The equilibrium effects of state-mandated minimum staff-to-child ratios*

Martin Garcia-Vazquez

October 27th, 2023

CLICK HERE FOR MOST RECENT VERSION

Abstract

Mandatory minimum staff-to-child ratios are a pervasive childcare market regulation in the US, and yet little is known on their effects on children's skills. This paper builds an equilibrium model of the childcare market and uses it to simulate the distribution of children's skills at preschool entry under various minimum mandatory staff-to-child ratios. The model allows for rich family heterogeneity, an endogenous distribution of childcare quality at each age, and endogenous wages that clear the market for teachers and childcare workers. I prove identification and estimate the model using both individual-level and state-level data. Counterfactual simulations show that increasing the stringency of minimum mandatory staff-to-child ratios increases the wages of childcare workers by up to 3% and wages of lead teachers by up to 2.5%. Increasing the minimum number of adults per child has different effects for one- and two-parent families. For one-parent families, it increases skills at the right tail of the skill distribution and decreases skills at the bottom. For two-parent families, gains are uniform across the skill distribution. Finally, these overall effects on the skill distribution mask large heterogeneity: Increases in ratios' stringency translate into big skill gains for some children and large drops for others, and the treatment effect distribution is more dispersed for children born to single mothers. Skill redistribution happens mostly across two types of poor families. Children born to poor families with higher substitution possibilities (more care from family relatives available and assets) experience higher skill losses, whereas children born to families less able to substitute away from paid care (less relative care available and lower assets) experience higher skill gains.

*I thank my advisors Mariacristina De Nardi, Alessandra Fogli, Jeremy Lise, and Joseph Mullins for their guidance, encouragement, their useful feedback, and stimulating discussions. I also thank Orazio Attanasio, Fil Babalievsky, Kyle Herkenhoff, Jose Montalban-Castilla, Fabrizio Perri, David Rahman, Aaron Sojourner, Jan Werner, and Matt Wiswall for useful discussions. I am also grateful to Sheri Fischer for providing me with historic data on childcare regulations on the US, and for all the effort and time that she put helping me make sense of that data.

1 Introduction

Individuals' skills are particularly malleable during early childhood and shape outcomes later in life (Cunha and Heckman (2007), Cunha, Heckman and Schennach (2010)). Moreover, returns to early childhood investment are very high (Heckman, Garcia, Leaf and Prados (2017)). An important fraction of early childhood investment is provided outside of the family, by market-based providers and public providers, and these providers are subject to regulations that vary by state. In this paper, I study how regulations in the childcare market affect the skills of children in different family types through the type and quality of environments that those children grow up in.

The regulations I focus on are mandatory minimum staff-to-child ratios, which are state-level regulations that determine the maximum number of children allowed per adult in the classroom. Mandatory staff-to-child ratios vary substantially by state: For instance, at 35 months old, the mandatory minimum staff-to-child ratio ranges from 1 adult per 4 children in Connecticut to 1 adult per 12 children in Mississippi.

From a normative perspective, the fact that policymakers in different states do not agree on the stringency of these regulations, suggests that mandatory-minimum ratios have important re-distributional impacts and state regulators in different states maximize different social welfare functions.[1]

From a positive perspective, the large variation in the minimum staffing requirements faced by childcare providers across states is likely to have an impact on the cost and quality of childcare provided in the market. This, in turn, can affect children born to different types of households differently. In fact, using a difference-in-difference strategy, Hotz and Xiao (2011) find that more stringent mandatory minimum staff-child ratios decrease the provision of childcare, especially for low-income markets, but that they also increase the quality of childcare provided.

Despite the evidence showing that mandatory minimum staff-to-child ratios have an impact on the market provision of childcare, their implications for the skill distribution of children are not well understood. On the one hand, their positive effect on the quality of childcare provided should translate into a positive effect in the skills of children. On the other hand, their negative impact on the quantity of childcare demanded can translate into a negative effect on the skills of children if families substitute market-provided childcare with options of inferior quality. Moreover, the extent to which either effect dominates may depend on family characteristics. For instance, some families may be intrinsically less prone to substitute away from market-provided care, maybe because they have less relative care available, or because their economic resources make it too costly for parents in those families to work less and look after their child. Children born to those families are likely to experience the positive effects of the increase in quality caused by regulations, and are less likely to experience the negative effect due to the substitution of care in favor of childcare arrangements of inferior quality. Instead, children who spend fewer hours in market-provided childcare as a consequence of the regulation may experience skill losses if the alternative childcare arrangements that

[1]It could also be that the primitives that are relevant to determine optimal regulations are very different across states

they experience are of inferior quality. Finally, since regulations affect the demand for quantity and quality of childcare, they can also affect the demand for inputs in childcare production, in particular the demand for teachers and other staff members. This may translate into an equilibrium effect on the wages of teachers and childcare workers, which in turn can amplify or attenuate the partial equilibrium effects of regulations (holding wages of teachers and childcare workers constant) on the demand for childcare quantity and quality for different family types.

In order to study the impact of mandatory minimum staff-to-child ratios, I build an equilibrium model of the US childcare market in which heterogeneous families make decisions about how much time their child spends in each type of childcare arrangement (maternal, paternal, relative, or paid childcare), the quality of paid childcare that they buy, and the type of paid childcare provider (center-based or home-based). Paid childcare is provided by center-based and home-based care providers, who face different regulations for each age, and take wages of lead teachers and childcare workers as given. The wages of Lead Teachers and Childcare Workers equate demand and supply in their respective labor markets.

As shown by Flood, McMurry, Sojourner and Wiswall (2021), the quality of parental and relative care available to families of different socieconomic status is very heterogeneous. Capturing this heterogeneity is crucial because the quality of non-market childcare arrangements available to families mediates the impact on skills of policies that induce reallocation from paid childcare to unpaid childcare. Moreover, the extent to which families can substitute paid for unpaid childcare services depends on the availability of free relative care and the opportunity cost of maternal and paternal care. Because of this, I let families in my model be heterogeneous in the quality of maternal, paternal, and relative care, the availability of relative care, the wages of mothers and fathers, and their initial level of assets.

The model that I build in this paper has several features that make it an attractive tool to analyze childcare market policies. First, it allows for rich heterogeneity. On the demand side, families are heterogeneous in terms of wages, assets, parenting quality, relative care quality, and relative care availability, all of which are relevant for understanding the impact of policies that induce reallocation across childcare arrangements. On the supply side, the model features an endogenous distribution of quality supplied by center-based and home-based care providers and an endogenous price-quality gradient that is affected by regulations. Moreover, in this model childcare market policies can also affect the price of childcare by affecting the wages of teachers, which are also endogenous. Second, I prove that the model is identified from a combination of individual-level and state-level data. The identification proof translates into a multi-step estimation procedure that makes estimation feasible. Third, I develop a computational strategy that makes solving the model under different policy scenarios feasible.

I use the model to simulate the effects of changing the stringency of mandatory minimum staff-to-child ratios on the wages of lead teachers and childcare workers and the distribution of cognitive skills of children. I find that more stringent regulations increase the wages of lead teachers and childcare workers by up to 3% for childcare workers and up to 2.5% for lead teachers. Moreover, this policy increases skills at most percentiles of the skill distribution, but decreases skills at the bottom of the skill distribution of children born to single-parent families. The overall effects in the distribution of skills hide very large heterogeneity. For instance, the 1% of children born to Single Mother families who gain the most from more stringent regulations increase their skills by 39% of a standard deviation, whereas the bottom 1% of children born to single mothers who lose the most see their skills decrease by 47% of a standard deviation. Increasing stringency of regulations induces the most skill redistribution from poor children to other poor children. Children who experience the largest skill gains are poor children born to families with little relative care available and lower assets, whereas children who experience the highest skill losses are poor children born to families with more hours of relative care available and higher assets.

The model in this paper is not designed to make normative statements about the skill distribution (i.e., what the skill distribution should be). This could be seen as a potential limitation of this paper, but it is a deliberate choice. This is because the only reason why the level of investment chosen by parents could be different from the social optimum is due to borrowing constraints.[2] However, the model doesn't allow for externalities, which are another potentially important reason for parental investment being socially sub-optimal. Potentially important externalities associated to skills are crime and innovation. For instance, Cunha, Heckman and Schennach (2010) show that low cognitive and non-cognitive skills are associated to crime, whereas Bell, Chetty, Jaravel, Petkova and Van Reenen (2019) show that cognitive skills are associated with innovation and becoming an inventor. Because of this, including all the sources of underinvestment that would make the model a suitable normative theory of the skill distribution, while retaining all the elements that make it a suitable positive theory of the skill distribution, is likely not to be tractable. In fact, including other sources of missing markets or the externalities discussed above involves modelling the continuation problem of the child when she becomes an adult, something outside the scope of this paper. Still, the model is well suited to analyze policies given a policy target for the skill distribution at kindergarten entry.

I make several contributions in this paper. First, I contribute a quantitative equilibrium model of the US childcare market that is point-identified and computationally tractable. The model is designed to predict for each state in the US the effect of childcare market policies on family decisions (such as childcare decisions and labor supply decisions), the skill distribution of children, the wages of teachers, the distribution of quality of childcare demanded at each age and the cost of childcare. Second, I use this model to study the effects of an important childcare market regulation, mandatory minimum staff-to-child ratios, on the distribution of children's skills. In doing so, I uncover that this regulation has very heterogeneous impacts on the skills of children,

[2] Borrowing constraints are not the only possible source of market incompleteness in models of early childhood investment. For instance, parents' inability to borrow against their child's future income or their inability to write investment contracts with their child are other common sources of market incompleteness. See Daruich (2018) for a more exhaustive explanation on the problem of missing markets in models with investment in children.

and I show that most of the skill redistribution happens between two types of poor children. Children that gain the most are on average poor children born to families with lower substitution possibilities (less care provided by relatives available and less assets), whereas children that lose the most are on average born to poor families with more substitution possibilities (more care provided by relatives available and more assets). Third, I prove that the average quality of different childcare arrangements is identified, even when the quality of different childcare arrangements is measured using different survey instruments. This is key when trying to predict the effect of childcare market policies that reallocate childcare across different types of childcare arrangements on the skill distribution of children.

2 Related Literature

This paper relates to four strands of literature. First, the important seminal literature that shows that early childhood interventions are a very powerful tool to improve economic mobility and individual outcomes. In this literature, the efficacy of early childhood investment in producing skills and improving adult outcomes is established in two ways. The first way is to study directly the effects of early childhood programs[3]. The second way is to estimate the deep parameters of the production function of skills and their relation to adult outcomes[4]. Some recent papers in this latter category dissagregate investment in children to take into account the quality and quantity of different childcare arrangements (see Griffen (2019); Chaparro, Sojourner and Wiswall (2020); McMurry (2021)). I adopt this approach and contribute to it by showing that the relative qualities of different types of childcare arrangements are identified even if different survey instruments are used to measure their qualities. That is, the quality of different childcare arrangements can be compared even if they are not measured in the same units.

Second, many papers study the effects of different policies from the one that I study on children's skill development. Examples of such policies are transfer programs, such as the EITC , parent interventions , and the introduction, expansion or universalization of early childhood education programs .[5]

Third, this paper relates the most to a flourishing literature that uses equilibrium

[3]See for example Campbell, Conti, Heckman, Moon, Pinto, Pungello and Pan (2014); Elango, García, Heckman and Hojman (2015); Heckman, Garcia, Leaf and Prados (2017); García, Heckman and Ronda (2021)

[4]See Todd and Wolpin (2003); Cunha, Heckman, Lochner and Masterov (2006); Cunha and Heckman (2007); Todd and Wolpin (2007); Cunha and Heckman (2008); Cunha, Heckman and Schennach (2010); Agostinelli and Wiswall (2016a,b); Attanasio, Meghir and Nix (2020); Attanasio, Cattan, Fitzsimons, Meghir and Rubio-Codina (2020); Attanasio, Bernal, Giannola and Nores (2020)

[5]For transfer programs see Bernal and Keane (2010, 2011); Dahl and Lochner (2012); Mullins (2022), for parent interventions see Sylvia, Warrinnier, Luo, Yue, Attanasio, Medina and Rozelle (2021); Gertler, Heckman, Pinto, Chang, Grantham-McGregor, Vermeersch, Walker and Wright (2021); Gomez, Bernal and Baker-Henningham (2022), and for the introduction, expansion, or universalization of early childhood programs see Heckman, Garcia, Leaf and Prados (2017); Daruich (2018); Chaparro, Sojourner and Wiswall (2020); Cascio (2023)

models of the childcare market for policy analysis. This paper contributes to this literature by combining a detailed model of the childcare sector with center-based and home-based providers, in which the cost of providing quality is endogenous to the labor market for teachers and the effects on the skills of children can be analyzed by taking into account the reallocation of care across childcare arrangements and the quality and availability of those childcare arrangements. In addition, this model adds heterogeneity to the quality of parental and relative care, endogenous asset accumulation, and accounts for the way the mandatory minimum staff-to-child ratio enters the quality-production process in paid care providers. Within this literature, Moschini (2023) analyzes the impact of childcare subsidies on children's skills in an OLG model with endogenous family formation in a model in which families buy a childcare good of homogeneous quality and the childcare price is a constant fraction of wages. Berlinski, Ferreyra, Flabbi and Martin (2023) look at the impact of various policies on the skills of children in a model in which there is imperfect competition in the childcare market. In their model, the cost of providing a given level of quality for a given provider is exogenous and does not depend on the price of inputs of childcare production. Borowsky, Brown, Davis, Gibbs, Herbst, Sojourner, Tekin and Wiswall (2022) examine the counterfactual impacts of adopting an important childcare market proposal. In order to do so, they build a model that endogenizes the cost of providing childcare via the labor market for teachers (so that increases in demand for childcare can increase the unit cost of childcare by raising teachers' wages). Relative to their paper, I include skill accumulation. Moreover, relative to both Berlinski, Ferreyra, Flabbi and Martin (2020) and Borowsky, Brown, Davis, Gibbs, Herbst, Sojourner, Tekin and Wiswall (2022) I add heterogeneity to the quality of relative and parental care, asset accumulation, and I model explicitly how mandatory minimum staff-to-child ratios distort the problem of paid childcare providers.

