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

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

October 29, 2023

Abstract

We formalize and estimate the dynamic marginal efficiency cost of redistribution (MECR) in the spirit of Okun's "leaky bucket". We analyze the MECR of an income-contingent childcare subsidy program and the income tax within the German context, using a dynamic structural heterogeneous-household model of childcare demand and maternal labor supply. This allows us to compare which of these two policies is more efficient in achieving redistributive goals. Our analysis identifies two competing forces. (i) Labor supply responses increase the MECR of the childcare subsidy relative to the income tax. (ii) Child development effects decrease the MECR of the childcare subsidy relative to the income tax. For reasonably large Pareto weights on children, we find that (ii) dominates (i) and therefore the childcare subsidy is the more efficient redistribution tool.

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

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

*We would like to thank Árpád Ábahám, Elizabeth Caucutt, Mark Colas, Mariacristina De Nardi, Philipp Eisenhauer, Eric French, Hans-Martin von Gaudecker, Nezih Guner, Peter Haan, Anne Hannusch, Emanuel Hansen, Mike Keane, Hamish Low, Joseph Mullins, Nicola Pavoni, Maria Polyakova, Daniel Reck, Víctor Ríos-Rull, Karl Schulz, Michèle Tertilt, Petra Todd, Gianluca Violante, Johanna Wallenius, and Hanna Wang for comments and suggestions. We thank Lea Fricke, Tim Hug, Nuan Stahl, and Victoria Szabo for excellent research assistance. We also thank conference participants at the CEPR Public Economics Symposium, CESifo Public Economics Week, SED, Barcelona Summer Forum, CESifo Public Economics Conference, ZEW Public Finance, FROGEE Workshop, MadMac, VfS, NASMES, IIPF, Conference on Dynamic Structural Estimation, CRC "Rationality and Competition" Retreat, CRC TR 224 Retreat as well as seminar participants at UC3 Madrid, CRED Paris, UAB Barcelona, Bank of Canada, IAB Nuremburg, University of Copenhagen for helpful comments and suggestions. David Koll gratefully acknowledges support by the German Research Foundation (DFG) through the CRC TR 224 (Project A03) and the Gottfried Wilhelm Leibniz-Prize. He also gratefully acknowledges funding through 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.

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

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

[§]Allianz

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

1 Introduction

Most OECD countries provide childcare subsidies, primarily to stimulate mothers' labor supply. Typically income-based, these subsidies favor poorer households, and thereby enhance income redistribution and social mobility. The policy question of the level of progressivity of subsidies is complex. It requires a careful balance between the promotion of equity as well as social mobility and the efficiency costs associated with these measures. These costs not only stem from the implied increase in effective marginal tax rates but also from the differentiated subsidy allocation that progressive subsidies create, i.e. higher childcare costs for high-income households and lower childcare costs for low-income households. Overall, this introduces an equity-efficiency trade-off, echoing the challenges faced when designing the income tax schedule.

To sidestep the normative element of this policy question, we evaluate and compare the marginal efficiency costs of redistribution (MECR) for childcare subsidies and for the income tax. We formalize Okun's (1975) intuitive "leaky bucket" concept and quantify it in a dynamic setting: for each dollar taken from the population of high-income families, how much is lost in the redistribution process due to current and future fiscal externalities caused by behavioral changes (leakage) and how much reaches the population of low-income families? By comparing the leakages of both instruments, we aim to determine whether the childcare subsidy schedule is more or less efficient than the income tax schedule in achieving a given amount of redistribution. Our approach offers a framework that can be adapted more widely to quantify and compare the MECR of other policies with redistributive elements.

We demonstrate how to integrate the MECR concept into a dynamic structural model, bridging public finance techniques with structural estimation. Specifically, we explore a model with endogenous dynamics of maternal earnings and public childcare demand in the German context. Our analysis allows us to break down the MECR into its different components in a transparent manner. We find that maternal labor supply responses to changes in the price of childcare per hour drive up the MECR of the childcare subsidy relative to the income tax. However, if we also account for the effects of changes in public childcare attendance on the future earnings of children, the childcare subsidy schedule emerges as the more efficient redistribution tool.

Formalizing Okun's leaky bucket The starting point of our paper is a static model, where we formally introduce the MECR. Heterogeneous families make decisions on labor supply, leisure, and public childcare use. These families face an income tax schedule and an income-contingent childcare subsidy schedule. To formalize Okun's leaky bucket, we construct budget-neutral reforms of both schedules which marginally increase redistribution from families above a certain income level y_p to families below this income level. We illustrate such a reform in Figure 1 for the tax schedule. The black bold line captures the baseline tax schedule which is illustrated as linear for simplicity. Tax payment is increased for households above y_p, kept constant at y_p, and decreased below y_p. For a given increase of the marginal tax rate above y_p, the increase of the marginal tax rate below y_p is chosen such that budget neutrality is achieved.[1] The calculation of the MECR (i.e. the leakage) then follows from relating the gains of the poor with the losses of the rich, both measured in terms of compensating variations. We show how this concept complements and relates to the marginal value of public funds (Hendren and Sprung-Keyser 2020) and to quantifications of the equity-efficiency trade-off considered by e.g. Immervoll et al. (2007). Our measure shares the spirit of the marginal value of public funds in that it provides an intuitive 'bang for the buck' comparison of different policy instruments.

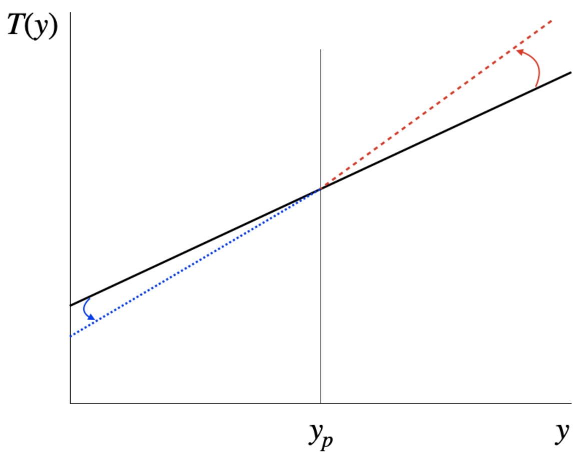


Figure 1: Illustration of Reform for Quantification of MECR

The theoretical analysis provides guidance and economic intuition for the key mechanisms. We show how the MECR depends on different own- and cross-price elasticities of labor supply and of public childcare demand along the household-income distribution. Rich empirical evidence about these elasticities is not readily available and motivates the use of a dynamic structural

[1]These reforms differ from the classical simple tax reform going back to Saez (2001), where the marginal tax (or subsidy) rate is changed in a very small interval. Instead, we consider an increase in marginal tax rates throughout the income distribution. The advantage of our reform is that it is not sensitive to how fine the income grid is, making it more suitable for large structural models.

model. In our quantitative analysis, we provide a decomposition of the MECR into its different components which is guided by this theoretical analysis.

Dynamic structural model We set up a dynamic discrete choice model with heterogeneous and unitary households, who have to decide on childcare provision for their children, leisure of the mother and whether the mother works full-time, part-time, or does not work at all.[2] During regular working hours, children can be cared for by the mother, through the use of informal childcare (e.g. grandparents) or in public childcare services.[3] In our model, labor supply decisions directly affect future wages: career penalties for working less than full-time reflect the empirical phenomenon that lower maternal labor supply today results in lower hourly wages in the future. Accounting for these penalties is crucial as they constitute a large part of the long-run fiscal effects of the redistribution tools we examine.

A notable feature of our model is the large amount of heterogeneity. First, we account for heterogeneity in the timing and spacing of births of up to three children, in education and in wages. The particular way in which the household's need for childcare and their preference for domestic childcare (carried out by the mother) evolve with the child age builds on Turon (2019).

Additionally, households differ from each other in three unobserved dimensions: (i) their preference for domestic childcare, (ii) their taste for the mother's leisure, and (iii) the family's access to free informal childcare. By incorporating (iii) and leisure choices, we depart from the often-made assumption that childcare hours equal work hours. Instead, we allow for hours of work to be greater or smaller than hours of childcare.[4]

Estimation and model fit The model is quantified using the rich panel data from the German Socio-Economic Panel (GSOEP). We estimate reduced-form relationships beginning with Mincerian wage equations which account for dynamic wage penalties and selection into work. In addition, we use the large cross-sectional data of the German Mikrozensus to estimate a non-parametric, stochastic fertility process conditional on age and education.

For the second part of the estimation, we use the structure of the model. After setting some standard parameters in line with the literature, we apply a maximum likelihood approach to estimate the joint distribution of (i) taste for domestic childcare, (ii) taste for leisure and (iii)

[2]We abstract from paternal labor supply decisions and assume that fathers always work full-time in line with the data: in 2016, 82% (86%) of fathers with a newborn (1 − 6 years old) child worked full-time (see report NL«H9§_nj _j W0}h9j´ ©·oP, p.64).

[3]Note that we will be referring to 'public' childcare for all types of market childcare, because the vast majority of market childcare is state-provided in Germany. Note also that subsidized childcare is generally available independent of the labor market status in Germany, which is different to the U.S., see e.g. Guner, Kaygusuz, and Ventura (2020).

[4]Bick (2016) emphasizes the differences observed in the data between labor supply and nursery attendance in the German context.

access to informal childcare. Our approach pins down the distributional parameters which maximize the likelihood of matching the observed dynamic household choices in terms of female labor supply and hours of public childcare services.

In addition, we cross-validate our model by comparing it to existing empirical evidence on labor supply and childcare responses. Our model predictions are in line with this evidence regarding labor supply elasticities along the intensive and the extensive margin as well as the responsiveness to wealth shocks. They also broadly concur with evidence regarding the impact of childcare subsidies on labor supply and childcare demand decisions (Busse and Gathmann 2020, Gathmann and Sass 2018).

Quantitative Results We first present our measure of the MECR when the potential effects of public childcare attendance on child development are ignored. This can either be considered as an intermediate step to understanding the different components of the MECR or as a context in which the social planner's horizon does not extend to the earnings of the next generation. Our benchmark results refer to redistribution from families with above-median income to families with below-median income. In this case, the MECR of the childcare subsidy schedule is 0.42 and that of the tax schedule is lower at 0.28. In other words, for each Euro taken from the population of above-median income households, 58 Cents (resp. 72 Cents) reach the population of families below the median. Both policies lead to a downward distortion in labor supply through higher effective marginal tax rates. The difference in MECR is mainly due to the cross-price effect of net childcare costs on labor supply, i.e. the effects that changes in the price per hour of childcare have on labor supply. A more progressive childcare subsidy schedule implies higher costs for high-wage mothers and lower costs for low-wage mothers. The resulting cross-price effects on labor supply are negative for high-wage mothers and positive for low-wage mothers. The fiscal impact of the former is significantly larger than the latter, which increases the leakage of the childcare subsidy schedule in comparison to the tax schedule. Dynamic wage effects play an important role in this result: for both policies, they amplify the fiscal effect of static labor supply responses by about 40%. These dynamic wage effects explain a bit more than 20% of the difference in the MECR.

We then extend our analysis to account for the long-term effects of public childcare attendance on children's outcomes and thereby integrate social mobility considerations into the redistribution analysis. We augment the model with results on the impact of changes in childcare attendance on children's long-term earnings by Havnes and Mogstad (2011, 2015). These authors obtain estimates across the parental income distribution, finding positive returns to public childcare for children with low to intermediate parental income, but negative effects for those with high parental income. This affects the MECR analysis in two ways: first, through fiscal externalities implied by different lifetime earnings of the children. Second, the changes in net-of-tax lifetime earnings of children have direct redistributive implications that should be accounted for.

Accounting for the implied long-term fiscal externalities lowers the MECR of the progressive childcare subsidy schedule from 0.42 to 0.35. The reason is that progressive childcare reforms boost childcare attendance of low-income children and lower it for high-income children. Further accounting for changes in the net-present value of children's after-tax lifetime income reduces the MECR of childcare subsidies to 0.12. Hence, for one Euro taken from richer families, not only 65 cents reach the parents of the poorer families through higher subsidies, but an additional 23 cents reach their children in the form of higher lifetime earnings. By contrast, the MECR of the income tax is almost unaffected by child development effects. The reason is that progressive tax reforms lower childcare attendance for both income groups, which has opposing effects that almost offset each other. As a consequence, the childcare subsidy schedule is the more efficient redistribution tool if one puts equal weight on children's utilities as on parents' utilities. In fact, the childcare subsidy schedule is the more efficient redistribution tool as long as the Pareto weight on children is at least 28% as large as those of the parents.

Finally, our formalization of Okun's leaky bucket also allows us to consider other thresholds for redistribution than the median income. We consider redistribution around the 30th percentile and the 70th percentile. While the results differ in size, the overall takeaway is very similar to our benchmark case of redistribution around the median income.

Contribution to the literature. This paper connects structural work on household decision-making with the more theoretical public finance literature on optimal redistribution.[5]

Regarding the former, a number of recent articles estimate the impact of different policies on households' dynamic labor supply choices.[6] Attanasio, Low, and Sánchez-Marcos (2008) find that the secular decline in childcare costs explains a large fraction of the increase in labor supply of married women in the U.S. over the last 30 years. Also for the U.S., Guner, Kaygusuz, and Ventura (2020) compare the welfare effects of child-related transfers and distinguish instruments along two dimensions: (i) whether transfers are conditional on work and (ii) whether transfers are means-tested. They find that means-tested transfers that do not condition on work yield the largest welfare gains. Blundell et al. (2016) find that tax credit policies (in-work benefits) in

[5]Our paper therefore shares the spirit of Blundell and Shephard (2012), Gayle and Shephard (2019), and Colas, Findeisen, and Sachs (2021).

[6]Some authors also examine the impact of policy on fertility decisions, e.g. Bick (2016), Wang (2022), Jakobsen, Jørgensen, and Low (2022) and Haan and Wrohlich (2011) or occupational choices, see Adda, Dustmann, and Stevens (2017). Hannusch (2022) adds a cross-country perspective on child-related transfers and maternal employment.

the UK increase the labor supply of lone mothers but decreases that of mothers with partners. For Germany, Bick (2016) estimates that a greater access to subsidized childcare would entice mothers of children under two to increase their labor supply along the intensive margin, while Wang (2022) examines a wide range of policy tools from parental leave, joint versus individual taxation and childcare subsidies. A key feature of this literature is the depreciation of human capital in non-participation and the lower returns to experience in part-time work relative to full-time employment.[7]

The other area of research which we contribute to, the optimal tax literature, has emphasized that childcare should be subsidized to counteract the negative incentive effects of taxes on labor supply. Domeij and Klein (2013) establish this in a Ramsey setting and show quantitatively that, for Germany, a linear subsidy of around 50% is optimal.[8] In a Mirrleesian setting, Bastani, Blomquist, and Micheletto (2020) allow for heterogeneous quality of childcare. They show that this weakens the subsidization argument because richer households buy higher quality childcare.[9]

Closest to us, Ho and Pavoni (2020) characterize optimal childcare subsidies which vary with income. Their quantitative analysis for the US shows that the optimal subsidy schedule decreases more strongly with income than the current policy, even if they constrain the reform to be Pareto improving.

Our paper extends the question of how childcare subsidies should vary with household income to a rich dynamic setting which accounts for dynamic wage effects for mothers and child development effects. We address the question of progressivity by incorporating a concept for the efficiency cost of redistribution going back to Okun (1975) into the structural model. To incorporate child development, we map our changes in public childcare attendance to changes in lifetime earnings of the children by augmenting our structural estimates with quasi-experimental evidence from Havnes and Mogstad (2015). This allows us to merge social mobility considerations with standard equity considerations and thereby provides a more comprehensive assessment about the MECR. Finally, the paper is closely related to Mullins (2022) who studies the optimal design of cash transfers to single mothers. First, as in our paper, the author studies redistributive policies in a structural household model. Second, both papers incorporate social mobility concerns into redistribution analysis by accounting for endogenous future earnings of

[7] Blundell et al. (2016) estimate these losses of potential earnings to be large. Adda, Dustmann, and Stevens (2017) stress that these losses vary across occupations. This relationship between current labor supply choices and future earnings plays a significant role in our evaluation of the long-term impact of policy changes on the fiscal budget.

[8] The theoretical reasoning for the childcare subsidy resembles the argument for education subsidies, see e.g. Bovenberg and Jacobs (2005), Krueger and Ludwig (2016) and Stantcheva (2017).

[9] They also consider the extension where the government offers public childcare with a given quality, where agents can opt in and out. This restores the subsidization result. However, they do not allow the price for public childcare to vary with income.

children. Given the different institutional backgrounds in the U.S. and Germany, the relative impact of maternal time and childcare use on children outcomes is modelled differently. In our German setting with regulated high-quality public childcare, we assume that children with low-parental income benefit if the mother works and this results in long-term benefits from public childcare use.[10]

2 Formalizing the MECR in a simple model

In this section, we introduce a static model to clarify the core principles that determine the marginal efficiency cost of redistribution for childcare subsidies and income taxes. We formally introduce our measure of the MECR and highlight the underlying trade-offs theoretically. This analysis uncovers the complexity of the factors contributing to the MECR, emphasizing the need for the dynamic structural model we use in subsequent sections.

2.1 Parents' preferences, constraints and decisions

Choices and constraints For this simple model, we examine a heterogeneous group of households, each with two parents and one young child. There is one unit of time. We assume that male labor supply is fixed at one unit, i.e. full time. For women, decisions are endogenous, requiring them to allocate their single unit of time across domestic childcare D, labor supply H, or leisure L. Formally, mothers face a time budget constraint

$$L + H + D = 1: \tag{1}$$

In this section 2, we assume that children require care for the full unit of time and that there is no private childcare available. Hence, children's time is shared between domestic childcare, D and time spent with a public childcare provider (nursery), N:

$$D + N = 1: \tag{2}$$

Preferences and heterogeneity Denote each household by i. We assume that utility is quasi-linear in consumption C and heterogenous across households i: $C + u_i(L; D)$.[11] The quasi

[10]For recent evidence for Germany that in particular children with weak parental background benefit from public childcare, see Busse and Gathmann (2020) and Cornelissen et al. (2018). Their findings on how short-term outcomes of public childcare attendance vary with parental income are consistent with the long-term effects of childcare subsidies in Havnes and Mogstad (2015). The latter authors also consider a setting where the quality of childcare institutions is rather homogeneous.

