative taste for female leisure over consumption. The parameter g is the preference for home-produced childcare relative to leisure and consumption. For now, the reader should simply think of them as preferences for time with children and leisure, respectively. In Section 3.5 we discuss how these parameters can be interpreted in a more general way.

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

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

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

3.3 Constraints

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

Informal childcare is free and only available to some households. The variable *oth* ranges between 0 and 40 hours a week. If available, households always prefer to use *oth* hours of costless informal childcare over *mcc* hours of costly public childcare. In that sense, *oth* is a reduced form for whether informal childcare is available and considered equally good as public childcare. We provide a longer discussion about the interpretation and role of *oth* in Section 3.5.

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

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

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

$$mcc(j) = \max\left\{0, \bar{t}_j - dcc - oth\right\} \tag{2}$$

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

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

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

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

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

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

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

$$lm_t + L_t + dcc_t \quad = \quad 1. \tag{3}$$

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

$$c_t + Ecc(K_t, y_t) \quad = \quad y_t - T(y_t), \tag{4}$$

where

$$y_t = 40 \cdot \left(w_{m,t}\left(w_{m,t-1}\right) + lm_t \cdot w_{f,t}\left(w_{f,t-1}, lm_{t-1}\right)\right). \tag{5}$$

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

Wage growth differs across age and across the wage distribution. Furthermore, the expected wage penalties for working lower or no hours also differ depending on the level of the wage as well as on age. Despite the Markov property, the wage process is thereby able to capture the key dynamics of more complex human capital accumulation frameworks (such as in, e.g. Blundell et al. 2016), as we demonstrate in Section 4.1.3.

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

3.4 Dynamic decision problem

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

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

At each age t, the household has to choose female labour supply (lm_t) and the use of home-produced childcare (dcc_t), which also imply the choices of consumption (c_t), female leisure (L_t), and the use of public childcare (mcc_t). The labour supply decision is discrete between full-time, part-time, and non-participation. The home-produced childcare decision is close to continuous with decisions in steps of 5 hours from 0 to 40 hours. The three constraints that the household faces are the need for childcare of current children (equation (2)), the time constraint for the female spouse (equation (3)), and the budget constraint (equation (4)).

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

$$V(\Omega_t) = \max_{lm_t, dcc_t} u(c_t, L_t, dcc_t|\Omega_t) + \beta\mathbb{E}[V(\Omega_{t+1}|\Omega_t, lm_t)], \tag{6}$$

subject to the constraints (2), (3), and (4). This problem can be broken down into an intratemporal choice and an intertemporal choice. The intratemporal decision reflects the optimal time allocation between leisure and home-produced childcare for a given female labour supply choice lm_t. The intertemporal choice is to choose the optimal labour supply, lm_t, given the conditional optimal choice of home-produced childcare $dcc_t(lm_t)$.

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

$$u^*(\Omega_t, lm_t) = \max_{dcc_t} u(c_t, L_t, dcc_t|\Omega_t, lm_t), \tag{7}$$

subject to the childcare constraint (2), the time constraint (3), and the budget constraint (4).

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

$$V(\Omega_t) = \max_{lm_t} u^*(\Omega_t, lm_t) + \beta V(\Omega_{t+1}|\Omega_t, lm_t) \tag{8}$$

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

3.5 Unobserved heterogeneity

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

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

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

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

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

4 Estimation of the Model

Our estimation can be decomposed into two broad parts. First, we estimate and calibrate various parameters without using the explicit structure of the model in Section 4.1. We set policy parameters such as tax rates, childcare fees, but also our assumptions on childcare needs in Section 4.1.1. Then, we estimate the fertility process and the wage process in Sections 4.1.2 and 4.1.3.

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

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

4.1 Auxiliary regressions

4.1.1 Policies and childcare need

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

Table 1: Weekly childcare need within normal working hours

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

Childcare need. Table 1 summarises the assumed childcare need for children of different ages: If a child is younger than 6, the childcare need is set to 40 hours per week, i.e. 100% of the time. To account for the fact that almost all 3 – 5 year olds attend kindergarten at least half-days (see Figure 2a), we impose that 20 of the 40 hours required for this age interval have to be covered by public childcare. Half-day public childcare attendance in this age range has become close to a social norm. For children aged 6 – 8, the need reduces to 15 hours per week because these children attend compulsory schooling for 25 hours per week. This yields for each child age j, the age-specific weekly hours of childcare needed, $\bar{t}_j$.

[19]For the estimation, we furthermore operationalize the large state space as follows: To capture the age range from 20 to 80, we set up $t = 20$ 3-year model periods. Heterogeneity in male and female wages is captured by 5 and 10 gridpoints, respectively, education by 2 different levels, and the family structure K as introduced in Section 3.1 requires 18 state space points. The unobserved heterogeneity in g and α is captured by 20 gridpoints each, while 17 gridpoints are sufficient for oth.

Public childcare cost structures. We approximate the cost structure of public childcare institutions by assuming the costs to be linear in the number of children. This abstracts from any possible non-linearities driven e.g. by capacity constraints, but provides a reasonable average value for public spending per child. We use the values in Table 2, which are provided by the German Statistical Office.

Table 2: Average government spending per child for 40h/week of public childcare

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

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

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

Given that we also observe a fraction of households paying zero fees, we estimate a Tobit model of childcare fees as a function of gross household income, which we also interact with the number of siblings (see Appendix C.1.2 for details). Figure 3 shows the estimated fee schedule. Childcare fees are slightly increasing in household income (between 2% and 3% at the margin) and decrease with the number of siblings. Furthermore, fees are higher for younger children.

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

[20] In terms of the sample construction, this estimation is based on the same sample as laid out in Section 2.2.

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

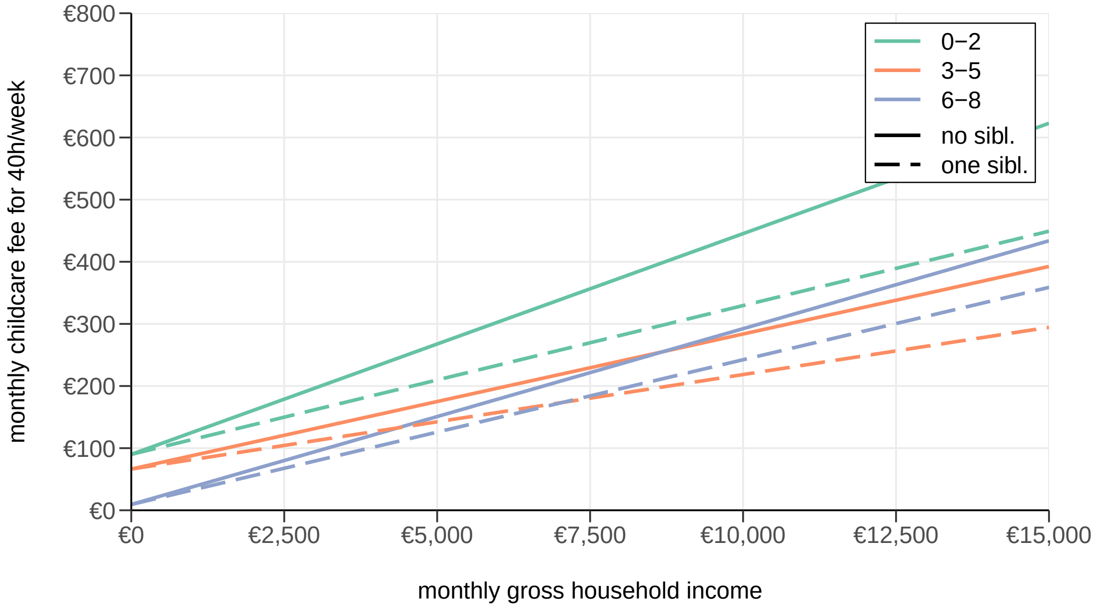


Figure 3: Estimated childcare fee schedules

Notes: See Appendix C.1.2 for details on the underlying Tobit regression. All values are adjusted to 2017 price levels. Source: 2013, 2015, 2017 GSOEP, FDZ-SOEP (2019).

rates faced by women differ with their spouses' income and child allowances reduce the taxable income of the household.

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

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

4.1.2 Estimation of the fertility process

As introduced in Section 3.1, children are assumed to be born one at a time in any 3-year model period to mothers aged 20 to 40. We also restrict households to have at most three children. The determinants of fertility are the age and education of the mother and the number and ages of children already present in the family. The transition probability between family composition K and family composition K' faced by a household aged t, with an education level *educ* captures the (deterministic) ageing of existing children and the fertility hazard over the next period. Our estimate of the birth rate for this household is simply the sample average of birth events conditional on $(t, educ, K)$.

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

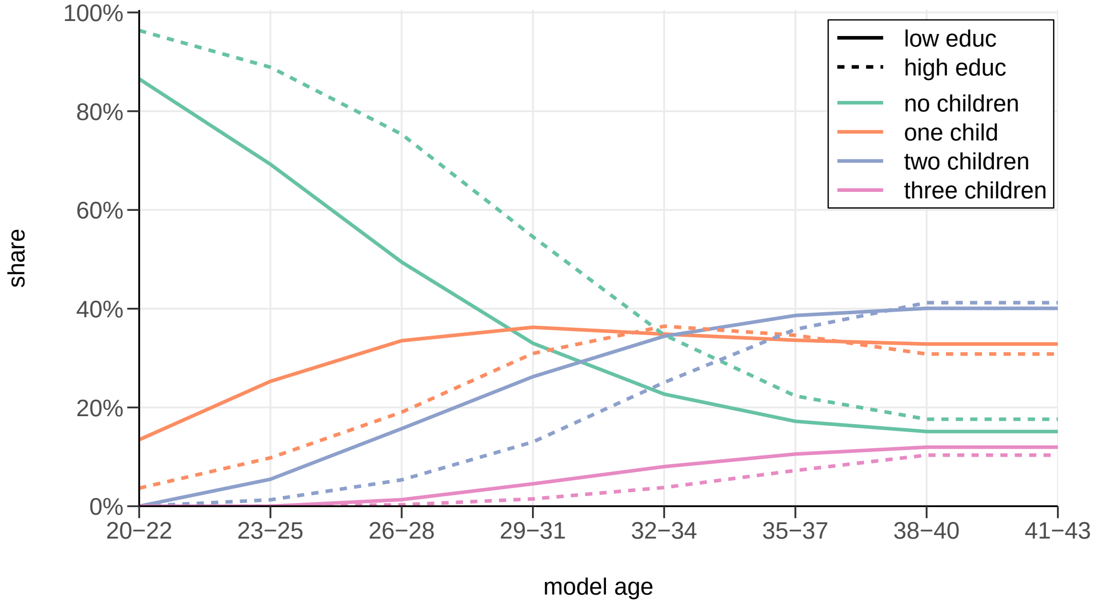


Figure 4: Family composition as implied by the fertility process

Notes: 'low educ' corresponds to no A-level, 'high educ' corresponds to having obtained an A-level. Sample restrictions as laid out in Section 2.2. Sources: FDZ-StABL (2020a), FDZ-StABL (2020b).

Figure 4 illustrates our estimates of the evolution of shares of families with zero to three children over the age of the mother and by education level, referring to having obtained an A-level or not. In terms of completed fertility, i.e. number of children at age 41, the figures are similar in both education groups: about 45% of households have two children, about 30% (respectively 10%) have one (respectively three) child(ren) and about 15% of households remain childless. The timing of births, however, differs markedly between education levels, with most curves for the high education group exhibiting a lag of three years, i.e. one model period, relative to the low education group. Only the share of families with three children grows nearly simultaneously for both groups. By age 34 (respectively 37) for the low (respectively high)

[22]We select the sample from the Microcensus data using the same restrictions as for our GSOEP survey data (see Section 2.2). Sources: FDZ-StABL (2020a), FDZ-StABL (2020b). This yields us 71,165 observations of households aged 20 to 40.

education group, the majority of households have completed their fertility. Half the households have had at least one child by age 26 (respectively 29) in the low (respectively high) education group.

4.1.3 Estimation of the wage process

Using the 2000 to 2017 GSOEP data, we observe monthly gross labour income as well as contracted working hours.[23] This allows us to directly compute hourly wages for every individual that is working. For females who choose not to work, on the other hand, we do not observe any labour income and therefore, we impute their *potential* gross hourly wages using a selection corrected wage model (see Appendix C.2.1 for details).

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

$$\begin{aligned} log(w_{f,it}) = \alpha + \beta_1 log(w_{f,it-1}) + \beta_2 \mathbb{1}\{lm_{it-1} = NP\} + \\ \beta_3 \mathbb{1}\{lm_{it-1} = PT\} + \beta_4 educ_i + \mathcal{A}(t) + \varepsilon_{it}^{w_f}, \end{aligned} \tag{9}$$

where $\mathbb{1}\{lm_{it-1} = NP\}$ and $\mathbb{1}\{lm_{it-1} = PT\}$ are dummy variables that indicate whether a woman i was either not working or working part-time in period $t-1$. The coefficients β_2 and β_3 are of particular interest for our analysis since they measure the dynamic wage penalty from working less than full-time. β_4 captures the wage increase due to having obtained an A-level and $\mathcal{A}(t)$ is a third-order polynomial in age. Note that the implied wage process is a Markov process, where the individual wage is drawn from a log-normal distribution that depends on the previous wage, previous employment decision, age, and education. The estimated coefficients are shown in Appendix-Table C.2.

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

In Figure 5, we illustrate our wage process by looking at the benefits of increasing labour supply relative to a typical labour supply pattern of mothers. Specifically, we consider a mother that conceives her child at 26. The benchmark is that she does not work while the child is 0 – 2,

[23]We extend the sample for the wage process estimation back until 2000 to ensure that we can robustly capture the key dynamics with a sufficient number of observations. Otherwise, we use exactly the same sample restrictions as described in Section 2.2.

[24]Based on the estimates in Appendix-Table C.2, transformed into percent changes.