Fourth, this paper also relates to the empirical literature on the effects of childcare market regulations, which use a reduced-form approach to identify the effects of childcare regulations on the provision of quantity and quality of childcare[6]. To the best of my knowledge, this paper is the first to study the effect of childcare regulations on the distribution of children's skills, and in estimating a structural model of childcare regulations.

3 Model

3.1 Families

Families are unitary households, and of two types, single-mother ($\mathtt{SM}$) and two-parent ($\mathtt{TP}$) households. Families are modeled for 3 periods, that is, when the child is 9 months, 2 years old and 4 years old. I choose to model those periods in order to match the data from ECLS-B (more on that dataset later).

[6]See Chipty (1995); Chipty and Witte (1997); Blau (2003); Currie and Hotz (2004); Blau (2007); Hotz and Xiao (2011)

3.1.1 Choices

In each period, Two-Parent families choose how much to consume c, and how much to save in a one-period risk-free asset a_{t+1}.

Two-parent families also make decisions about how parents spend their time: How much leisure the mother (m) and the father (f) enjoy (l_t^m, l_t^f), how much time each of them spends with their child (τ_t^m, τ_t^f), how much each of them work (n_t^m, n_t^f)

Importantly, Two-Parent families make decisions about non-parental childcare arrangements of their child: How much time the child spends with relatives (τ_t^r), whether to use Center-Based paid care ($D_t = CB$), Home-Based care ($D_t = HB$), or no paid care at all ($D_t = HB$), how much time the child spends in paid care τ_t^p, and the quality of this paid care q_t^p.

Single mothers are similar but they can only choose father-specific variables (the ones with a superscript f) to be 0.

3.1.2 Preferences

Single-Mother households at $t = 1,2,3$ derive utility from consumption (c_t), leisure of the mother (l_t^m), time of the mother with her child (τ_t^m), and the skills of the child according to:

$$\log c_t + \delta_l^m \log l_t^m + \delta_\tau^m \log \tau_t^m + \delta_{\theta,t} \log \theta_t \ .$$

Two-parent households at $t = 1,2,3$ derive utility from consumption (c_t)[7], leisure of the mother (l_t^m), leisure of the father (l_t^f), time of the mother with her child (τ_t^m), time of the father with the child (τ_t^f), and skills of the child θ_t according to:

$$\log c_t + \delta_l^m \log l_t^m + \delta_l^f \log l_t^f + \delta_\tau^m \log \tau_t^m + \delta_\tau^f \log \tau_t^f + \delta_{\theta,t} \log \theta_t \ ,$$

with $\delta_{\theta,1} = 0$. Note that $\delta_{\theta,1} = 0$ is a normalization, in the sense that in this model θ_1 is given and I have no data on pre-natal investment behavior that could be informative about $\delta_{\theta,1}$.

The continuation utility at period 4 is given by:

$$\delta_a \log a_4 + \delta_\theta \log \theta_4 \ .$$

Note that I impose common parameters to be the same across family types. In other words, I am imposing that $\delta_l^m, \delta_\tau^m, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a$ to take the same numerical value for Two-Parent families and Single-Mother Households

3.1.3 Time use constraints

As total time use for each family member cannot exceed total available time. The time use constraint for parent j is:

$$l^j + \tau^j + n^j = \overline{T}^j \ , \tag{TC_j}$$

[7]Note that I am not dividing c by an equivalence scale to capture economies of scale. This is because families do not change size in the model and the equivalence scale would appear as an additive constant in utility that does not affect behaviour.

for $j = m, f$, where n^j is time spent working on the labor market for parent j.

This time use constraint means that parents have to split their time between working, spending time with their child, and leisure.

The time-use constraint for the mother (m) holds for both two-parent (TP) and single-mother (SM) households.

Moreover, the child has to be supervised at every point in time:

$$\tau^m + \tau^f + \tau^r + \tau^p = \overline{T}\,, \tag{SC}$$

where τ^r and τ^p denote relative and professional care respectively.

For single parent-households, τ^f is equal to 0. That is, in families with a single mother the supervision constraint for the child reduces to:

$$\tau^m + \tau^r + \tau^p = \overline{T}\,,$$

3.1.4 Supply of relative care

Relative care is free but limited to $\overline{T}^r$, that is:

$$\tau^r \in [0, \overline{T}^r]\,. \tag{RCC}$$

where $\overline{T}^r$ is heterogeneous across families.

3.1.5 Production of cognitive skills

The production function for skills follows closely Chaparro, Sojourner and Wiswall (2020) and McMurry (2021) . Child skills are produced according to:

$$\log\theta_{t+1} = \log A_t + \gamma_{\theta,t}\log\theta_t + \gamma_{m,t}\frac{\tau_t^m}{\overline{T}}\log q^m + \\ \gamma_{f,t}\frac{\tau_t^f}{\overline{T}}\log q^f + \gamma_{p,t}\frac{\tau^p}{\overline{T}}\log q_t^p + \gamma_{r,t}\frac{\tau^r}{\overline{T}}\log q_t^r + \eta_{t+1}\,, \tag{PF Skills}$$

with η_{t+1} iid accross periods and where q^m, q^f, q_t^r, q_t^p denote the qualities of maternal, paternal, relative, and paid care respectively. I assume that the quality of maternal and paternal care are exogenous and time-invariant. The quality of relative care is exogenous and time-varying, and its family-specific time-path it's known to each family. This assumption captures that the comparative advantage or disadvantage of relatives at producing cognitive skills with respect to parents can depend on the age of the child. Importantly, this comparative advantage or disadvantage with respect to parents is identified (see Appendix J.5). Note that what it is assumed to change with time is the relative advantage of relative care at producing skills with respect to parents, given that the absolute location of parental care is not identified. This assumption does not imply that parenting skills do not change on average as the child ages, it only allows for the quality of relative and parental care to change at different rates [8]. Finally, quality of center-based care can be purchased on the market.

For households with a single mother $\tau^f = 0$, the production function for skills reduces to:

$$\log \theta_{t+1} = \log A_t + \gamma_{\theta,t} \log \theta_t + \gamma_{m,t} \frac{\tau_t^m}{T} \log q^m + \gamma_{p,t} \frac{\tau^p}{T} \log q^p + \gamma_{r,t} \frac{\tau^r}{T} \log q_t^r + \eta_{t+1} \,.$$

Again, the parameters that are shared by skill production in Two-Parent and Single-Parent households are required to take the same numerical value.

3.1.6 Budget constraint

Assets tomorrow plus expenditures in consumption and formal care cannot exceed total household income:

$$c_t + 1\{D = HB\} D_t^{HB}(q^p)\tau^p + 1\{D = CB\} D_t^{CB}(q^p)\tau^p + a_{t+1} = w^f n_t^f + w^m n_t^m + a_t(1+r) \,. \quad \text{(BC)}$$

where D is a categorical variable that takes values H, C, N corresponding to Home-Based Care (HB) , Center-Based Care (CB) and No Professional Care (N). For single-parent households, the same budget constraint applies with $n_t^f = 0$.

3.1.7 Borrowing constraint

I assume that assets have to exceed $\underline{a}$, which is a parameter:

$$a_{t+1} \geqslant \underline{a} \,. \quad \text{(AC)}$$

In other words, there are constraints to borrowing, which are important because they could be a source of underinvestment in children cognitive skills, specially in lower income families.

3.1.8 Fixed cost of choosing paid care

Families are ex-post heterogeneous in their fixed costs of choosing Home-Based and Center-Based care (o_t^{CB}, o_t^{HB}). The cost o_t^D is independent of $o_t^{D'}$, and $o_{t'}^D, o_{t'}^{D'}$ for $t \neq t'$ and $D \neq D'$, $D, D' = CB, HB$, and independent across families. The random utility cost o_t^P is assumed to be exponentially distributed with parameter λ_t^D.

These exponential utility costs play many roles. First, they help to match the choice probabilities of not using paid care, using center-based care, and home-based care at each age. Second, they imply that some families at a given age do not find optimal using paid care, which helps with the identification of the distribution of the relative care endowment (see Appendix H). Third, since they are a fixed cost, they help rationalize why most families that use paid care do not use it for just a few hours per week.

[8] In fact, the assumption in this paper is slightly more general, because the relative advantage of relatives versus parents is allowed to change across families according to observables.

3.2 Summing-up family heterogeneity

At $t = 1$, families within each location are ex-ante heterogeneous in their initial level of assets a_1, the initial cognitive skills of their child θ_1, the wages of the mother and the father w^m, w^f, quantity of relative care they can use $\overline{T}^r$, and the quality of maternal, paternal and relative care $q^f, q^m, \{q_t^r\}_{t=1}^3$. For each t, families are ex-post heterogeneous in their fixed-costs of using each type of care (o_t^{CB}, o_t^{HB}).

It will be useful later to separate the ex-ante heterogeneity of households into time-invariant heterogeneity (H) and assets a_t, that is:

$$H = (w^m, w^f, q^m, q^f, \{q_t^r\}_{t=1}^3, \overline{T}^r) \,.$$

3.3 Recursive formulation

Here I present the problem of the Two-Parent Household in recursive form. The problem of Single-Mother families is similar and is therefore omitted. Let the vector of continuous choices at age t be given by Y_t, that is:

$$Y_t = (c_t, n_t^m, n_t^f, l_t^m, l_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p, q_t^p) \,.$$

The problem of the Two-Parent Family can be summarized by the following functional equation:

$$\begin{aligned} V_t^{TP}(a_t, H, \theta_t, o_t^{CB}, o_t^{HB}) = \max_{Y_t, D \in \{CB,HB,N\}} & \log c + \delta_l^m \log l^m + \delta_l^f \log l^f + \delta_\tau^m \log \tau^m + \delta_\tau^f \log \tau^f + \delta_{\theta,t} \log \theta_t \\ & - o_t^{CB} 1\{D = CB\} - o_t^{HB} 1\{D = HB\} + \beta \mathbb{E} V_{t+1}^{TP}(a_{t+1}, H, \theta_{t+1}, o_{t+1}^{CB}, o_{t+1}^{HB}) \\ & \text{s.t BC, TC}_j \; j = m, f \text{, SC, PF Skills , AC ,} \\ & \text{for } t = 1,2 \end{aligned}$$

For $t = 3$ we have a similar functional equation with

$$\mathbb{E}V_4^{TP}(a_4, H, \theta_4, o_4^{CB}, o_4^{HB}) = \delta_a \log a_4 + \delta_{\theta,4} \mathbb{E}_3 \log \theta_4 \,.$$

In Appendix B I prove that the value functions are log-additive in skills, which simplifies further the recursive formulation. This result also implies that the policy functions do not depend on the child's skills, which simplifies the computational solution of the individual problem.

Moreover, in order to find the probability distribution of optimal household choices, it is enough to calculate the choice probabilities over discrete choices $\mathbb{P}_t^{FT}(D_t = D | a_t, H)$ for $FT = TP, SM$ and $D = N, CB, HB$ and the choice-specific policy functions $g_\omega^{FT,D}(a_t, H)$ for

$$\omega \in \Omega = \{c, a', n^m, n^f, l^m, l^f, \tau^m, \tau^f, \tau^r, \tau^p, q^p\} \,.$$

In other words, instead of working with $V_t^{TP}(a_t, H, \theta_t, o_t^{CB}, o_t^{HB})$ and the corresponding policy functions, we can work with the choice probabilities $\mathbb{P}^{FT}(D_t = D | a_t, H)$ the choice-specific value functions $\tilde{V}_t^{FT,D}(a_t, H)$ and their corresponding policy functions.

More details on this on Appendix C

3.4 Formal childcare providers

Formal childcare providers can be of two types: Center-based providers (CB) or Home-Based providers (HB). The market for formal childcare has free entry and is perfectly competitive. More precisely, upon entering the market, a childcare provider of type P offering quality q can sell each hour of childcare at the competitively determined price $P^D(q)$, where $D = CB, HB$. Formal childcare providers produce quality of childcare per hour q combining efficiency units of the lead teacher per number of children in the classroom during that hour $\frac{\hat{E}}{k}$ and the number of caregivers in the classroom per number of children during that hour $\frac{\hat{C}}{k}$ according to:

$$q = F^j_t(\hat{E}, \hat{C}, k) = A^j_t \left(\frac{\hat{E}}{k}\right)^{\alpha^j_{E,t}} \left(\frac{\hat{C}}{k}\right)^{1-\alpha^j_{E,t}}.$$

Note that I am allowing the parameters of the production function to change with the age of the children that are being cared for, and also note that different types of formal childcare are allowed to operate different technologies.

This technology captures that the quality of formal childcare is produced by combining the talent of the lead teacher (who plans curriculum, communicates with parents, designs play areas and reacts to specific needs of children) with the number of caregivers, which includes the lead teacher herself, but also other adults in the classroom like assistant teachers and childcare workers.