[11]The term $u_i(L; D)$ may capture the parents' utility derived from leisure L and time with their children D and an 'altruistic' term representing the household caring for the child's enjoyment of their time with their

linearity implies that childcare and labor supply decisions only depend on relative prices but not on wealth. We make this assumption here for simplicity to better single out the most important forces, as is often done in the optimal tax literature. In our dynamic structural model, however, we will consider a utility function with income effects.

Besides preferences, households differ in their wages $(w_{f,i}, w_{m,i})$, where subscripts f and m denote female and male. Household income is defined as $y_i = w_{f,i}H_i + w_{m,i}$.

Public policies Public childcare is available at price $K - s(y_i)$ per unit. K represent the cost per unit of childcare and $s(y_i)$ is the subsidy per unit of childcare, which depends on household income y_i. Further, households pay taxes according to a nonlinear tax schedule based on household income $T(y_i)$.

Decision problem Household i solves the following problem:

$$\max_{H_i, N_i} C_i + u_i(L_i, D_i) \quad \text{subject to} \quad \begin{cases} C_i = y_i - T(y_i) - [K - s(y_i)]\, N_i, \\ y_i = w_{f,i}H_i + w_{m,i}\text{ , (1) and (2).} \end{cases}$$

The first-order condition for H_i is given by:

$$[1 - T'(y_i) + s'(y_i)N_i]\, w_{f,i} = \frac{\partial u_i}{\partial L_i}.$$

This shows the trade-off between work and leisure, holding D constant. We see that labor supply is distorted by both the marginal tax rate T' and the marginal childcare subsidy s'. We define the "labor wedge" τ_i^H and the implied net wage $w_{f,i}^{net}$ as:

$$\tau_i^H = T'(y_i) - s'(y_i)N_i \quad \text{and} \quad w_{f,i}^{net} = (1 - \tau_i^H)w_{f,i}. \tag{3}$$

Next, we turn to the first-order condition for N:

$$\frac{\partial u_i}{\partial D_i} + K\left[1 - \frac{s(y_i)}{K}\right] = \frac{\partial u_i}{\partial L_i}.$$

This illustrates the trade-off between leisure L and domestic childcare D holding labor supply H constant. The left-hand side shows the utility gains from consuming what one would have spent on one hour of public childcare, $K\left[1 - \frac{s(y_i)}{K}\right]$, and the marginal utility from domestic childcare $\frac{\partial u_i}{\partial D}$. The utility costs in terms of foregone leisure are on the right-hand side. Similar to

mother, D, and at nursery, $N(D)$. This could be formalized as $u_i(L, D) = v_i(L, D) + \alpha_c u_c(D)$, where v_i is the parents' utility, u_c the child's utility and α_c the altruism parameter.

the labor wedge τ_i^H, we now define the "childcare wedge" τ_i^N, which is the rate of subsidization for one hour of public childcare, and the implied net childcare cost K_i^{net}:

$$\tau_i^N = \frac{s(y_i)}{K} \quad \text{and} \quad K_i^{net} = (1 - \tau_i^N) K. \tag{4}$$

The net-prices $w_{f,i}^{net}$ and K_i^{net} affect both the labor supply and the childcare demand decisions. (3) and (4) show that the subsidy schedule affects both of these net-prices: the marginal subsidy $s'(y)$ affects the net wage (3) and the absolute subsidy $s(y)$ affects the net childcare cost (4). This contrasts with the second policy instrument, the tax schedule, which only affects the net-wage through the marginal tax rate $T'(y)$. This asymmetry is a key reason for the difference in efficiency costs of redistribution between the two instruments, as we show below.

2.2 Measuring the marginal efficiency cost of redistribution with parametric reforms

To quantify the marginal efficiency cost of redistribution, we introduce parametric perturbations for both policy instruments. The reforms are budget neutral by construction and redistribute from households above a given percentile p of the income distribution to households below that percentile. To understand our notion of budget neutrality first note that we define net revenue NR, i.e. tax revenue net of subsidy spending, as:

$$NR = \int_i T(y_i)di - \int_i N_i s(y_i)di. \tag{5}$$

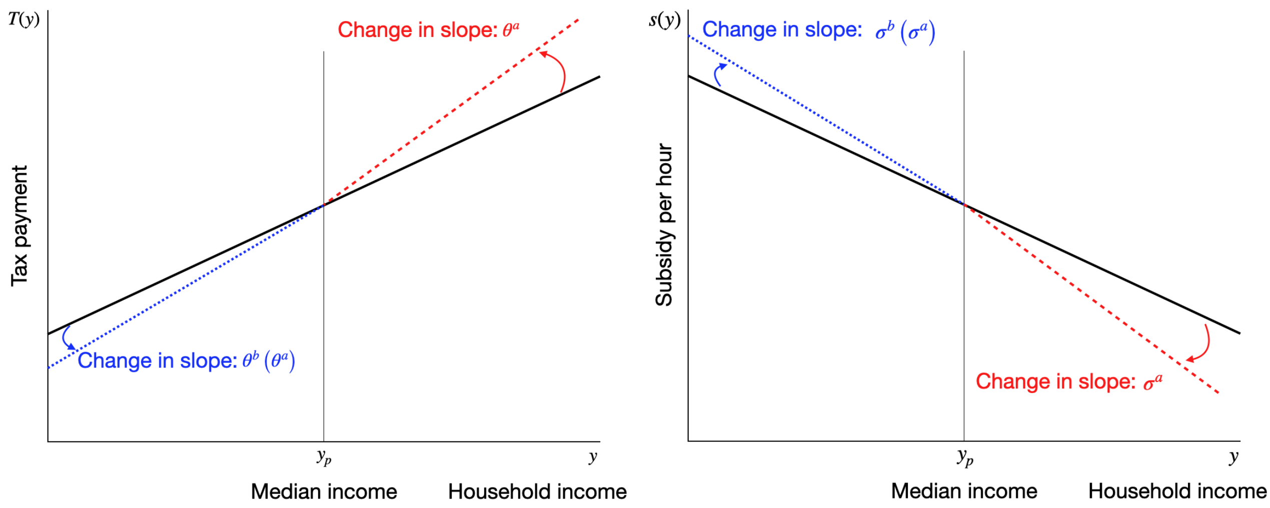
We then define budget neutral reforms as reforms that imply $d(NR) = 0$ after accounting for all behavioral changes.

Perturbation of the income tax schedule Before formally defining the reform, we first provide an illustration in Figure 2a. The black solid line represents the initial tax schedule, i.e. tax payment as a function of household income. Note that we depict the initial tax schedule as linear merely for simplicity – our analysis does not necessitate a linear baseline tax schedule. The reform increases the marginal tax rate and the absolute tax payment for households with incomes above y_p, as shown by the red dashed line. For households with incomes below y_p, the marginal tax rate is also increased, but the absolute tax payment decreases, as indicated by the blue dotted line. Given a specific increase in tax payments above the median income (a given red dashed line), the blue dotted line is chosen so that, after accounting for all household responses, the entire reform remains budget-neutral. This reform redistributes resources from households with incomes above y_p to those below y_p. Moreover, the change in tax payment is greater the further away from y_p.

We now provide a formal definition of this reform:

$$\hat{T}_p(y) = \begin{cases} \theta^a [y - y_p] & \text{for } y > y_p \\ -\theta^b(\theta^a) [y_p - y] & \text{for } y \leq y_p \end{cases} \tag{6}$$

θ^a is the increase in the marginal tax rate above y_p and $\theta^b(\theta^a)$ is the increase in the marginal tax rate below y_p. We consider small reforms with $\theta^a \to 0$ so that we can focus on first-order effects in our analysis. As indicated above, for a given value of θ^a, $\theta^b(\theta^a)$ is defined such that the reform is budget neutral ($dNR = 0$), once agents have adapted their behavior. The after reform tax schedule is then given by $T(y) + \hat{T}_p(y)$.



(a) Income tax reform

(b) Childcare subsidy reform

Figure 2: Illustrations of tax and subsidy reforms

Perturbation of the childcare subsidy schedule The perturbation of the childcare subsidy schedule, which we illustrate in Figure 2b, is defined analogously:[12]

$$\hat{s}_p(y) = \begin{cases} -\sigma^a (y - y_p) & \text{for } y > y_p \\ \sigma^b(\sigma^a) (y_p - y) & \text{for } y \leq y_p \end{cases} \tag{7}$$

where superscripts a and b denote above and below y_p. The after-reform hourly subsidy schedule is given by $s(y) + \hat{s}_p(y)$. The reform implies that the subsidy per hour is decreased above income level y_p and increased below. Marginal subsidies – i.e. how the subsidy per hour decreases with

[12]Note that the functions $\theta^b(\cdot)$ and $\sigma^b(\cdot)$ also depend on the percentile p. This is omitted here for ease of notation.

household income – are increased (in absolute terms) throughout the income distributions. The subsidy reform (7) is very similar to the tax reform (6) in its distributional effects.[13] However, as noted above already, the incentive effects of the reform are different: while the tax reform only implies changes in the net wage, the subsidy reform changes both the net wage and the net childcare cost.

Relation to elementary tax reforms Following the Mirrleesian approach of optimal nonlinear taxation, a so-called "elementary tax reform" could be considered the natural reform for our purpose.[14] In elementary tax reforms, the marginal tax rate (or the marginal rate of subsidization) is altered only within a very narrow interval surrounding some income level y_p. The additional tax revenue generated can then be redistributed in a lump-sum fashion, which implies – similar to our reforms – redistribution from those above y_p to those below y_p.

As put forward e.g. by Saez and Stantcheva (2018), the marginal efficiency costs of redistribution implied by these reforms are highly sensitive with respect to the value of the local Pareto parameter of the continuous income distribution. This sensitivity arises because marginal tax rates are essentially increased at a single point in the income distribution. In dynamic structural models, the earnings distribution is typically discrete, with a limited grid size due to the curse of dimensionality, making policy implications very sensitive with respect to how the small interval around y_p is chosen. Our proposed reforms $\hat{T}_p(y)$ and $\hat{S}_p(y)$ circumvent this complication and are thus more easily implementable in structural models with coarser income grids.

2.3 Marginal efficiency cost of redistribution: definition

We now formalize the popular idea of Okun's leaky bucket based on our parametric reforms (6) and (7). Our definition is based on money-metric utility changes, i.e. compensating variations (Hicks 1939).[15] These are not equal to changes in consumption: they are only the mechanical changes in consumption that would arise for fixed behavior have first-order effects on individual utility. Changes in income and consumption induced by changes in choices have only second-order effects on utility due to the envelope theorem.

[13] In fact, the reforms differ slightly because the level of childcare demand N_i also matters for how much households are affected by reform $\hat{S}_p$. Hence, two households i, j with the same household income ($y_i = y_j$) are affected in the same way in terms of the subsidy per hour, however, if they chose different amount of childcare $N_i \neq N_j$, they are affected differently in how much more or less total subsidies they receive.

[14] The idea of using such reforms to derive the optimal tax schedule goes back to Piketty (1997) and was then formalized more thoroughly by Saez (2001). It was extended to a dynamic setting by Golosov, Tsyvinski, and Werquin (2014) and to general equilibrium settings by Sachs, Tsyvinski, and Werquin (2020). The latter two papers also used this reform approach to evaluate reforms of current tax systems.

[15] Since we only look at small perturbations, we do not distinguish between equivalent variation and compensating variation. The literature also uses the term willingness to pay (e.g. Hendren and Sprung-Keyser (2020)), which could be used here as well.

Definition 1. ¦\0 h9}X_j9e 0¹H_0jH´ Hn § nT }0L_ §}_@«§_nj nT 9 §9² }0Tn}h $\hat{\tau}_p$ 9 L0»j0L _j (6) n} 9 «@ _L´ }0Tn}h $\hat{s}_p$ 9 L0»j0L _j (7) _ X_®0j @´I

$$MECR(\hat{\tau}_p) = 1 \quad \frac{CV^b \quad \hat{\tau}_p}{jCV^a \quad \hat{\tau}_p \quad j} \tag{8}$$

$$MECR(\hat{s}_p) = 1 \quad \frac{CV^b(\hat{s}_p)}{jCV^a(\hat{s}_p)j} \tag{9}$$

°\0}0

$$CV^a(\hat{\tau}_p) = \quad ^a \int_{i:y_i > y_p} [y_i \quad y_p]\, di \quad 9jL \quad CV^b(\hat{\tau}_p) = \; ^b(\; ^a) \int_{i:y_i \quad y_p} [y_p \quad y_i]\, di:$$

9}0 §\0 9XX}0X9§0L Hnhq0j 9§_jX ®9}_9§_nj nT \n« 0\neL °_§\ _jHnh0 9@n®0 9jL @0en° y_p Tn} §\0 §9² }0Tn}h $\hat{\tau}_p$u S«}§\0}J

$$CV^a(\hat{s}_p) = \quad ^a \int_{i:y_i > y_p} N_i\,[y_i \quad y_p]\, di: \quad 9jL \quad CV^b(\hat{s}_p) = \; ^b(\; ^a) \int_{i:y_i \quad y_p} N_i\,[y_p \quad y_i]\, di:$$

9}0 §\0 9XX}0X9§0L Hnhq0j 9§_jX ®9}_9§_nj Tn} §\0 «@ _L´ }0Tn}h $\hat{s}_p$u

The marginal efficiency cost of redistribution as defined in (8) and (9) can be interpreted as follows: for each unit of aggregate money-metric utility taken from the group earning above y_p, $(1 \quad MECR)$ unit of aggregate money-metric utility gain can be achieved for the group earning below y_p and $MECR$ is lost through the leakage of Okun's bucket.

If households did not change their behavior as a response to the reforms, we would have $MECR(\hat{\tau}_p) = MECR(\hat{s}_p) = 0$. In other words, each mechanical Euro taken from those above the p-th percentile would reach those below p. If however households respond to the reform, the $MECR$ are no longer zero because behavioral changes affect net revenue (5). These changes in net revenue due to changes in behavior are referred to as » H9e 0²§0}j9e_§_0 in the public finance literature (Hendren 2016) and determine the $MECR$. We now introduce four terms that capture these externalities.

Definition 2. 8j _jH}09 0 _j §\0 j0§ °9X0 $w^{net}_{f;i}$ _hqe_0 9 » H9e 0²§0}j9e_§´ §\}n«X\ 9j n°j]q}_H0 0¸0H§ O^i_w nj \n«} °n}c0L 9jL 9 H}n]q}_H0 0¸0H§ X^i_w nj H_eLH9}0 L0h9jLI

$$O^i_w \;=\; \frac{^H_i}{1 \quad ^H_i} \quad y_{f;i} \quad "^i_{H;w^{net}_f}: \tag{10}$$

$$X^i_w \;=\; \frac{^N_i}{1 \quad ^H_i} \quad N_i \quad "^i_{N;w^{net}_f} \tag{11}$$

8j _jH}09 0 _j §\0 j0§ Hn § K_i^{net} _hqe_0 9 » H9e 0²§0}j9e_§´ §\}n«X\ 9j n°j]q}_H0 0¸0H§ O_K^i nj H_eLH9}0 L0h9jL 9jL 9 H}n]q}_H0 0¸0H§ X_K^i nj \n«} °n}c0LI

$$O_K^i \;=\; \frac{1}{K} \; \frac{{}_i^N}{1 \quad {}_i^N} \; N_i \quad "^i_{N;K^{net}} \tag{12}$$

$$X_K^i \;=\; \frac{1}{K} \; \frac{{}_i^H}{1 \quad {}_i^N} \; y_{f;i} \quad "^i_{H;K^{net}} \tag{13}$$

°\0}0 $"^i_{Z;I} = \frac{@log(Z_i)}{@log(I_i)}$ Tn} H\n_H0 $Z_i = H_i; N_i$ 9jL _j §}«h0j§ $I_i = w^{net}_{f;i}; K_i^{net}$u

The derivation of these four objects can be found in Appendix A.[º£] Note that all four fiscal externalities are proportional to the product of the relevant elasticity and the baseline value of N_i and $y_{f;i}$ respectively.

Since a higher net wage implies higher labor supply and therefore higher tax revenue, the own-price wage effect, O_w^i is positive. On the other hand, the cross-price wage effect, X_w^i is negative since an increase in the net wage increases childcare demand, which itself increases government spending on subsidies.

The fiscal externality coming from the own-price childcare effect, O_K^i is positive since the increase in net cost implies a reduction in childcare demand, which reduces subsidy spending. Finally, the cross-price effect of the net cost change, X_K^i, is negative because a higher net cost of childcare implies lower labor supply and therefore lower tax revenue.

To obtain closed-form expressions for $MECR(\hat{T}_p)$ and $MECR(\hat{S}_p)$ in terms of (10)-(13), we need to solve for the budget-neutral values of $^b(\ ^a)$ and $^b(\ ^a)$. We derive such analytical expressions for $MECR(\hat{T}_p)$ and $MECR(\hat{S}_p)$ in Appendix A. We now turn to the most important aspects of these results.

2.4 Marginal efficiency cost of redistribution: theoretical results

We start with a tax reform and first of all summarize how the tax reform affects the net wage (3) and the net childcare cost (4).

Lemma 1. 8 §9² }0Tn}h $\hat{T}_p(y)$ 9 L0»j0L _j r£s

0u en°0} §\0 j0§ °9X0 $w^{net}_{f;i}$ Tn} 9ee \n« 0\neL u

©u Ln0 jn§ 9¸0H§ §\0 j0§ H_eLH9}0 Hn § K_i^{net} Tn} 9j´ \n« 0\neLu

Hence, to calculate the $MECR$ of the tax reform, we only need to take into account that households adjust their behavior due to a change in their net wage $w^{net}_{f;i}$.

[16]See the steps prior to equations (22), (23), (25) and (26).

Proposition 1. ¦\0 h9}X_j9e 0 ¹H_0jH´ Hn § nT }0L_ §}_@«§_nj nT }0L_ §}_@«§_nj $MECR(\hat{\tau}_p)$ _

or9s _jH}09 _jX _j $E[O_w|y > y_p]$ 9jL $E[O_w|y < y_p]$u

or@s L0H}09 _jX _j $E[X_w|y > y_p]$ 9jL $E[X_w|y < y_p]$u

p}nnTu This immediately follows from the formula for $MECR(\hat{\tau}_p)$ in (24) in Appendix A. □

1(a) refers to the own-price effect of a change in w_f^{net}: all households work less due to the lower net-wage. This results in a negative fiscal externality, which increases the MECR of the tax reform. 1(b) refers to the cross-price effect of this change in w_f^{net}: households will also demand less childcare, which results in a positive fiscal externality leading to lower MECR.