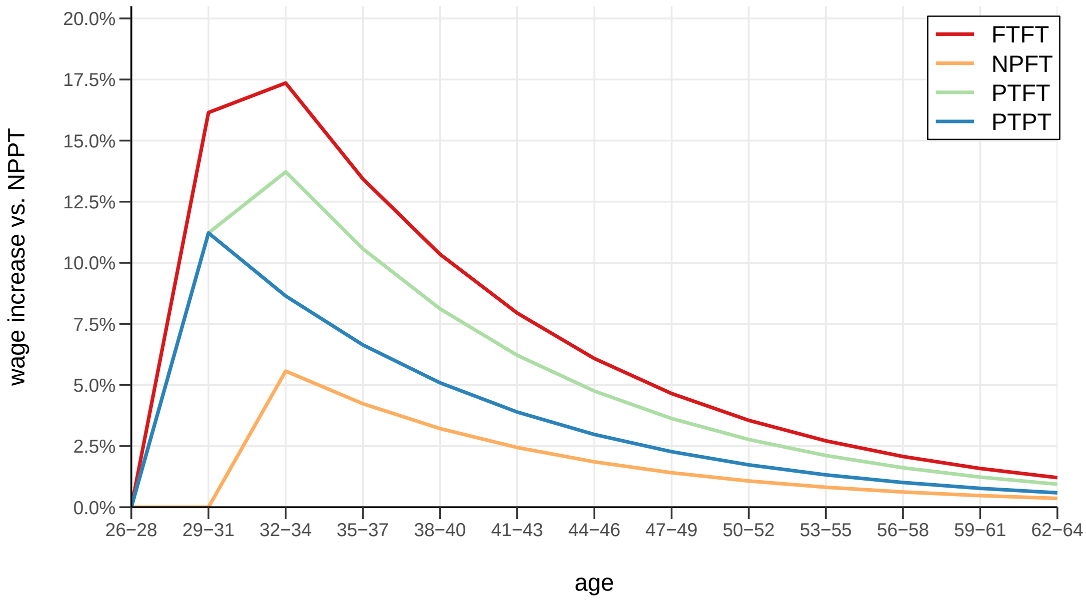


Figure 5: Illustration of the wage process dynamics

Notes: Relative increase in wages of different labour supply patterns, always compared to not working at age 26 – 28 and working part-time at age 29 – 31. NP, PT, and FT denote not working, part-time work, and full-time work, respectively. PTFT denotes part-time work at age 26 – 28 and full-time work at age 29 – 31. All other patterns are defined analogous. Simulations based on female wage process estimates from Appendix-Table C.2.

works part-time when the child is 3 – 5, and works full-time afterwards. The graph illustrates the dynamic wage gains that the mother would obtain if she increased her labour supply. The blue line shows the case when the mother already starts working part-time when the child is 0 – 2. The red line shows the case where the mother switches to full-time work both when the child is 0 – 2 and also when the child is 3 – 5.

Asides from the substantial wage gains from increasing labour supply, the graph clearly illustrates that the potential wage gains are quite persistent. This will play a key role for the dynamic fiscal effects that we present later: if public childcare allows a mother to increase her labour supply, this does not only affect her tax payments while the child is young, but also her tax payments in the future.[25]

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

Finally, we also estimate the male wage process in a similar fashion (see Appendix-Table C.2 for details). Since we only consider males that always work full-time, the wage equation does not contain part-time or non-employment penalties.

4.2 Structural estimation

4.2.1 Calibration of homogeneous preference parameters

In line with Blundell et al. (2016), we set the discount factor β to 0.94 per 3-year model period.[26] We calibrate the homogeneous parameters of the utility function in equation (1) to match, jointly with the estimated distribution of unobserved heterogeneity, data moments of labour supply and public childcare take-up as well as benchmark labour supply elasticities from the literature. We discuss this further in Section 5, where we present the model fit.

Table 3 provides an overview of all calibrated parameters. We use a log specification for consumption ($\gamma_c = 1$) and set the CRRA coefficient on leisure (γ_L) to 2 and on home-produced childcare (γ_{dcc}) to 1.125. The leisure endowment $\overline{L}$ is calibrated to 1 and the home-produced childcare endowment $\overline{dcc}$ will be set to 4, if children below 9 are present. Furthermore, the shifter for the preference for home-produced childcare κ is 0.075. It implies that households have a much stronger preference to spend time with children below age 3 compared to 3 to 8 year old children. This allows us to capture social norms in a reduced form way: many women spend the first years after birth out of the job caring for their children.[27]

Table 3: Calibrated homogeneous model parameters

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

4.2.2 Maximum likelihood estimation of heterogeneous preferences

In our model, the preference for home-produced childcare g, the availability of informal childcare oth, and the taste for female leisure α are driving forces of female labour supply and childcare

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

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

decisions during parenthood. Their unobserved joint distribution determines how female labour supply and therefore women's career paths react to changes in childcare policies. In this subsection, we present our estimation to identify this joint distribution of permanent unobserved heterogeneity.

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

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

The time-varying state variables are summarized in the vector s and are also observed over two model periods. s includes the wages of the male and female spouse (w_m and w_f), the family composition K, and age t.[28] The unobserved heterogeneity $h = (g, oth, \alpha)$ and the constant characteristics x, on the other hand, are assumed to be constant. Given s and h, our model predicts female labour supply and total public childcare for each period p in a deterministic manner. Let us denote these predicted choices as $\widehat{lm_p}$ and $\widehat{Tcc_p}$. The likelihood of the unobservable 'true' choices (lm_p, Tcc_p) being predicted for a household with characteristics (s_p, h) is thus as shown in equation (10) below. The terminology 'true' choices is used to distinguish these from the choices observed in the data, which include measurement errors.

$$\ell(lm_p, Tcc_p | s_p, h) = \begin{cases} 1 & \textit{iff} \quad \widehat{lm_p}(s_p, h) = lm_p \text{ and } \widehat{Tcc_p}(s_p, h) = Tcc_p, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

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

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

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

$$\ell(\widetilde{w_{p,q}}|w_{p,q}) = \mathcal{F}(\log(\widetilde{w_{p,q}}) - \log(w_{p,q})) \text{ where } |\epsilon_{p,q}| = |\log(\widetilde{w_{p,q}}) - \log(w_{p,q})| \sim EV_{\text{II}}, \tag{11}$$

$$\ell(\widetilde{Tcc_p}|Tcc_p) = \digamma(\widetilde{Tcc_p} - Tcc_p) \text{ where } |u_p| = |\widetilde{Tcc_p} - Tcc_p| \sim EV_{\text{II}}, \tag{12}$$

where $p = 1, 2$ denotes the time period, $q = m, f$ denotes the spouse, and $(\epsilon_{p,q}, u_p)$ denote the measurement errors in wages and total public childcare, respectively. We omit the conditioning on the full set of time-varying characteristics s in (11) and (12), as well as in (13) below, to ease the notation, but these expressions should still be understood as conditional on it.

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

$$\ell(lm_p, \widetilde{Tcc_p}|\widetilde{w_p}, h) = \int\int\int \ell(lm_p, Tcc_p|w_p, h) \cdot \mathcal{F}(\epsilon_{p,m}) \cdot \mathcal{F}(\epsilon_{p,f}) \cdot \digamma(u_p)\ d\epsilon_{p,m} d\epsilon_{p,f} du_p. \tag{13}$$

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

$$\ell(lm, \widetilde{Tcc}|\widetilde{s}, h) = \prod_{p=1}^{2} \ell(lm_p, \widetilde{Tcc_p}|\widetilde{s_p}, h). \tag{14}$$

Our object of interest is the joint distribution of unobserved heterogeneity $\ell(h|x)$, which we aim to recover from the observed household choices. We estimate this distribution conditional on a set of constant household characteristics, denoted x, whose choice we discuss later. The

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

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

likelihood of observing a household's sequence of choices $(lm, \widetilde{Tcc})$ conditional on observed characteristics is given by the following expression:

$$\ell(lm, \widetilde{Tcc}|\widetilde{s}, x) = \int_{\boldsymbol{h}} \ell(lm, \widetilde{Tcc}|\widetilde{s}, h) \cdot \ell(h|x) \; dh. \tag{15}$$

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

$$\mathcal{L} = \prod_{n=1}^{N} \ell(lm^n, \widetilde{Tcc^n}|\widetilde{s^n}, x^n). \tag{16}$$

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

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

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

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

$$het^{\mu} = \gamma^{het} + \boldsymbol{x}^{het}\boldsymbol{\beta}^{het} + u^{het}, \tag{17}$$

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

$$het^{\mu}|\boldsymbol{x}^{het} \sim \mathcal{N}\left(\gamma^{het} + \boldsymbol{x}^{het}\boldsymbol{\beta}^{het}, \left(\sigma^{het}\right)^2\right). \tag{18}$$

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

Our maximum likelihood procedure will consequently estimate the parameters (γ^g, $\boldsymbol{\beta}^g$, σ^g, γ^{oth}, $\boldsymbol{\beta}^{oth}$, σ^{oth}, γ^{α}, $\boldsymbol{\beta}^{\alpha}$, σ^{α}) which maximize the sample likelihood function given in equation (16).[31]

Sample. We conduct the maximum likelihood estimation with data from the GSOEP covering 2012 to 2017 (see Section 2.2 for details). For every household, we convert the data from the six years into two corresponding model periods. Further details on this process can be found in Appendix C.3.1 and summary statistics on the sample are presented in Appendix-Table C.3.

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

Identification. In the absence of a formal proof, we provide an intuition for the identification of the time-invariant parameters that govern the joint distribution of unobserved heterogeneity. The identification is conditional on the calibrated and reduced form regression inputs (see Section 4.1), the homogeneous preference parameters (see Section 4.2.1), and the previously described assumptions for our maximum likelihood procedure. The distribution of $h = (g, oth, \alpha)$ will be jointly and set identified. There are three interacting ingredients that identify the time-invariant unobserved heterogeneity: i) cross-sectional variation in choices conditional on the same observed states, ii) the longitudinal dimension of our panel data, iii) using data not only from households with small children, but also from those with older children. The following paragraphs describe the three ingredients in more detail.

First, we observe households making different choices conditional on the same observed states s and constant characteristics x. Within our model, these differences in choices are therefore driven by differences in unobserved heterogeneity h. For illustration purposes, consider the

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

example of a household with a single child aged $0-2$ and a part-time working mother that buys 20 hours of public childcare. From this household's choices in isolation, oth is identified to be ≤ 0.5 (≤ 20 hours), as otherwise the household would buy less public childcare. Nonetheless, oth is only set identified: For a given preference for leisure α, the above choices could result from different combinations of g and oth. A low preference for home-produced childcare g relative to leisure α would be consistent with oth close to 20 hours, i.e. the mother consumes leisure and does not provide much home-produced childcare. On the contrary, a high preference for home-produced childcare g relative to leisure α would be consistent with oth close to 0 hours, i.e. the mother spends a lot of time on home-produced childcare and little on leisure.

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

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

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

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

Constant characteristics x. The fact that we allow the joint distribution of unobserved heterogeneity h to differ by constant characteristics x is important in multiple ways: The characteristics in x allow us to capture that subgroups in our data may have very different preferences, thereby improving the capability of our model to predict the behavioural patterns in the data. In addition, we select constant characteristics x to address the initial conditions problem, i.e. that the time-invariant joint distribution of unobserved heterogeneity might have affected the initial values of our time-varying state variables. We include all covariates in x as indicator variables. Thereby, they act as mean shifters in the data-generating process of each type of heterogeneity as introduced in equation (17). Summary statistics of the variables in x can be found in Appendix-Table C.3.

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

Second, we estimate the distribution of the availability and preference for informal childcare ℓ^{oth} separately for East and West Germany. Social norms about childcare differ significantly between both regions, which likely affects the availability of grandparental childcare.[33] Therefore

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

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

and for computational reasons, we only estimate the mean of ℓ^{oth} for East Germany and fix the standard deviation to 1, as this allows for greater numerical stability in the face of very little variation.[34] For households living in West Germany, we do estimate the standard deviation and additionally include maternal education and whether the household lives in an urban area in $\boldsymbol{x}^{oth}$.

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

Results. The estimated coefficients from our maximum likelihood estimation (MLE) are shown in Table 4. Figure 6 shows the implied cumulative distribution functions of g, oth, and α. For details on the optimisation routine and on the sensitivity of the estimates, see Appendix C.3.

Table 4: Maximum likelihood estimates

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

Notes: 'east' indicates having lived in former East Germany at some point; 'demanding occupation' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine; 'catholic' indicates being Catholic at age 20; 'high education' indicates having obtained at least an A-level; 'urban' indicates living predominantly in an urban area. See Appendix C.3.2 for details on the optimization and Appendix C.3.3 for an illustration of the sensitivity of the estimates. † fixed to 1 for computational reasons.

First, considering the preference for home-produced childcare (g), Table 4 shows that women who live in former East Germany have lower preferences for home-produced childcare than those

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

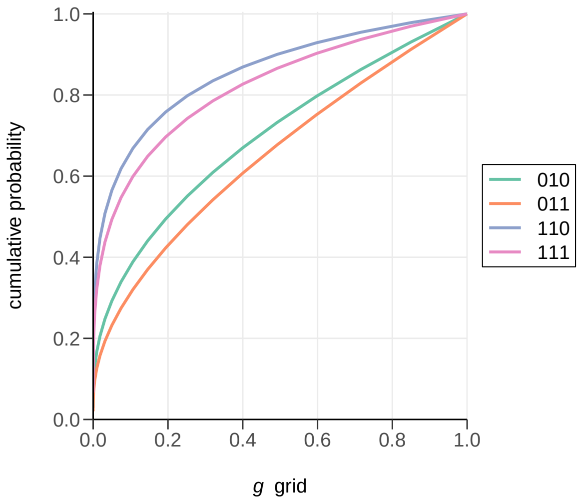
in former West Germany, as reflected by $\beta_{\text{east}<0}$. Additionally, if the mother was Catholic at age 20, we observe a higher preference for home-produced childcare, in line with the prior that Catholics have more conservative views regarding childcare. This is also visible in Figure 6a: the cumulative distribution function of Catholic mothers living in West Germany is the orange line. It features the lowest mass at low values of g, i.e. on a low preference for home-produced childcare. Overall, the distinction between East and West Germany is the more important factor and both distributions for West Germany first-order stochastically dominate the East distributions.

Second, we find large differences in the distribution of the availability of informal childcare (oth) between East and West Germany. As visible in Figure 6b, only very few households in East Germany rely on informal childcare (light green line). Generally, the large differences between East and West Germany point to cultural differences in the use (and availability) of grandparental and other informal childcare. Focusing on West Germany, the coefficients in Table 4 show that lower educated mothers as well as those not living in urban areas have a higher availability of informal childcare. Hence, as also illustrated in Figure 6b, households with highly educated mothers living in urban areas rely only little on informal childcare.

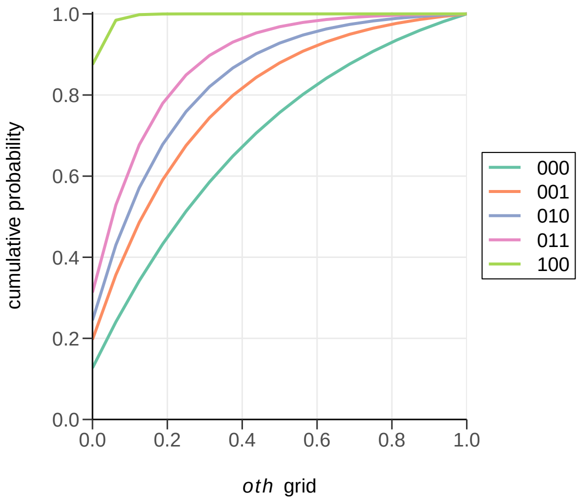
Third, we turn to the marginal distribution of the preference for leisure (α), which we estimate conditional on maternal education and occupational characteristics. Both covariates matter almost equally, as shown in Table 4: being highly educated or having primarily worked in a demanding occupation both imply a lower preference for leisure. The resulting cumulative distribution functions in Figure 6c reflect this as well. Those who are highly educated and have worked primarily in a demanding occupation have the highest mass at low values of α.