Childcare providers hire efficiency units of the lead teacher E and caregiving time C in a competitive labor market, with factor prices denoted by w^E and w^C respectively. Since there is a labor market for each location, wages are location-specific

Moreover, the state government legislates the minimum staff-to-child ratio, which is age-dependent. Call this legislated ratio $\underline{R}_{l,t}$, where l denotes the location (state) and t the age.

3.4.1 Factor demands

The factor demands conditional on producing quality q for k children for τ hours are given in the following lemma.(where $h = \tau k$, the total hours of care). E and C denote the total number of hour-efficiency units of the lead teacher (that is, the efficiency units of the lead teacher per hour, times her total number of hours), and the total number of hours of caregiving (the number of caregivers in the classroom time the number of hours) This is summarized in the following Lemma:

Lemma 1 (Factor demands)**.**

$$E^j(h,q) = \begin{cases} \left(\frac{\alpha^j_{E,t}}{1-\alpha^j_{E,t}} \frac{w^C}{w^E}\right)^{1-\alpha^j_{E,t}} \frac{qh}{A^j_t} & \textit{if } q > q^* \\ \left(\frac{q}{A^j_t}\right)^{\frac{1}{\alpha^j_{E,t}}} \left(\frac{1}{\underline{R}_{j,l,t}}\right)^{\frac{1-\alpha^j_{E,t}}{\alpha^j_{E,t}}} h & \textit{if } q \leqslant q^* \end{cases}$$

$$C^j(h,q)=\begin{cases}\left(\frac{1-\alpha^j_{E,t}}{\alpha^j_{E,t}}\frac{w^E}{w^C}\right)^{\alpha^j_{E,t}}\frac{qh}{A^j_t} & \text{if } q>q^* \\ \underline{R}_l h & \text{if } q\leqslant q^*\end{cases}$$

where q^* *is given by:*

$$q^*=A\underline{R}^j_{l,t}\left(\frac{\alpha^j_{E,t}}{1-\alpha^j_{E,t}}\frac{w^C}{w^E}\right)^{\alpha^j_{E,t}}.$$

Proof. The result comes out of solving the cost-minimization problem of the paid care center. Full proof in Appendix A.1 □

Note that we can write:

$$E^j(h,q)=\hat{E}^j(q)h\,,$$
$$C^j(h,q)=\hat{C}^j(q)h\,.$$

3.4.2 Costs

From the factor demands above we can get the cost functions, which are given by the following lemma.

Lemma 2. *Given factor prices* w^E, w^C*, the cost function of a childcare provider of type* j *operating in location* l *and serving age* t $c_{j,l,t}(q,h)$ *is given by:*

$$c_{j,l,t}(q,h)=\begin{cases}\left[w^E\left(\frac{\alpha^j_{E,t}}{1-\alpha^j_{E,t}}\frac{w^C}{w^E}\right)^{1-\alpha^j_{E,t}}+w^C\left(\frac{1-\alpha^j_{E,t}}{\alpha^j_{E,t}}\frac{w^E}{w^C}\right)^{\alpha^j_{E,t}}\right]\frac{qh}{A} & \text{if } q>q^*_{j,l,t} \\ \left[w^E\left(\frac{q}{A}\right)^{\frac{1}{\alpha^j_{E,t}}}\left(\frac{1}{\underline{R}^j_l}\right)^{\frac{1-\alpha^j_{E,t}}{\alpha^j_{E,t}}}+w^C\underline{R}^j_{l,t}\right]h & \text{if } q<q^*_{j,l,t}\end{cases}$$

where $q^*_{j,l,t}$ *is defined as in Lemma 1.*

Proof. Full proof in Appendix A.1 □

Again, note from the previous lemma that the cost of offering childcare quality q for h total child-hours is linear in h, so we can again write, in a slight abuse of notation:

$$c_{j,l,t}(q,h)=c_{j,l,t}(q)h$$

3.4.3 Price of paid care

Lemma 3 (Pricing schedule of paid care)**.** *Given factor prices* w^E, w^C*, if a positive amount of childcare hours are offered in equilibrium in the market for childcare of type* j *at age* t*, the price of childcare has to be given by:*

$$P_{j,l,t}(q)=\begin{cases}\overline{P}q \text{ if } q>q^*_{j,l,t} \\ \left[\underline{P}q^{\frac{1}{\rho_P}}+\kappa_P\right] & \text{if } q\leqslant q^*_{j,l,t}\end{cases}$$

where

$$\overline{P} = \left[w^E \left(\frac{\alpha_E}{1-\alpha_E} \frac{w^C}{w^E} \right)^{1-\alpha_E^j} + w^C \left(\frac{1-\alpha_E^j}{\alpha_E} \frac{w^E}{w^C} \right)^{\alpha_E} \right] \frac{1}{A}$$

$$\underline{P} = w^E \left(\frac{1}{\underline{R}_l} \right)^{\frac{1-\alpha_E}{\alpha_E}} \left(\frac{1}{A} \right)^{\frac{1}{\alpha_{E,t}}}$$

$$\rho_P = \alpha_E$$

$$\kappa_P = w^C \underline{R}_l$$

$$q_l^* = A_t \underline{R}_l \left(\frac{\alpha_E}{1-\alpha_E} \frac{w^C}{w^E} \right)^{\alpha_E}$$

Proof. Comes directly out of the zero profit condition for firms and the expression for costs from before. Full proof in Appendix A.1 □

From the previous expression for the equilibrium price schedule we can see that two things happen upon an increase in the stringency of staff-to-child ratios in partial equilibrium (with factor prices constant). First, an increase in the mandatory minimum staff-to-child ratio increases the price level for qualities for which the minimum staff-to-child is binding (due to an increase in κ_P). Second, higher minimum staff-to-child ratios distort lower qualities more than higher qualities, which implies that the price schedule flattens below q^*, that is, in the region in which the staff-to-child ratio binds. This can be seen by looking at $\underline{P}$ and seeing that it is a decreasing function of $\underline{R}_l$. An increase in the price level is likely to push some families out of the market. The flattening of the price schedule is likely to induce families that buy qualities for which the staff-to-child ratio is binding to increase their demand for quality (because the savings from buying lower quality are now lower). Figure 1 plots the price schedule, for a particular age and a particular type of care, before and after an increase in the staff-to-child ratio $\underline{R}_l$, but keeping factor prices w^E, w^C constant.

Lemma 3 and Figure 1 capture the direct effect of regulations, that is, keeping factor prices (the wage of childcare workers and the lead-teacher premium) constant.

While the effects of changing the staff-to-child ratio on the price schedule of quality given factor prices is known, the effect on the demand for lead-teacher talent and total caregiver hours is ambiguous. On the one hand, the fact that the price schedule shifts upward is likely to reduce demand for hours of care, either through the intensive margin (families reducing their hours), the intensive margin (families exiting the childcare market altogether), or more likely, a combination of both. On the other hand, the fact that the price schedule flattens for lower qualities may induce an increase in the demand for lead teacher talent. Furthermore, if families buying qualities such that the regulation is binding keep their demand for quality and quantity constant, this results in an increase in the demand for total caregiving time C (because the stricter

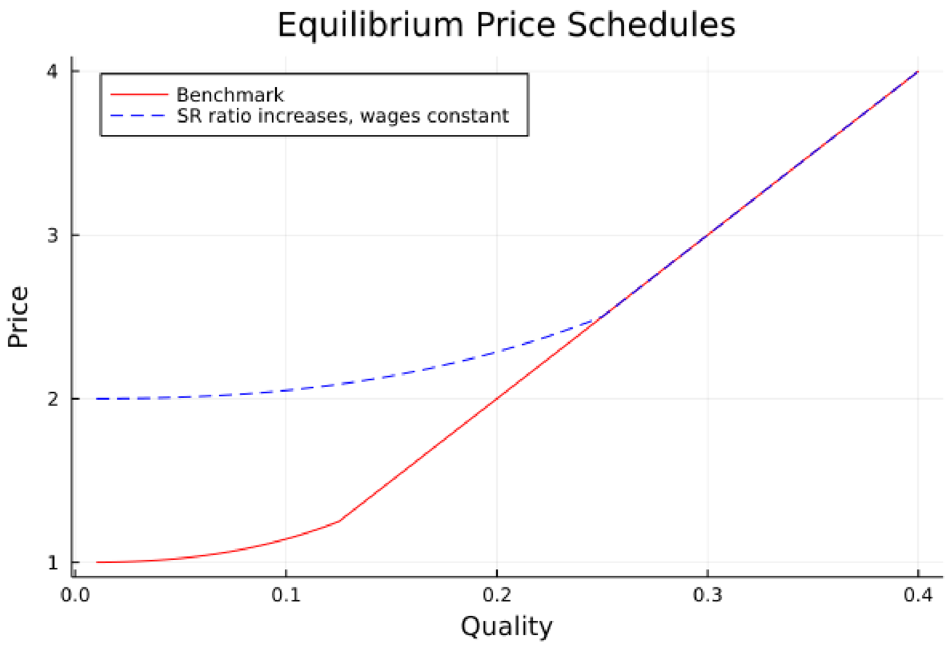


Figure 1: Price schedule before and after an increase on the stringency of the staff-to-child ratio. Factor prices are kept constant

ratio must be met), but in a decrease in the demand for lead teacher talent E (because quality is kept constant). Hence, the effect on factor prices w^E, w^C of an increase in the mandatory minimum staff-to-child ratio is ambiguous. Figure 2 plots the pricing schedule after the regulation change in two scenarios: One scenario in which as a consequence of the regulation change factor prices increase, and another one in which factor prices decrease. As it can be seen in Figure 2 the redistributive effects on the price of childcare associated to changes in the minimum staff-to child ratio could be very different depending on the associated change in factor prices.

Without a change in factor prices, families that buy lower qualities (who are more likely to be poorer) pay a higher price, whereas families that buy high enough qualities (who are likely to be richer) are not affected. However, if factor prices increase as a consequence of the change of the regulation, all the families pay a higher price for childcare.

However, if factor prices go down as a consequence of the change in the regulation, higher qualities are more affordable than in the benchmark, which can benefit higher socioeconomic status families.

3.5 Labor supply of teachers and childcare workers

The labor supply of lead teachers and childcare workers is stylized and follows a constant elasticity of labor supply:

$$LT_l = \overline{LT}_l (w_l^{LT})^{\eta_{LT}} \ ,$$

$$CCW_l = \overline{CCW}_l (w_l^{CCW})^{\eta_{CCW}} \ .$$

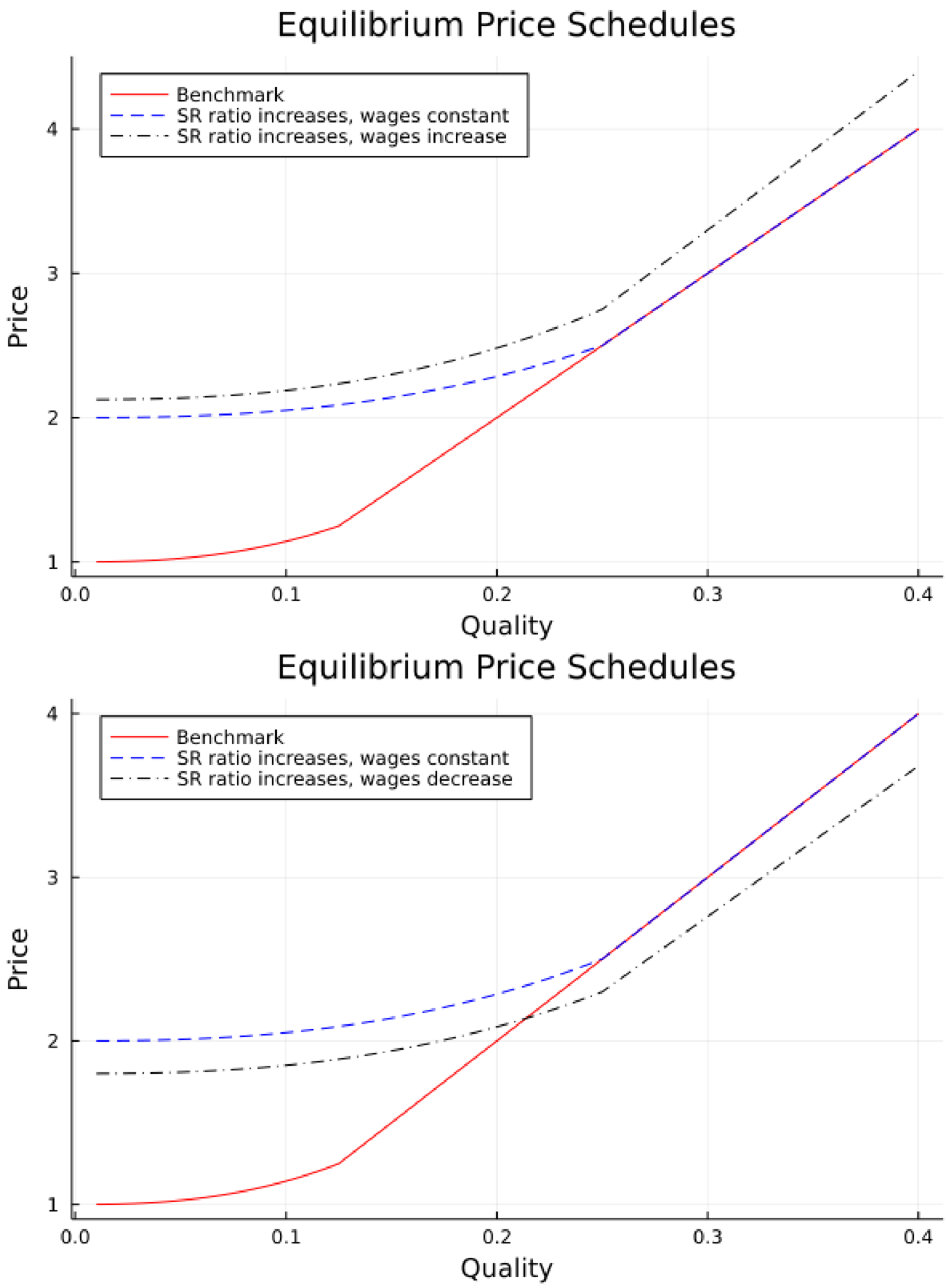


Figure 2: Price schedule for a benchmark level of the mandatory minimum staff-to-child ratio for a benchmark level of regulation and wages (solid line), more stringent regulation but wages as in the benchmark (dashed) and more stringent regulation and different wages (dash-dotted line). The upper pannel shows a scenario in which the change on the regulation leads to an increase in wages and the lower pannel shows a situation in which the change on regulation leads to a decrease in wages.