We now turn to a childcare subsidy reform.

Lemma 2. 8 «@ _L´ }0Tn}h $\hat{s}_p(y)$ 9 L0»j0L _j r¢sI

ou en°0} §\0 j0§ °9X0 $w_{f;i}^{net}$ Tn} 9ee \n« 0\neL u

©u _jH}09 0 r}0 qu L0H}09 0 s §\0 j0§ H_eLH9}0 Hn § K_i^{net} Tn} \n« 0\neL °_§\ _jHnh0 9@n®0 r}0 qu @0en°s y_pu

As for the tax reform, a steeper slope of the childcare subsidy schedule lowers the net wage, this is captured by point 1 in Lemma 2. The second point captures the change in the net childcare cost incurred by the reform above and below percentile p. Note that the change in the price per hour is larger the further away the household's income is from y_p. The following proposition states how these changes in net wages and net childcare costs affect $MECR(\hat{s}_p)$:

Proposition 2. ¦\0 h9}X_j9e 0 ¹H_0jH´ Hn § nT }0L_ §}_@«§_nj $MECR(\hat{s}_p)$ _ I

or9s _jH}09 _jX _j $E[O_w \quad N|y > y_p]$ 9jL $E[O_w \quad N|y < y_p]$u

or@s L0H}09 _jX _j $E[X_w \quad N|y > y_p]$ 9jL $E[X_w \quad N|y < y_p]$u

©r9s L0H}09 _jX _j $E[O_K \quad (y_i \quad y_p)|y > y_p]$ 9jL _jH}09 _jX _j $E[O_K \quad [y_i \quad y_p]|y < y_p]$u

©r@s _jH}09 _jX _j $E[X_K \quad (y_i \quad y_p)|y > y_p]$ 9jL L0H}09 _jX _j $E[X_K \quad (y_i \quad y_p)|y < y_p]$u

1(a) and 1(b) mirror the statements in Proposition 1 and capture the own-price and cross-price effects of changes in net wages. A difference is that the fiscal externality terms are multiplied by N. This captures the fact that the implied increase in the effective marginal tax rate is larger the greater the household demand for public childcare. 2(a) and 2(b) reflect responses to changes in net prices K^{net}, which increase for households with $y > y_p$ and decreases for those with $y < y_p$. The own-price effect of these changes, O_K, reflected in 2(a), is a decrease in nursery demand for richer households and an increase in nursery demand for the less well-off. The net impact of these two changes on the government budget and therefore on the MECR is ambiguous. Finally, 2(b) captures the cross-price effect of changes in K^{net} on the labor supply of all households. Again here we have two effects of opposite signs for households above and below the p-th percentile and the net effect is ambiguous. Finally, note that the changes in the net cost that trigger effects described 2(a) and 2(b) are larger the further away the households are from y_p.

Under which conditions is $MECR(\hat{s}_p) > MECR(\hat{\tau}_p)$? Since the net wage effects 1(a) and 1(b) are almost the same for both reforms, the most important question is whether the net childcare price effects 2(a) and 2(b) add up to an increase or a decrease in the $MECR$. Both 2(a) and 2(b) yield an ambiguous contribution to the MECR since the effects on the populations below and above the p-th percentile are of opposite signs. The net effects will hinge on the relative magnitudes of the behavioral adjustments and on the relative costs/benefits for the government budget per unit of adjustment for rich and poor households.

The own-price effect on the use of nursery services, embodied in point 2(a), depends on how sensitive different households are to childcare prices. The relative costs/benefits of unit changes in nursery use for the government budget depend on how progressive childcare prices are: if these are the same for all households, unit adjustments above and below p have the same impact on the budget. When the price of childcare is progressive, i.e. is higher for richer households, a unit decrease in nursery use by a rich household will save less to the budget than a unit increase in nursery use by a poor household will cost (since the government subsidy is greater for poorer households).

The cross-price effect of a childcare price reform on labor supply, embodied in point 2(b), depends on the complementarity of nursery use and labor supply for different households. Richer households facing more expensive nursery fees will decrease their labor supply while poorer households are likely to increase their labor supply in the face of cheaper nursery use. The net effect of these two behavioral adjustments in terms of hours worked in the economy is ambiguous and depends on the relative labor supply elasticities of the two groups (above and below the p-th percentile) to the price of childcare. For given values of the cross-price elasticity, the fiscal externality O_K is increasing in income and the labor wedge ^{H}. Therefore, it is likely that the negative fiscal effect of households with income above y_p dominates the positive fiscal effect of households with below y_p income, since not only is income larger above the median, but labor wedges as well due increasing marginal tax rates of the baseline tax schedule.

Which of these effects dominates in the real world is a non-trivial question since the labor and childcare wedges and the various elasticities all vary across the income distribution. There is no direct empirical evidence for the quantification of the terms in Propositions 1 and 2. We therefore develop and estimate a dynamic structural model based on dynamic decision making observed in panel data. We will relate our quantitative findings to Propositions 1 and 2 in a transparent manner. Before we turn to our structural quantitative analysis, we now discuss how the comparison of $MECR$s is further affected by dynamic effects such as future maternal earnings and child development considerations.

2.5 Dynamic effects

2.5.1 Parental decisions

We considered a static version of the model. The logic simply extends to a dynamic setting. If there were more than one period, the reforms $\hat{T}_p$ and $\hat{s}_p$ may affect future earnings or childcare demand of parents even if the reform is only implemented in one period. An important channel is the dynamics of mothers' wages: less labor supply today results in lower wages and lower tax payment in the future. Incorporating this into the analysis is straightforward. As opposed to considering a static budget neutrality, one can consider a net-present value dynamic budget neutrality. These changes in mothers' future earnings then affect the MECR simply through their impact on $^{b}(\ ^{a})$ and $^{b}(\ ^{a})$. The fiscal externalities will get enlarged by a dynamic component. This will be included in our structural model.

2.5.2 Child development

We now show how to integrate the impact of nursery attendance on children's future earnings into the MECR analysis.

Fiscal externalities Potential changes in children's earnings caused by the reforms under study create fiscal externalities, which matter for the MECR if we take a longer-run perspective. They affect $^{b}(\ ^{a})$ and $^{b}(\ ^{a})$ which in turn affects the MECRs as defined in (8) and (9) through $CV^{b}(\hat{T}_p)$ and $CV^{b}(\hat{s}_p)$. For example, if low-income children benefit from attending public childcare in terms of their lifetime earnings, the MECR of childcare subsidies will be lower, H0§0}_ q9}_@« , since the childcare fee reform increases their attendance at a nursery. On the contrary, this increases the MECR of the tax reform H0§0}_ q9}_@« as this reform lowers their nursery attendance.

Children as distinct individuals One could also argue that there is no need to include child-development returns on the net-earnings of children because parents internalise the impact of their decisions on children's future income and therefore the envelope theorem applies.[17] As emphasized by Farhi and Werning (2010), if altruistic parents make decisions for their children, one should nevertheless also account for utility of children as they are distinct individuals. They should explicitly be accounted for in the social planner's objective.

We now introduce new notation for the $MECR^c$ which accounts for childrens' earnings being endogenous to today's nursery attendance.

$$MECR^c(\hat{\tau}_p;\ ^c) \quad = \quad 1 \quad \frac{CV^b_{parent}\ \ \hat{\tau}_p \ \ + \ ^c\, CV^b_{child}\ \ \hat{\tau}_p}{CV^a_{parent}\ \ \hat{\tau}_p \ \ + \ ^c\, CV^a_{child}\ \ \hat{\tau}_p} \tag{14}$$

$$MECR^c(\hat{s}_p;\ ^c) \quad = \quad 1 \quad \frac{CV^b_{parent}(\hat{s}_p) + \ ^c\, CV^b_{child}(\hat{s}_p)}{CV^a_{parent}(\hat{s}_p) + \ ^c\, CV^a_{child}(\hat{s}_p)} \tag{15}$$

where ^{c} denotes the Pareto weight the social planner puts on children relative to the generation of their parents. The CV of the parents are as in Definition 1, the subscript "parent" is only added for clarity. Note that even when $^c = 0$ these $MECR$ will be different from $MECR(\hat{s}_p)$ and $MECR(\hat{\tau}_p)$ defined above because they account for the fiscal externalities from child development.

2.6 Relation to similar concepts in the literature

The idea of quantifying the trade-off between equity and efficiency at the margin has been pursued by e.g. Immervoll et al. (2007).[18] They compare the efficiency cost of providing transfers to the working poor versus to the unemployed. Their measure could be defined as follows for our tax reform:

$$(\hat{\tau}_p) = \frac{jCV^a\ \ \hat{\tau}_p\ \ j}{CV^b\ \ \hat{\tau}_p}:$$

Besides the difference in the exact interpretation of the measures, which in our case relates to the idea of Okun (1975), we apply the concept to evaluate two redistribution instruments that redistribute between the same groups. Hence, policy implications can be drawn independently from preferences for redistribution between these groups. In Immervoll et al. (2007), the transfer recipients differ.

[17]One could question the application of the envelope theorem here because empirical research finds that parents – in particular low-income parents – underestimate the gains from early childhood education (Boneva and Rauh 2018, Cunha, Elo, and Culhane 2022).

[18]Browning and Johnson (1984) and Ballard (1988) follow a similar approach as Immervoll et al. (2007). Adam (2005) asks a very similar research question as Immervoll et al. (2007) in the UK context.

We also relate to a recent and widely applied concept of the marginal value of public funds (MVPF) introduced by Hendren and Sprung-Keyser (2020). This measure provides an intuitive 'bang for the buck' interpretation. It measures the aggregate compensating variation of a policy divided by the net cost of the policy, where the measure of costs accounts for the fiscal externalities of the policy. To understand the relation between our measures (which apply to budget-neutral policies) and theirs, it is useful to split our reform into two parts. E.g. for the tax reform, denote the tax increase above y_p by $\hat{\tau}_p^a$ and the tax decrease below y_p by $\hat{\tau}_p^b$. Denote the respective MPVFs as $MVPF(\hat{\tau}_p^a)$ and $MVPF(\hat{\tau}_p^b)$. It is then simple to show that:[19]

$$MECR(\hat{\tau}_p) = 1 \quad \frac{MVPF(\hat{\tau}_p^b)}{MVPF(\hat{\tau}_p^a)}:$$

We consider our MECR as complementary to the MVPF. It allows to compare different redistribution tools regarding their efficiency in achieving redistribution. We now turn to our application of childcare subsidies and thereby provide a showcase for the implementation in a structural model.[20]

Last but not least, the concept of the MECR implicitly is a crucial element in the literature on optimal nonlinear income taxation going back to Mirrlees (1971) and Saez (2001). For the optimal tax schedule, at each income level, the marginal tax rate is chosen such that the MECR equals the distributional gains from households above that income level to households below that income level. This is particularly transparent in Heathcote and Tsujiyama (2021), who express the optimality condition in exactly that way.

3 Dynamic structural model

As shown in Section 2, the magnitude of the MECR of both policy instruments depends on the distribution of households' elasticities of labor supply and public childcare demand with respect to w^{net} and K^{net}. Our structural model aims to provide a rich picture of households faced with labor supply and childcare choices within a dynamic framework.

Our environment is composed of households with 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

[19]Since our reform $\hat{\tau}_p$ is budget neutral, the net-fiscal effects in the denominators of $MVPF(\hat{\tau}_p^b)$ and $MVPF(\hat{\tau}_p^a)$ cancel. Hence, what remains of the ratio $\frac{MVPF(\hat{\tau}_p^b)}{MVPF(\hat{\tau}_p^a)}$ are the compensating variations, which then gives (8).

[20]Our MECR measure could, however, similarly as the MVPF be quantified based on given empirical estimates if they are available.

stochastic process, which captures the substantial empirical heterogeneity in family composition and in the age of parents at first birth.

Households with young children make two decisions each period: how to provide care for their children and how much maternal labor to supply. Regarding childcare, they decide between the mother caring for the children at home, which we call 'domestic childcare', and externally provided childcare. The latter can either be informal childcare by, e.g., grandparents, or the use of public childcare services, which we call 'nursery'.[21] Labor supply choices are discrete: The female spouse can work full-time, part-time, or choose not to participate, while the male spouse is assumed to always work full-time.[22] An important dynamic component of our framework comes from the positive impact of current working hours on the expected growth rate of future wages.

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, and in the number and ages of the children they currently have. Besides education, female wages, male wages and children demographics, households are heterogeneous in three further (unobserved) dimensions: their preference for domestic childcare, their taste for the female spouse's leisure, and their access to free informal childcare. As we argue below, accounting for this unobserved heterogeneity is key to capturing the large heterogeneity in childcare and labor supply choices of households.

3.1 Children

Children are born, one by one, to parents between the ages of 20 and 40. Subsequent siblings can only be born one or two 3-year interval(s) later, 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 and 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 a 4-element vector indicating the presence of a child in each of these age brackets. For example, a family whose composition is represented by the vector $= (0; 1; 1; 0)$ 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 can only be 0 or 1 since only one child can be born in each period. Transitions between different values of are governed by (stochastic)

[21]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 Figure B.1.

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

fertility events and the (deterministic) ageing of the household's children. Finally, we assume that households cannot have more than three children.[23]

3.2 Preferences

As in our simple model presented in Section 2, households value female leisure time L, household consumption C, and domestic childcare D. Household consumption is made comparable across different household sizes k by applying a square root equivalence scale. Preferences are reflected in the following instantaneous utility function:

$$u(C;L;D) = \left(1 - G(g;\gamma)\right)\left((1-\alpha)\frac{\left(\frac{C}{\sqrt{k}}\right)^{1-\sigma_c}-1}{1-\sigma_c} + \alpha\frac{L^{1-\sigma_L}-1}{1-\sigma_L}\right) + G(g;\gamma)\frac{D^{1-\sigma_D}-1}{1-\sigma_D};$$

where the three CRRA coefficients σ_c, σ_L amd σ_D are homogeneous across all households.

Preference heterogeneity. Households' preferences are heterogeneous in two dimensions: α represents the relative taste for female leisure over consumption, and g is the relative preference for domestic childcare. We allow the taste for domestic childcare to vary with the age of the child by introducing G as follows:

$$G(g;\gamma) = \begin{cases} g & \text{if youngest child's age } \in [0;3); \\ \gamma \cdot g & \text{if youngest child's age } \in [3;9). \end{cases}$$

This allows us to capture the sharp difference in public childcare enrolment between 0 − 2 year olds and 3 − 5 year olds, see Section 5.1.

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, t_j, within normal working hours (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. Apart from domestic childcare, households may fulfil the childcare need by calling on informal childcare providers, e.g. grandparents, denoted I, or public childcare services, i.e. a nursery, denoted N.

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

Informal childcare is free and only available to some households. The variable I ranges between 0 and 40 hours a week. If available, households always prefer to use I hours of costless informal childcare over N hours of costly public childcare. In that sense, I captures both whether informal childcare is available and if it is considered equally good as public childcare.

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

$$p(j; \ ; y).$$

In terms of our simple model in Section 2, p corresponds to $K \quad s_i(y)$, i.e. the subsidized price of childcare. For a given amount of domestic childcare and informal childcare use, the resulting amount of public childcare necessary for a child of age j is thus given by:

$$N(j) = \max \ 0; t_j \quad D \quad I \tag{16}$$

We also define the share of childcare needs that a household covers with public childcare:

$$m(\ ) = \frac{\sum_{j=1}^{3} \ (j) \ N(j)}{\sum_{j=1}^{3} \ (j) \ t_j} \tag{17}$$

where, as defined above, (j) is the j-th element of the vector indicating if a child of age j currently lives in the household. The household expenditure on public childcare of all its children is thus given by S:

$$S(\ ; y) = \sum_{j=1}^{3} \ (j) \ N(j) \ p(j; \ ; y).$$

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 with domestic or informal childcare, where the same unit of time can be used to look after one to three children.

Parental time constraint. At each age t, the household has to choose between female labor supply (H_t), female leisure (L_t), and the provision of domestic childcare (D_t). Hence the time constraint is written as:

$$H_t + L_t + D_t = 40. \tag{18}$$

where 40 captures the usual weekly full-time hours.

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:

$$C_t + S(\ _t, y_t) = y_t - T(y_t), \tag{19}$$

where

$$y_t = 40 \quad w_{m,t}(w_{m,t-1}) + H_t \quad w_{f,t}(w_{f,t-1}, H_{t-1}).$$

$T(\ )$ captures the tax and transfer system and $H_t \in \{0, 20, 40\}$ represents non-participation, part-time, and full-time work respectively. Male wages $w_{m,t}$ and female wages $w_{m,t}$ are assumed to follow first-order Markov processes. For women, transition rates between wage grid points depend on current labor supply H_t. Finally, once the spouses retire, they get a fraction B of their last period's full-time earnings potential as retirement benefits.

3.4 Dynamic decision problem

We summarize all heterogeneity in the following vector:

$$\ _t = (s_t, h) \quad \text{with} \quad s_t = (t, w_{m,t}, w_{f,t}, \ _t, educ) \quad \text{and} \quad h = (g, l, \ )$$

At each age t, the household has to choose female labor supply (H_t) and the amount of domestic childcare (D_t). These imply the values of consumption (C_t), female leisure (L_t), and the use of public childcare (N_t). The three constraints that the household faces are the need for childcare (16), the time constraint for the female spouse (18), and the budget constraint (19).

The full dynamic household problem is defined for a given state space vector $\ _t$ as:

$$V(\ _t) = \max_{H_t, D_t} u(C_t, L_t, D_t | \ _t) + \ \mathbb{E}[V(\ _{t+1} | \ _t, H_t)], \text{ s.t. } (16), (18) \text{ and } (19) \tag{20}$$

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 anymore.

3.5 Unobserved heterogeneity

We now provide a more thorough discussion of the role that the unobserved heterogeneity parameters $h = (g, l, \ )$ play. The distribution of $(g, l, \ )$ conditional on observables is key to capturing the observed behavior of households. First, we allow for heterogeneity in leisure preferences to account for the sizeable variation in labor supply conditional on wages. Such heterogeneity in leisure preferences (or equivalently, disutility of work) is a common component in structural models to match hours worked (such as in, e.g., Blundell et al. 2016).