5 Model Fit

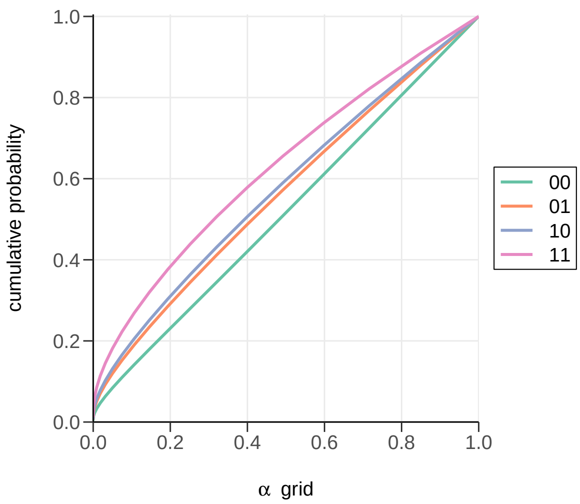
Based on the just presented estimated joint distribution of unobserved heterogeneity and the calibrated parameters of the utility function, we now turn to the model fit. Specifically, we evaluate the ability of the estimated structural model to match data moments for the two choices, female labour supply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



(a) home-produced childcare preference
covariates: east, demanding occupation, catholic



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



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

Figure 6: Marginal cumulative distribution functions of the unobserved heterogeneity

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

First, Table 5 shows the model and data moments for female labour supply by the age of the youngest child. Overall, we are able to achieve a good fit of the labour supply patterns of mothers conditional on child age. In particular, the model matches the observed increase in participation and hours once the youngest child turns 3. Our model is also able to match the labour supply pattern of mothers with completed fertility, i.e. children aged 9 or older. Splitting up the fit by the hourly wage of the mother and father in Appendix-Figures D.8a and D.8b affirms that the conclusions above also hold across the wage distribution.

Table 5: Model fit for labour supply

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

Notes: PT and FT denote the female working part-time and full-time, respectively. NP denotes not working. Sample as defined in Section 4.2.2. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

Next, we evaluate the fit with respect to the total public childcare take-up of households by the age of their youngest child. Families with the same age of the youngest child might face a different total childcare need as they might or might not have older children. To take this into account, we normalise each household's total public childcare take-up by the household's total childcare need. This yields the share of childcare need that a household covers with public childcare, which we denote as $m(Tcc)$.[35]

Table 6 summarises three public childcare take-up moments by the youngest child's age: i) the share of households who cover less than 33% of their total childcare need with public childcare ($m(Tcc) \leq \overline{m}_1$), ii) those who cover between 33% and 75% with public childcare ($\overline{m}_1 < m(Tcc) \leq \overline{m}_2$), and iii) those who cover more than 75% with public childcare ($\overline{m}_2 < m(Tcc)$). In summary, we achieve a good fit for all child age brackets. The fit is even almost perfect for children between 3 and 5. Again, splitting up the fit by the hourly wage of the mother and

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

father in Appendix-Figures D.9a and D.9b affirms that we match the overall dynamics across the wage distribution.

Table 6: Model fit for total public childcare take-up

	Children 0 – 2		
	$m(Tcc) \leq \overline{m}_1$	$\overline{m}_1 < m(Tcc) \leq \overline{m}_2$	$\overline{m}_2 < m(Tcc)$
Model	0.48	0.32	0.20
Data	0.44	0.40	0.16
	Children 3 – 5		
	$m(Tcc) \leq \overline{m}_1$	$\overline{m}_1 < m(Tcc) \leq \overline{m}_2$	$\overline{m}_2 < m(Tcc)$
Model	0.00	0.53	0.47
Data	0.02	0.52	0.46
	Children 6 – 8		
	$m(Tcc) \leq \overline{m}_1$	$\overline{m}_1 < m(Tcc) \leq \overline{m}_2$	$\overline{m}_2 < m(Tcc)$
Model	0.33	0.17	0.50
Data	0.29	0.23	0.48

Notes: $m(Tcc)$ denotes the share of total childcare need that is covered through public childcare. $\overline{m}_1$, respectively $\overline{m}_2$, indicates that the household would cover a share of 33%, respectively 75%. Sample as defined in Section 4.2.2. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

Third, we evaluate two labour supply elasticities: the participation elasticity and the Hicksian (compensated) elasticity of total working hours. Getting the female labour supply responses right is crucial for the credibility of our policy experiments in Section 6. To obtain the model counterparts for the empirical estimates, we simulate a permanent increase in the gross female wage rate by 1% and calculate the labour supply responses along the participation and total hours margin. The resulting participation elasticity is 0.16, which is within the range of estimates for women in the quasi-experimental literature.[36] In terms of the Hicksian (compensated) elasticity of total working hours, our model prediction is 0.32. This estimate is smaller than the meta-study aggregate of 0.59 reported in Chetty, Guren, et al. (2011), but the aggregate stems from a wide range of estimates of which some are well in line with our results.[37]

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

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

6 Policy Experiments

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

We start by evaluating the net fiscal effects of changes in childcare policies. The idea is that some (or all) of the initial cost of subsidizing public childcare is offset by additional revenue from income tax payments of mothers who decide to work more as a result of the policy ('marginal' mothers). Some of this additional income tax revenue will be perceived in the current period. Thereafter, some more revenue will be generated in the next periods since our model includes a positive return on labour supply in terms of earnings potential and a positive impact of earnings potential on the incentives to work.

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

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

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

i) We focus on the amount of subsidy spent and income taxes received in the model period of the policy introduction and call this perspective *Impact period*. This perspective includes both direct as well as anticipation effects, but no dynamic wage effects (as there are no past behavioural changes yet).

ii) We simulate the remaining life cycle starting from the observed age, wages, and family composition of the household and derive the discounted net fiscal effect. As opposed to i), this accounts for dynamic wage effects and we refer to this perspective as *All periods*. Note that this perspective includes the effects from the *Impact period* and also contains additional direct and anticipation effects from childbirths after the model period of the policy introduction.

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

6.1 Public childcare expansion for 0 – 2 year old children

As laid out in Section 2.1, the mid-2000s were characterized by substantial rationing of public public childcare slots for 0 – 2 year old children in former West Germany.[38] A large expansion policy was started in 2005 and effectively ended the rationing by approximately 2012.[39] To evaluate this policy, we limit the sample to West Germany and introduce in our model that a share of households does not have access to any public childcare for 0 – 2 year olds. Comparing this counterfactual to the non-rationed observed state of the economy therefore allows us to calculate the self-financing degree of the public childcare expansion.

[38]There was no rationing in former East Germany due to persistent institutional structures from the socialist era, see Section 2.1 for details.

[39]See Appendix B.3 for a discussion of the remaining gap between childcare demand and childcare enrolment for 0 – 2 year olds.

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

This implies that, when public provision of childcare was rationed, a sizeable share of mothers of 0 – 2 year olds were relying on informal childcare, as no private market for childcare existed.[40] When more public childcare slots became available as rationing came to an end, many mothers started sending their children to childcare facilities instead of relying on informal care arrangements. These mothers therefore did not contribute additional tax revenue since they were already working, but still increased their take-up of subsidized public childcare. This leads to a lower degree of self-financing of the policy compared to the case if we were to assume that only our estimated (low) levels of informal childcare were available.

However, the switch from informal to public childcare just described is a transition that goes beyond the capabilities of our model, which is best suited to capture the current institutional environment.[41] Nevertheless, we believe that it is insightful to study the ending of rationing with a slightly modified version of our model. This perspective allows us to contrast the effects of expanding access to public childcare with the effects of changes in the current fee schedule, which we focus on in Sections 6.2 and 6.3.

Therefore, we add the following component to our model: Under rationing, each mother has access to a fallback option for childcare, additionally to the estimated *oth* distribution. This fallback childcare is free and uniformly distributed, but is fully crowded out once public childcare is available.[42] Thereby, it allows mothers to work in the rationed environment, but it does not play a role for the decision making in the non-rationed environment.

Within this modified framework, we calibrate the availability of fallback childcare to 17.5h/week and the share of households without access to public childcare to 85%. This

[40]See Section 2.1 for details on the German childcare institutions.

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

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

allows us to closely match the 2005 employment and childcare enrolment rates.[43] Adding the fallback childcare component therefore represents a minimally invasive extension to be able to capture the observed transition between modes of childcare from a heavily rationed to a non-rationed environment. The simplistic extension thereby achieves our main goal of avoiding an overestimation of the fiscal effect of the expansion due to an overestimation of the labour supply effects. Furthermore, the implied increases in the employment rate (+15pp) and public childcare use (+33pp) from the rationed to the non-rationed scenario are also consistent with quasi-experimental evidence: we find that about 45% of mothers who start using public childcare also start to work. This matches well with quasi-experimental evidence from Bauernschuster and Schlotter (2015) for an earlier reform.[44]

Table 7: Self-financing degree of the public childcare expansion

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

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

Results. The main results on the self-financing degree of the public childcare expansion are presented in the first column of Table 7. For the *Impact period*, the increased tax revenue from mothers increasing their labour supply due to the availability of public childcare already makes up for 71.1% of the increased spending on childcare subsidies. Considering *All periods*, the self-financing degree increases to 103.3%, implying that the public childcare expansion for 0 – 2 year olds was fully self-financing when taking into account the effects from future periods. Given that 55% of mothers increased their public childcare usage without reacting on the labour

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

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

supply margin, this implies that the average mother who started to work paid taxes that are around twice the average subsidy paid per childcare slot.

Splitting up the effects by initial female wages, the additional columns in Table 7 reveal substantial effect heterogeneity across the wage distribution. Focusing on mothers with wages of ≤€15 per hour, the expansion of childcare was only 38.2% self-financing in the *Impact period*. In comparison, it was already 82.3% self-financing for those who earn around €20 per hour and even 170.1% self-financing for mothers with wages above €25 per hour. These effects increase to 63.9%, 120.9%, and 212.5%, respectively, for *All periods*, illustrating that even for those earning a wage of around €20, the increased government revenue from taxes recoups more than 100% of the spending on subsidies.

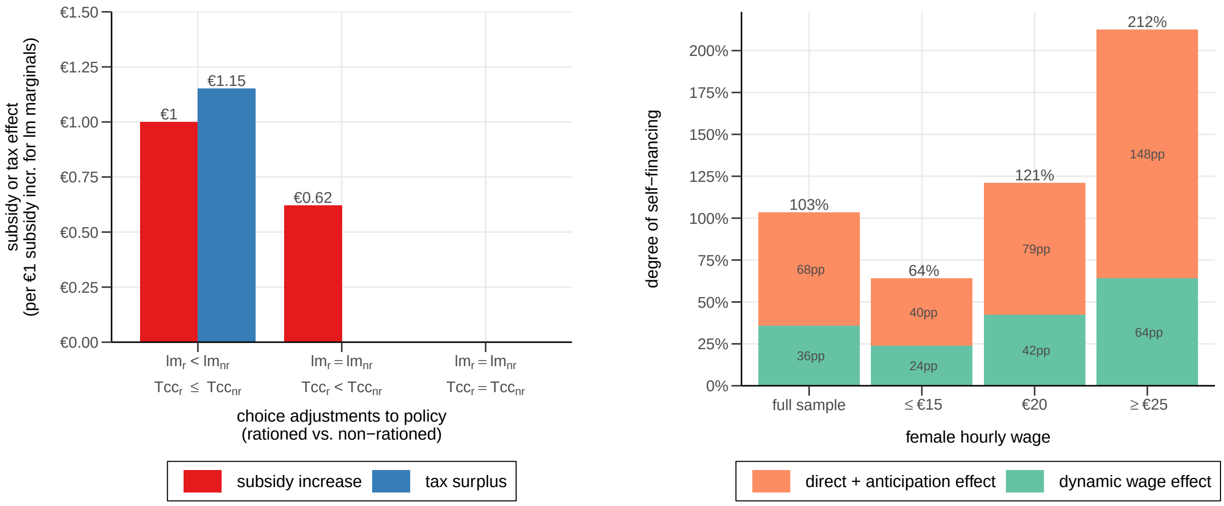
In Figure 7b we present a decomposition based on the three types of effects introduced at the beginning of this section, namely the direct and anticipation effect versus the dynamic wage effect. Based on the *All periods* perspective, we capture the dynamic wage effect as the share of the total self-financing degree that can be attributed to changes in the wage distribution. Throughout the wage distribution, Figure 7b illustrates that the dynamic wage effect makes up around one-third of the total net fiscal effect of the public childcare expansion. Overall, the childcare expansion only becomes fully self-financing once dynamic wage effects are taken into account.

These results also underscore that expanding public childcare generates high fiscal returns throughout the wage distribution and that the effects are not just driven by high wage individuals. This affirms that our main conclusions on the fiscal effects of the expansion are unlikely to be driven by the specific assumptions of our model extension to incorporate rationing.[45]

To better understand the underlying drivers of the self-financing rates, Figure 7a illustrates the contributions of different groups of individuals to the 71.1% *Impact period* self-financing degree. While some mothers do not respond to the policy, others change labour supply and use of public childcare. We refer to the group who changes their labour supply as 'marginal' mothers. Those mothers who use public childcare and thereby consume subsidies but do not change their labour supply due to the reform are referred to as 'inframarginal'.

The bars in Figure 7a show the magnitude of the increase in childcare subsidy spending and income tax revenue for each subgroup, normalized in relation to a one Euro subsidy increase

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



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

Figure 7: Results decompositions for the public childcare expansion

Notes: Decomposition of the effects from Table 7. lm_r and Tcc_r denote the labour supply and public childcare decision in the rationing scenario. lm_{nr} and Tcc_{nr} denote the labour supply and public childcare decision in the non-rationing scenario. *All periods* decomposed into: i) *direct effects* + *anticipation effects*, fiscal effects from behavioural adjustments in response to changes in the public childcare availability or fees while the household has children/may have children in the future in a policy affected age range. ii) *dynamic wage effects*, fiscal effects from higher wages and behavioural adjustments in response to higher wages, where higher wages are the result of past behavioural changes. Decomposition by initial wages at policy introduction. Values within the bars in Figure 7b denote percentage point contributions.

for those who are marginal with respect to their labour supply (red bar on the left). Focusing on the marginal mothers (blue bar, $lm_r < lm_{nr}$), the policy generated €1.15 in tax surplus per Euro spent on subsidies, which makes it clearly self-financing for this subgroup even when focusing only on the *Impact period*. However, the overall self-financing degree is negatively impacted by the subgroup which only increases their public childcare take-up, but does not increase their labour supply ($lm_r = lm_{nr}$, $mcc_r < mcc_{nr}$). For every Euro spent on subsidies for the marginals, the government has to spend 62 cents on subsidies for the inframarginals, which leads to the *Impact period* self-financing degree of 71.1%.