where LT_l and CCW_l are the number of lead teachers and childcare workers in location l, w_l^{LT} and w_l^{CCW} are their average wages, and η_{LT} and η_{CCW} are their labor supplies. In order to map this labor supplies of lead teachers and childcare workers to factor supplies of efficiency units E and number caregivers C, I make the following assumptions:

First, I assume that for each hour a lead teacher works, she is paid for one hour of caregiving and her efficiency units E. Hence, the hourly wage of a lead teacher with efficiency units E is given by:

$$w^{LT}(E) = w^{E}E + w^{C} \ ,$$

whereas childcare workers are only compensated for the caregiving time that theyoffer, so their wages w^{CCW} are given by the price of hour our of caregiver time:

$$w^{CCW} = w^{C} \ .$$

Second, I assume that the number of hours that lead teachers and childcare workers work is fixed, so changes in hours of caregiving and efficiency units are only given by changes on the extensive margin. Third, I assume that teachers do not know their efficiency units before they join the teaching profession, and I normalize the average efficiency units to 1. Under this assumption an increase in the wages of lead teachers does not induce positive or negative selection into the teaching profession, so a change in the total number of efficiency units supplied is simply the change in the supply of workers becoming lead teachers, times the hours that lead teachers work (which is fixed by assumption), times the average efficiency units of lead teachers (which is fixed and normalized to 1). Under these assumptions, the aggregate supply of paid childcare production factors becomes:

$$E^{S}(w^{E}, w^{C}) = \overline{H}_{LT}\overline{LT}_{l}(w^{E} + w^{C})^{\eta_{LT}} \ , \tag{Supply of E}$$

$$C^{S}(w^{E}, w^{C}) = \overline{H}_{LT}\overline{LT}_{l}(w^{E} + w^{C})^{\eta_{LT}} + \overline{H}_{CCW}\overline{CCW}_{l}(w^{C})^{\eta_{CCW}} \ , \tag{Supply of C}$$

where $\overline{H}_{LT}$ and $\overline{H}_{CCW}$ are the number of hours that lead teachers and childcare workers work respectively.

Note that the supply of efficiency units depends also on the price of an hour of caregiving because part of the remuneration of lead teacher corresponds to caregiving. At the same time, the supply of caregiving hours also depends on the price of efficiency units because lead teachers provide caregiving apart from their efficiency units.

3.6 Equilibrium

The solution concept of this model is competitive equilibrium. In a competitive equilibrium in this economy, center-based and home-based childcare providers enter freely and maximize profits given factor prices (the lead-teacher premium and the wage of childcare workers), and given regulations. In particular, their factor demands are optimal. Given those prices, families decide optimally childcare arrangements, asset accumulation, consumption, and labor supply decisions. The childcare decisions of families include from which type of childcare provider to buy, if any, and how much quality and quantity to buy from that type of childcare provider. These demands need to be satisfied by childcare providers in equilibrium (market clearing). Finally, the demand for lead teachers's efficiency units and caregiver-hours by childcare providers equates the supply for those factors at the equilibrium wages.

Now I define more formally the notion of equilibrium for a location l. I omit the location subscript for economy of notation.

Definition 1 (Equilibrium). *Given mandatory minimum staff to child ratios* $R^{CB} = (R_1^{CB}, R_2^{CB}, R_3^{CB})$, $R^{HB} = (R_1^{HB}, R_2^{HB}, R_3^{HB})$ *and an initial distribution of family types and asset levels* $G_1^{TP}(a_1, H)$, $G_1^{SM}(a_1, H)$ *an Equilibrium in this environment is given by factor prices* (w^C, w^E), *Pricing schedules for paid-care* $\{P_t^{CB}(q,\tau)\}_{t=1}^3$, $\{P_t^{HB}(q,\tau)\}_{t=1}^3$, *labor demand functions for child-care centers per hour of care offered* $\{\hat{C}_t^P(q), \hat{E}_t^D(q)\}_{D=CB,HB}$, *factor supplies* $\overline{C}^{\mathcal{S}}, \overline{E}^{\mathcal{S}}$, *Value functions for families* $\{\{\tilde{V}_t^{TP,D}(a_t,H), \tilde{V}_t^{SM,D}(a_t,H)\}_{D=N,CB,HB}\}_{t=1}^3$ *policy functions for families* $\{\{\{g_{\omega,t}^{TP,D}(a_t,H)\}_{\omega\in\Omega}\}_{D=N,CB,HB}\}_{t=1}^3$, *and* $\{\{\{g_{\omega,t}^{SM,D}(a_t,H)\}_{\omega\in\Omega}\}_{D=N,CB,HB}\}_{t=1}^3$, *endogenous distributions of family types at* $t = 2,3$ $G_2^{TP}(a_2,H)$, $G_2^{SM}(a_2,H)$, $G_3^{TP}(a_2,H)$, $G_3^{SM}(a_2,H)$, *choice probabilities* $\{\mathbb{P}_t^{FT,CB}(a_t,H), \mathbb{P}_t^{FT,HB}(a_t,H)\}_{t=1}^3$ *for* $FT = TP, SM$, *endogenous measures of families over quality of paid care type* D *at time* t $\{\{\mathcal{F}_t^{TP}, \mathcal{F}_t^{SM}\}_{D=CB,HB}\}_{t=1}^3$ *and endogenous measures of paid care providers* $\{\{\mathcal{P}_t\}_{D=CB,HB}\}_{t=1}^3$ *(where measures of families and providers map the Borel sets on* $\mathcal{R}_+$ $\mathcal{B}(\mathcal{R}_+)$ *to* $\mathcal{R}_+$*) such that:*

- *Given factor prices and regulations (Mandatory minimum staff-to-child ratios) the price schedule of paid care providers of type* P *at time* t *is given by Lemma 3.*
- *Factor demands* $\{\{\hat{C}_t^D(q), \hat{E}_t^D(q)\}_{D=CB,HB}\}_{t=1}^3$ *are given by*

$$\hat{E}_t^D(q) = E_t^D(q,1)\,,$$
$$\hat{C}_t^D(q) = C_t^D(q,1)\,,$$

  where $E_t^D(q,h)$ *and* $C_t^D(q,h)$ *are given by Lemma 1*

- $\overline{C}^{\mathcal{S}}, \overline{E}^{\mathcal{S}}$ *are consistent with Supply of* C *and Supply of* E *given wages, that is:*

$$\overline{C}^{\mathcal{S}} = C^{\mathcal{S}}(w^E, w^C)\,,$$
$$\overline{E}^{\mathcal{S}} = E^{\mathcal{S}}(w^E, w^C)\,.$$

- $\tilde{V}^{FT,D}(a_t,H)$ *solve* $\tilde{V}^N$ *and* $\tilde{V}^P$.
- $g_{\omega,t}^{FT,D}(a_t,H)$ *belong to the argmax of* $\tilde{V}^N$ *and* $\tilde{V}^P$ *respectively.*
- *Choice probabilities* $\{\mathbb{P}_t^{FT,CB}(a_t,H), \mathbb{P}_t^{FT,HB}(a_t,H)\}_{t=1}^3$ *are consistent with* $\{\tilde{V}_t^{FT,D}\}_{D=N,CB,HB}$, *that is:*

$$\mathbb{P}_t^{FT,D}(a_t,H) = \mathbb{P}\left(\tilde{V}_t^{FT.D}(a_t,H) - o_t^D \geqslant \max\{\tilde{V}_t^{FT.j}(a_t,H) - o_t^j, \tilde{V}_t^{FT,N}(a_t,H)\}\right)\,,$$
$$\text{for } D = CB, HB \text{ and } j \neq D$$

- $G_{t+1}^{FT}(a_{t+1},H)$ *is consistent with the optimality of saving decisions of households, that is:*

$$G_{t+1}^{FT}(a_{t+1},H) = \int_{a,\tilde{H}\leqslant H}\Big(\sum_{D=N,CB,HB}\mathbb{P}^{FT,D}(a,H)1(g_{a',t}^{FT,D}(a,H) \leqslant a_{t+1})\Big)dG_t(a,\tilde{H})\,.$$

- *The measure of families with a child of age* t *demanding qualities* $q \in Q$ *of type* D *is consistent with the optimal family decisions:*

$$\mathcal{F}_t^{FT,D}(Q) = \mathcal{M}_{FT}\int_{a_t,H}\mathbb{P}_t^{FT,D}(a_t,H)1(g_{q_p,t}^{FT,D}(a_t,H) \in Q)g_{\tau_p,t}^{FT,D}(a_t,H)dG_t(a_t,H) \text{ for all } Q \in \mathcal{B}(\mathcal{R}_+)\,.$$

- $q \in Q$ *at age* t *in market* D *for* $t = 1,2,3$ *and for* $D = CB, HB$ *clears:*

$$\mathcal{F}_t^{SM,D}(Q) + \mathcal{F}_t^{TP,D}(Q) = \mathcal{P}_t^D(Q) \text{ for all } Q \in \mathcal{B}(\mathcal{R}_+) .$$

- *The factor markets clear:*

$$\sum_{t=1}^{3} \sum_{D=CB,HB} \int_q \hat{C}_t^D(q) d\mathcal{P}_t^D(q) = \overline{C}^{\mathcal{S}} ,$$
$$\sum_{t=1}^{3} \sum_{D=CB,HB} \int_q \hat{E}_t^D(q) d\mathcal{P}_t^D(q) = \overline{E}^{\mathcal{S}} .$$

4 Computational Tractability

The model described above is an equilibrium model, where given household structure (Two-Parent vs Singe-Parent households) families are heterogeneous in 7 time invariant characteristics (4 continuous and 3 discrete[9]) and there are 2 dynamic states. Moreover, families make 10 continuous choices and 1 discrete one. There is a market for each type of care (Center-Based and Home-Based), for each age, and for each quality. Apart from these type-age-quality-specific markets for childcare, there are also 2 factor markets (Lead teacher efficiency units and caregivers). Hence, solving for the equilibrium requires solving for 6 equilibrium pricing schedules $P_t^D(q)$, each of them infinite dimensional, and 2 factor prices.

Here I describe the features of the model and numerical techniques that make the computation of the model tractable.

4.1 Household problem

First, instead of discretizing the time-invariant family characteristics H, I sample from the joint distribution of initial assets and family characteristics $G_1(a_1, H)$. High-dimensional integrals are more precisely calculated using Monte-Carlo techniques than quadrature techniques for a given budget of computational resources. This is because relying on simulation automatically zooms in regions of the space of family characteristics that are more likely, increasing precision.

Despite the fact that the household problem has 2 dynamic continuous states, I show that value functions are log-additive in cognitive skills (see Appendix B), which also implies that policy functions do not depend on skills.

A priori, solving for household choices seems computationally very burdensome for a particular household structure $FT = TP, SM$, a vector of family characteristics H and assets a_t, given that households in this model make 10 continuous choices and 1 discrete choice. In order to alleviate that problem, I do the following:

[9]The three discrete are relative care at $t = 1$ and 3 (at $t = 2$ relative care is a deterministic function of relative care at $t = 1$), and the availability of relative care. The support for each of these three variables is finite but contains many values.

- First, I split the problem of making optimal choices in three discrete-choice-conditional problems, one for each $D = N, CB, HB$

- Second, I further split this problem between an asset conditional choice and the choice of choosing assets optimally.

- This step is where the computational gains come. Given future assets a_{t+1} and the discrete choice D we still have 9 continuous choices to make. I show that choosing those optimally amounts to essentially [10] a sequence of one-dimensional root-finding problems, which is more tractable than a 9-dimensional search.

On top of this, the presence of the exponential utility costs allows me to:

- Find the expected future value function in closed form given the discrete-choice-conditional value functions

- Find choice probabilities in closed form for the discrete choice $D = N, CB, HB$ given the choice-conditional value functions

- Use fewer grid points for future assets, given that the future expected value function is differentiable thanks to the exponential costs

4.2 Equilibrium

Looking for an equilibrium of this model seems intractable at first, even given what we know now about the computation of households' optimal choices. This is because we still need to find 6 pricing functions (that is, 6 infinite dimensional objects), and 2 scalar prices. Lemma 3 allows as to reduce this for a search for 2 scalar prices (because given those Lemma 3 allows us to calculate the pricing functions in closed form).