A more distinctive feature of our model is the heterogeneity in g and I. It 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 which may capture i) the true preference heterogeneity for spending time with the child; ii) heterogeneity in how much parents (dis)like their child being in nursery or informal childcare, e.g., due to social norms or trust in the quality of the childcare institutions; iii) the fixed utility cost of bringing children to nursery.[24] Finally, the heterogeneous use of informal childcare I across households represents a combination of the household's access to informal care (e.g. availability of grandparents) and on the household's view on this type of care being an acceptable alternative.

4 Estimation methodology

Our estimation procedure can be decomposed into two parts. First, in Section 4.1, we estimate and calibrate various parameters without using the explicit structure of the model. In a second step, in Section 4.2, we quantify the remaining parameters by using the model structure.

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 for the government of providing a full-day public childcare slot and estimate the childcare fee schedule as a function of parental income. Finally, we calibrate the government policies required as exogenous inputs for our model.

Childcare need. The age-specific weekly hours of childcare needed, t_j, are calibrated as follows, for each child age j. If a child is younger than 6, the childcare need is set to 40 hours per week, i.e., 100% of the usual working week. To account for the fact that nearly all $3-5$ year olds attend kindergarten at least half-days , we impose that 20 of the possible 40 hours for this age group have to be covered by public childcare. For children aged $6-8$, the need reduces to 15 hours per week because these children attend compulsory schooling for 25 hours per week.

[24]Distance to the childcare facility could be one potential reason why parents do not send their children to public childcare.

Public childcare cost structures. We approximate the cost structure of public childcare institutions by assuming the costs to the government to be linear in the number of children. We use the values in Table 1, which are provided by the German Statistical Office.

Table 1: Average annual cost per child for 40h/week of public childcare

Children's age interval	0 − 2	3 − 5	6 − 8
Annual cost	€11,837	€7,927	€6,733

i n§0 : See Statistisches Bundesamt (2012), converted to 2017 prices.

Childcare fees. The 2013, 2015, and 2017 waves of our GSOEP data contain information on public childcare hours per day and monthly fees paid. We use the same sample restrictions as in our structural sample that we describe below in Section 4.2.1. In Appendix C.1.2 we estimate childcare fees as a function of gross household income, which we interact with the number of siblings. Hourly fees are found to increase with income and decrease with the number of siblings (per child). E.g. for a full-time slot for a 0-2 year-old child, the marginal price is 3%: for a 100 Euro increase in earnings, the monthly full-time childcare fee increases by 3 Euro.[25]

Taxes. We use the Matlab implementation of the German tax and transfer code provided by Bick et al. (2019) to map gross to net income and calculate tax revenues. 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.[26]

Pensions. We approximate the German pension system by assuming that households receive 40% of both partners' last period's potential gross full-time earnings (OECD 2017).

Interest rate. We set the real interest rate of the government to 6% per 3-year model period, which corresponds approximately to 2% per annum.

[25] Since subsidies also vary between regions, we ran these regressions with state dummies and dummies for living in an urban region as a robustness check. The slope coefficients on income were very similar.

[26] When calculating the fiscal effects of changes in labor supply, we account for the sum of income tax payments, social security contributions for public sickness and care insurance, and solidarity surcharge payments. We disregard social-security spending because the German Bismarckian pension system implies pension benefits that are proportional to social security contributions paid; there is no concavity in the benefit formula as e.g. in the U.S. Aside from a precise implementation of the non-linearities of the tax code, it includes joint taxation of couples as well as child benefits for each child in the household. Marginal tax rates faced by women vary with their spouses' income and child allowances reduce the taxable income of the household.

4.1.2 Estimation of the fertility process

We estimate the fertility process in Appendix C.2 as transition probabilities consistent with our model assumptions set out in Section 3.1. We use the 2014 and 2018 Microcensus waves and focus on births taking place from 2012 to 2017 (see Appendix C.2 for more details on the sample). Figure 3 illustrates our estimates in terms 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, 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 low-educated women having children earlier. By age 34 (respectively 37) for the low (respectively high) education group, the majority of households have completed their fertility.

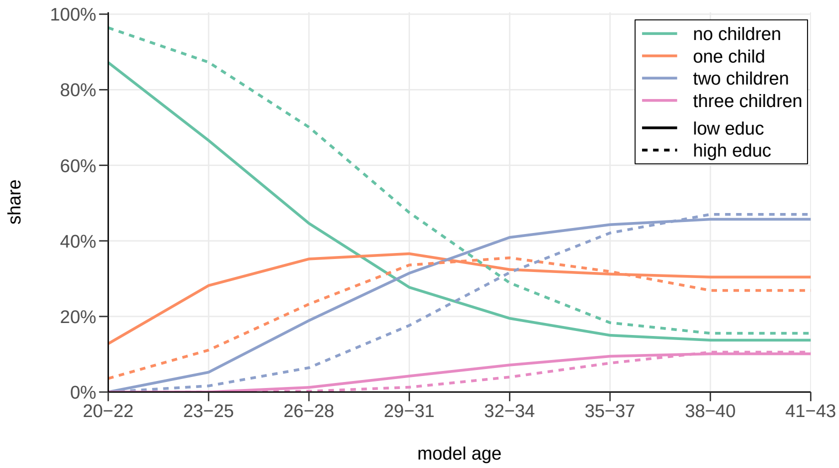


Figure 3: Family composition as implied by the fertility process

Note: 'low educ' corresponds to no A-level, 'high educ' corresponds to having obtained an A-level. Sources: FDZ-StABL (2020a), FDZ-StABL (2020b).

4.1.3 Estimation of the wage process

We estimate the following equation for the wage process of women:

$$
\begin{aligned}
log(w_{f;it}) = \alpha + \beta_1 log(w_{f;it-1}) + \beta_2 \mathbb{1}\{flm_{it-1} = NP\} + \\
\beta_3 \mathbb{1}\{flm_{it-1} = PT\} + \beta_4 educ_i + A(t) + \varepsilon^{w_f}_{it};
\end{aligned}
$$

where $\mathbb{1}\{flm_{it-1} = NP\}$ and $\mathbb{1}\{flm_{it-1} = PT\}$ are dummy variables that indicate whether a woman i was either not working or working part-time in period $t-1$ and $A(t)$ is a third-order polynomial in age.[27] The GSOEP sample is an extended version of our structural sample that we describe in detail in Section 4.2.1. To increase the power of this regression, we consider a larger time span, namely 2000-2017, see Appendix C.3.2.

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.[28] In Figure 4,

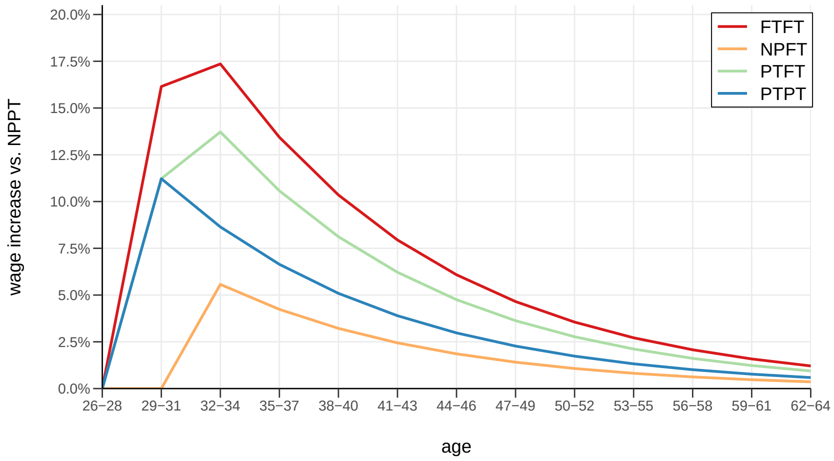


Figure 4: Illustration of the female wage process

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

we illustrate our wage process estimates by looking at the benefits of increasing labor supply relative to a typical labor supply pattern of mothers. Specifically, we consider a mother who has her first child at 26. The benchmark is that she does not work while the child is 0 – 2, works part-time when the child is 3 – 5, and works full-time afterwards. The graph illustrates the dynamic wage gains that the mother would obtain if she increased her labor supply relative to the benchmark. 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 when the child is 3 – 5. Aside from the substantial wage gains from increasing labor supply, the graph clearly illustrates that the potential wage gains are quite persistent. Finally, we also estimate the male wage process in a similar fashion (see Appendix-Table C.2), but without part-time or non-employment penalties since we focus on full-time working males.

[27] We have omitted here the selection term, but we describe in Appendix C.3.1 the detail of our joint estimation of wages and participation into work à la Heckman.

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

4.2 Structural estimation

4.2.1 Estimation sample

We focus on German mothers aged between 20 and 65 who are currently not in education and share a household with a full-time working partner. We track this group over the time span 2012 to 2017 in a representative longitudinal survey data set, the German Socio-Economic Panel (GSOEP). We allocate all children into the corresponding model age brackets and the household into the corresponding child-age structure . We only keep households with complete information for two model periods and average all household variables of interest within each assigned period. This leaves us with an estimation sample of 2,182 households. 1,076 of these households face some childcare needs in at least one of the two periods. The other half of the sample does not face childcare needs in either period as their children are aged 9 or older in both periods. Nevertheless, we keep these in the estimation sample as they help to identify heterogeneity in leisure preferences . Further details on the data and the assignment procedure can be found in Appendix D.1 and summary statistics on the sample are presented in Table 2. To understand the role of last four rows, we refer the reader to the paragraph ‘Constant characteristics x’ in Section 4.2.3.

Table 2: Summary statistics for the MLE sample

	mothers of 0−8 year olds	mothers of 9+ year olds
age	36.10	50.54
$w^{h\,9e0}$	23.56	23.94
$w^{T0\,h\,9e0}$	17.26	16.16
share high education	51%	32%
number of 0-9 children	1.28	
age of youngest child	3.41	
share living in former East	20%	27%
share demanding occupation	43%	36%
share catholic	29%	32%
share urban	64%	60%
N	1,076	1,106

i n§0 : ‘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).

4.2.2 Parameters set externally

Table 3: Parameters set externally

parameter		$_c$	$_D$	$_L$	$\overline{L}$	$\overline{D}$	
value	0.94	1	1	1.75	1	1	0.075

In line with Blundell et al. (2016), we set the discount factor to 0.94 per 3-year model period. For $_c$ we follow common practice in macroeconomics and assume that consumption enters the utility function in a logarithmic manner (Bick and Fuchs-Schündeln 2018, Guner, Kaygusuz, and Ventura 2020). We also use a logarithmic functional for domestic childcare, i.e. set $_D = 1$. To avoid Inada conditions we set a floor value of 1 weekly hour for both L and D.

This leaves us with two further parameters to set: $_L$ and . We calibrate these parameters in the following fashion: we chose the value of $_L$ to obtain a Hicksian intensive margin elasticity that comes close to the value of 0.33 that is widely used in the public finance literature and goes back to Chetty (2012). Furthermore, the shifter for the preference for domestic childcare is set such that the model predicts well how childcare demand differs by the age of the youngest child in the family.[29] The value of 0.075 implies that households have a much stronger preference to spend time with children below age 3 compared to 3 to 8 year old children. Besides having a stronger preference to be with the child while it is young, this could also be considered as a reduced form for social norms.[30]

4.2.3 Maximum likelihood estimation of heterogeneous preferences

Our data comprises observations of female labor supply H_p and total public childcare take-up of the household $m_p^h(_p)$ for two model periods $p = 1; 2$. We estimate the distributions of unobserved heterogeneities $h = (g; l; \ )$ to maximize the likelihood of these dynamic choices. In practice, we estimate distributions of $(g; l; \ )$ conditional on observables x, which we introduce below. As a first step, we now build up the likelihood function step by step and start with measurement-error components.

Measurement Error. We allow for some measurement error in the wages of both spouses and in the amount of public childcare consumed. We denote observed wages and total public

[29]For computational reasons we did not include the estimation of into our maximum likelihood estimation that we describe below. Instead we conducted the MLE conditional on different values of . We iterated over several values and picked the one which implied the best model fit in terms of childcare.

[30]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).

childcare as $(\tilde{w}, \tilde{N}^h)$, in contrast to the 'true' quantities (w, N^h).[31] Since it is measured in a discrete manner, the labor supply of mothers is assumed to be error-free. We denote the errors as follows:

$$\begin{aligned} \log(\tilde{w}_{p,q}) &= \log(w_{p,q}) + \epsilon_{p,q} \text{ with } |\epsilon_{p,q}| \sim EV_{\lambda\alpha} \\ \tilde{N}^h_p &= N^h_p + u_p \text{ with } |u_p| \sim EV_{\lambda\alpha} \end{aligned}$$

for $q = f, m$ and $p = 1, 2$. All measurement errors are assumed to be distributed as type II extreme value distributions.[32] We further assume that they are independent so that the joint distribution of errors is:

$$f(\epsilon_{p,f}, \epsilon_{p,m}, u_p) = ev_w(\epsilon_{p,f}) \, ev_w(\epsilon_{p,m}) \, ev_T(u_p)$$

where $ev_d(.)$, for $d = w, T$ denotes the density of the type II extreme value distribution for wages and childcare hours respectively. The likelihood of observing the choices of a household in period p can then be written as:

$$\ell(H_p, \tilde{N}^h_p | \tilde{s}_p, h) = \int\int\int \ell(H_p, N^h_p | s_p, h) \, f(\epsilon_{p,f}, \epsilon_{p,m}, u_p) \, d\epsilon_{p,m} d\epsilon_{p,f} du_p.$$

where $\tilde{s}_p$ denotes the time varying state space including the observed wages $\tilde{w}_{p,f}$ and $\tilde{w}_{p,m}$. Note that for this intermediary step, we condition on unobservables h.

The likelihood of the 'true' choices (H_p, N^h_p) matching the model predictions $\hat{H}_p$ and $\hat{N}^h_p$ for a household with characteristics (s_p, h) is:

$$\ell(H_p, N^h_p | s_p, h) = \begin{cases} 1 & \text{if} \quad \hat{H}_p(s_p, h) = H_p \text{ and } \hat{N}^h_p(s_p, h) = N^h_p, \\ 0 & \text{otherwise.} \end{cases}$$

The likelihood of observing a household's sequence of choices $(H, \tilde{N}^h)$ given the full set of time-varying characteristics $\tilde{s} = (\tilde{s}_1, \tilde{s}_2)$ and unobserved heterogeneity h is thus:

$$\ell(H, \tilde{N}^h | \tilde{s}, h) = \prod_{p=1}^{2} \ell(H_p, \tilde{N}^h_p | \tilde{s}_p, h).$$

[31]For our model, we discretize weekly hours of public childcare in 2.5 hours steps, i.e. $N \in \{0, 2.5, 5, ..., 40\}$. We assign the values observed in the data to these discretized values.

[32]For the measurement error in wages, we set the scale and shape parameters to $\lambda = 0.026$ and $\alpha = 0.5$ to ensure that 90% (respectively 95%) of errors are no more than 20% (respectively 40%). This is in line with the literature (e.g. Blundell et al. (2016)). The calibration of the measurement error in nursery hours is such that 90% of the errors are no more than 5.6 hours ($\lambda = 1.05$ and $\alpha = 0.5$).

Our object of interest is the joint distribution of unobserved heterogeneity $\ell(h|x)$ conditional on a set of constant household characteristics, denoted x. The likelihood of observing a household's sequence of choices $(H, \tilde{N}^h)$ conditional on observed characteristics is given by the following expression:

$$\ell(H, \tilde{N}^h | s, x) = \int_h \ell(H, \tilde{N}^h | s, h) \cdot \ell(h|x) \, dh.$$

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

$$L = \prod_{n=1}^{N} \ell(H^n, \tilde{N}^{h^n} | s^n, x^n). \tag{21}$$

Joint distribution of unobserved heterogeneity. We now zoom into the joint distribution of unobserved heterogeneity $\ell(h|x)$. We assume the marginal distributions of g, l, and , to be independent conditional on constant characteristics x:

$$\ell(\underbrace{g, l, \ }_{=h} | x) = \ell^g(g|x^g) \cdot \ell^l(l|x^l) \cdot \ell \ (\ | x \ );$$

where $x^g, x^l,$ and x are subsets of x that are allowed to overlap. Such overlap creates correlations between the marginal distributions $\ell^g, \ell^l,$ and ℓ without assuming an explicit correlational structure. We assume each type of heterogeneity, $het \in \{g, l, \ \}$, to be normally distributed conditional on x^{het}:[33]

$$het \, | x^{het} \quad N\left({}^{het} + x^{het} \, {}^{het}, 1 \right).$$

Our maximum likelihood procedure will estimate the parameters (^{g}, ^{g}, , ^{l}, ^{l}, ,) which maximize the sample likelihood function given in equation (21).

Constant characteristics x. The fact that we allow the joint distribution of unobserved heterogeneity h to vary with characteristics x is important in multiple ways: The characteristics in x allow us to capture that subgroups in our data may have very different preferences, thereby improving the capability of our model to predict the behavioral patterns in the data. These constant characteristics also help us address the initial conditions problem, i.e., that the time-invariant joint distribution of unobserved heterogeneity might have affected the initial values

[33]Each dimension of heterogeneity, g, l, and , is defined on the closed interval $[0, 1]$ as set up in Section 3. Therefore, we truncate the normal distribution of $het \, | x^{het}$ at 0 and 1.

of our time-varying state variables. Summary statistics of the variables in x can be found in Table 2.

The variables selected in x^g are indicators variables for living in East Germany, for being Catholic at age 20, and for having primarily worked in a 'demanding occupation'.[34] The variables selected in x^I are indicators of East/West Germany, maternal education and whether the household lives in an urban area.[35] Finally, the variables included in x are maternal education and 'demanding occupation'.

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.

There are three interacting ingredients that identify the time-invariant unobserved heterogeneity: i) cross-sectional variation in choices conditional on observed states, ii) the longitudinal dimension of our panel data, iii) using data for both households with small children and those with older children. Appendix D.2 describes these three ingredients in more detail and argues that our model is credibly identified.

5 Estimation results

5.1 Results

Since our estimated coefficients do not carry much intuitive meaning, we show them in Appendix D.4, where details on the optimization routine and on the sensitivity of the estimates are also discussed. We comment here on the main features of our estimates.