6.2 Childcare subsidy increases

After evaluating the expansion of public childcare slots for a given degree of subsidization of the unit price of childcare, we now turn to quantifying the fiscal externalities of small increases in the subsidy. This exercise investigates to which degree self-financing attribute of childcare subsidies holds for increases from the current level, keeping in mind that in the status quo childcare is already heavily subsidized (see Section 4.1.1).

Table 8: Self-financing degree of childcare subsidy increases

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

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

Untargeted childcare subsidy increase. We first study an untargeted increase in the subsidization of the hourly childcare fee that is equivalent to increasing monthly subsidies for full-time public childcare (40h/week) by €50. Exemplified for a household with only a $0-2$ year old and gross earnings of €$4,000$, this translates into a 21.5% decrease in the hourly childcare fee from €1.4 to €1.1.[46]

Panel (a) in Table 8 shows that for such an untargeted increase in childcare subsidies increased tax revenues only make up for 5.9% (4.1%) of the increased subsidy spending considering *All periods* (*Impact period*). We find the fiscal effects to be similar across both time perspectives for all three wage levels and illustrate in Appendix-Figure E.11a that the dynamic effects contribute approximately one-third.[47]

[46] Plugging in €$4,000$ and 40h of childcare need for a $0-2$ year old into the childcare fee schedule from Appendix-Table C.1.

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

The underlying driver for the low self-financing rate is the very high share of inframarginal mothers, who receive a windfall gain through higher subsidies but whose labour supply is unaffected.[48] Decomposing the *Impact period* effect by the choice margins in Appendix-Figure 7a, we find that for every Euro spent on marginal mothers, the government has to pay €40.64 to inframarginal mothers. This is barely counterbalanced by €1.68 in additional tax revenue from marginal mothers who change their labour supply. The small and untargeted increase in the subsidy is therefore self-financing for those who are induced to increase their labour supply, but the substantial spending on inframarginals makes the policy very costly. The latter is also the main difference to the results for the public childcare expansion in Section 6.1: Figure 7a showed a rather low spending on inframarginals for the ending of rationing, which was a fully self-financing policy. Appendix-Figure E.10a, on the other hand, illustrates that the spending on inframarginals outweighs the spending on marginals by more than a factor 40 for an untargeted increase in subsidies from the current level.

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

Panel (b) in Table 8 shows that in the *Impact period* 8.7% of childcare subsidies are refinanced through additional income tax revenues. This number increases to 12.4%, when we simulate household behaviour over *All periods*, with dynamic effects contributing again around one-third (see Appendix-Figure E.11b). Splitting up the effect by female wage levels, the €20 category stands out slightly, with effects that are 3 to 4 percentage points higher than the other two wage levels.

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

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

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

Full-time-contingent childcare subsidy increase. Third, we study an increase in the subsidization of the hourly childcare fee targeted to households in which the mother works full-time (i.e. an increase equivalent to increasing monthly subsidies for full-time public childcare by €50). This policy effectively cuts out all households who may increase their use of public childcare to consume additional leisure and creates additional incentives to work full-time.

Panel (c) in Table 8 clearly shows the effectiveness of the targeting: The degree of self-financing increases to 34.1% when we consider the *Impact period* and even further to 50.1% for *All periods*. Comparing the effects across wage levels, the effects are largest at the medium wage level, reaching up to 54%, and the relative contribution of the dynamic effect is again around one-third (see Appendix-Figure E.11c).

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

6.3 Redistribution through childcare fees

The existing childcare fee schedule is increasing in household income (see Figure 3). This implies that subsidies are falling with income and that the childcare fee schedule effectively redistributes within the childcare using population. Based on this observation, a number of policy relevant questions arise: How progressive should a childcare fee schedule be? Is it too progressive in Germany? Or should it rather be more progressive since the current schedule still implies substantial subsidies for high-income households? These are generally thorny normative questions that require to trade-off the utility of high and low-income households. For such purposes, it is typically required to specify a social welfare function. A common alternative is to work with social marginal welfare weights (Saez and Stantcheva 2016), which capture how much the government values one more (marginal) Euro in the hands of different income groups. However, the choice of these welfare weights is again a normative question.

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

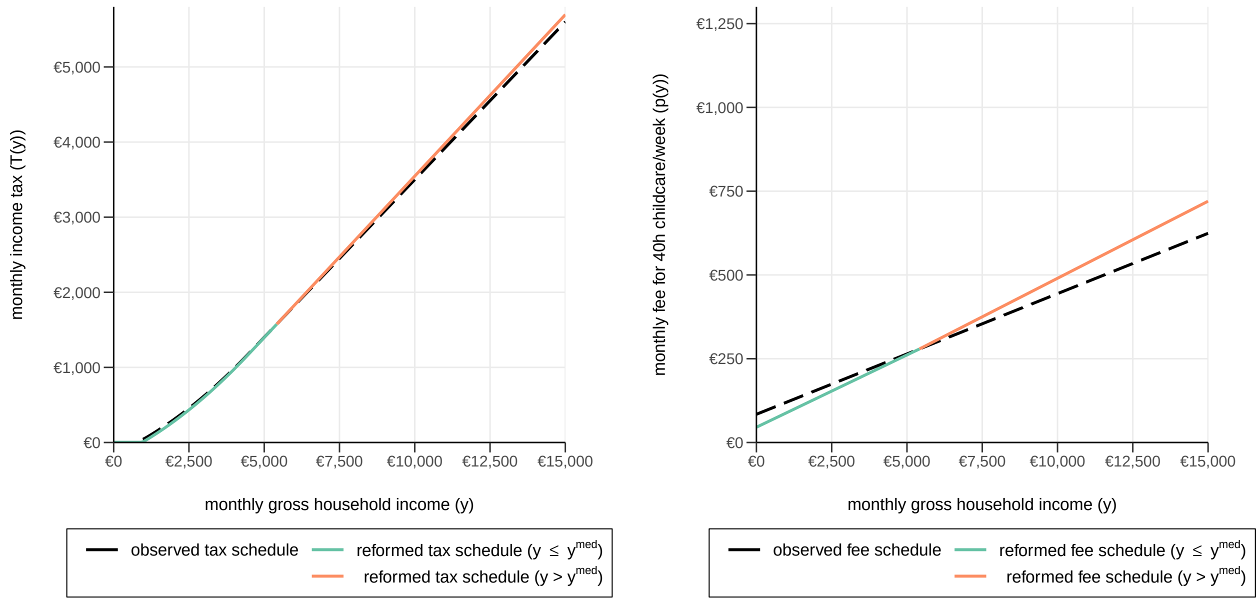
Most of the inverse-optimum literature focusses on rather simple static models and can therefore consider fully non-linear tax reforms in the spirit of Mirrlees (1971) to quantify the welfare weights for all income levels. In our large dynamic model, such a granular analysis is not tractable. We therefore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

6.3.1 Progressive reforms

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

$$T^*(y) = \begin{cases} T(y) + \tau_1^T \left(y - y^{med}\right) & \text{\textit{for} } y > y^{med} \\ T(y) - \tau_2^T(\tau_1^T) \left(y^{med} - y\right) & \text{\textit{for} } y \leq y^{med} \end{cases} \tag{19}$$

We fix τ_1^T, the tax increase for above-median income households ($y > y^{med}$), to 0.01 and choose τ_2^T, the tax decrease for below-median income households ($y \leq y^{med}$), such that the reform is government-budget neutral. We implement this parametric reform as a one-off reform: The reformed tax schedule is only applied in the *Impact period* and households are fully aware that the observed tax schedule will be applied in all future periods. To account for the dynamic effects of behavioural adjustments in response to the reform, we conduct all simulations until the end of the life cycle (*All periods* perspective). The resulting reformed income tax schedule is illustrated in Figure 8a.



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

Figure 8: Illustration of budget-neutral reforms to quantify the marginal excess burden

Notes: The income tax schedule representation in Figure 8a abstracts for illustration purposes from social security contributions and child benefits/allowances. For the childcare fee schedule in Figure 8b, we focus, for illustration purposes, on the case with a $0-2$ year old and no siblings. The budget-neutral parameter combinations are $\tau_1^T = 0.01$ and $\tau_2^T = 0.00744$ as well as $\tau_1^c = 0.01$ and $\tau_2^c = 0.00714$.

We then calculate the marginal excess burden of the tax schedule based on the reform. It quantifies the relation of the average inverse-optimum weights between both groups in the following intuitive fashion: For each Euro that the government takes from above-median income households, how much reaches those with below-median income? In the absence of behavioural responses, the full Euro would reach below-median income households. However, individual decisions are endogenous with respect to the reform. As the reformed tax schedule is steeper throughout (see Figure 8a), this implies lower labour supply incentives for all households. Consequently, households reduce their labour supply which lowers their tax payments at the margin. This leads to the marginal excess burden.

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

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

$$p^*(y) = \begin{cases} p(y) + \tau_1^c \left(y - y^{med}\right) & \textit{for } y > y^{med} \\ p(y) - \tau_2^c(\tau_1^c) \left(y^{med} - y\right) & \textit{for } y \leq y^{med} \end{cases} \tag{20}$$

We again set $\tau_1^c = 0.01$ and choose τ_2^c such that the reform is government-budget neutral. The resulting reformed childcare fee schedule, which is illustrated in Figure 8b, is also steeper throughout. Marginal childcare fees increase for all income groups, which lowers the incentives to work just as the tax reform did. However, note that Figure 8b also shows that childcare fees per hour (holding income constant) decrease for below-median income and increase above. As will become clear below, this implies more complex behavioural reactions than the tax reform.

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

6.3.2 Economic mechanisms

To illustrate the underlying mechanics that drive the differences in the marginal excess burdens, we start with a look at the *Impact period* perspective (different to the just described results). This perspective allows us to carefully decompose the effects by the changes in labour supply choices in Table 9, similar to Figures 7a and E.10. Afterwards, we show that accounting for the dynamic effects, as done above, is crucial as altered labour supply choices have sizeable long-run effects via their persistent impact on the wage distribution.

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

Table 9: Decomposition of the *Impact period* marginal excess burden

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

Notes: Behavioural reactions in response to budget-neutral reforms as described in Section 6.3.1, but simulated only for the model period of the reform (*Impact period*). Values in parentheses denote the contribution to the marginal excess burden.

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

To understand the intuition behind these results, first note that both reforms have in common that they increase the effective marginal tax rate on labour income: the tax reform increases the marginal tax rate $T'(y)$ and the childcare fee reform increases the marginal hourly price of childcare $p'(y)$ (see Figure 8). The increase in the effective marginal tax rate for a given amount of public childcare translates into decreased labour supply incentives for all income levels in both reforms. This effect in isolation leads to labour supply reductions and therefore losses in tax revenue. In the case of the tax schedule, it is the sole driver of the marginal excess burden, as illustrated in Table 9.

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

Our results in Table 9 show that the contribution of changed labour supply from below-median income households is much lower for the childcare fee reform compared to the tax reform (0.7 cents vs. 7.9 cents in Table 9). This illustrates that the two opposing labour supply effects for below-median income households almost cancel each other out. The revenue lost due to labour supply changes of households with above-median income, on the other hand, is very similar in both cases (around 12 cents): changes in the absolute price of childcare per hour do not play much of a role for above-median income households.[49] In summary, this shows that below-median income households are more likely to be marginal in their decision to work with respect to a change in the absolute price of childcare $p(y)$ than above-median households.

A final important difference between both schedules is that decreasing the absolute price of childcare per hour also affects the leisure vs. home-produced childcare trade-off. Leisure becomes more expensive relative to home-produced childcare for above-median income households and vice versa for below-median income households. As public childcare is heavily subsidized, any change its take-up also affects the government budget and thereby the efficiency cost of redistribution. Aggregating over above and below-median income, we observe that poorer households increase their Tcc take-up to consume leisure more than richer households decrease it, respectively. This yields the additional 4.4 cents in marginal excess burden from households with unchanged labour supply behaviour, shown at the bottom of Table 9.

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

Dynamic effects. Based on these decompositions, we now turn to the role of the dynamic wage effects for the marginal excess burden. Recall that the marginal excess burdens including dynamic effects are substantially larger than the *Impact period* ones shown in Table 9: 31.3 cents vs. 20.5 cents for the tax schedule and 21.2 cents vs. 17.2 cents for the childcare fee schedule. This highlights the importance of accounting for dynamic wage effects when evaluating the efficiency cost of redistribution of both schedules.

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

Furthermore, Table 9 illustrates that especially lower labour supply from below-median households is an important contributor in the case of the tax schedule, but not in the childcare fee schedule. As the dynamic wage effects are especially large for low-income households,[50] this is an additional contributor to the large dynamic component of the marginal excess burden of the tax schedule.

6.3.3 Policy implications.

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

Second, if the government has the same desire for redistribution in the childcare-using population as it has in the general population, then the childcare fee schedule should be more progressive. One potential reform would be to increase the progressivity of the childcare fee schedule using the simple parametric approach from equation (20). In Figure 9, we implement such a reform that increases the income dependency of the hourly price of public childcare. Specifically, we increase the income dependency at the margin from around 3% to around 10% by

[50] See the effect decompositions from the previous sections in Figures 7b and E.11.

setting $\tau_1^c = 0.07$ and redistribute the additional revenue via the corresponding budget-neutral τ_2^c .[51] The childcare fee schedule becomes considerably steeper, i.e. prices become considerably more progressive.[52] Nevertheless, the reformed schedule still implies a lower marginal excess burden of 29.6 cents than the tax schedule (31.3 cents). Hence, applying the inverse-optimum weights found for the tax schedule would suggest to implement this reform.

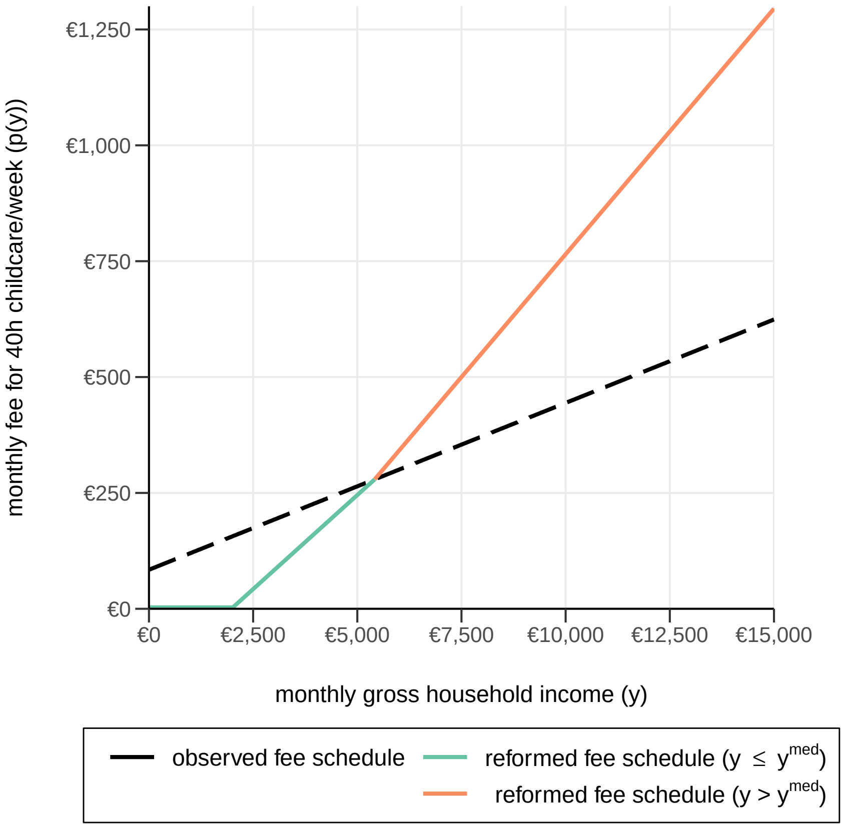


Figure 9: Reform of the childcare fee schedule

Notes: Parametric budget-neutral reform of the childcare fee schedule as set up in equation (20) with $\tau_1^c = 0.07$ and $\tau_2^c = 0.04518$. For illustration purposes, we focus on the case with a $0 - 2$ year old and no siblings.