Moreover, once we sample a_t, H, the fact that the choice probabilities are in closed form given the value functions allows me to simulate fewer individuals (essentially, I can compute the probability distribution of choices exactly without relying in montecarlo simulation for each realization of a_1, H, see Appendix D). This allows to check for market clearing in the factor markets more efficiently. Check Appendix E for more details on this.See Appendix C for more details.

5 Data

This paper uses data from a variety of sources. In particular, I combine information from the individual-level surveys Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) and the Survey of Consumer Finances (SCF), with aggregate state-level data on mandatory minimum staff-to-child ratios and wages of teachers and childcare workers.

[10]The word "essentially" here indicates that for some searches in that sequence, we actually have a one-dimensional root-finding problem nested in another one-dimensional root-finding problem

5.1 ECLS-B

The main dataset is the ECLS-B. The ECLS-B is a nationally representative survey of children born in the United States in 2001, and contains information on roughly 10700 children. The ECLS-B contains detailed information on the childcare arrangements and the skills of the children sampled, and socioeconomic information of the families that those children are born into.

5.2 SCF

Because the ECLS-B contains only limited information on assets, I use information from the 2001 wave of the SCF to impute wealth for each ECLS-B family in my analysis at each point in time.

5.3 Wages of lead teachers and childcare workers

I take wages of lead teachers (Occupational code 25-2011) and childcare workers (Occupational Code 39-9011) for the years 2001-2006 from the BLS Occupational Employment and Wage Statistics[11] .

5.4 Staff-to-child ratios

I use data on mandatory minimum staff-to-child ratios for center-based providers for the years 2002 and 2005, and for home-based care providers for years 2001 and 2007. The data for the years 2001 and 2002 was compiled by Sheri Fischer for a project at Wheelock College in Boston. The data for years 2005 and 2007 was also compiled by Sheri Fischer for the 2005 Childcare Licensing Study respectively and the 2007 Childcare Licensing Study respectively.

In the data, regulations are more complex than in the model. For instance, while the model periods correspond to the ages of the child 9 months to 2 years old, 2 years old-4 years old, and 4 years old to 5 years old, mandatory minimum staff-to-child ratios can change within those periods. For example, in 2005 in Alabama, the ratio for centers changes from 4 children per adult before 18 months, to 6 children per adult at 18 months. Moreover, because Home-Based care providers usually mix children of different ages in the same room, mandatory minimum ratios for Home-Based care providers usually depend on the age distribution of children in the room. I transform the raw data on mandatory minimum staff-to-child ratios in a way that I can use in estimation and when solving the model in two ways. First, mandatory minimum staff-to-child ratios change for some states between ages 9 months to 2 years old and 2 years old to 4 years old (corresponding to model periods 1 and 2), but not between 4 and 5 years old (corresponding to model period 3). Hence, for those model periods in which the ratios change I consider the ratio that applies at that model period to be the ratio that applies at the midpoint of the period. That is, when estimating parameters and when solving

[11]See https://www.bls.gov/oes/tables.htm

the model, the staff-to-child ratio at $t = 1$ corresponds to the staff-to-child ratio at age 18 months in the data. Moreover, in the data staff-to-child ratios for home-based care providers can depend on the mix of ages of children in the room and the number of providers, but in the model I assume that neither home-based care providers nor center-based care providers cannot mix ages, and the ratio does not change with the number of providers. The way I construct age-specific staff-to-child ratios for home-based care providers at age t in state l is by reading the regulation of state l and finding out what would be the ratio that applies if home-based care providers in state l were to only provide care to age t. If small home-based care providers (1 staff member in the room) and large home-based care providers (2 or more) I pick the staff-to-child ratio to be the most lenient[12] (the lowest staff-to-child ratio) For the sake of illustration, consider the case of Missouri in 2007. According to the childcare licensing study, Missouri had different licensing requirements in 2007 for small home-based providers (with one adult in the room) and for large home-based providers (with 2 or more adults in the room). A Small Home in Missouri would comply with the mandatory minimum staff-to-child ratio if the one adult in the classroom was caring for at most:

1. 10 children, if 2 are younger than 2 years.
2. **OR** 6 children, if 3 are younger than 2 years.
3. **OR** 4 children, if all are younger than 2 years.

Hence, if a small home-based provider were to provide care in Missouri to only children aged 18 months old, the mandatory minimum ratio that would apply to that home-based care provider would be 1:4. For children aged 3 years old, it would be 1:10. If we look now at Large Home-Based care providers, the 2007 childcare licensing study tells us that 20 children could be cared for by 2 providers without specifying the age of the children, so the ratio that applies at all ages for large home-based care providers is 2:20. Hence, the staff-to-child ratio for home-based care providers in Missouri at age 18 months that I use to estimate and solve the model is 1:10

5.5 Other aggregate data

In order to account for state-level differences in prices, I construct a state-level Price Index for each state in the years 2001-2008 by taking the the State-Level Regional Price Parities of the BEA, and assuming that inflation was the same in each state in the period 2001-2008 and equal to the US overall inflation, which I take from FRED.

I also use state-level fertility data, which I use as an instrument in the estimation of the labor supply elasticity of lead teachers and childcare workers. I get the data on Fertility for each state in the US for the years 1999-2003 from the CDC [13]

[12] Note that if I were to allow in the model for home-based care providers of different sizes facing different regulations, only the size corresponding to the most lenient regulation would survive in equilibrium, given the constant return to scale and free entry assumptions.

[13] https://www.cdc.gov/nchs/data/statab/natfinal2003.annvol1_08.pdf

6 Model Identification

One of the nice features of the model is that we can prove that it is identified given the data that we have available. Because the model has many ingredients and identification proofs for each of these ingredients are different, I relegate to the Appendix the formal identification discussions. Hence, I will only present here a high-level heuristic identification discussion for each group of model parameters and distributions, but the interested reader can refer to the Appendix for more details.

The objects to be identified can be split into 4 different groups, with some subgroups:

1. Parameters of the measurement system for non-parental, maternal, and paternal quality.

2. Household-side Objects

   - Production Function of child skills
   - Initial distribution of households over their time-invariant characteristics H and their initial level of assets a_1
   - Preference parameters
     - Distribution of fixed utility costs of using Home-Based and Center-Based providers
     - Rest of the preference parameters

3. Technology of quality production for Home-Based and Center-Based providers

4. Labor supply of Lead Teachers and Childcare Workers

6.1 Parameters of the measurement system for non-parental, maternal, and paternal quality.

Parental and maternal care quality are latent in nature, and so they are not readily available from ECLS-B data. Because of that, the mapping between latent qualities and data objects needs to be identified. I assume that the relationship between latent care quality j (where j refers to maternal, paternal, relative or paid care) and its s-th noisy measure is given by a linear measurement system:

$$\widetilde{\log q}^{j,s} = \mu^{j,s} + \alpha^{j,s} \log q^{j} + \epsilon^{j,s} \ ,$$
$$\text{for } j = m, f, r, p \text{ and } s = 1, \ldots, N_s \ ,$$

where $\epsilon^{j,s}$ are independent of $\log q^{o}$ for $o = m, f$, $\epsilon^{j,s'}$ for $s \neq s'$ and $\epsilon^{j',p}$ for $j \neq j$ all p (that is, measurement errors are independent of all latent qualities and all the other measurement errors). The noisy measures of parental care are constructed from survey responses to questions about parental attitudes and behaviors. For the mother, scores capturing the quality of mother-child interactions constructed from direct observations are also available, and hence I incorporate them when measuring maternal quality. The two measures of maternal and paternal quality constructed from the same set of questions are normalized by subtracting the mean and dividing by the standard deviation of the raw measure for maternal care. The measure of non-parental quality that I use is the Arnett score. Let's describe first the identification of the measurement systems of paternal and maternal quality. A subset of questions asked by the ECLS-B to mothers and fathers coincides. If the mapping from responses to those questions to latent parental quality is gender-independent, it makes sense to assume that paternal and maternal care qualities are measured in the same units. More formally, letting $\widetilde{\log q}^{m,1}, \widetilde{\log q}^{f,1}$ be the measures of maternal and paternal quality constructed from applying the same transformations to the same questions, we have that:

$$\widetilde{\log q}^{j,1} = \mu^{par,1} + \alpha^{par,1} \log q^j + \epsilon^{j,1} \text{ for } j = m, f\,,$$

where note that $\mu^{par,1}$ and $\alpha^{par,1}$ do not depend on j, that is, they are the same for mothers and fathers (hence the choice of superindex par for parental). Because parental quality is latent and has no natural scale, we can normalize:

$$\begin{aligned} \mu^{par,1} &= 0\,, \\ \alpha^{par,1} &= 1\,. \end{aligned}$$

At this point it is important to note one thing. First, we should only normalize one measure for all types of care. To see why, suppose that we normalize $\mu^{par,1} = \mu^{ARNETT} = 0$ where μ^{ANRETT} stands for the shifter in the measurement system for the standardized Arnett score. Because both the standardized $\widetilde{\log q}^{m,1}$ and the standardized Arnett score both have mean zero by construction, imposing

$$\mu^{par,1} = \mu^{ARNETT} = 0$$

amounts to imposing

$$\mathbb{E}[\log q^m] = \alpha_t^{ANETT} \log q_t^r\,,$$

which implies taking a stand on the location of relative care quality with respect to maternal quality that is not necessary for identification. Doing that may bias the predicted effect on children's skills of policies that reallocate care from relatives to parents or vice-versa. Going back to the identification of measurement parameters, and following arguments similar to those in Cunha, Heckman and Schennach (2010), $\alpha^{j,2}$ can be identified from:

$$\alpha^{j,2} = \frac{cov(\widetilde{\log q}^{j,2}, Z)}{cov(\widetilde{\log q}^{j,1}, Z)}\,,$$

Where Z is an instrument that is uncorrelated with measurement error (in estimation, I use the predicted Arnett score for relative care).

Given $\alpha^{j,2}$, $\mu^{j,2}$ is identified from:

$$\mu^{j,2} = \mathbb{E}[\widetilde{\log q}^{j,2}] - \alpha^{j,2}\mathbb{E}[\widetilde{\log q}^{j,1}] \, .$$

Now, let

$$\log q^{par} = \begin{pmatrix} \log q^{m} \\ \log q^{f} \end{pmatrix} ,$$

$$\widetilde{\log q}^{par,1} = \begin{pmatrix} \widetilde{\log q}^{m,1} \\ \widetilde{\log q}^{f,1} \end{pmatrix} ,$$

and

$$\widetilde{\log q}^{par,2} = \begin{pmatrix} \frac{\widetilde{\log q}^{m,1} - \mu^{m,2}}{\alpha^{m,2}} \\ \frac{\widetilde{\log q}^{f,1} - \mu^{f,2}}{\alpha^{f,2}} \end{pmatrix} .$$

By Theorem 1 in Cunha, Heckman and Schennach (2010), the distribution of $\log q^{par}$ is identified. Moreover, since the measurement errors are fully independent, the distribution of the measurement error is identified by a standard deconvolution argument. See sub-appendix J.3 for more formal identification results on the measurement system, details on the construction of parental quality measures, details on the estimation of the measurement system for parental quality, and estimation results.

It remains to be argued that the Arnett score measurement system parameters are identified. In principle, this claim may seem surprising, because the usual latent factor models tell us that with only one measure we cannot identify the parameters of a measurement system. However, the structure of the model imposes further restrictions that can be exploited to identify α_t^{ARNETT} and μ_t^{ARNETT}. In particular, note that combining the measurement equation for the Arnett score and the production function of quality of paid providers implies the following:

$$\begin{aligned} ARNETT_{i,t} = \mu_t^{Arnett} + \alpha_t^{Arnett}\Big(\log A_t^{HB} + 1(D_{i,t} = CB)(\log A_t^{CB} - \log A_t^{HB})\Big) + \\ \alpha_t^{Arnett}\left(\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right)\right) + \epsilon_t^{Arnett} , \end{aligned}$$

where the term

$$\Big(\log A_t^{HB} + 1(D_{i,t} = CB)(\log A_t^{CB} - \log A_t^{HB})\Big) ,$$

is there to allow for center-based and home-based care providers to have different productivities. From the previous equation it is apparent that if we can construct the term

$$\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t}) ,$$

then α_t^{ARNETT} is identified from the projection of $ARNETT_{i,t}$ on a constant, a dummy for the provider being center-based, and the input composite term

$$\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right) .$$

Because $\left(\frac{E}{k}\right)$ and $\left(\frac{C}{k}\right)$ are data, and $\alpha_{E,t}$ is identified independently of α_t^{ARNETT} (as we will see later), then α_t^{ARNETT} is identified. That is, the factor loading of the Arnett score α_t^{ARNETT} is identified from the constant returns to scale Cobb-Douglas assumption in the production of quality. Note that if the Cobb-Douglas technology of quality production was not constant returns to scale, the relationship between the Arnett score and the input composite term would be

$$\begin{aligned}ARNETT_{i,t} = \mu_t^{Arnett} + \alpha_t^{Arnett}\Big(\log A_t^{HB} + 1(D_{i,t} = CB)(\log A_t^{CB} - \log A_t^{HB})\Big) + \\ \nu\alpha_t^{Arnett}\left(\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right)\right) + \epsilon_t^{Arnett},\end{aligned}$$

where ν is the returns to scale parameter. In this case, ν and α^{ARNETT} wouldn't be separately identified. The location parameter μ^{ARNETT} is identified jointly with the parameters of the Production Function for skills (more on this later). See sub-appendix J.1.1 for a more formal identification argument for α^{ARNETT} and estimation results.