First, considering the preference for domestic childcare (g), women who live in former East Germany have lower preferences for domestic childcare than those in former West Germany. Additionally, for Catholic mothers, we observe a higher preference for domestic childcare. Second, we find large differences in the distribution of the availability of informal childcare (I) between East and West Germany. Only very few households in East Germany rely on informal childcare, see also Footnote 35. Focusing on West Germany, the coefficients show that lower educated mothers as well as those not living in urban areas have a higher availability of informal childcare.

[34] This 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. 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).

[35] More specifically, we estimate the distribution of the availability of informal childcare x^I separately for East and West Germany. Social norms about childcare differ significantly between both regions, which likely affects the availability of grandparental childcare. See Hank, Tillmann, and Wagner (2001) for a discussion. For households living in East Germany, we only estimate the intercept.

Third, highly educated women and those having primarily worked in a demanding occupation tend to have a lower preference for leisure ( ).

5.2 Model fit

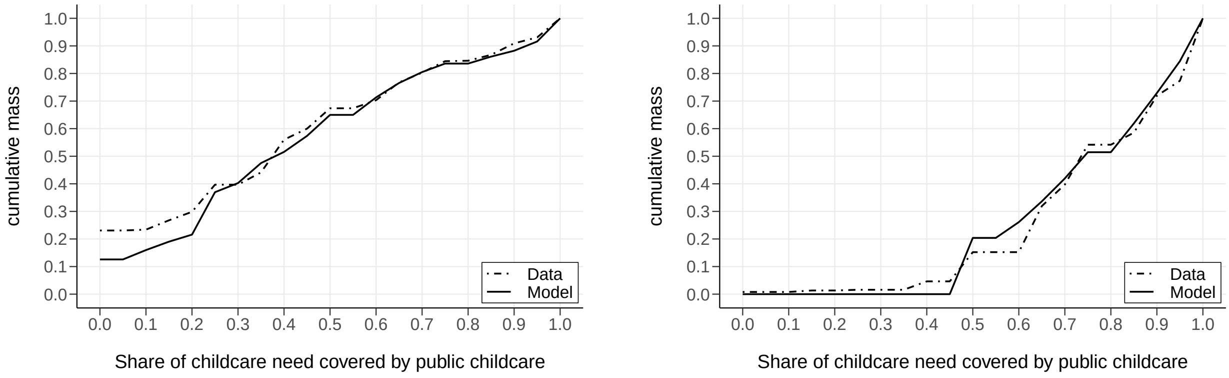
We now turn to the model fit. Specifically, we evaluate the ability of the estimated model to match data moments for the two choices, female labor supply H and household use of public childcare N^h, as a function of the age of the youngest child. As shown in Table 4, overall, we are able to achieve a good fit of the labor supply patterns of mothers conditional on children's ages. In particular, the model matches the observed increase in labor supply once the youngest child turns 3. Our model is also able to match the labor supply pattern of mothers with completed fertility, i.e., children aged 9 or older.

Table 4: Model fit for labor supply

	Children 0 – 2			Children 3 – 5		
	NP	PT	FT	NP	PT	FT
Model	0.49	0.44	0.07	0.14	0.64	0.22
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	0.14	0.64	0.22	0.12	0.64	0.24
Data	0.13	0.68	0.19	0.16	0.54	0.30

i n§0 : PT and FT denote the female working part-time and full-time, respectively. NP denotes non participation. Sample as defined in Section 4.2.3. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

Figure 5 illustrates the fit of our model in terms of public childcare demand by the age of the 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 account for this, we plot the share of childcare needs that a household covers with public childcare $m(\ )$. In general, the model fits the data well. It slightly underpredicts the share of households with zero public childcare if their youngest child is below 3, see Figure 5a. The reason is that our model also allows for quite small amounts of public childcare (starting with 2.5 hours per week) and therefore few households choose a corner solution of 0 hours. Figure E.5 presents the fit for families with the youngest child between 6 and 8.



(a) Youngest child in the family aged $0-2$ (b) Youngest child in the family aged $3-5$

Figure 5: Model fit of families' public childcare demand

i n§0 : Sample as defined in Section 4.2.3. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

5.3 External validity

Our policy experiments in Section 6 below rely on our model predicted responses in terms of female labor supply and public childcare take-up to policy changes. To bring external validity to our policy counterfactuals, we compare the behavioral responses produced by our model with estimates found in the literature.

First, we use our estimates to compute compensated labor supply elasticities at the intensive margin and obtain a value of 0.23. Separating the sample by household income, we find a value of 0.18 for women in below-median income households and 0.30 for above-median income households. As laid out in Section 4.2.2, we chose the curvature parameter on leisure to obtain values that come as close possible to the literature value of 0.33 (Chetty 2012, Chetty et al. 2011). Given that we constrain the intensive margin response to be only about the part-time vs. full-time margin, it appears reasonable that we stay a little below that number. We also simulate participation elasticities with respect to wage changes and obtain numbers of 0.26 below the median and 0.08 above the median. The average participation elasticity is 0.16. This is in line with the empirical estimates of the participation elasticity found in the quasi-experimental literature and surveyed in Chetty et al. (2011) which are in a range of 0.15 to 0.24 for (non-single) women. Finally, we simulate the propensity to earn out of one unearned Euro. We find that a 1000 Euro increase in unearned income decreases earnings in that period by 213 Euro. This number is higher than the values found in Cesarini et al. (2017) for lottery winners in Sweden and lower than the numbers estimated by Golosov et al. (2021) in the U.S. context.

Second, we consider responses to changes in childcare prices, for which there is much less clear-cut evidence. Gathmann and Sass (2018) and Busse and Gathmann (2020) both provide recent evidence in the German context. Gathmann and Sass (2018) consider the introduction of a so-called homecare subsidy, whereas Busse and Gathmann (2020) use a staggered introduction of free childcare in some German states. We implement such reforms in our model and compare the implied response along the childcare and labor supply margin. As we now argue, our model is broadly consistent with their evidence.

We find a decrease in labor force participation of 6.4 percent for the affected population if we introduce the homecare subsidy in our model, which is close to the estimate of 5.1 percent by Gathmann and Sass (2018). Regarding the Busse and Gathmann (2020), they find that, when they control for the number of offered slots, the labor force participation response is 10% but with large standard errors.[36] Our model predicts an increase of 3.4%, which is well within the confidence interval of their estimate.

Turning to responses in childcare demand, we find an 8.3% decrease in childcare use due to the introduction of the home subsidy. This is significantly lower than the 23% reduction in childcare demand from the Gathmann and Sass (2018) estimates. Our model performs much better for the abolition of fees. Busse and Gathmann (2020) find that the introduction of free childcare increases childcare take up by 9.6% for parents with a child between 2 and 3 years. Making childcare free in our model, we find increases of 6.8%. While this figure is smaller than the one found by Busse and Gathmann (2020), we note that our findings below would be accentuated rather then reversed if our prediction was closer to theirs.[37]

6 Quantifying the leaks

We are now in a position to quantify the different components of the MECR for both policy instruments. As discussed in Section 2 these depend on the responses of households across the income distribution to the altered incentives to supply labor and demand childcare services. Our estimates of the structural model presented in Section 5 give us a rich picture of the heterogeneity in these responses, which we now integrate over households below and above the p-th percentile to quantify the leaks of the "redistribution buckets" (8) and (9) and decompose them into the different components described in Propositions 1 and 2.

In practice, we simulate the tax and subsidy reforms as described in (6) and (7) in our quantitative model. We implement them as temporary reforms (i.e. they apply for one three-year model period) and apply the reform of the childcare subsidy schedule to childcare for all three child-age groups, $0-2$, $3-5$, and $6-8$. We design them such that they are dynamically budget neutral: we account for long-run budgetary effects and take a net-present value perspective on

[36]We use the estimates from Table 4 of their paper, where they isolate price effects from rationing.

[37]To be precise, we define the public childcare demand extensive margin as the fraction of households increasing their demand from covering less than half of their needs to over half of their needs. The reason is that in our model there are few true zeros because we allow for very small amounts of childcare starting from 2.5 hours per week, see Figure 5a.

the government budget constraint. For this purpose, we simulate the model until the end of the lifecycle for all households in our sample and account for all earnings and childcare demand changes in the future.[38]

In Section 6.1, our benchmark is to conduct these reforms around the median household income. In Section 6.2 we take into account child-development effects of public childcare. Finally, in Section 6.3, we show how our model can be used to assess redistributive reforms around other income percentiles than the median.

6.1 Redistribution from above to below median income

Our main finding is as follows: if we abstract from child development, we find $MECR(\hat{\tau}_{50}) = 0.28$ and $MECR(\hat{s}_{50}) = 0.42$: for each Euro taken from the top half of the income distribution, 72 Cents reach the bottom of the distribution for the tax reform, and 58 Cents do for the childcare subsidy reform. Hence, the marginal efficiency cost of redistribution is 50% larger for the childcare subsidy schedule than for the income tax schedule. To understand the sources of the leakages, we decompose the MECR into its different components in Table 5.

Table 5: Decomposition of MECR

MECR Decomposition	**Tax Reform**			**Subsidies Reform**		
	$> y_{50}$	$\leq y_{50}$	$\sum$	$> y_{50}$	$\leq y_{50}$	$\sum$
Labor Supply	0.21	0.13	0.34	0.43	0.05	0.47
Childcare Demand	-0.04	-0.03	-0.06	-0.14	0.09	-0.05
Total	0.17	0.10	**0.28**	0.28	0.14	**0.42**

The first row captures the leaks which are due to changes in labor supply. For the tax reform, we find that this number is 0.34 – with 0.21 (respectively 0.13) coming from labor supply effects above (respectively below) the median. This quantifies the own-price effect of the change in the net wage O_w, see part 1(a) of Proposition 1. The steeper tax schedule acts as a disincentive to supply labor throughout the distribution.[39] For the childcare subsidy reform, the leak due to labor supply responses is significantly larger at 0.47. Most of this, 0.43, can be attributed to above-median households. This number is higher than for the tax reform because in addition to the own-price effect on labor supply O_w, these households also work less due to a cross-price effect X_K, see part 2(b) of Proposition 2: childcare becomes more expensive. The labor supply

[38]Note that we are not considering a balanced cohort. In our policy simulation, we include all households who have a 0-8 years-old child between 2012-2017, i.e. those in the left column of Table 2. We then implement our reform for the period 2012-2014.

[39]Note that in contrast to the model in Section 2, income effects are at work. They mitigate the own-price effect on labor supply above the median and reinforce it below the median.

leakage among households below the median income is small but positive. Hence, the incentives to work more provided by the cheaper childcare price X_K are more than offset by the disincentive to work due to the steeper childcare price schedule O_w. The larger magnitude of the leaks for the above-median income household comes mostly from the fact that both incomes and tax rates are larger in this group, hence fiscal externalities are larger even when cross-price elasticities are of similar size (see Section 5.3).

The second row shows the magnitude of the leaks occurring as a result of households' changes in their demand for public childcare. For the tax reform, these are the result of the cross-price effect of a change in the net wage, X_w: as households decrease their labor supply, they also decrease their use of nursery services – although some of the change in working hours is absorbed by changes in leisure. This is a saving for the government budget and partially mitigates the leaks occurring via the labor supply channel in the first row. Now looking at the right-hand side panel of Table 5 for the reform to childcare subsidies, the -0.14 decrease in the MECR coming from the childcare adjustments of households above the median income represents the sum of the own-price effect of the price of childcare O_K (better-off households consume less childcare as it is more expensive) and of the cross-price effect of the net wage X_w (as the childcare fee schedule has become steeper in the reform, the net wage is lower at the margin). Both these adjustments create savings in terms of childcare subsidies and a reduction in the MECR. For households below the median, childcare has become cheaper and the own-price effect O_K constitutes a leak, which is only partly offset by the cross-price effect of the net wage, X_w, so that adjustments to childcare demand for this group add 0.09 to the MECR.

Comparing the contributions to the MECR coming from behavioral adjustments by households above and below the median in the last row of Table 5, we see that the bulk of the difference in MECRs between the two reforms comes from adjustments by the better-off households: although these consume less public childcare, the fiscal externality caused by their decreased labor supply is large.

Table 6: Decomposition of labor supply effects into static and dynamic components

MECR Decomposition	**Income Tax**			**Childcare Subsidies**		
	$> y_{50}$	y_{50}	$\sum$	$> y_{50}$	y_{50}	$\sum$
Static labor supply	0.15	0.09	0.24	0.30	0.03	0.34
Dynamic labor supply	0.06	0.04	0.11	0.12	0.01	0.14
Total	0.21	0.13	0.34	0.43	0.05	0.47

Another angle we can take to better understand the sources of the leaks is to decompose the labor supply effects into static and dynamic components.[40] This is what we show in Table 6. We define as static the part of the leak that happens in the period of the reform: i.e. how do changes in labor supply in that period directly affect the leakage. Given the dynamic wage effects discussed above, these changes will also affect earnings in the future: both through wage effects and through labor supply effects implied by those wage changes. This will affect the budget in the future and hence also imply a leakage, since our reforms are budget neutral in the dynamic sense, i.e. in terms of the long-term government budget. Examining all the columns in Table 6, we see that the static effect is a bit more than twice as large as the dynamic effect.

6.2 Accounting for child development

We now incorporate effects of the mode of childcare on child outcomes into the MECR analysis as described in Section 2.5. While our data does not allow us to measure child outcomes, our model delivers counterfactual predictions on childcare demand responses to reforms – and as discussed in Section 5.3, the model is broadly consistent with quasi-experimental evidence in this regard. To translate these changes in childcare demand into changes in child outcomes, we extrapolate on the influential analysis by Havnes and Mogstad (2011, 2015) who consider a large public childcare expansion in Norway. In a first step, we translate their results into returns per year of full-time childcare: these range from 6% for the poorest children to about -2% for the richest children, as illustrated in Figure 6. These different returns are in line with the hypothesis that public child acts as an equalizer (Cornelissen et al. 2018).

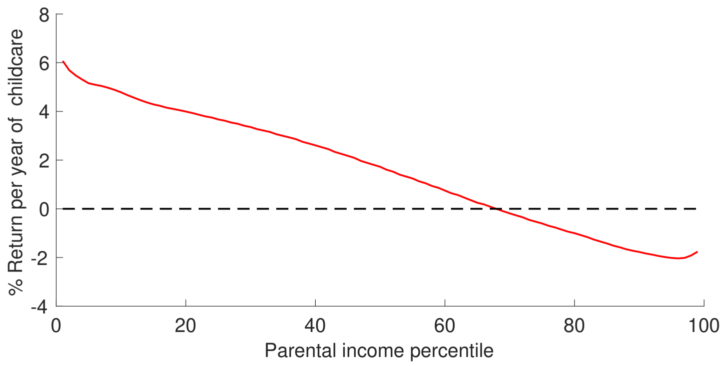


Figure 6: Long-term returns to childcare attendance

Notes: This figure shows the return to children's earnings to one year of full-time public childcare attendance. The numbers are based on Havnes and Mogstad (2015), see Appendix F for details.

[40]We refrain from decomposing childcare responses into dynamic and static components. The reason is that the dynamic component is very small.

Table 7: MECR for different considerations of child development

Type of MECR	^{c}	Tax Reform	Subsidy Reform
$MECR$ with no child development	-	$0{:}28$	$0{:}42$
$MECR^c$ with F.E.	0	$0{:}28$	$0{:}35$
$MECR^c$ with F.E. and N.I.	1	$0{:}31$	$0{:}12$
$MECR^c$ with F.E. and N.I.	$0{:}28$	$0{:}29$	$0{:}29$

i n§0 : F.E. = fiscal externality, N.I. = net-income increases, both due to child development.

Within our setting, we assume that these returns apply to children attending public childcare. To incorporate this into the MECR analysis, we first need to obtain the implied increases in lifetime earnings of children. For this purpose, we use recent estimates for Germany from Dodin et al. (2022) to obtain average child earnings as a function of the parental income rank when the parents were young. We augment these numbers with lifecycle profiles from Bönke, Corneo, and Lüthen (2015). This yields lifecycle earnings profiles as a function of the parental income rank. Combining these with the returns illustrated in Figure 6, we obtain the increase in the net-present value of lifetime earnings due to public childcare attendance. All steps are explained in detail in Appendix F.[41]

In line with Section 2.5, we now discuss two ways to incorporate child development. Our first step is to consider only the fiscal externalities imposed by child development.[42] As can be seen in Table 7, this reduces $MECR(\hat{s}_{50})$ from 0.42 to 0.35 and leaves $MECR(\hat{\tau}_{50})$ almost unaffected 0.30. The reason the $MECR(\hat{s}_{50})$ decreases is that low-income children gain from increased public childcare attendance. This gets amplified since some of the above-median income children also gain from the implied reduction in public childcare attendance. For the tax reform, both income groups decrease childcare attendance, and the implied negative effect on child development of low-income children roughly offsets the effect on high-income children. In sum, the difference between $MECR(\hat{s}_{50})$ and $MECR(\hat{\tau}_{50})$ shrinks from 0.14 to 0.10 due to the fiscal externalities from child development.

In the next step, we account for the change in child utility through the change in the net-present value of after-tax lifetime earnings. We compute the augmented definitions of the MECR defined in (14) and (15), and first consider the same Pareto weight on children as on parents, i.e. with $_{child} = 1$. As can be seen in the third row of Table 7, this slightly increases the MECR for the tax reform to 0.31 and significantly reduces the MECR of the subsidy reform

41 Arguably, these returns to nursery attendance along the distribution of parental income are subject to debate. Our computation of the MECR based on our structural estimates could easily be tailored to alternative values for the returns to nursery attendance in terms of children's lifetime income. E.g. we could consider a flatter curve in Figure 6 or shift it up or down.

42 We take the same tax function as we assume for parents. We assume that children are singles or singe-earners in a married couple. Results only barely change if we assume that they are married to a partner with positive income.

to 0.12. This sharp reduction comes from the fact that the lifetime utility of the children born in households below the median is increased following their increased nursery attendance. This offsets 0.23 of the leaks attached to the childcare subsidy reform and make this reform the more efficient redistribution instrument at the margin, with a total MECR of 0.12 versus 0.31 for the tax reform.