[51]Under the observed childcare fee schedule, one marginal Euro in monthly gross household income increases hourly childcare fees by around 3 cents (see Appendix-Table C.1). With $\tau_1^c = 0.07$, one marginal Euro in monthly gross household income for above-median income households increases hourly childcare fees by around 10 cents.

[52]We do not allow for negative childcare fees and therefore cap them below at 0. We also cap them at the top to the actual cost of a slot (see Table 2).

7 Conclusion

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

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

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

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

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

References

Adda, J., C. Dustmann, and K. Stevens (2017). “The Career Costs of Children”. *Journal of Political Economy* 125.2, 293–337.

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*.

Baker, M., J. Gruber, and K. Milligan (2008). “Universal Child Care, Maternal Labor Supply, and Family Well-Being”. *Journal of Political Economy* 116.4, 709–745.

Baker, M., J. Gruber, and K. Milligan (2019). “The Long-Run Impacts of a Universal Child Care Program”. *American Economic Journal: Economic Policy* 11.3, 1–26.

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 (2020). “Child Care Subsidies, Quality, and Optimal Income Taxation”. *American Economic Journal: Economic Policy* 12.4, 1–37.

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.

Bettendorf, L. J., E. L. Jongen, and P. Muller (2015). “Childcare subsidies and labour supply? Evidence from a large Dutch reform”. *Labour Economics* 36, 112–123.

Bick, A. (2016). “The Quantitative Role of Child Care for Female Labor Force Participation and Fertility”. *Journal of the European Economic Association* 14.3, 639–668.

Bick, A., B. Brüggemann, N. Fuchs-Schündeln, and H. Paule-Paludkiewicz (2019). “Long-term changes in married Couples’ labor supply and taxes: Evidence from the US and Europe since the 1980s”. *Journal of International Economics* 118, 44–62.

Bick, A. and N. Fuchs-Schündeln (2018). “Taxation and labour supply of married couples across countries: A macroeconomic analysis”. *The Review of Economic Studies* 85.3, 1543–1576.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). “Female Labor Supply, Human Capital, and Welfare Reform”. *Econometrica* 84.5, 1705–1753.

BMFSFJ - Bundesministerium für Familie, Senioren, Frauen und Jugend (2016). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2015*.

BMFSFJ - Bundesministerium für Familie, Senioren, Frauen und Jugend (2020). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2019*.

Bourguignon, F. and A. Spadaro (2012). "Tax–benefit revealed social preferences". *The Journal of Economic Inequality* 10.1, 75–108.

Busse, A. and C. Gathmann (2020). "Free daycare policies, family choices and child development". *Journal of Economic Behavior & Organization* 179, 240–260.

Chetty, R. (2012). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80.3, 969–1018.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins". *American Economic Review* 101.3, 471–475.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2013). "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities". *NBER Macroeconomics Annual* 27, 1–56.

Colas, M., S. Findeisen, and D. Sachs (2021). "Optimal Need-Based Financial Aid". *Journal of Political Economy* 129.2, 492–533.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). "Who benefits from universal child care? Estimating marginal returns to early child care attendance". *Journal of Political Economy* 126.6, 2356–2409.

Dengler, K., B. Matthes, and W. Paulus (2014). "Occupational Tasks in the German Labour Market - An alternative measurement on the basis of an expert database". FDZ-Methodenreport 12/2014. Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Domeij, D. and P. Klein (2013). "Should Day Care be Subsidized?" *The Review of Economic Studies* 80.2, 568–595.

Eckhoff Andresen, M. and T. Havnes (2019). "Child care, parental labor supply and tax revenue". *Labour Economics* 61, 101762.

FDZ-DJI - Forschungsdatenzentrum Deutsches Jugendinstitut (2017). *DJI-Kinderbetreuungsstudie - KiBS*. DOI: 10.17621/kibs2017.

FDZ-SOEP - Forschungsdatenzentrum Sozioökonomisches Panel am Deutschen Institut für Wirtschaftsforschung (2019). *Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2017*. DOI: 10.5684/soep.v34.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2020a). *Mikrozensus 2014, On-Site-Zugang*. DOI: 10.21242/12211.2014.00.00.1.1.1.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2020b). *Mikrozensus 2018, On-Site-Zugang*. DOI: 10.21242/12211.2018.00.00.1.1.3.

Felfe, C. and R. Lalive (2018). "Does early child care affect children's development?" *Journal of Public Economics* 159, 33–53.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program". *Journal of Political Economy* 128.7, 2502–2541.

Gathmann, C. and B. Sass (2018). "Taxing Childcare: Effects on Childcare Choices, Family Labor Supply, and Children". *Journal of Labor Economics* 36.3, 665–709.

Geis, W. (2018). "Kinderbetreuung: Es fehlen immer noch fast 300.000 U3-Plätze". IW-Kurzbericht 11/2018. Institut der deutschen Wirtschaft.

Geis-Thöne, W. (2020). "Kinderbetreuung: Über 340.000 Plätze für unter Dreijährige fehlen". IW-Kurzbericht 96/2020. Institut der deutschen Wirtschaft.

GESIS - Leibniz-Institut für Sozialwissenschaften (2017). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2016*. DOI: 10.4232/1.12796.

Geyer, J., P. Haan, and K. Wrohlich (2015). "The effects of family policy on maternal labor supply: Combining evidence from a structural model and a quasi-experimental approach". *Labour Economics* 36, 84–98.

Givord, P. and C. Marbot (2015). "Does the cost of child care affect female labor market participation? An evaluation of a French reform of childcare subsidies". *Labour Economics* 36, 99–111.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). "The German Socio-Economic Panel (SOEP)". *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics* 239.2, 345–360.

Guner, N., R. Kaygusuz, and G. Ventura (2020). "Child-Related Transfers, Household Labour Supply, and Welfare". *The Review of Economic Studies* 87.5, 2290–2321.

Haan, P. and K. Wrohlich (2011). "Can child care policy encourage employment and fertility?: Evidence from a structural model". *Labour Economics* 18.4, 498–512.

Haeck, C., P. Lefebvre, and P. Merrigan (2015). "Canadian evidence on ten years of universal preschool policies: The good and the bad". *Labour Economics* 36 (Supplement C), 137–157.

Hank, K., K. Tillmann, and G. G. Wagner (2001). “Institutional child care in eastern Germany before and after Unification. A comparison with western Germany in the years 1990-1999”. *Zeitschrift für Bevölkerungswissenschaft* 26.1, 55–65.

Havnes, T. and M. Mogstad (2011). “Money for nothing? Universal child care and maternal employment”. *Journal of Public Economics* 95.11-12, 1455–1465.

Hendren, N. and B. Sprung-Keyser (2020). “A Unified Welfare Analysis of Government Policies*”. *The Quarterly Journal of Economics* 135.3, 1209–1318.

Ho, C. and N. Pavoni (2020). “Efficient Child Care Subsidies”. *American Economic Review* 110.1, 162–199.

Jacobs, B., E. L. Jongen, and F. T. Zoutman (2017). “Revealed social preferences of Dutch political parties”. *Journal of Public Economics* 156, 81–100.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019). “Child Penalties across Countries: Evidence and Explanations”. *AEA Papers and Proceedings* 109, 122–126.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2020). “Do family policies reduce gender inequality? evidence from 60 years of policy experimentation”. NBER Working Paper 28082. National Bureau of Economic Research.

Kleven, H., C. Landais, and J. E. Søgaard (2019). “Children and Gender Inequality: Evidence from Denmark”. *American Economic Journal: Applied Economics* 11.4, 181–209.

Kulms, J. and A. Nehls (2018). *Warum in Deutschland 300.000 Kitaplätze fehlen.* Deutschlandfunk. URL: `https://www.deutschlandfunk.de/kindertagesstaetten-warum-in-deutschland-300-000.724.de.html?dram:article_id=424386` (visited on 08/31/2021).

Laun, T. and J. Wallenius (2021). “Having It All? Employment, Earnings, and Children*”. *The Scandinavian Journal of Economics* 123.1, 353–381.

Lockwood, B. B. and M. Weinzierl (2016). “Positive and normative judgments implicit in U.S. tax policy, and the costs of unequal growth and recessions”. *Journal of Monetary Economics*. “Inequality, Institutions, and Redistribution”held at the Stern School of Business,New York University,April 24-25, 2015 77, 30–47.

Lorenz, N. and D. Sachs (2016). “Identifying Laffer Bounds: A Sufficient-Statistics Approach with an Application to Germany”. *The Scandinavian Journal of Economics* 118.4, 646–665.

Mirrlees, J. A. (1971). “An Exploration in the Theory of Optimum Income Taxation”. *The Review of Economic Studies* 38.2, 175–208.

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.

OECD (2017). *Pensions at a Glance 2017*.

Saez, E. and S. Stantcheva (2016). “Generalized Social Marginal Welfare Weights for Optimal Tax Theory”. *American Economic Review* 106.1, 24–45.

Semykina, A. and J. M. Wooldridge (2010). “Estimating panel data models in the presence of endogeneity and selection”. *Journal of Econometrics* 157.2, 375–380.

Statistisches Bundesamt (2012). *Finanzen der Kindertageseinrichtungen in freier Trägerschaft 2010*.

Turon, H. (2019). “Home production of childcare and labour supply decisions in a collective household model”. IZA Discussion Paper 12148. IZA Institute of Labor Economics.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors (2019). “SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python”. *arXiv e-prints* 1907.10121.

Wales, D. J. and J. P. K. Doye (1997). “Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms”. *The Journal of Physical Chemistry A* 101.28, 5111–5116.

Wang, H. (2019). “Fertility and Family Leave Policies in Germany: Optimal Policy Design in a Dynamic Framework”.

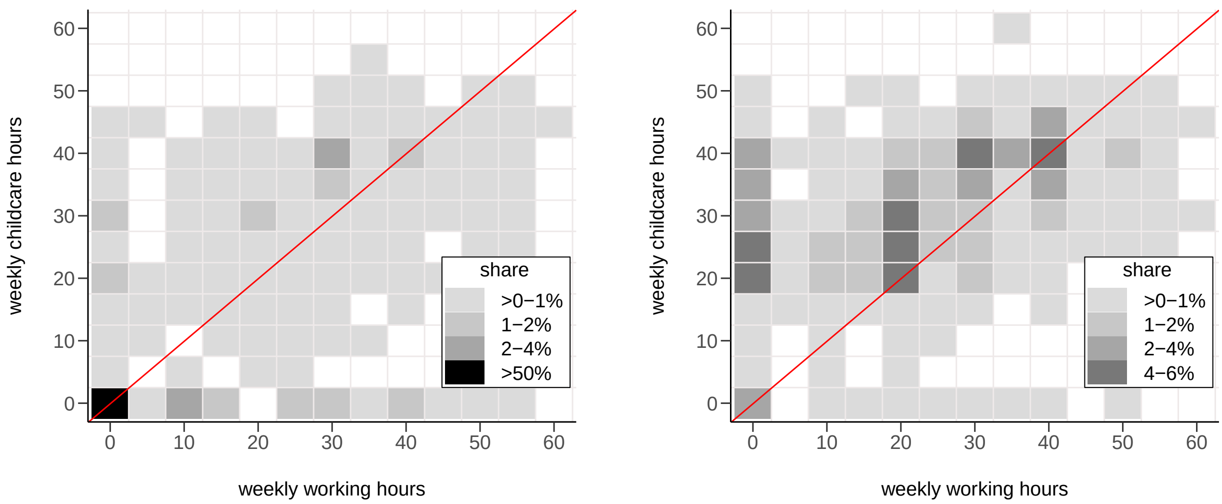
Appendix

A Details on the Estimation Sample

In addition to the sample restrictions described in Section 2.2, we condition on observing the following covariates for every female: hourly wage if working, hourly wage of the partner, education (A-level or not), religion (Catholic at age 20), state of residence, predominantly living in an urban or rural area, demanding occupation (having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine). Furthermore, we drop households that are either in the top 1% or bottom 1% of the male or female wage distribution to avoid distortions.

B Additional Stylised Facts

B.1 Childcare hours vs. working hours



(a) Children aged 0 – 2 (b) Children aged 3 – 5

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

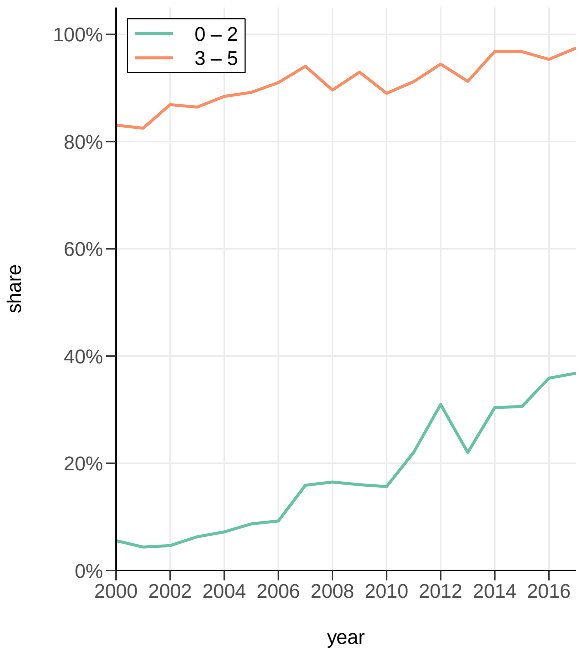
Notes: Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, conditional on having a child aged 0 – 2 for Figure B.1a or 3 – 5 for Figure B.1b. Source: 2009 to 2017 GSOEP, FDZ-SOEP (2019).