6.2 Household-side objects

6.2.1 Production Function of skills

Let $\widehat{ARNETT}^j$ be the value of the Arnett score of type of care j predicted by observables. Because true quality is assumed to be a **deterministic** function of observables (family characteristics in the case of relative care, production inputs in the case of paid care) $\widehat{ARNETT}^j$ can be written as:

$$\widehat{ARNETT}^j = \frac{\log q^j - \mu^{ARNETT}}{\alpha^{ARNETT}},$$

where the dependence on t is omitted for economy of notation. Substituting the relationship between $\widehat{ARNETT}$ and true quality into the production function for child skills we get:

$$\begin{aligned}\log\theta_{t+1} = \log A_t + \gamma_{\theta,t}\log\theta_t + \gamma_{m,t}\frac{\tau_t^m}{T}\log q^m + \gamma_{f,t}^j\frac{\tau_t^f}{T}\log q^f + \frac{\gamma_{p,t}}{\alpha_t^{ARNETT}}\frac{\tau^p}{T}\widehat{ARNETT}_t^p + \\ \frac{\gamma_{r,t}}{\alpha_t^{ARNETT}}\frac{\tau^r}{T}\widehat{ARNETT}_t^r - \frac{\gamma_{p,t}}{\alpha_t^{ARNETT}}\frac{\tau_t^p}{T}\mu_t^{ARNETT} - \frac{\gamma_{r,t}}{\alpha_t^{ARNETT}}\frac{\tau^r}{T}\mu_t^{ARNETT} + \eta_{t+1}.\end{aligned}$$

Note that η_{t+1} does not affect childcare decisions independently of timing assumptions about when η_{t+1} is realized, because of the log-additivity of skills in the value function. Hence the only unobservables causing an endogeneity problem when estimating the previous equation are the measurement errors in maternal and paternal quality. This source of endogeneity can be tackled by using repeated measures of maternal and paternal care (hence the need for constructing at least two such measures), so identification is restored. Hence, since α_t^{ARNETT} is identified, the production function of skills and the additive shifter of the Arnett score measurement system are identified for $t = 2, 3$.

For $t = 1$ this argument doesn't quite work because we have a missing data problem: At 9 months, ECLS-B did not observe non-parental childcare arrangements, so the Arnett score is not available for $t = 1$. This is not only a problem from the point of view of estimating skill-production function parameters, but also from the point of view of knowing the distribution of $\log q^{r,1}$. In order to tackle these two issues, I make two assumptions: Firs, I assume that $\gamma_{r,1} = \gamma_{p,1}$. This is a sensible restriction because relative care and paternal care are both measured by the Arnett score, so they should be in the same units. Second, I assume that the change in relative care quality between waves 1 and 2 is constant across individuals, that is:

$$\log q_2^r = \log q_1^r + g_1^{q,r} .$$

Under these assumptions, the evolution of skills between $t = 1$ and $t = 2$ for children spending no time in paid care can be written as:

$$\log\theta_2 = \log A_1 + \gamma_{\theta,t}\log\theta_1 + \gamma_{m,1}\frac{\tau_1^m}{\overline{T}}\log q^m + \gamma_{f,1}\frac{\tau_1^f}{\overline{T}}\log q^f +$$
$$\frac{\gamma_{r,1}}{\alpha_2^{ARNETT}}\frac{\tau_1^r}{\overline{T}}\widehat{ARNETT}_2^r - \frac{\gamma_{r,1}}{\alpha_2^{ARNETT}}\frac{\tau_1^r}{\overline{T}}(\mu_2^{ARNETT} + \alpha_2^{ARNETT} g_1^{q,r}) + \eta_2 ,$$

and note that we are using the predicted Arnett score at $t = 2$, since the one at $t = 1$ is not available. Also note that now the term accompanying the fraction of time spent in relative term is not only μ_2^{ARNETT}, the extent to which the standardized Arnett score overstates true quality with respect to parental quality, but also $g_1^{q,r}$, which is the extent to which we would overstate relative quality at 1 if we used predicted quality at 2. It is important to note that we are conditioning in $\tau^P = 0$. If we didn't do that and instead considered the whole population of children, we would have an endogeneity problem coming from the fact that the quality of paid care is not observed at 1. Also, the fact that we are conditioning on $\tau^P = 0$ does not create selection bias, because the only unobservables that appear when estimating the previous equation are η_2 and the measurement errors in parental quality. Since the only source of endogeneity is again measurement error in maternal and paternal quality, and we have repeated measures of maternal and paternal quality, the previous equation is identified. Hence, production function parameters at $t = 1$ and the growth rate of relative care quality between periods 1 and 2 are identified. See sub-appendix J.5 for more formal identification arguments, details on the estimation, and estimation results.

6.2.2 Initial Distribution of household characteristics and assets

Suppose that we want to predict the effects of a policy that is likely to increase the price of childcare for some levels of quality on the behavior of households and the skills of their children. To what extent households are able to substitute paid childcare with informal childcare depends on the availability of relative care, which in the model is given by the household-specific endowment $\overline{T}^r$. Moreover, the extent to which the skills of the children of families that use more relative care as a consequence of this policy change, depends on the quality of the relative care available to these families at each age of the child, which in the model is given by $\{\log q^r_t\}^3_{t=1}$. It is clear then why, in order to predict the distributional effect of childcare market policies in general, and changes to mandatory minimum staff-to-child ratios in particular, it is important to identify the distribution of households over initial levels of wealth, time-invariant family characteristics, and initial skills $G_{a,H,\theta}(a_1, H, \theta)$. I will argue why this distribution is non-parametrically identified under the assumptions of the model for Two-Parent families. The argument for Single Mothers is similar. First, note that the time vector of time-invariant family characteristics contains the wages of the mother, the wages of the father, the quality of maternal care, the quality of paternal care, the path of quality of relative care, and the endowment of relative care, that is:

$$H = (w^m, w^f, q^m, q^f, \{q^r_t\}^3_{t=1}, \overline{T}^r) \ .$$

The wages of the father and the mother are observed for fathers and mothers that work in some wave, and imputed for those that never work in the sample period. Hence, they can be treated as observable.

Quality of relative care is assumed to be a deterministic function of observables for $t = 2, 3$. That is:

$$\log q^r_t = X'_{q,r} \beta^{q,r}_t \ .$$

This implies the following relationship between the Arnett score and observables in $X_{q,r}$:

$$\mathrm{ARNETT}^r_{i,t} = \mu^{\mathrm{ARNETT}}_t + \alpha^{\mathrm{ARNETT}}_t X'_{q,r} \beta^{q,r}_t + \epsilon^{\mathrm{ARNETT}}_{i,t} \ .$$

Even though the households for which the Arnett score of relative care is observed are endogenously selected, identifying the constant and the slope of the projection of the Arnett score on $X_{q,r}$ because the only unobservable present is measurement error in the Arnett score. Hence, the predicted Arnett score $\widehat{\mathrm{ARNETT}}^r_{i,t}$ is observed for all households, which means that $\log q^r_t$ for $t = 2, 3$ is observed for all households, since the measurement parameters of the Arnett score are identified. See sub-appendix J.2 for a more formal identification argument and for estimates of $\alpha_t \beta^{q,r}_t$. Moreover, since the growth of relative care quality between periods 1 and 2 is identified, we can construct $\log q^r_1$ from $\log q^r_2$. This establishes that $\{\log q^r_t\}^3_{t=1}$ can be treated as observed.

The endowment of relative care is assumed to be a deterministic function of family characteristics, that is:

$$\overline{T}^r = \overline{T}^r(Z^{T,r}) < \overline{T} \ ,$$

where $Z^{T,r}$ is a vector of discrete-valued family covariates. Under the assumption that $\log q^m, \log q^f$ can be made arbitrarily small regardless of other family characteristics, the function $\overline{T}^r(Z^{T,r})$ is non-parametrically identified. Intuitively, if parents care about their children, for every possible value of $Z^{T,r}$ we can always find families that draw utility costs of using paid care high enough and that are bad enough parents such that they want to exhaust relative care. Hence $\overline{T}^r(Z^{T,r})$ is identified from the maximum amount of relative care used by families with covariates $Z^{T,r}$. Appendix H provides a formal proof of this result, details on the estimation, and estimation results. Given that $(w^m, w^f, \log q^r, \overline{T}^r)$ can be treated as observable, it suffices to show that the distribution of $(a_1, \log q^m, \log q^f, \log\theta_1)$ given $(w^m, w^f, \log q^r, \overline{T}^r)$ is identified. First, note that because assets are observed in the SCF, the distribution of measurement error in assets is identified from SCF data. Second, the distribution of measurement error for maternal, paternal quality, and initial skills are identified (see Appendix J for the identification argument, details on the estimation, and estimation results). Under the assumption that measurement error in assets is independent of a_1, H and the measurement error in maternal and paternal quality, and measurement error in skills,[14], the distribution of $(a_1, \log q^m, \log q^f, \log\theta_1)$ given $(w^m, w^f, \log q^r, \overline{T}^r)$ is identified from a standard deconvolution argument (given that $(\tilde{a}_1, \widetilde{\log q}^{m,1}, \widetilde{\log q}^{f,1})$ are observed, and the distribution of $(\epsilon_1^a, \epsilon^{m,1}, \epsilon^{f,1})$ is known). Note that the same arguments can be used to identify non-parametrically G_2, G_3.

6.2.3 Preference Parameters

Preference parameters can be divided into parameters of the distribution of the fixed utility costs of using center-based and home-based care, and the rest of preference parameters. Let's start with the rest. These parameters are $\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t}\}_{t=2}^4, \delta_a$ The parameters governing the marginal utility of leisure δ_l^m, δ_l^f are identified by the ratio of the value of consumption to the value of leisure. If the expenditure share of leisure is larger than the expenditure share of consumption for both mothers and fathers, that means that they are likely to value leisure more. $\{\delta_{\theta,t}\}_{t=2}^4$ is identified from observing how much families pay for paid care, how much of an increase in quality they get with respect to their relatives, and how much they consume. If families are willing to pay a lot for center-based care, consume little, and get only a modest increase in quality with respect to free care provided by relatives, that means that $\delta_{\theta,t}$ is high. Note that in the first wave of ECLS-B data on the quality of paid care is not available. I circumvent this limitation by exploiting that the structure of the model implies that the quality of paid care bought by a family in an interior solution for relative care from a provider for which the staff-to-child ratio doesn't bind is known once the quality of relative care is known. The quality of relative care in the first wave given that the quality of relative care in the second wave is known, and the change in relative care quality between the first and the second wave is constant and identified. $\delta_\tau^m, \delta_\tau^f$ are identified once $\{\delta_{\theta,t}\}_{t=0}^3$ and δ_l^m, δ_l^f are known from observing the time use decisions of fathers and mothers. For instance, if we know that mothers value leisure and children's skills a lot, but we observe mothers that are worse at fostering skill development than the relatives they

[14]The full independence assumption on the measurement error in assets is indeed an assumption, that is, it doesn't happen by construction. The imputation procedure relies on projecting net worth on SCF on observables that are both available in SCF and ECLS-B, and then use the coefficient of that projection to impute net worth in ECSLS-B. Such a procedure ensures that the measurement error has zero covariance with the observables used in the imputation, and by extension with the imputed measure of assets, but it does not ensure full independence.

have available spending a lot of time with their children, that means that they should intrinsically value time with their children a lot, which translates into a high τ^m. Finally, δ_a is identified from the following relation:

$$a_4 = \delta_a c \ .$$

Both assets and consumption are measured with error, and the measure of contemporaneous consumption is constructed from the measure of contemporaneous assets. This creates an endogeneity issue, so a regression of assets in wave 4 on consumption on wave 3 does not identify δ_a. Fortunately, consumption at wave 1 can be used as an instrument to recover δ_a. See Appendix L for formal identification results. See Appendix M for details on the estimation procedure and estimation results.

Now let's turn to the parameters of the distribution of the utility costs of using center-based and home-based care. Given data on wages, regulations, and technology parameters for center-based and home-based parameters providers, we can find the price schedules for quality that families face in the model according to the expressions given in Lemma 3. Given knowledge of the rest of the preference parameters, we can find the choice-specific value functions of choosing no paid care, center-based, and home-based care at age 4 for a household with assets and time-invariant characteristics given by (a_3, H). Because the distribution of utility costs is exponential, the probability of a household choosing center (home) care is monotonically increasing provided that that household prefers center (home) care absent the utility costs. Hence, as long as the group of households that prefer center-based care to no paid care and the group of households that prefer home-based care to no paid care both have positive masses, we have that the unconditional choice probabilities are monotonic in the exponential parameters of the utility cost distributions. Using backward induction we can repeat this argument for $t = 2$, and $t = 1$, which implies that $(\lambda_{CB,t}, \lambda_{HB,t})$ are identified for $t = 1, 2, 3$. See Appendix L for formal identification results. See Appendix M for details on the estimation procedure and estimation results.

6.3 Technology of quality production for Home-Based and Center-Based providers

For childcare providers for which the staff-to-child ratio is not binding, $\alpha_{E,t}$ can be identified from the share of income used to remunerate the efficiency units of the lead teacher in the classroom, that is:

$$\alpha_{E,t} = \frac{w^E E_{i,t}}{w^C C_{i,t} + w^E E_{i,t}} \ .$$

The factor prices w^E, w^C are available from state-level data, $C_{i,t}$ is the number of caregivers in the classroom of the child i at age t and $E_{i,t}$ can be constructed from the wage of the lead teacher. Because identifying $\alpha_{E,t}$ requires knowing for which providers the mandatory minimum staff-to-child ratio doesn't bind, and because of the complexity of the regulations for home-based care providers, I use only center-based care providers to estimate $\alpha_{E,t}$. Moreover, at $t = 1$ information of the wage of the lead teacher is not available, so I extrapolate $\alpha_{E,1}$ from $\alpha_{E,2}$ and $\alpha_{E,3}$.