As a last step, we compute the relative Pareto weight on children, as defined in (14) and (15), for which the reforms imply the same MECR. We find this value to be $_{child} = 0{:}28$. Hence, as long as $_{child}$ is larger than 28% of the parent's weight, the policy implications are that the childcare subsidy schedule is the more efficient redistribution instrument at the margin.

6.3 Redistribution at other percentiles

We now use the flexibility of our approach and consider other threshold values for redistribution: we discuss here the MECR of redistributive reforms around the 30th and the 70th percentile. The left panel of Figure 7 illustrates the case of $p = 30$ and the right panel the case of $p = 30$. Both can be directly compared to the case of $p = 50$ (seen above) in the middle panel of Figure 7. The patterns are very similar. If child development is not accounted for (blue bar), the MECR of the tax system is lower also for $p = 30$ and $p = 70$. The difference, however, is much larger for $p = 70$. If child development is accounted for through fiscal externalities and through the impact on the compensating variation of the children, the ranking changes and the childcare subsidy turns out to be the more efficient instrument for redistribution at the margin for both, $p = 30$ and $p = 70$.[43]

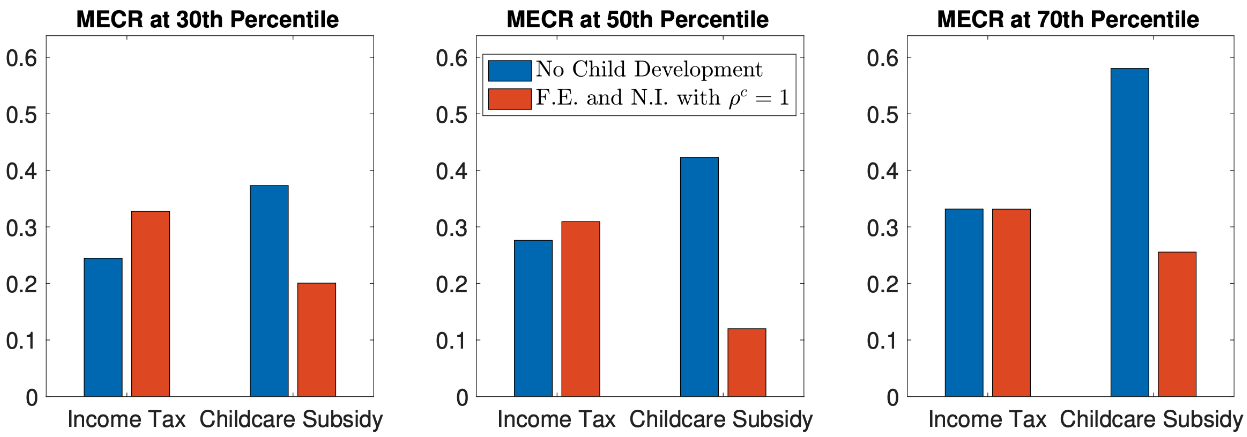


Figure 7: MECR for reforms at different percentiles

i n§0 : The blue bars illustrate $MECR(\hat{\tau}_p)$ and $MECR(\hat{s}_p)$ as defined in (8) and (9). The red bars illustrate $MECR(\hat{\tau}_p; 1)$ and $MECR(\hat{s}_p; 1)$ as defined in (14) and (15). The left panel refers to $p = 30$, the middle panel to $p = 50$ and the right panel to $p = 70$. F.E. = fiscal externality, N.I. = net-income increases, both due to child development.

[43]Note that we illustrate the case with equal Pareto weights on children and parents here.

7 Conclusion

In this paper, we have incorporated the public finance approach of quantifying the efficiency costs of redistribution into a dynamic structural model of labor supply. Our measure of efficiency costs formalizes the intuitive notion of Okun's leaky bucket (Okun 1975) in that behavioral adjustments in response to a reform cause some leaks in the redistribution process. We have compared the MECR of income-contingent childcare subsidies and of the income tax.

We have identified competing effects coming from maternal labor supply and child development. The maternal labor supply channel increases the MECR of childcare subsidies relative to the income tax. The child development channel decreases these MECR relative to the income tax. If one puts reasonably high Pareto weight on children and assumes returns to public childcare attendance that are in line with the quasi-experimental literature, childcare subsidies turn out to be the more efficient tool to redistribute from high-income to low-income families: for one Euro taken from higher-income families, a greater amount reaches lower-income families.

Our decomposition of the efficiency costs makes it clear that – besides its immediate fiscal externalities – redistribution has a medium-term impact on the fiscal budget in the next 10-20 years via the impact of labour supply adjustments on mothers' future wages and a long-term impact on on the income distribution of the children and the fiscal budget in 20-50 years' time via the impact of nursery demand adjustments on children's future earnings. Which of these effects is taken into consideration by the policymakers depends on the horizon over which they weigh the costs and benefits of potential reforms. And this will affect which tool is considered more efficient in achieving redistribution.

The general approach laid out in this paper could also be used to evaluate the dynamic efficiency costs of redistribution of other government policies such as the pension system, social housing, health insurance, etc. More conceptually, our paper also shows that one can harmonize equity concerns and social mobility concerns into one measure and therefore take a comprehensive perspective on policies that have distributional consequences on both the parent and the child generations.

References

Adam, S. (2005). "Measuring the marginal efficiency cost of redistribution in the UK". Tech. rep. IFS Working Papers.

Adda, J., C. Dustmann, and K. Stevens (2017). "The Career Costs of Children". *`n«}j9e nT pne_§_H9e NHnjnh´* 125.2, 293–337.

Arnoud, A., F. Guvenen, and T. Kleineberg (2019). "Benchmarking global optimizers". Tech. rep. National Bureau of Economic Research.

Attanasio, O., H. Low, and V. Sánchez-Marcos (2008). "Explaining changes in female labor supply in a life-cycle model". *8h0}_H9j NHnjnh_H |0®_0°* 98.4, 1517–1552.

Authoring Group Educational Reporting (2018). *NL«H9§_nj _j W0}h9j´ ©·oP*.

Ballard, C. L. (1988). "The marginal efficiency cost of redistribution". *¦\0 8h0}_H9j NHnjnh_H |0®_0°*, 1019–1033.

Bargain, O., K. Orsini, and A. Peichl (2014). "Comparing Labor Supply Elasticities in Europe and the United States New Results". *`n«}j9e nT [«h9j |0 n«}H0* 49.3, 723–838.

Bastani, S., S. Blomquist, and L. Micheletto (2020). "Child Care Subsidies, Quality, and Optimal Income Taxation". *8h0}_H9j NHnjnh_H `n«}j9eI NHnjnh_H pne_H´* 12.4, 1–37.

Bick, A. (2016). "The Quantitative Role of Child Care for Female Labor Force Participation and Fertility". *`n«}j9e nT §\0 N«}nq09j NHnjnh_H 8 nH_9§_nj* 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". *`n«}j9e nT ^j§0}j9§_nj9e NHnjnh_H* 118, 44–62.

Bick, A. and N. Fuchs-Schündeln (2018). "Taxation and labour supply of married couples across countries: A macroeconomic analysis". *¦\0 |0®_0° nT NHnjnh_H ~§«L_0* 85.3, 1543–1576.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). "Female Labor Supply, Human Capital, and Welfare Reform". *NHnjnh0§}_H9* 84.5, 1705–1753.

Blundell, R. and A. Shephard (2012). "Employment, hours of work and the optimal taxation of low-income families". *¦\0 |0®_0° nT NHnjnh_H ~§«L_0* 79.2, 481–510.

Boneva, T. and C. Rauh (2018). "Parental beliefs about returns to educational investments—the later the better?" *`n«}j9e nT §\0 N«}nq09j NHnjnh_H 8 nH_9§_nj* 16.6, 1669–1711.

Bönke, T., G. Corneo, and H. Lüthen (2015). "Lifetime earnings inequality in Germany". *`n«}j9e nT d9@n} NHnjnh_H* 33.1, 171–208.

Bovenberg, A. L. and B. Jacobs (2005). "Redistribution and education subsidies are Siamese twins". *`n«}j9e nT p«@e_H NHnjnh_H* 89.11-12.

Browning, E. K. and W. R. Johnson (1984). "The trade-off between equality and efficiency". *`n«}j9e nT pne_§_H9e NHnjnh´* 92.2, 175–203.

Busse, A. and C. Gathmann (2020). “Free daycare policies, family choices and child development”. *`n«}j9e nT NHnjnh_H ?0\9®_n} : m}X9j_¶9§_nj* 179, 240–260.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2017). “The Effect of Wealth on Individual and Household Labor Supply”. *¦\0 8h0}_H9j NHnjnh_H |0®_0°* 107 (12), 3917–3946.

Chetty, R. (2012). “Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply.” *NHnjnh0§}_H9* 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”. *8h0}_H9j NHnjnh_H |0®_0°* 101.3, 471–475.

Colas, M., S. Findeisen, and D. Sachs (2021). “Optimal Need-Based Financial Aid”. *`n«}j9e nT pne_§_H9e NHnjnh´* 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”. *`n«}j9e nT pne_§_H9e NHnjnh´* 126.6, 2356–2409.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). “Interpreting the evidence on life cycle skill formation”. *[9jL@nnc nT §\0 NHnjnh_H nT NL«H9§_nj* 1, 697–812.

Cunha, F., I. Elo, and J. Culhane (2022). “Maternal subjective expectations about the technology of skill formation predict investments in children one year later”. *`n«}j9e nT NHnjnh0§}_H* 231.1, 3–32.

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).

Dodin, M., S. Findeisen, L. Henkel, D. Sachs, and P. Schüle (2022). “Social Mobility in Germany”. *¯n}c_jX p9q0}*.

Domeij, D. and P. Klein (2013). “Should Day Care be Subsidized?” *¦\0 |0®_0° nT NHnjnh_H ~§«L_0* 80.2, 568–595.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015). “Early childhood education”. *NHnjnh_H nT h09j]§0 §0L §}9j TO} q}nX}9h _j §\0 ª j_§0L ~§9§0 J ®ne«h0 ©*. University of Chicago Press, 235–297.

Farhi, E. and I. Werning (2010). “Progressive Estate Taxation*”. *¦\0 w«9}§0}e´ `n«}j9e nT NHnjnh_H* 125.2, 635–673.

FDZ-SOEP - Forschungsdatenzentrum Sozioökonomisches Panel am Deutschen Institut für Wirtschaftsforschung (2019). *~n¶_n]n0cnjnh_ H\0 p9j0e r~mNpsJ K9§0j L0} `9\}0 okPV]©·o¢*. DOI: 10.5684/soep.v34.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2020a). *g_c}n¶0j « ©·oVJ mj]~_§0]µ«X9jX*. DOI: 10.21242/12211.2014.00.00.1.1.1.

FDZ-StABL - Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (2020b). *g_c}n¶0j « ©·oPJ mj]~_§0]µ«X9jX*. DOI: 10.21242/12211.2018.00.00.1.1.3.

Gathmann, C. and B. Sass (2018). "Taxing Childcare: Effects on Childcare Choices, Family Labor Supply, and Children". *`n«}j9e nT d9@n} NHnjnh_H* 36.3, 665–709.

Gayle, G.-L. and A. Shephard (2019). "Optimal taxation, marriage, home production, and family labor supply". *NHnjnh0§}_H9* 87.1, 291–326.

GESIS - Leibniz-Institut für Sozialwissenschaften (2017). *8eeX0h0_j0 ?0®Ǹec0}«jX «hT}9X0 L0} ~n¶_9e°_ 0j H\9T§0j 8dd?ª~ ©·o£*. DOI: 10.4232/1.12796.

Golosov, M., M. Graber, M. Mogstad, and D. Novgorodsky (2021). "How Americans respond to idiosyncratic and exogenous changes in household wealth and unearned income". Tech. rep. National Bureau of Economic Research.

Golosov, M., A. Tsyvinski, and N. Werquin (2014). "A variational approach to the analysis of tax systems". Tech. rep. National Bureau of Economic Research.

Guner, N., R. Kaygusuz, and G. Ventura (2020). "Child-Related Transfers, Household Labour Supply, and Welfare". *¦\0 |0®_0° nT NHnjnh_H ~§«L_0* 87.5, 2290–2321.

Haan, P. and K. Wrohlich (2011). "Can child care policy encourage employment and fertility?: Evidence from a structural model". *d9@n«} NHnjnh_H* 18.4, 498–512.

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". *µ0_§ H\}_T§ Tƅ} ?0®Ǹec0}«jX °_ 0j H\9T§* 26.1, 55–65.

Hannusch, A. (2022). "Taxing Families: The Impact of Child-related Transfers on Maternal Labor Supply". *¯n}c_jX p9q0}*.

Havnes, T. and M. Mogstad (2011). "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes". *8h0}_H9j NHnjnh_H `n«}j9eI NHnjnh_H pne_H´* 3 (2), 97–129.

Havnes, T. and M. Mogstad (2015). "Is universal child care leveling the playing field?" *`n«}j9e nT p«@e_H NHnjnh_H* 127, 100–114.

Heathcote, J. and H. Tsujiyama (2021). "Optimal income taxation: Mirrlees meets Ramsey". *`n«}j9e nT pne_§_H9e NHnjnh´* 129.11, 3141–3184.

Hendren, N. (2016). "The policy elasticity". *¦9² pne_H´ 9jL §\0 NHnjnh´* 30.1, 51–89.

Hendren, N. and B. Sprung-Keyser (2020). “A Unified Welfare Analysis of Government Policies*”. *¦\0 w«9}§0}e´ `n«}j9e nT NHnjnh_H* 135.3, 1209–1318.

Hicks, J. R. (1939). “The Foundations of Welfare Economics”. *¦\0 NHnjnh_H `n«}j9e* 49.196, 696–712.

Ho, C. and N. Pavoni (2020). “Efficient Child Care Subsidies”. *8h0}_H9j NHnjnh_H |0®_0°* 110.1, 162–199.

Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez (2007). “Welfare reform in European countries: a microsimulation analysis”. *¦\0 NHnjnh_H `n«}j9e* 117.516, 1–44.

Jakobsen, K. M., T. H. Jørgensen, and H. Low (2022). “Fertility and family labor supply”. *¯n}c_jX p9q0}*.

Krueger, D. and A. Ludwig (2016). “On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium”. *`n«}j9e nT gnj0§9}´ NHnjnh_H* 77, 72–98.

Mirrlees, J. A. (1971). “An Exploration in the Theory of Optimum Income Taxation”. *¦\0 |0®_0° nT NHnjnh_H ~§«L_0* 38.2, 175–208.

Mullins, J. (2022). “Designing cash transfers in the presence of children’s human capital formation”. *¯n}c_jX p9q0}*.

OECD (2017). *p0j _nj 9§ 9 We9jH0 ©·o¢*.

Okun, A. M. (1975). *Nx«9e_§´ 9jL N¹H_0jH´I ¦\0 ?_X ¦}9L0n¸*. Brookings Institution Press.

Piketty, T. (1997). “La redistribution fiscale face au chômage”. *|0®«0 T}9jč9_ 0 L¿ĹHnjnh_0* 12 (1), 157–201.

Sachs, D., A. Tsyvinski, and N. Werquin (2020). “Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium”. *NHnjnh0§}_H9* 88 (2), 469–493.

Saez, E. (2001). “Using Elasticities to Derive Optimal Income Tax Rates”. *¦\0 |0®_0° nT NHnjnh_H ~§«L_0* 68 (1), 205–229.

Saez, E. and S. Stantcheva (2018). “A simpler theory of optimal capital taxation”. *`n«}j9e nT p«@e_H NHnjnh_H* 162, 120–142.

Semykina, A. and J. M. Wooldridge (2010). “Estimating panel data models in the presence of endogeneity and selection”. *`n«}j9e nT NHnjnh0§}_H* 157.2, 375–380.

Stantcheva, S. (2017). “Optimal taxation and human capital policies over the life cycle”. *`n«}j9e nT pne_§_H9e NHnjnh´* 125.6, 1931–1990.

Statistisches Bundesamt (2012). *S_j9j¶0j LO} b_jL0}§9X0 0_j}_H\§«jX0j _j T}0_0} ¦}ÞX0} H\9T§ ©·o·*.

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". *9}±_® 0]q}_j§* 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". *¦\0 `n«}j9e nT p\´ _H9e G\0h_ §}´ 8* 101.28, 5111–5116.

Wang, H. (2022). "Fertility and Family Leave Policies in Germany: Optimal Policy Design in a Dynamic Framework". *¯n}c_jX p9q0}*.

Appendix

A Proofs and Derivations of Section 2

In this appendix, we derive the fiscal externalities and the resulting formulas for $MECR(\hat{\tau}_p)$ and $MECR(\hat{s}_p)$. Propositions 1 and 2 then directly follow. We start with the tax reform $\hat{\tau}_p$ and then consider the subsidy reform $\hat{s}_p$.