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

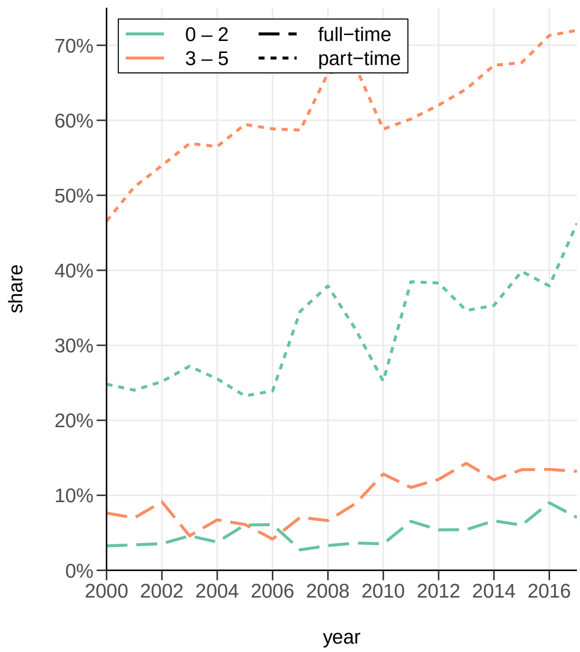
As Figure B.1a shows, more than half of the mothers of 0 – 2 year olds neither work nor send their children to public childcare. Of those who do work, not everybody sends their child to public childcare at the same time (mass at zero public childcare hours and positive working hours). Furthermore, we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

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

B.2 Developments in West Germany



(a) Public childcare enrolment



(b) Employment of mothers

Figure B.2: Development of public childcare enrolment and maternal employment in West Germany

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, conditional on having a child aged 0 – 2 or 3 – 5. Source: 2000 to 2017 GSOEP living in West Germany, FDZ-SOEP (2019).

B.3 Background on childcare demand and enrolment of 0 – 2 year olds

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

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

To understand the role that both explanations play for the gap between reported demand and observed take-up, we start by reconstructing the numbers from the government reports within a single data set, the DJI-KiBS panel.[53] This is different from the approach taken in the reports, which do use the same survey data to calculate the demand for public childcare, but then rely on administrative childcare enrolment data and birth records to calculate take-up. Instead, we calculate both demand and enrolment from the survey data to ensure that we accurately capture the gap within the sample at hand. In terms of the sample restrictions, we focus on children below age 3 and limit the analysis to married and cohabiting couples (as for the estimation).[54]

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

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

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

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

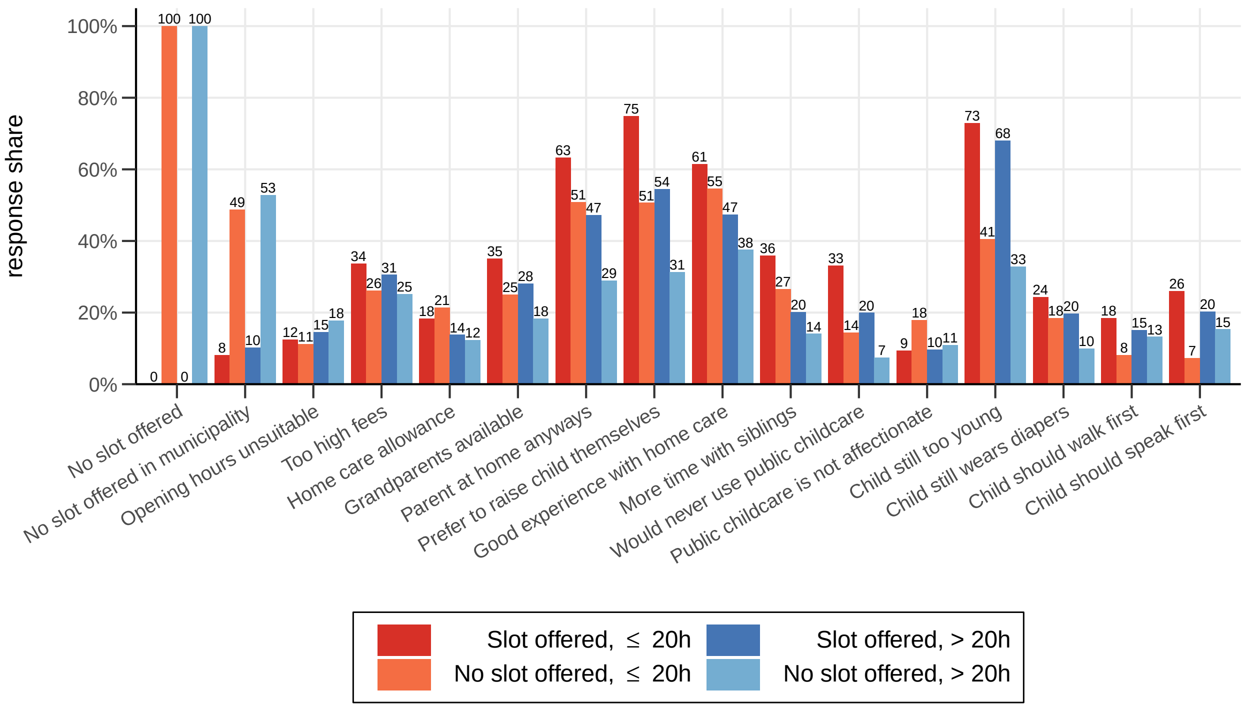


Figure B.3: Reasons for not using public childcare

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

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

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

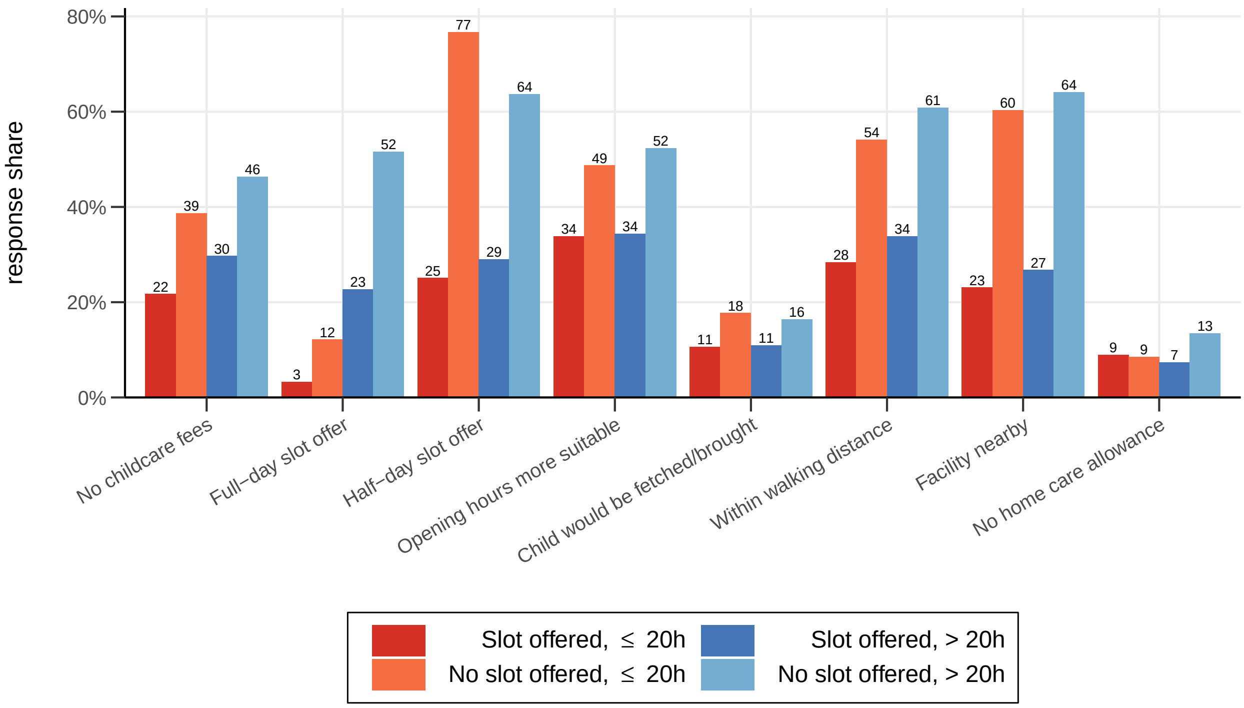


Figure B.4: Conditions under which public childcare would be used

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

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

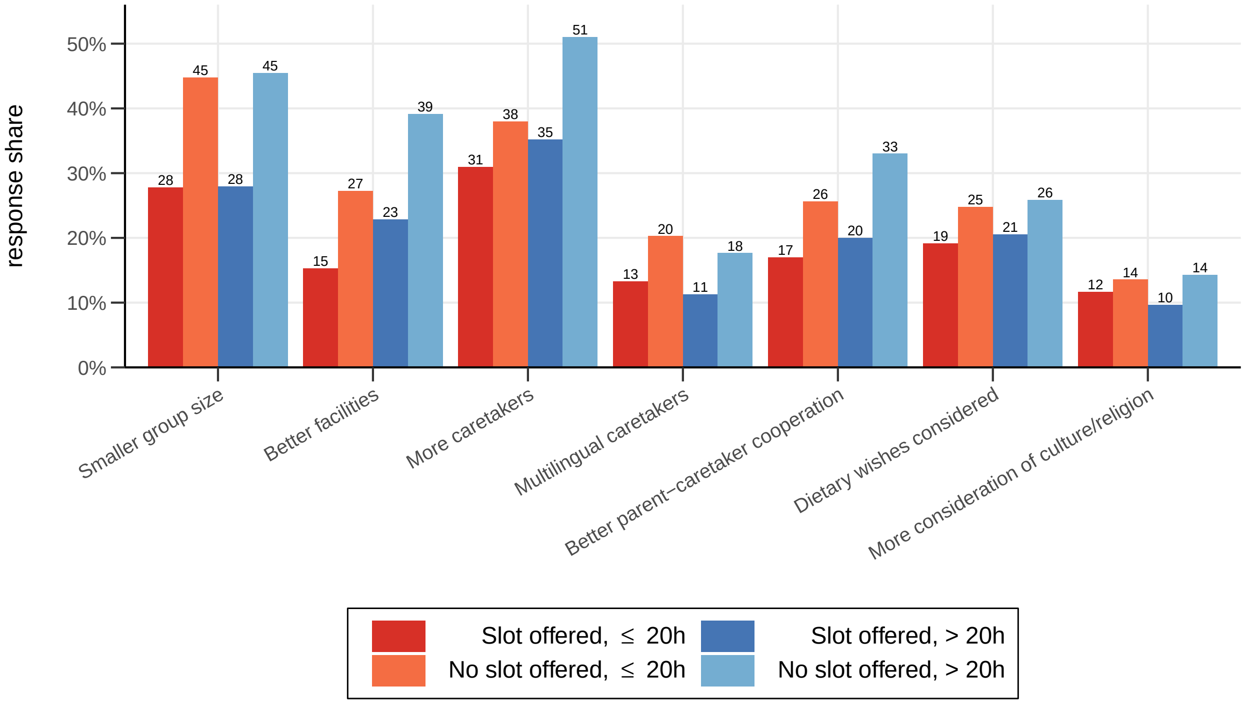


Figure B.5: Conditions under which public childcare would be used (quality dimensions)

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

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

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

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

C Additional Details on the Auxiliary Regressions and the MLE

C.1 Childcare fees

C.1.1 Determinants of childcare fees

Child age. One of the important determinants of childcare fees in Germany is the age of the child. This is due to the fact that children of different ages visit different childcare institutions, as laid out in Section 2.1. Fees are usually higher for younger children since the costs of operating nurseries are higher than those of kindergartens.

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

Further determinants of childcare fees. Despite their autonomy, different states define in their legislation vastly similar determinants of childcare fees besides child age:[55]

1. *Household income:* In 11 out of 16 states the household income has to be used as a determinant and in two additional states it can be used.[56]

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

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

[56] 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).

C.1.2 Estimation of the childcare fee schedule

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

$$\begin{aligned} p_{nt} = \; & \alpha + \beta_1 y_{nt} + \beta_2 \left(y_{nt} \times \mathbb{1}\left\{ \text{one sibling with age} < 17 \text{ in HH} \right\}_{nt} \right) \\ & + \beta_3 \left(y_{nt} \times \mathbb{1}\left\{ \text{two siblings with age} < 17 \text{ in HH} \right\}_{nt} \right) + \epsilon_{nt}. \end{aligned} \tag{21}$$

The dependent variable p_{nt} is the monthly fee that household n would pay for full-time childcare (40h/week) for a child aged j in year t. The interaction terms of gross household income with indicators for the number of siblings capture discounts granted to families with multiple children.[57] Our empirical model thereby closely reflects the current childcare fee schedule regulation as laid out in the previous section.

We estimate equation (21) as a Tobit regression with censoring at €0 and €725, the lowest and highest observed monthly childcare payments in our data.[58] We abstract from regional variation to keep the state space of the structural model tractable.

Results. The results of the Tobit regressions are summarised in Table C.1. Monthly childcare fees increase significantly in gross household income for all age brackets. Average fees are estimated to be highest for the youngest children, who require the most intensive care. The presence of siblings implies a significant reduction of the income gradient for 0 – 2 and 3 – 5 year olds, decreasing it by more than half if two siblings live in the household.

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

[58] 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 C.1: Tobit estimation of the childcare fee schedule

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

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

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

C.2 Details on the wage process

C.2.1 Potential wages for non-working females

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

$$log(w_{f,it}) = \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}, \tag{22}$$

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

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

$$\mathbf{Z}_{it}\boldsymbol{\zeta} + \nu_{it} > 0, \tag{23}$$

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

In line with the selection correction procedure proposed by Semykina and Wooldridge (2010), we run a Probit version of equation (23) for each time period. In these, we also include the individual specific means of all covariates in $\mathbf{Z}$ across 2000 to 2017, denoted by $\overline{\mathbf{Z}}$:

$$Pr(\text{participation}_i = 1) = \Phi\left(\mathbf{Z}_i\boldsymbol{\zeta} + \overline{\mathbf{Z}}_i\boldsymbol{\xi}\right). \tag{24}$$

After estimating (24) for each year, we obtain the inverse Mills ratios λ_{it}, which we then use as control functions in the selection corrected version of the wage equation (22):

$$log(w_{f,it}) = \mathbf{X}_{it}\boldsymbol{\rho} + \overline{\mathbf{Z}}_i\boldsymbol{\xi} + \gamma\lambda_{it} + u_{it}. \tag{25}$$

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

C.2.2 Details on the wage process estimation

Table C.2: Estimation of female and male wage dynamics

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

Notes: See Section 4.1.3 for the regression setup. 'educ' indicates having obtained at least an A-level, NP and PT denote not working and working part-time, respectively. Sample: mothers aged 20 to 65 that are not in education and live with a full-time working partner, non-participation wages imputed as described in Appendix C.2.1. Source: 2000 to 2017 GSOEP, FDZ-SOEP (2019).