The TFP of childcare providers can be identified from two sources. The first one is by linearly projecting the Arnett score for centers on the input composite

$$\alpha_{E,t} \log\left(\frac{E}{k}\right) + (1 + \alpha_{E,t}) \log\left(\frac{C}{k}\right) .$$

The intercept of this projection combines the efficiency of quality production (log-TFP) with the extent to which the standardized Arnett score overstates the quality of relative care when compared to the measures of parental care (which is measured by the intercept of the measurement system of the Arnett score μ_t^{ARNETT}). Prices also contain information of the efficiency of quality production. Intuitively, if the technology of childcare quality is more efficient, producing the same quality should be cheaper. This is apparent in the expression for $\overline{P}$ in 3. The only reason that the expression for $\overline{P}$ cannot be used directly is that q^P is measured with error and that the measurement error is multiplicative in q^p but additive in $P(q)$ (the price of childcare) [15] However, one can again (as in the identification of preference parameters) use the fact that the quality of paid care is known given for a family who is using relative care (but not exhausting it) and who is buying childcare from a provider for whom the mandatory minimum staff-to-child ratio doesn't bind. See Appendices I and K for formal identification arguments, details on the estimation, and estimation results.

6.4 Labor Supply Lead Teachers and Childcare Workers

In the model, the labor supply of teachers and childcare workers is given by the following constant elasticity of labor supply specifications:

$$LT_l = \overline{LT}_l (w_l^{LT})^{\eta_{LT}} ,$$

$$CCW_l = \overline{CCW}_l (w_l^{CCW})^{\eta_{CCW}} .$$

Using the previous two equations for identification and estimation amounts to assuming that there are no state-specific idiosyncratic reasons that affect the Lead Teachers and Childcare Workers, which can bias the estimates of the labor supply elasticities. We can allow for those unobservable factors by augmenting the estimation equations to be given by:

$$LT_{l,t} = \overline{LT}_l (w_l^{LT})^{\eta_{LT}} \exp(\xi_{l,t}^{LT}) ,$$

$$CCW_{l,t} = \overline{CCW}_l (w_l^{CCW})^{\eta_{CCW}} \exp(\xi_{l,t}^{LT}) .$$

[15] We need the measurement error in $P(q)$ to be additive for the identification argument for $\delta_{\theta,t}$ to go through.

Including these shifters create a familiar simultaneity problem, that prevents us from identifying the labor supply elasticities just from observing aggregate labor supply decissions of lead teachers and childcare workers for each state and their wages. However, we can overcome this problem by using a familiar strategy: We can instrument wages using some labor demand shifter. The instrument that I use in this paper is 2 year-lagged fertility. The rationale for using this instrument is that, if lagged fertility predicts labor demand (as it would be the case if more children in a state translate into a higher demand for childcare in that state) and lead teachers and childcare workers joint the labor force quickly enough after having children (so that current fertility may affect labor supply of teachers and childcare workers, but lagged fertility doesn't), then 2-periods lagged fertility is a valid instrument. Once we have the elasticities, we can identify the shifters by ensuring that the overall level of wages of teachers and childcare workers across states predicted by the model is consistent with the level of wages observed in the data. Intuitively, if we observe wages, and given knowledge of the household-side and provider-side parameters, we can predict whether childcare demand is high or low at those prices by solving the model in partial equilibrium (that is, feeding in the wages into the model and solving for all the other equilibrium objects). If the observed wages are low, and demand is high, then that means that lead teachers and childcare workers are abundant, which translates into a high level for their shifters. See Appendix N for formal identification arguments, details on the estimation, and estimation results.

7 Equilibrium effects of Staff-to-child ratios

7.1 Effects on teachers' earnings

Here compare the factor prices effects of moving each state to the least stringent staff-to-child ratios, the average staff-to-child ratios, and the most stringent staff-to-child ratios. I define the least (most) stringent staff-to-child ratio at age t and for type of care j as the minimum (maximum) staff-to-child ratio across all states at age t for type of care j. The average regulation at age t for type of care j is given by the average staff-to-child ratio at age t for type of care j across states.

The following table shows the least stringent, average, and most stringent regulations across states in my sample for each type of care and for each age:

Before presenting the results, it is worth discussing why a priori the factor price response of an increase in the stringency of staff-to-child ratios is ambiguous: If the minimum ratio for both types of providers becomes more stringent, Lemma 3 tells us that the overall level of prices increases by the change in the ratio times the wage of childcare providers. Because of this increase in price, some families may demand less quantity of childcare τ^P, or exit the market altogether. The degree to which they are able or willing to do that depends on their family characteristics. For instance, if they have a lot of relative care available, they might find it easy to substitute. However, if the quality of their relatives is very low, they might not want do so. This effect depresses

Table 1: Least, average, and most stringent regulations across ages and types of care

	18 months old	3 years old	4 years old
Least stringent, Centers	9	15	20
Least stringent, Homes	10	15	18
Average, Centers	5.28	10.7	12.41
Average, Homes	4.60	6.90	7.07
Most stringent, Centers	3	7	8
Most stringent, Homes	2	3	3

To facilitate interpretation, I report the inverse of the staff-to-child ratios, that is, the number of children per adult in the classroom.

demand for childcare, and all else equal, it depresses the wages of Lead Teachers and childcare workers. Moreover, note that an increase in the stringency of the minimum ratio makes the pricing schedule flatter for qualities for which the minimum ratio is binding. This can be seen again in the expression for $\underline{P}$ in Lemma 3. Intuitively, the mandatory minimum staff-to-child ratio distorts more the cost minimization problem of providers offering lower quality than the cost minimization problem of the ones offering higher quality. Because of that, it reduces the cost difference of providing higher vs lower quality, and compresses the price schedule. This force encourages families to buy higher quality. Because of this increase in quality demanded, childcare providers demand more $\frac{E}{k}$. However, the flattening of the pricing schedule occurs only for qualities for which the mandatory minimum ratio is binding, so demand for $\frac{C}{k}$ does not increase. This effect alone increases the demand for E and has no effect on the demand for C. Moreover, this increase in quality demanded may come together with a decrease in the quantity of childcare for some families, which decreases the demands of both E and C. This reduction on the quantity demanded decreases w^E and w^C. The third partial-equilibrium force is reallocation from the efficiency units of lead teachers E to hours of care C. If the mandatory minimum staff-to-child ratio increases, providers offering qualities for which the minimum ratio is binding need to increase $\frac{C}{k}$ to meet licensing requirements. In order to keep quality constant, they need to decrease $\frac{E}{k}$. This force alone increases w^C and decreases w^E. Overall, the effect on demand of the factors C and E of increasing mandatory minimum staff-to-child ratios is ambiguous, and therefore the response of their factor prices is ambiguous. Appendix O shows quantitative evidence that reinforces that message.

Although the distribution of initial heterogeneity across households is identified separately for each state, the number of observations per state is small. Because of that, I assume that all states within the same US Census Region have the same distribution of household characteristics and initial assets, but regions differ from one another. Given that assumption, states within the same region only differ in their regulations. Because here I am giving all the states the same regulations (the least, most lenient, and the average regulation), I can just report the outcomes for each region (because all the states within the same region have the same outcomes)[16].

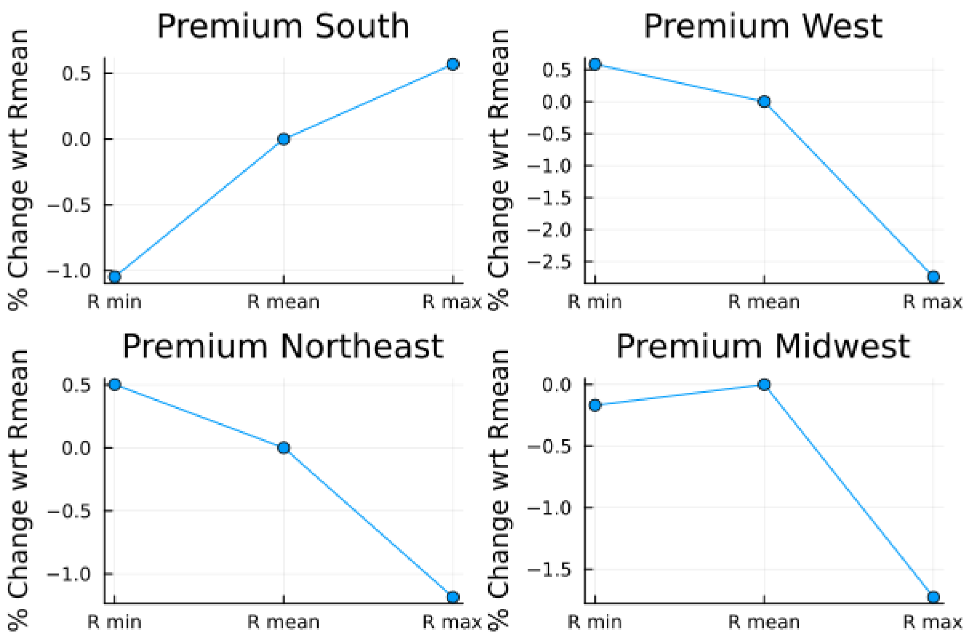


Figure 3: Comparison of average Lead Teacher Premia for different regulation stringencies

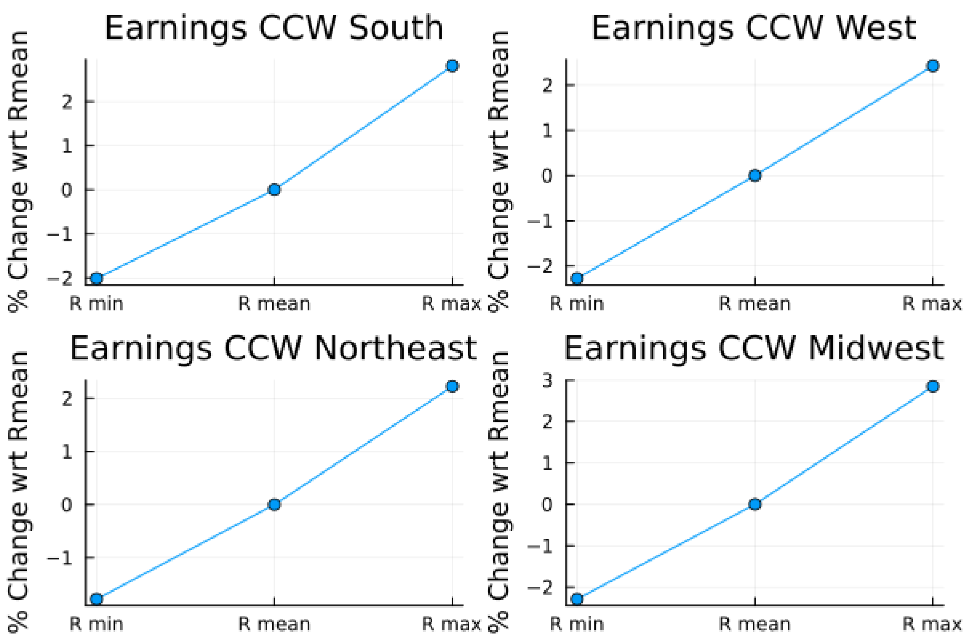


Figure 4: Comparison of Child Care Worker Earnings for different regulation stringencies

As it can be seen, the impact of changing the leniency of regulations on the wages of teachers and the lead teacher premium is not negligible. First, note from Figure 3 that the effect of regulations' stringency on the lead teacher premium is positive in the South,

[16]I am maintaining this assumption throughout the whole results section, and also when estimating the labor supply shifters for Lead Teachers and Childcare Workers (see Appendix N. Moreover, I am dropping states for which I do not observe any family with all the necessary observables to simulate their initial state.