A.1 Tax reform

We now consider all revenue effects of the tax reform $\hat{\tau}_p$ in order to derive the budget neutral value for $\delta_b(\delta_a)$ First of all, the tax reform has mechanical effects on government revenue given by:

$$\delta_a \int_{i:y_i>y_p} (y_i - y_p)\, di$$

and

$$\delta_b \int_{i:y_i \le y_p} (y_i - y_p)\, di.$$

Fiscal externalities In addition, it has budgetary effects through fiscal externalities. First of all, female labor supply for households above income y_p changes as follows due to increase of the marginal tax rate by δ_a:

$$\frac{\partial H_i}{\partial (1-\tau_i^H)}(-\delta_a) = -\frac{\partial H_i}{\partial w_{f,i}^{net}} w_{f,i}\delta_a = -\varepsilon^i_{H,w_f^{net}} \frac{y_{f,i}}{w_{f,i}^{net}}\delta_a.$$

Multiplying with the female wage and the labor wedges, yields the fiscal externality of this own-price effect on labor supply:

$$-\frac{\tau_i^H}{1-\tau_i^H}\varepsilon^i_{H,w_f^{net}} y_{f,i}\delta_a =: -O_w^i\delta_a. \tag{22}$$

Similarly, for households i with income below y_p the fiscal externality is:

$$O_w^i\delta_b = -\frac{\tau_i^H}{1-\tau_i^H}\varepsilon^i_{H,w_f^{net}} H_i\delta_b.$$

The change in the net-wage that is caused by the reform as changes childcare demand through a cross-price effect. Households i with $y_i > y_p$ change their childcare demand as follows:

$$\frac{\partial N_i}{\partial w_{f,i}^{net}} w_{f,i}\delta_a = -\varepsilon^i_{N,w_f^{net}} \frac{N_i}{1-\tau_i^H}\delta_a.$$

The implied fiscal externality then reads as:

$$\frac{\tau_i^N}{1-\tau_i^H} N_i \varepsilon^i_{N;w_f^{net}} \delta\tau_a =: X_w^i \delta\tau_a. \tag{23}$$

Similarly for households with $y_i < y_p$, it reads as

$$X_w^i \delta\tau_b = \frac{\tau_i^N}{1-\tau_i^H} N_i \varepsilon^i_{N;w_f^{net}} \delta\tau_b.$$

Budget neutrality To obtain $\delta\tau_b(\delta\tau_a)$, we add up all those fiscal effects and equate them to zero:

$$\begin{aligned}
&\delta\tau_a \int_{i:y_i>y_p} (y_i - y_p)\, di + \delta\tau_b \int_{i:y_i \leq y_p} (y_i - y_p)\, di \\
&- \delta\tau_a \int_{i:y_i>y_p} O_w^i di - \delta\tau_b \int_{i:y_i \leq y_p} O_w^i di \\
&+ \delta\tau_a \int_{i:y_i>y_p} X_w^i di + \delta\tau_b \int_{i:y_i \leq y_p} X_w^i di = 0.
\end{aligned}$$

This implies

$$\delta\tau_b(\delta\tau_a) = \delta\tau_a \frac{\int_{i:y_i>y_p} (y_i - y_p)\, di - \int_{i:y_i>y_p} O_w^i di + \int_{i:y_i>y_p} X_w^i di}{\int_{i:y_i \leq y_p} (y_p - y_i)\, di + \int_{i:y_i \leq y_p} O_w^i di - \int_{i:y_i \leq y_p} X_w^i di}.$$

Inserting this into (8) yields:

$$MECR(\hat{\tau}_p) = 1 - \delta\tau_a \frac{\int_{i:y_i>y_p} (y_i - y_p)\, di - \int_{i:y_i>y_p} O_w^i di + \int_{i:y_i>y_p} X_w^i di}{\int_{i:y_i \leq y_p} (y_p - y_i)\, di + \int_{i:y_i \leq y_p} O_w^i di - \int_{i:y_i \leq y_p} X_w^i di} \frac{\int_{i:y_i \leq y_p} [y_p - y_i]\, di}{\int_{i:y_i>y_p} [y_i - y_p]\, di}$$

and hence

$$MECR(\hat{\tau}_p) = 1 - \frac{1 - \frac{\int_{i:y_i>y_p} O_w^i di - \int_{i:y_i>y_p} X_w^i di}{\int_{i:y_i>y_p} (y_i - y_p) di}}{1 + \frac{\int_{i:y_i<y_p} O_w^i di - \int_{i:y_i<y_p} X_w^i di}{\int_{i:y_i \leq y_p} (y_p - y_i) di}}. \tag{24}$$

This result implies the statements in Proposition 1.

A.2 Subsidy reform

We now consider all revenue effects of the tax reform $\hat{s}_p$ in order to derive the budget neutral value for $\delta\tau_b(\delta\tau_a)$. First of all, the subsidy reform has mechanical effects on government revenue given by:

$$\delta\tau_a \int_{i:y_i>y_p} (y_i - y_p)\, N_i di$$

and
$$\tau_b \int_{i: y_i \le y_p} (y_i - y_p) N_i di.$$

Note that here the difference is the multiplication with N_i since the burden or relief implied by this reforms depends on how much public childcare N_i the household demands. In addition, it has budgetary effects through fiscal externalities.

Fiscal externalities due to changes in net wage First of all, female labor supply for households above income y_p changes as follows due to increase of the effective marginal tax rate by $\tau_a N_i$:

$$\frac{\partial H_i}{\partial (1 - \tau_i^H)} (-\tau_a N_i) = \frac{\partial H_i}{\partial w_{f,i}^{net}} w_{f,i} (-\tau_a N_i) = -\varepsilon^i_{H, w_f^{net}} \frac{y_{f,i}}{1 - \tau_i^H} (\tau_a N_i)$$

The fiscal externality is then given by

$$O_w^i \, \tau_a N_i = -\frac{\tau_i^H}{1 - \tau_i^H} \varepsilon^i_{H, w_f^{net}} y_{f,i} \, \tau_a N_i.$$

Similarly, for households with $y_i \le y_p$, the fiscal externality is:

$$O_w^i \, \tau_b N_i = -\frac{\tau_i^H}{1 - \tau_i^H} \varepsilon^i_{H, w_f^{net}} y_{f,i} \, \tau_b N_i.$$

The change in the net-wage that is caused by the reform as changes childcare demand through a cross-price effect. Households i with $y_i > y_p$ change their childcare demand as follows:

$$\frac{\partial N_i}{\partial w_{f,i}^{net}} w_{f,i} \, \tau_a N_i = -\varepsilon^i_{N, w_f^{net}} \frac{N_i}{1 - \tau_i^H} \, \tau_a N_i.$$

The implied fiscal externality then reads as:

$$X_w^i \, \tau_a N_i = -\frac{\tau_i^N}{1 - \tau_i^H} \varepsilon^i_{N, w_f^{net}} N_i \, \tau_a N_i.$$

Similarly for households with $y_i < y_p$, it reads as

$$X_w^i \, \tau_b N_i = -\frac{\tau_i^N}{1 - \tau_i^H} \varepsilon^i_{N, w_f^{net}} N_i \, \tau_b N_i.$$

Fiscal externalities due to changes in net cost The reform $\hat{s}_p$ increases net costs of households with $y_i > y_p$ by $\tau_a (y_i - y_p)$. This as an own-price effect on childcare demand. Childcare demand of households with $y_i > y_p$ changes as follows:

$$\frac{\partial N_i}{\partial K_i^{net}} \tau_a (y_i - y_p) = \varepsilon^i_{N;K_{net}} \frac{N_i}{K_i^{net}} \tau_a (y_i - y_p)$$

The resulting fiscal externality is given by:

$$\frac{\tau_i^N}{K_i^{net}} \varepsilon^i_{N;K_{net}} N_i \tau_a (y_i - y_p) =: O_K^i \tau_a (y_i - y_p). \tag{25}$$

For households with $y_i \leq y_p$, the net-cost LOH}09 0 $\tau_b (y_p - y_i)$by it is given by:

$$O_K^i \tau_b (y_i - y_p) = \frac{\tau_i^N}{K_i^{net}} \varepsilon^i_{N;K_{net}} N_i \tau_b (y_i - y_p).$$

Finally, we turn to related cross-price effects. Female supply of households with $y_i > y_p$ changes as follows

$$\frac{\partial H_i}{\partial K_i^{net}} \tau_a (y_i - y_p) = \varepsilon^i_{H;K_{net}} \frac{H_i}{K_i^{net}} \tau_a (y_i - y_p)$$

This results in the following fiscal externality

$$\frac{\tau_i^H}{K_i^{net}} \varepsilon^i_{H;K_{net}} H_i \tau_a (y_i - y_p) =: X_K^i \tau_a (y_i - y_p). \tag{26}$$

For households with $y_i \leq y_p$, it is given by:

$$X_K^i \tau_b (y_i - y_p) = \frac{\tau_i^H}{K_i^{net}} \varepsilon^i_{H;K_{net}} H_i \tau_b (y_i - y_p).$$

Budget neutrality To obtain $\tau_b(\tau_a)$, we add up all those fiscal effects and equate them to zero:

$$\begin{aligned}
&\tau_a \int_{i:y_i>y_p} N_i (y_i - y_p)\, di + \tau_b \int_{i:y_i \leq y_p} N_i (y_i - y_p)\, di \\
&- \tau_a \int_{i:y_i>y_p} O_w^i N_i di - \tau_b \int_{i:y_i \leq y_p} O_w^i N_i di \\
&- \tau_a \int_{i:y_i>y_p} X_w^i N_i di - \tau_b \int_{i:y_i \leq y_p} X_w^i N_i di \\
&+ \tau_a \int_{i:y_i>y_p} O_K^i (y_i - y_p)\, di + \tau_b \int_{i:y_i \leq y_p} O_K^i (y_i - y_p)\, di \\
&+ \tau_a \int_{i:y_i>y_p} X_K^i (y_i - y_p)\, di + \tau_b \int_{i:y_i \leq y_p} X_K^i (y_i - y_p)\, di = 0.
\end{aligned}$$

$$\tau_b(\tau_a) = -\tau_a \frac{\int_{i:y_i>y_p} N_i (y_i - y_p)\, di - \int_{i:y_i>y_p} (O_w^i + X_w^i) N_i di + \int_{i:y_i>y_p} (O_K^i + X_K^i)(y_i - y_p)\, di}{\int_{i:y_i \leq y_p} N_i (y_p - y_i)\, di + \int_{i:y_i \leq y_p} (O_w^i + X_w^i) N_i di - \int_{i:y_i \leq y_p} (O_K^i + X_K^i)(y_i - y_p)\, di}$$

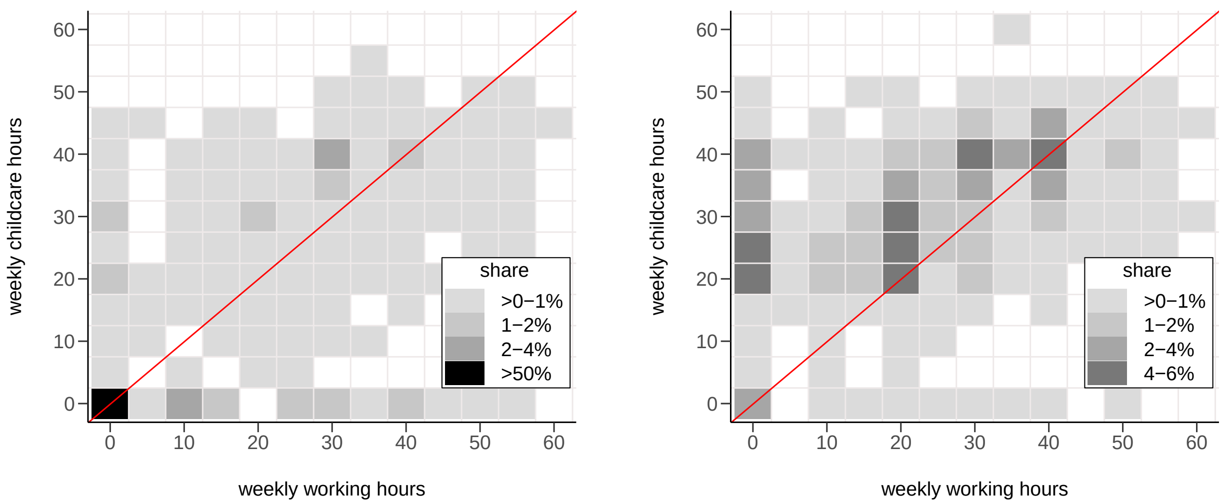
$$MECR(\hat{s}_p) = 1 - \frac{b(1-a)\int_{i:y_i \le y_p} N_i\,[y_p - y_i]\,di}{a\int_{i:y_i > y_p} N_i\,[y_i - y_p]\,di}$$

$$= \frac{1 - \frac{\int_{i:y_i > y_p} \left(O_w^i + X_w^i\right) N_i\,di - \int_{i:y_i > y_p} \left(O_K^i + X_K^i\right)(y_i - y_p)\,di}{\int_{i:y_i > y_p} N_i (y_i - y_p)\,di}}{1 + \frac{\int_{i:y_i \le y_p} \left(O_w^i + X_w^i\right) N_i\,di - \int_{i:y_i \le y_p} \left(O_K^i + X_K^i\right)(y_i - y_p)\,di}{\int_{i:y_i \le y_p} N_i (y_p - y_i)\,di}}. \qquad (27)$$

This result implies the statements in Proposition 2.

B Additional Stylized 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).

C Details on the Auxiliary Regressions

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: from 0-2, they visit the nursery, and from 3-5 they visit kindergarten. 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:[44]

1. *[n« 0\neL _jHnh0I* 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.[45]

2. *i«h@0} nT H_eL}0j _j §\0 \n« 0\neLI* 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.

C.1.2 Estimation of the childcare fee schedule

We use data from the 2013, 2015, and 2017 GSOEP waves, which contains information on public childcare hours per day and monthly fees paid.[46] We normalize the monthly fees by the reported daily public childcare hours to extract the monthly fee of full-time public childcare, defined by an attendance of 8 hours per day or 40 hours per week. For this purpose, we assume linearity of childcare fees in hours.

[44] See Authoring Group Educational Reporting (2018): *NL«H9§_nj _j W0}h9j´ ©·oP*, Section C2, p. 70–71 and Table C2-15web.

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

[46] In terms of the sample construction, this estimation is based on the same sample as laid out in Section 4.2.1 and D.1.

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 for details).

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:

$$
\begin{aligned}
p_{nt} = \quad & + {}_1 y_{nt} + {}_2 \; y_{nt} \quad \mathbb{1} \text{ one sibling with age} < 17 \text{ in HH } {}_{nt} \\
& + {}_3 \; y_{nt} \quad \mathbb{1} \text{ two siblings with age} < 17 \text{ in HH } {}_{nt} + {}_{nt}.
\end{aligned} \tag{28}
$$

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.[47] Our empirical model thereby closely reflects the current childcare fee schedule regulation as laid out in the previous section.

We estimate equation (28) as a Tobit regression with censoring at €0 and €725, the lowest and highest observed monthly childcare payments in our data.[48] We abstract from regional variation to keep the state space of the structural model tractable. But also if we extend the above regression (28) by state fixed effects and a dummy for living in an urban region to capture different levels of subsidies across regions, the resulting slopes of childcare subsidies with income remain unchanged compared to the following baseline results in Table C.1.

Results. The results of the Tobit regressions are summarized 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.

Figure C.2 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.

Finally, we take tax deductibility of childcare expenditures into account. We adjust the childcare fee schedule by the tax change implied by the childcare expenditures of the respective household income assuming full-time use of childcare.

[47] 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.

[48] 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

K0q0jL0j§ ®9}_9@e0I
gnj§\e´ T00 Tn} q«@e_H H_eLH9}0 9§§0jL9jH0 nT V·\¤°00c

		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	1950	626

i n§0 : 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).

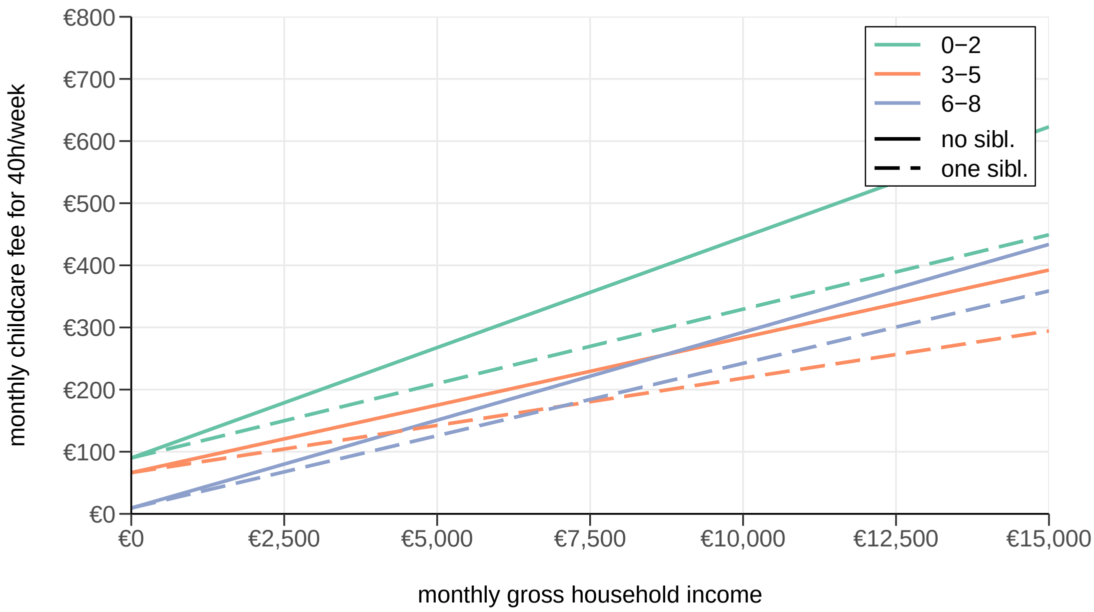


Figure C.2: Estimated childcare fee schedules

i n§0 : All values are adjusted to 2017 price levels. Source: 2013, 2015, 2017 GSOEP, FDZ-SOEP (2019).

C.2 Details on 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^0 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.[49]

C.3 Details on the wage process

C.3.1 Potential wages for non-working females

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

$$log(w_{f;it}) = X_{it} \quad + u_{it}; \tag{29}$$

where $w_{f;it}$ is the wage of female i in period t and 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, X also contains the number of children below age 5, the overall number of children, an urban indicator, an indicator for living in former East Germany, and a full set of year indicators.

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

$$Z_{it} \quad + \quad _{it} > 0; \tag{30}$$

[49]We select the sample from the Microcensus data using the same restrictions as for our GSOEP survey data (see Section 4.2.1 and D.1). Sources: FDZ-StABL (2020a), FDZ-StABL (2020b). This yields us 71,165 observations of households aged 20 to 40.

where Z contains 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 (30) for each time period. In these, we also include the individual specific means of all covariates in Z across 2000 to 2017, denoted by $\overline{Z}$:

$$Pr(\text{participation}_i = 1) = \quad Z_i \quad + \overline{Z}_i \quad . \tag{31}$$

After estimating (31) for each year, we obtain the inverse Mills ratios $\;_{it}$, which we then use as control functions in the selection corrected version of the wage equation (29):

$$log(w_{f,it}) = X_{it} \quad + \overline{Z}_i \quad + \quad _{it} + u_{it}. \tag{32}$$

With the estimated coefficients and at hand, we impute the wages of the non-working females.

C.3.2 Details on the wage process estimation

Using the 2000 to 2017 GSOEP data, we observe monthly gross labor income as well as contracted working hours.[50] This allows us to directly compute hourly wages for every individual that is working. For females who choose not to work, on the other hand, we do not observe any labor income and therefore, we impute their qn§0j§_9e gross hourly wages using a selection corrected wage model ().