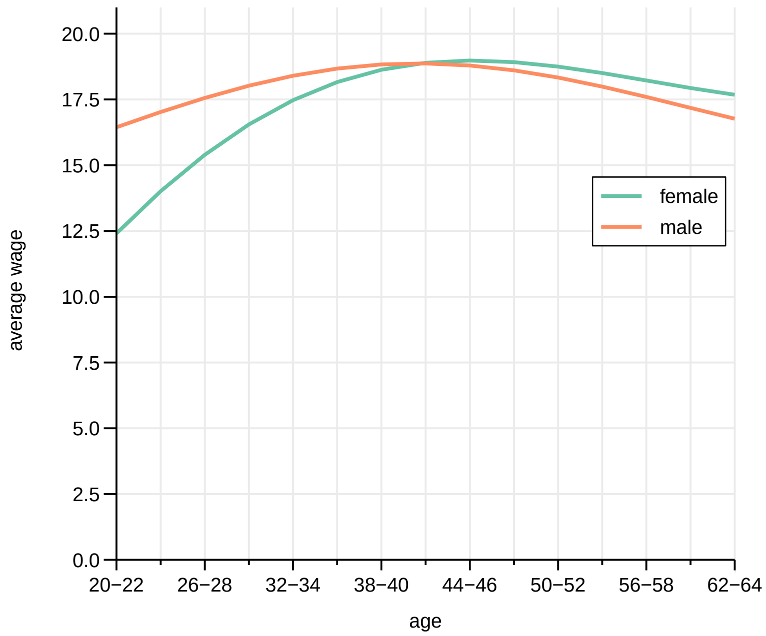


Figure C.6: Wage dynamics across the life cycle

Notes: Simulated average wages starting with the observed wage distributions for individuals without an A-Level aged 20 – 22 and assuming continuous full-time work for the remaining life cycle. Based on wage process estimates from Table C.2.

C.3 Additional details on the maximum likelihood estimation

C.3.1 MLE sample

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

Table C.3: Summary statistics for the MLE sample

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

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

C.3.2 Optimisation routine

To solve the optimisation problem numerically, we use the basin-hopping algorithm in combination with a Matlab built-in minimisation routine for constrained target functions (*fmincon*). The basin-hopping algorithm is a stochastic global optimisation algorithm used in various fields (Chemistry, Applied Mathematics, ...), which was first introduced by Wales and Doye (1997).[59]

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

Detailed description of the basin-hopping algorithm. As described above, the basin-hopping algorithm is a stochastic global optimisation algorithm. Starting from an arbitrary pre-specified initial parameter combination, it employs a minimisation routine (in our case Matlab's *fmincon*) to find a (possibly local) first minimum. This first minimum is the first candidate for the global minimum. Using the parameter combination of the minimum found after the first 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 minimum candidate, this parameter combination becomes the current global minimum candidate. The *step*-taking, which is discussed in detail below, is then repeated for either a pre-specified number of times or terminated if the global minimum candidate has not changed for a pre-specified

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

number of *steps*. After termination, the final global minimum 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. 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 step-size – which we set to 0.2 –, each of the parameters of the first minimum is separately perturbed by a random shock whose size is at most equal to the (initial) step-size. This implies that the perturbed version of each parameter is within a ± 0.2 interval around the respective parameter of the first minimum. After running the 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 minimum candidates, the algorithm retains some flexibility to explore the target function in the proximity of the current global minimum candidate.

The procedure furthermore includes an adaptive adjustment of the search radius (as determined by the size of the perturbations, i.e. the *step*-size). After every n-th *step*, the algorithm compares the rate at which *steps* are *accepted* to a target rate.[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. Consequently, the step-size is adjusted downwards by 10%, narrowing the search radius. If the *acceptance* rate is above the target rate, this implies that the perturbations are too small, i.e. the algorithm is likely stuck in a basin of attraction. Consequently, the step-size is adjusted upwards by 10%, widening the search radius to be able to escape the current basin of attraction and explore the target function outside of it.

[60] A *basin of attraction* is the set of possible starting points of a minimisation routine that leads to the same minimum (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.

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

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

C.3.3 Sensitivity of MLE results

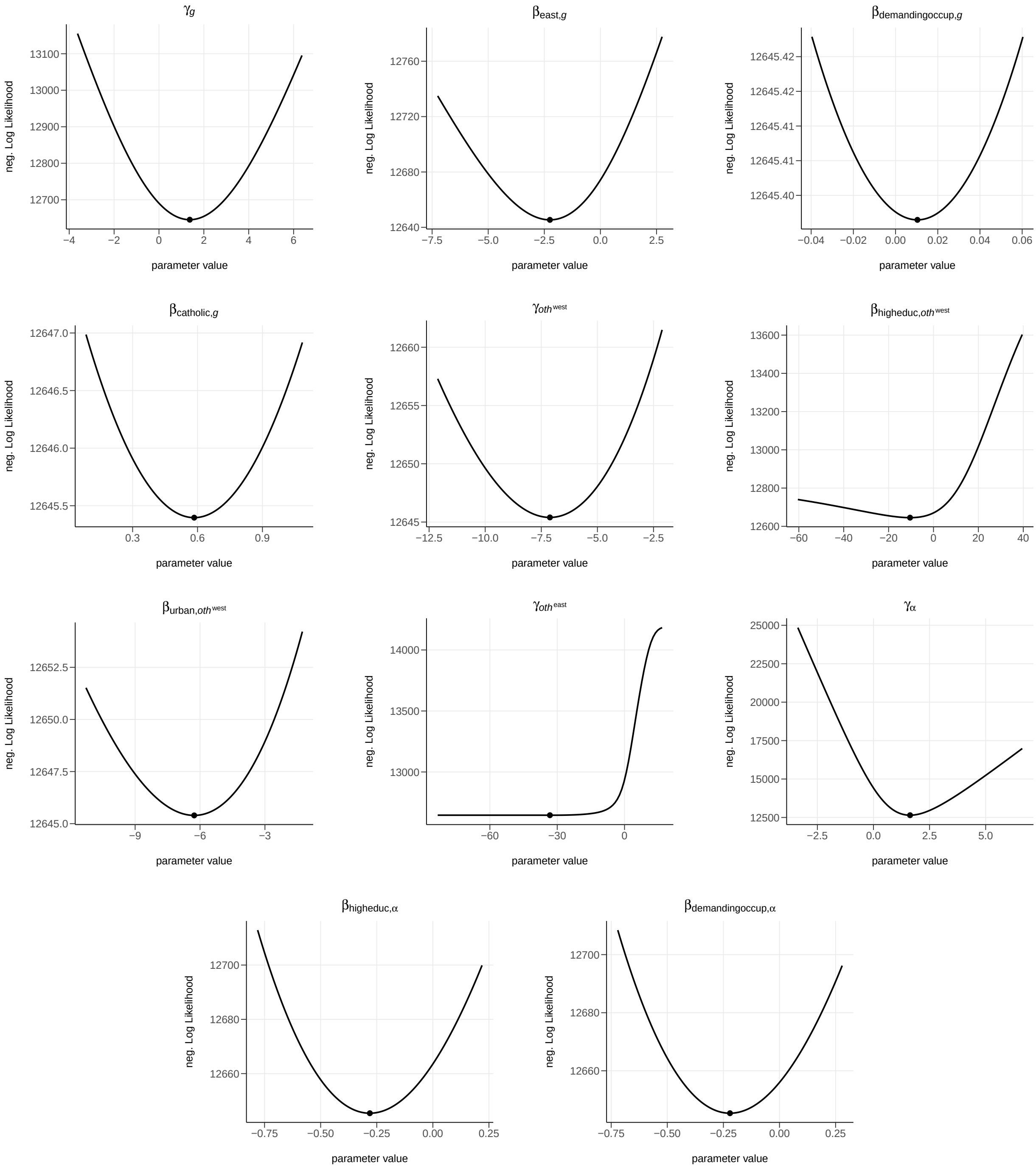


Figure C.7: Sensitivity of estimated structural parameters

Notes: Illustration of changes in the negative Log-Likelihood in response to small deviations of each parameter from its estimated value (labelled by black dot), keeping all other parameters at their point estimates. We constrain the estimation of σ_{oth} within the interval [0,2] to facilitate the computation, but the resulting CDF (as displayed in Figure 6) remains virtually identical if we allow for a larger interval. 'demanding occup' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine. 'high educ' indicates having obtained at least an A-level.

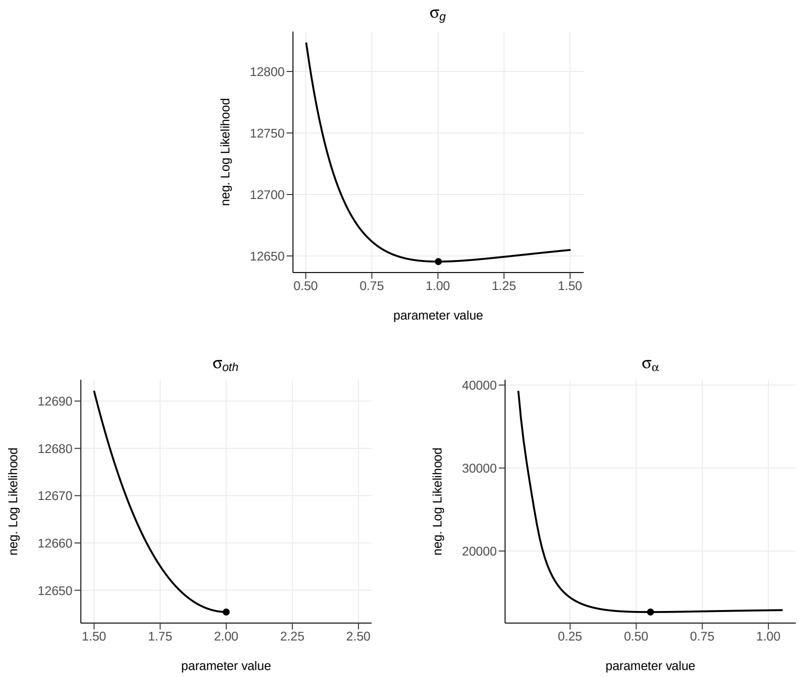
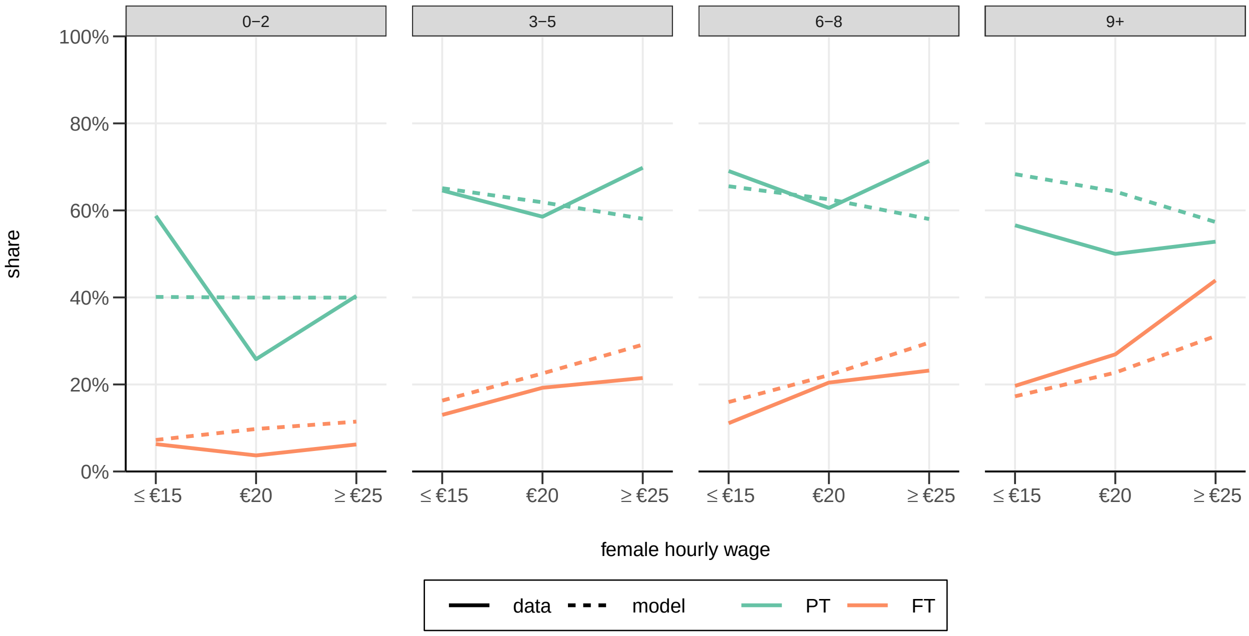


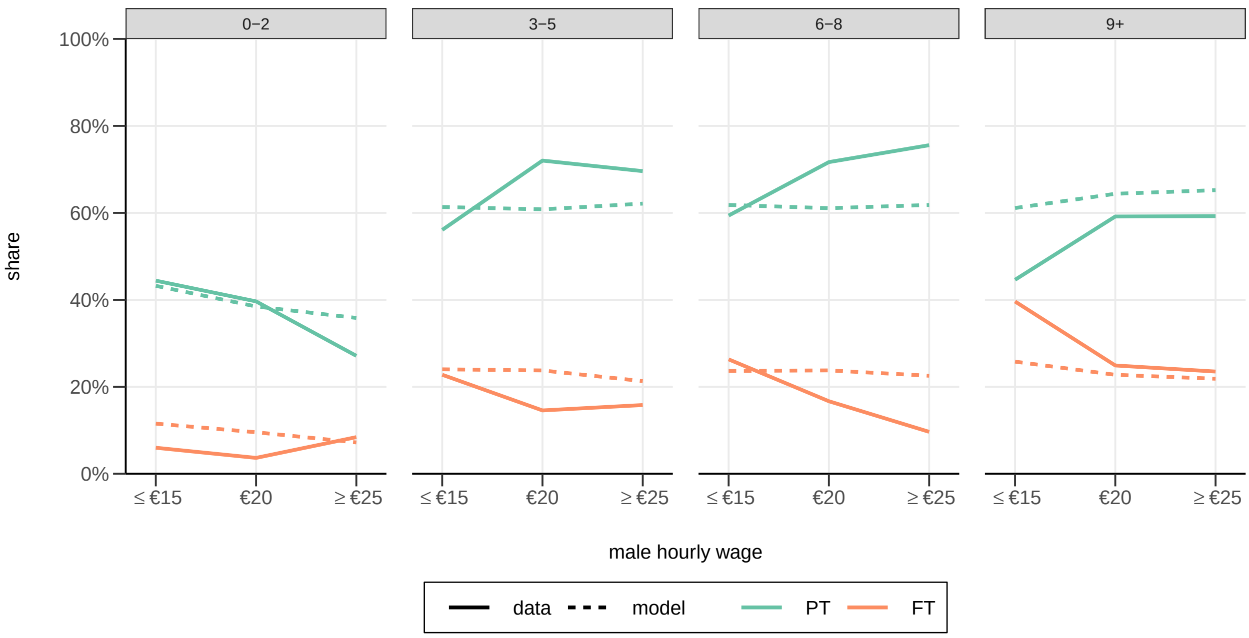
Figure C.7 (cont.): Sensitivity of estimated structural parameters

Notes: Illustration of changes in the negative Log-Likelihood in response to small deviations of each parameter from its estimated value (labelled by black dot), keeping all other parameters at their point estimates. We constrain the estimation of σ_{oth} within the interval [0,2] to facilitate the computation, but the resulting CDF (as displayed in Figure 6) remains virtually identical if we allow for a larger interval. 'demanding occup' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine. 'high education' indicates having obtained at least an A-level.