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

$$\begin{aligned} log(w_{f,it}) = \quad + \;_1 log(w_{f,it-1}) + \;_2 \mathbb{1}\{flm_{it-1} = NP\} + \\ \;_3 \mathbb{1}\{flm_{it-1} = PT\} + \;_4 educ_i + A(t) + \varepsilon^{w_f}_{it}; \end{aligned} \tag{33}$$

where $\mathbb{1}\{flm_{it-1} = NP\}$ and $\mathbb{1}\{flm_{it-1} = PT\}$ are dummy variables that indicate whether a woman i was either not working or working part-time in period $t-1$.

The coefficients $\;_2$ and $\;_3$ are of particular interest for our analysis since they measure the dynamic wage penalty from working less than full-time. $\;_4$ captures the wage increase

[50]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 4.2.1 and D.1.

due to having obtained an A-level and $A(t)$ 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 Table C.2.

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

	$log(w_{f,t})$	$log(w_{m,t})$
t	0.020	0.010
	(0.013)	(0.0093)
t^2	-0.0026	-0.0018
	(0.0017)	(0.0012)
t^3	0.000082	0.000059
	(0.000067)	(0.000048)
$higheduc$	0.076	0.027
	(0.0052)	(0.0036)
$\mathbb{1}\{lm_{i,t-1} = NP\}$	-0.18	
	(0.0065)	
$\mathbb{1}\{lm_{i,t-1} = PT\}$	-0.057	
	(0.0057)	
$log(w_{f,t-1})$	0.75	
	(0.0057)	
$log(w_{m,t-1})$		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.3.1. Source: 2000 to 2017 GSOEP, FDZ-SOEP (2019).

D Details on the maximum likelihood estimation

D.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.

Further, 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). Finally, we drop households that are either in the top 1% or bottom 1% of the male or female wage distribution to avoid distortions.

For the estimation, we 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 11 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 is captured by 20 gridpoints each, while 17 gridpoints are used for l reflecting a 2.5h grid $\{0; 2.5; 5; \ldots; 40\}$.

D.2 Identification

The identification is conditional on the calibrated and reduced form regression inputs (see Section 4.1), the homogeneous preference parameters (see Section 4.2.2), and the previously described assumptions for our maximum likelihood procedure. The distribution of $h = (g; l; \;)$ will be jointly and set identified.

There are three interacting ingredients that identify the time-invariant unobserved heterogeneity: i) cross-sectional variation in choices conditional on the same observed states, ii) the longitudinal dimension of our panel data, iii) using data not only from households with small children, but also from those with older children. The following paragraphs describe the three ingredients in more detail.

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

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

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 , which explains the variation in their labor market choices conditional on wages and other observed characteristics. Hence, this group adds significantly to the identification of the distribution of independent of g and l.

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

D.3 Optimization routine

To solve the optimization problem numerically, we use the basin-hopping algorithm in combination with a Matlab built-in minimization routine for constrained target functions (Th_jHnj). 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).[51]

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 Th_jHnj. 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 Th_jHnj 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.

We verified our global optimization routine by implementing the TikTak algorithm, a recent multistart global optimization algorithm put forward by Fatih Guvenen and co-authors (see Arnoud, Guvenen, and Kleineberg (2019) for further information). Initially, the TikTak algorithm explores the parameter space uniformly (by setting so called Sobold points) and then, based on the gathered shape of the objective function, narrows its search to the most favorable areas. It then initiates local searches from specifically favorable points within the parameter space. The structure of our model allows us to set a high number of Sobol seed points (100,000), i.e. points in the multi-dimensional parameter space at which the function will be initially evaluated.

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

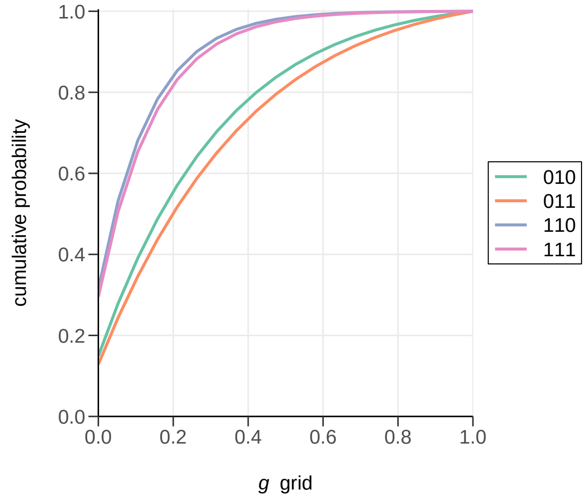
D.4 Results

The estimated coefficients are shown in Table D.3.

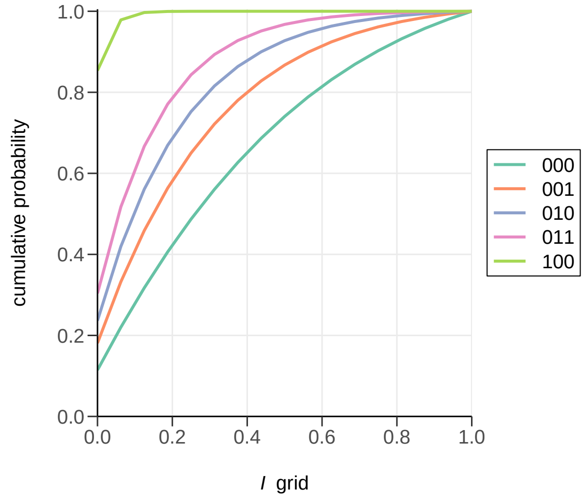
Table D.3: Maximum likelihood estimates

	domestic childcare (g)	avail. of informal childcare (I °0 §)	avail. of informal childcare (I09 §)	leisure ( )
	-3.04	-1.27	-30.73	0.18
09 §	-4.37			
L0h9jL_jX nHH«q	0.37			-0.80
H9§\ne_H	0.54			
_X\ 0L«H		-2.80		-0.76
«}@9j		-1.55		

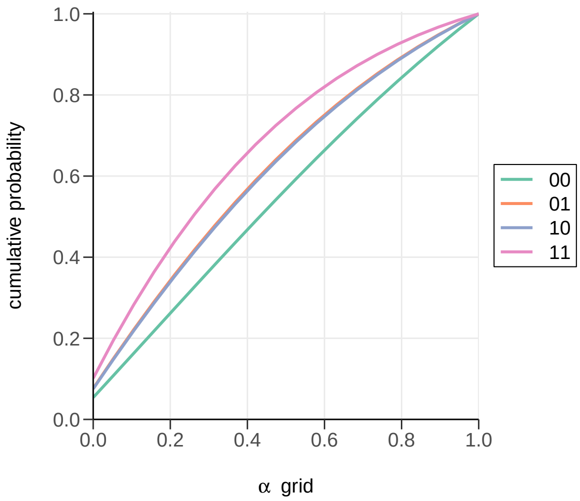
i n§0 : 'east' indicates having lived in former East Germany at some point; 'demanding occupation' indicates having primarily worked in an occupation where at least one-third of the tasks can be classified as analytic non-routine; 'catholic' indicates being Catholic at age 20; 'high education' indicates having obtained at least an A-level; 'urban' indicates living predominantly in an urban area. See Appendix D.3 for details on the optimization and Appendix D.5 for an illustration of the sensitivity of the estimates. Note that we fix the variance of all normal distributions to 1.



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



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



(c) leisure preference
covariates: high experience, education

Figure D.3: Cumulative distributions of unobserved heterogeneity

i n§0 : The legend of each subfigure indicates if the respective dummy – in the same order as the covariates listed below the subfigure – is 0 or 1. In case of the preferences for domestic childcare, we omit the plots for *j n* skilled occupation because the implied difference by skilled occupation is small and would render the graph unreadable.

D.5 Sensitivity of MLE results

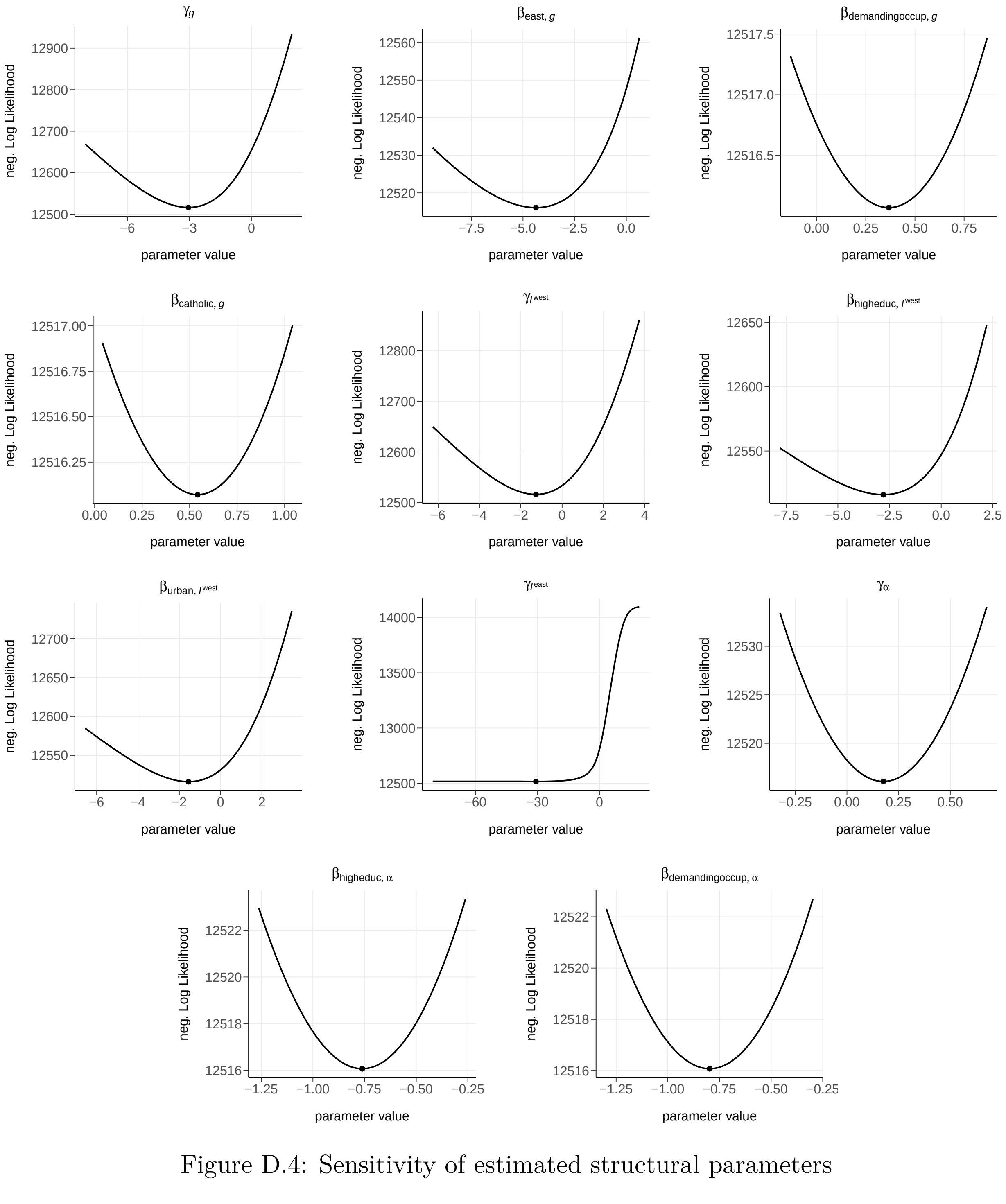
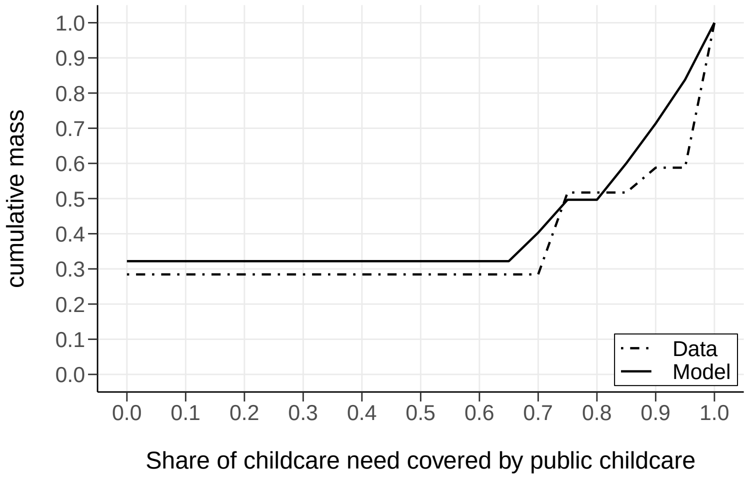


Figure D.4: Sensitivity of estimated structural parameters

i n§0 : 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. ‘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. The intercept ^{l} is not well identified. This likely has to do with the fact that in East Germany parents rely barely on grandparents for childcare. Note that the implied marginal distribution of l is effectively identical for other values of ^{l}.

E Additional Model Fit Illustrations



(a) Youngest child in the family aged 6 − 8

Figure E.5: Model fit of families' public childcare demand

i n§0 : Sample as defined in Section 4.2.3. Data source: 2012 to 2017 GSOEP, FDZ-SOEP (2019).

F Child development analysis

F.1 Returns to childcare attendance

A large body of literature studies the effect of childcare on cognitive and non-cognitive skills and schooling outcomes of children.[52] However, evidence on long-term outcomes, particularly on labor income, is scarce. In this Appendix, we detail how we extrapolate the effect of attending one year of childcare on children's earnings during adulthood from reduced-form estimates of the impact of universal childcare on adult earnings in Norway by Havnes and Mogstad (2015).

Havnes and Mogstad (2015) exploit time and geographical variation in childcare provision in Norway induced by the Kindergarten Act in 1975. They obtain the effect of childcare attendance for children aged 3 to 6 on future adult earnings. They show how these returns vary with family income. Their regression estimates the reduced-form impact on all children from post-reform cohorts. Thus, their effects need to be interpreted as an intention-to-treat effect (ITT). In order to retrieve the impact of the treatment on the treated (TT), the ITT estimate is divided by the probability of treatment. Havnes and Mogstad (2015) define this as the percentage point difference between the increase in childcare coverage in the treatment and control municipalities

[52] See the surveys from Cunha et al. (2006) and Elango et al. (2015).

in 1979, which equals 17.85%. Their estimate of TT thus suggests that each additional childcare slot induced by the policy reform increases adult earnings for children from low-income families on average by NOK $52'028$.[53]

This number of 17.85% reflects a treatment effect per childcare slot offered but does not account for the gradual increase in childcare coverage over the years 1976 to 1979. Since we aim to quantify the yearly return of spending one year in public childcare, we need to adjust this estimate to reflect the actual increase in time spent in childcare caused by the reform. For example, a child born in 1976 enters childcare in 1979 and experiences the full treatment intensity of 17.85% during all three years of eligibility for public childcare. However, a child born in 1975 experiences a smaller treatment intensity during their first year of eligibility due to the gradual increase in childcare coverage. To account for this fact that the treatment intensity is hence not 3 years for all these children, we assume that the difference in childcare coverage between treatment and control municipalities equals 0.00% and then increases linearly until it reaches 17.85% in 1979; this is in line with Figure 3 on page 105 in Havnes and Mogstad (2015). The implied average treatment intensities (defined as an increase in the probability of attending childcare for 3 years) per birth year read as follows:

$$
\begin{aligned}
Prob_{1973}[Treat] &= \frac{0.00\% + 5.95\% + 11.90\%}{3} = 5.95\% \\
Prob_{1974}[Treat] &= \frac{5.95\% + 11.90\% + 17.85\%}{3} = 11.90\% \\
Prob_{1975}[Treat] &= \frac{11.90\% + 2 \times 17.85\%}{3} = 15.87\% \\
Prob_{1976}[Treat] &= \frac{3 \times 17.85\%}{3} = 17.85\%
\end{aligned}
$$

We further assume that equal numbers of children are born in each year and obtain an average treatment intensity of 12.89%. Based on the estimates by Havnes and Mogstad (2015),[54] we compute the TT of attending up to three years in childcare for income-poor and income-rich families. Taking a median family income to NOK $341'330$.[55] and weighing estimates by population density, we obtain a TT for children born in income-poor families of NOK $41'068$ and an effect of NOK $-6'376$ for children born in income-rich families.

Subsequently, we express these treatment effects as returns to one year of public childcare by assuming constant compounded returns implying that $1 + r_{TT,1} = (1 + r_{TT,3})^{\frac{1}{3}}$. The return to

[53] $NOK/EUR \approx 8$.

[54] See Figure 7 on page 110.

[55] Below-median families earn on average NOK $273'860$, while above-median families earn NOK $440'420$ ($NOK/EUR \approx 8$).

one year in childcare for children born in families with below-median income equals 3.67%. For children from families with above-median income, the number equals to 0.52%. More generally, this approach yields the returns to one year of childcare as a function of parental income as illustrated in Figure 6. We assume that these returns apply to one year of attending childcare for 40 hours per week and assume that these returns apply equally for child age between 0 and 5.

F.2 Translating returns into net-present value increases

The next step consists of translating the returns to one year of full-time childcare into increases in the net-present value of lifetime earnings. To obtain these numbers, in a first step, we use recent estimates for Germany from Dodin et al. (2022) to obtain average child earnings as a function of the parental income rank. We take their rank-rank coefficients from Table 5 where child income is measured as individual labor earnings and parent income as gross family income.[56] This provides us with the estimated child rank at age 29-33 as a function of the parent rank. Based on the sample of the SOEP data used in this paper, we can then assign the child income in Euros that corresponds to these predicted child income ranks. In the final step, we extrapolate the lifecycle earnings profiles for each child earnings rank. We use the lifecycle profiles from Bönke, Corneo, and Lüthen (2015) estimated with German administrative pension data.[57] Assuming that children work from 25-60, this yields lifecycle earnings profiles from age 25-60 as a function of the parental income rank. Combining these with the returns illustrated in Figure 6, we obtain the increase in the net-present value of lifetime earnings due to public childcare attendance.

[56] The intercept of this regression, which is not provided in this paper, equals 36.14.

[57] We assume the intermediate education level,high school + vocational training, for this extrapolation since Bönke, Corneo, and Lüthen (2015) estimate these profiles separately for high school only, high school + vocational training, and for college. Finally, we take the average of the male and female implied earnings growth rates over the lifecycle.