D Additional Model Fit Illustrations



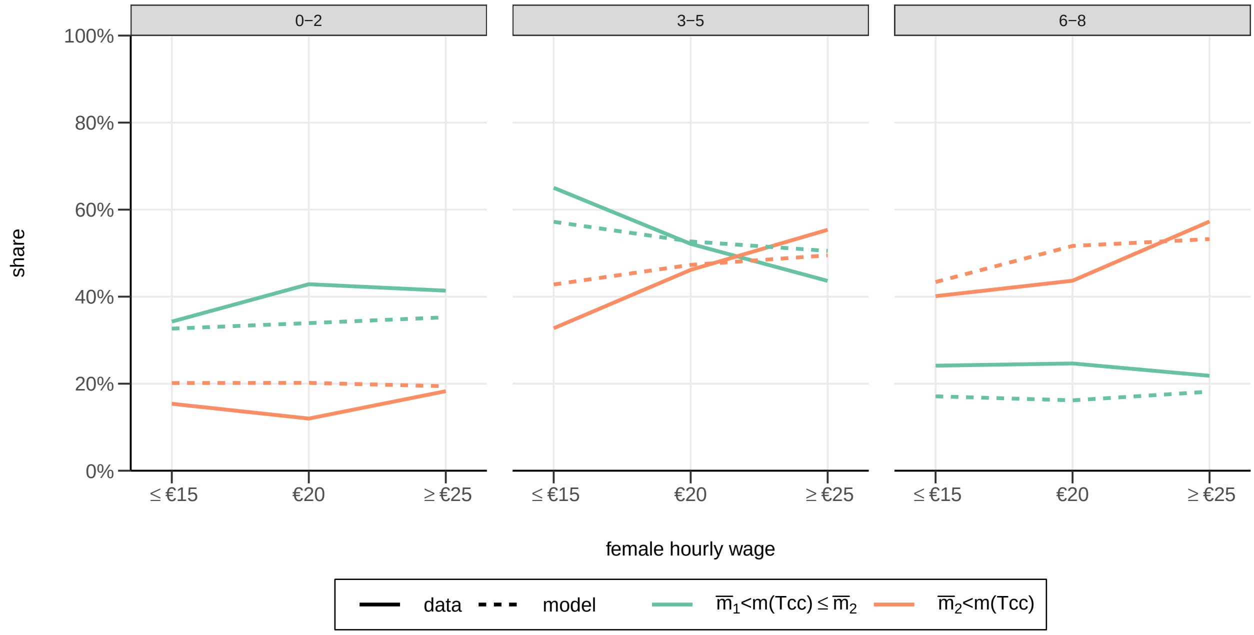
(a) Model fit for labor supply by female wages



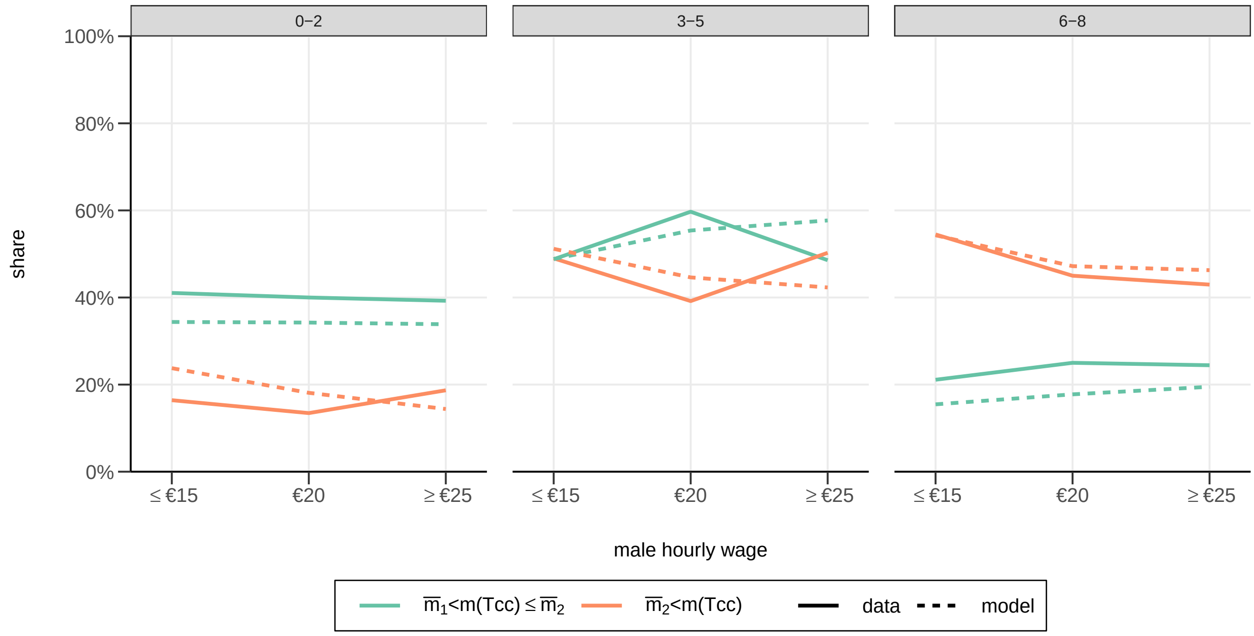
(b) Model fit for labor supply by male wages

Figure D.8: Model fit for labour supply by male and female wages

Notes: PT and FT denote the female working part-time and full-time, respectively. Sample as defined in Section 4.2.2. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).



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



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

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

Notes: $m(Tcc)$ denotes the share of total childcare need that is covered through public childcare. $\overline{m}_1$, respectively $\overline{m}_2$, indicates that the household would cover a share of 33%, respectively 75%. Sample as defined in Section 4.2.2. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

E Additional Policy Experiment Results

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

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

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

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

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

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

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

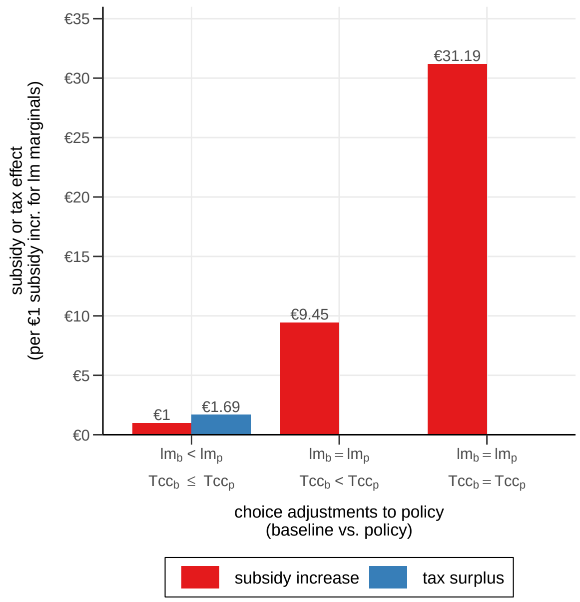
		female hourly wage		
	total	≤ €15	€20	≥ €25
(a) *Impact period*				
tax surplus	79.16	40.49	29.74	8.94
subsidy increase	911.36	549.68	250.16	111.52
self-financing	8.7%	7.4%	11.9%	8.0%
(b) *All periods*				
tax surplus	204.03	107.61	76.45	19.98
subsidy increase	1,649.07	980.35	496.04	172.68
self-financing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

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

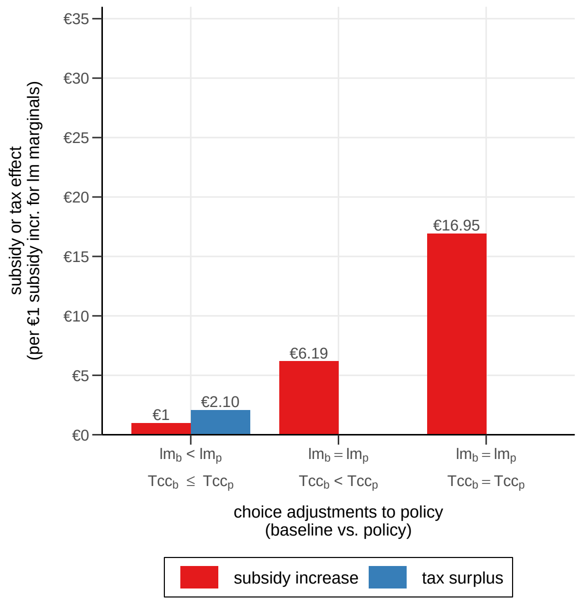
Table E.7: Detailed results for a full-time-contingent childcare subsidy increase

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

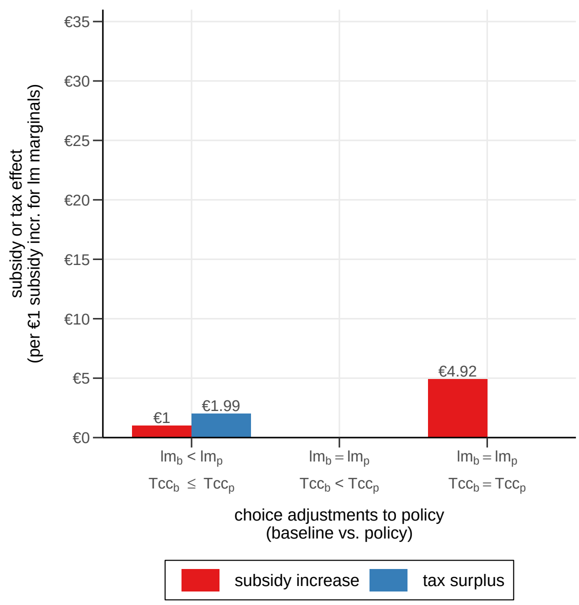
Notes: Self-financing degree as the ratio of tax surplus over subsidy increase generated by the policy. *Impact period*: model period of the policy introduction. *All periods*: full simulation of the remaining life cycle. Decomposition by initial wages at policy introduction. Policy experiment: full-time-contingent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



(a) *untargeted +€50*



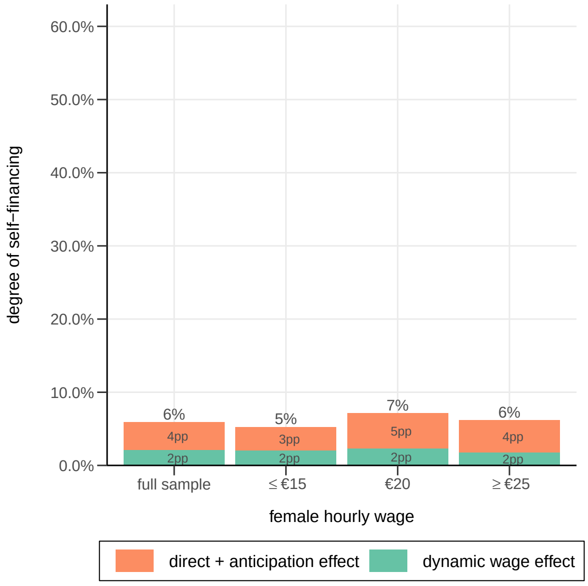
(b) *work-contingent +€50*



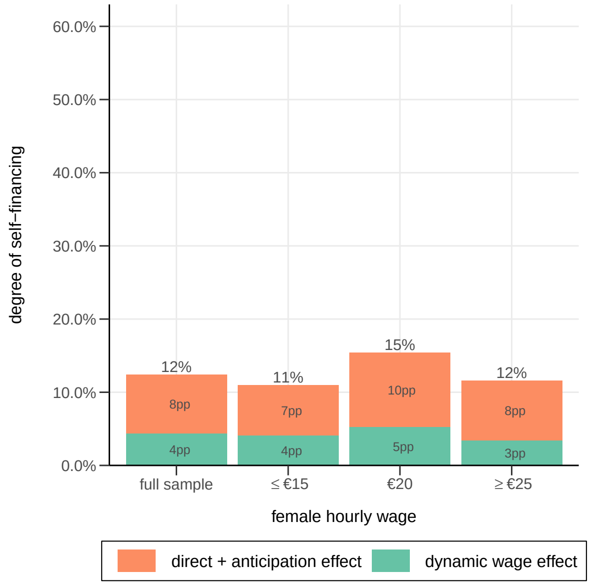
(c) *full-time-contingent +€50*

Figure E.10: *Impact period* effect decompositions for childcare subsidy increases

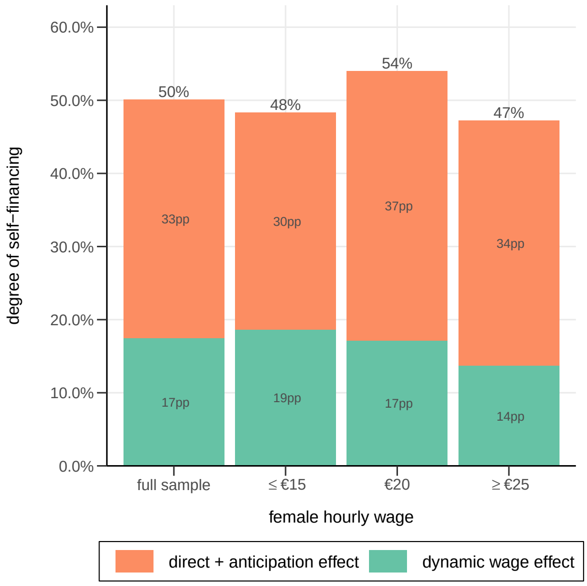
Notes: *Impact period* effects from Table 8 decomposed by subsidy increase and tax surplus, normalized by sample size. lm_b and Tcc_b denote the labour supply and public childcare decision in the baseline scenario. lm_p and Tcc_p denote the labour supply and public childcare decision in the respective policy scenario.



(a) *untargeted +€50*



(b) *work-contingent +€50*



(c) *full-time-contingent +€50*

Figure E.11: *All periods* effect decompositions for childcare subsidy increases

Notes: *All periods* effects from Table 8 decomposed into two main components: i) *direct effects + anticipation effects*, fiscal effects from behavioural adjustments in response to changes in the public childcare availability or fees while the household has children/may have children in the future in a policy affected age range. ii) *dynamic wage effects*, fiscal effects from higher wages and behavioural adjustments in response to higher wages, where higher wages are the result of past behavioural changes. Decomposition by initial wages at policy introduction. Values within the bars denote percentage point contributions.