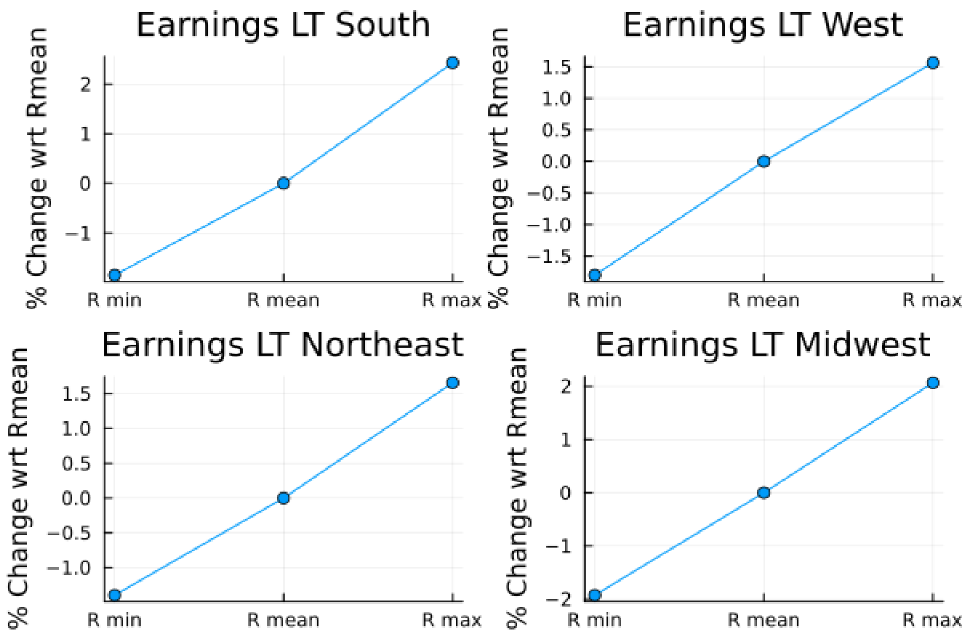


Figure 5: Comparison of Lead Teacher average Earnings for different regulation stringencies

negative in the West and Northeast, and non-monotonic in the Midwest. This is because the different forces described above through which regulations' stringency affects factor prices are more or less important in different regions and within the same region at different levels of stringency. This confirms the observation that the effect of regulations stringency on factors' prices is theoretically ambiguous in sign. The quantitative effect of regulations' stringency on the wage of childcare workers is positive for every region. Going from the average regulation to the most stringent increases childcare worker wages by up to 2-3%. Moreover, going from the average regulation to the least stringent decreases wages by around 2% in all regions. Despite the heterogeneous effects of ratios' stringency on the lead teacher's *premium* across regions, the effect on lead teachers' *wages* is monotonic. This is because the positive effect of regulations on the price of an hour of caregiving work w^C dominates the effects on the premium per hour for the average lead teacher w^E. In fact, increasing the stringency of regulations in the same fashion increases the wages of lead teachers by 1.5%-2%, and decreasing the stringency in the same fashion decreases the wages of lead teachers by 1%-2%.

7.2 Effects on children skills

In this section, I explore the effects that changing from lenient staff-to-child ratios to more stringent staff-to-child ratios has on the skills of children. In order to illustrate these effects, I focus again on an extreme change in the leniency of the regulations. In particular, I focus on how the skills of children at kindergarten entry would change when going from the most lenient regulation (corresponding to the lowest mandatory minimum staff-to-child ratios for each age and type of care) to the most stringent one (corresponding to the highest mandatory minimum staff-to-child ratio for each age and type of care).

Figure 6 shows the overall effects on the distribution of skills of children born to two-parent and single-parent households. Increasing the stringency of the regulation increases skills overall the distribution of children born to two-parent families, and the increase in skills (measured in standard deviations of a math test score taken at kindergarten entry) is roughly constant across the distribution. For children born to one-parent families, the picture is pretty different. Skills at the top of the distribution increase more than at the middle, and skills at the bottom decrease. See Appendix P for an analysis of skill-maximizing ratios at each percentile.

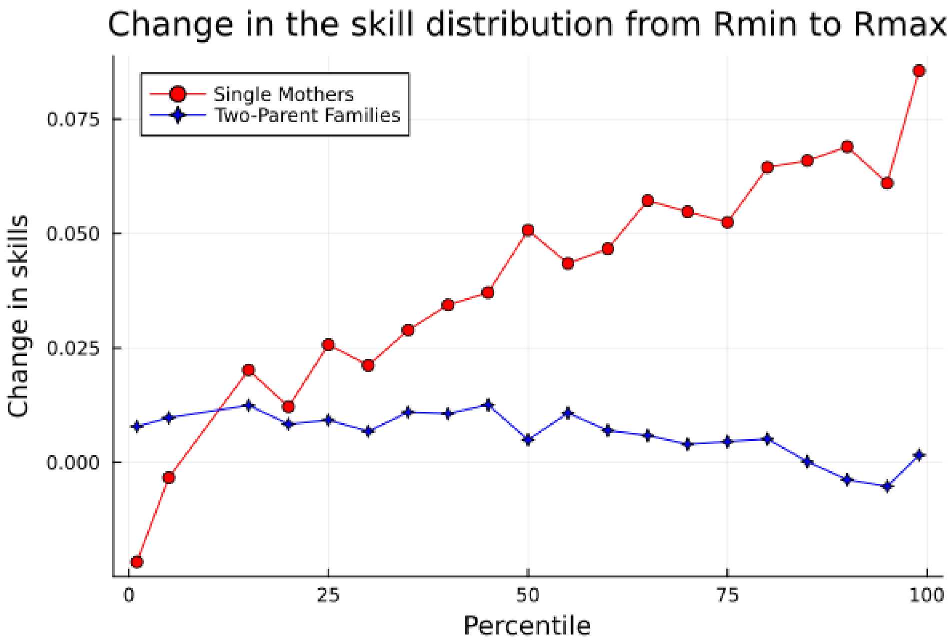


Figure 6: Change associated to going from the least to the most stringent set of regulations in the 1st,10th,20th,30th,40th,50th,60th,70th,80th,90th and 99th percentiles of the distribution of skills for children born to Two-Parent and Single-Parent families.

These relatively muted effects on skills overall distribution can be a product of moderate gains for most children or the product large skill drops compensated by large skill gains. In order to see which of these two stories dominates, I look at the distribution of General Equilibrium Treatment Effects

Define:

$$\Delta \log \theta_{i,4} = \log \theta_{i,4}(R_{max}) - \log \theta_{i,4}(R_{min}) ,$$

to be the change in child skills at kindergarten entry (measure in standard deviations of a math test score) for child i between a scenario in which the US is subject to regulations R_{min} to a scenario in which the US is subject to regulations R_{max}.

Figure 7 shows that the treatment effects $\Delta \log \theta_{i,4}$ are heterogeneous across children, and the heterogeneity is much larger for children of Single Mothers Table 2 illustrates the same point using percentiles of the distribution of $\log \theta_{i,4}$ for children in Two-Parent households and Single-Mother households.

Hence, the relatively small changes overall the skill distribution in Figure 6 is the product of some children experiencing large skill gains, and other children experiencing

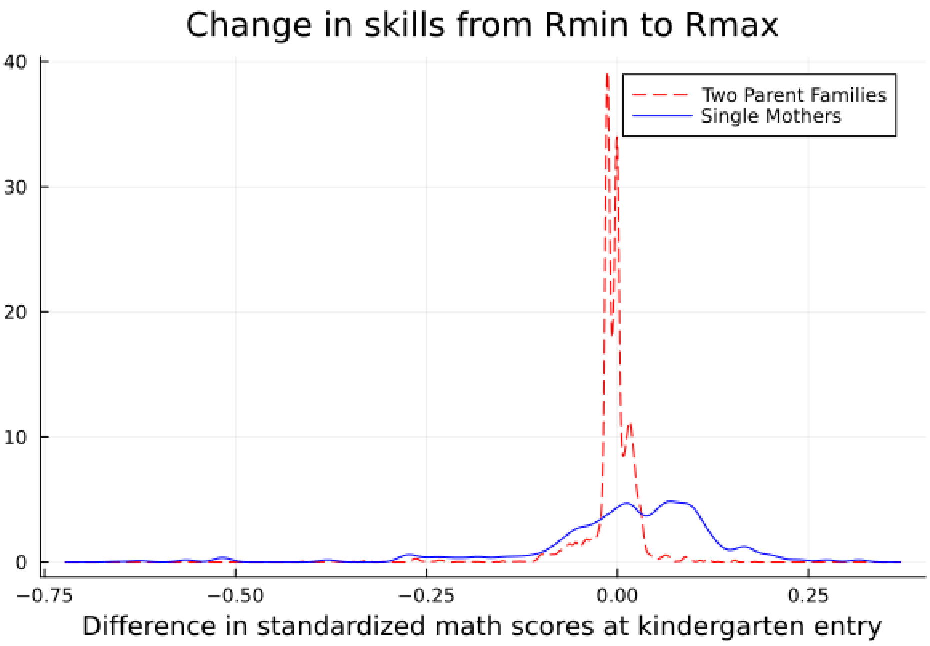


Figure 7: Distribution of GE treatment effects on children skills at pre-school entry. Treatment consists of changing the regulations from the least to the most stringent and letting wages adjust

Percentiles of $\Delta \log \theta_{4,i}$ for TP and SM families

Percentile	Two Parent Families	Single Mothers
1	-24.07%	-47.09%
5	-6.45%	-29.36%
10	-3.77%	-14.76%
25	-1.42%	-0.07%
50	-0.07%	1.08%
75	1.64%	14.11%
90	7.23%	24.87%
95	14.50%	29.47%
99	24.39%	38.86%

Table 2: Percentiles of the treatment effect distribution (measured in standard deviations of a math test score at kindergarten entry) for Two-Parent households and Single-Mother households. X% corresponds to X% of a standard deviation in the data. The treatment consists of changing families from a scenario with lenient regulations to a scenario with stringent regulations. Treatment effects are general equilibrium treatment effects and incorporate the change in wages of Lead Teachers and Child Care Workers between scenarios

large skill losses. This opens the question: Who are the children that win, and who are the children that lose as a consequence of more stringent regulations? Tables **??** and **??** answer that question:

From tables 3 and 4 uncovers different skill-redistribution patterns First, children gaining skills from the increase in stringency are on average born to richer families, both in terms of assets and parental wages Second, children that experience high skill gains are born to poorer families than children that experience moderate gains, and

Winners and losers for TP families

	$\Delta \log \theta > 0$	$\Delta \log \theta < 0$	$\Delta \log \theta > p_{95}$	$\Delta \log \theta < p_5$
a_1	76000	367000	51000	73000
w^m	11	18	9	11
w^f	14	24	11	15
$\log q^m$	-0.03	0.09	-0.15	-0.05
$\log q^f$	-0.11	0.01	-0.27	-0.15
$\log q_2^r$	0.94	1.36	0.13	1.05
$\log q_3^r$	2.89	3.69	1.38	3.13
$\overline{T}^r$	3831	4112	3553	4213

Table 3: Average characteristics of winners and losers for TP families.$p5$ and $p95$ are the 5th and 95th percentiles of the treatment effect distribution respectively.
Assets and wages expressed in 2001 dollars, and the endowment of relative care expressed in annual hours. Assets are rounded to the nearest 1000 and wages to the nearest dollar.

Winners and losers for SM families

	$\Delta \log \theta > 0$	$\Delta \log \theta < 0$	$\Delta \log \theta > p_{95}$	$\Delta \log \theta < p_5$
a_1	26000	65000	7000	33000
w^m	10	10	8	7
$\log q^m$	-0.09	-0.14	-0.15	-0.23
$\log q_2^r$	0.85	0.53	0.49	-0.10
$\log q_3^r$	2.79	1.97	2.34	0.07
$\overline{T}^r$	3817	4670	2465	4967

Table 4: Average characteristics of winners and losers for TP families.$p5$ and $p95$ are the 5th and 95th percentiles of the treatment effect distribution respectively.
Assets and wages are expressed in 2001 dollars, and the endowment of relative care is expressed in annual hours. Assets are rounded to the nearest 1000 and wages to the nearest dollar.

children that experience high skill losses are born to poorer families than children that experience moderate losses. And third, skill gains are decreasing in available hours of relative care.

These patterns deserve some comment. First, remembering Figure 1 we can see that an increase in mandatory minimum ratios increases the price level for lower qualities (for high enough qualities mandatory-minimum ratios do not bind), and flattens the price schedule. The increase in the price level induces families that can substitute paid care to reduce hours of paid care, or even exit the childcare market altogether. The families that can do that are families with higher relative (they can use more relative care without exhausting it), or with more assets (For example, one of the parents can reduce their working hours and spend more time with their child). On the other hand, families with lower relative care available or lower assets are less able to substitute paid care when the price increase. Moreover, because of the flattening effect of regulations (the price of lower qualities increases more than the price of higher qualities), those families may end up buying higher quality. The children of those families experience skill gains. Finally, the equilibrium effects of more stringent regulations increases the wages of lead teachers and childcare workers, which means that even quality levels for which the ratios are not binding are more expensive now. Therefore, families buying quality levels high enough that the ratio is not binding after the increase in the stringency of the regulation still face higher prices, and as a consequence of that they reduce their quality of childcare, their quantity, or both. As a consequence of that, their children suffer skill losses.

8 Conclusion

Mandatory minimum staff-to-child ratios are a common licensing requirement for childcare providers across states. Despite the empirical evidence showing that they have an effect on the market provision of childcare, their effect on children's skills was unexplored. This paper fills this gap by building and estimating an equilibrium model of the childcare market in which paid providers are subject to mandatory minimum staff-to-child ratios and the wages of teachers adjust in response to changes in regulations. This paper shows that changes in the leniency of the minimum ratios have non-negligible impacts on the wages of teachers. Moreover, their impact on the overall distribution of children's skills is modest, but these muted effects hide large heterogeneity. Some children experience large skill gains from the increase in stringency, whereas others experience large skill losses. Both the highest and the lowest skill losses are experience by children born to poor families. Children that experience large skill gains are born to poor families with lower assets and less relative care available, whereas children experiencing large skill losses are born to poor families with higher assets and more relative care available.

This paper is a step toward understanding the effect of different childcare market regulations on the development of young children and their impact on the labor market of early childhood educators. A fruitful avenue for future research is examining the equilibrium impacts of other widespread regulatory requirements for childcare providers, such as minimum educational requirements for teachers.

Appendices

A Proofs of Lemmas 1-3

A.1 Proof of Lemma 1

Consider the cost minimization problem of a childcare provider offering h child-hours[17] at quality q:

$$\begin{aligned} \min_{E,C} w^E E + w^C C \\ \text{s.t } A\left(\frac{E}{h}\right)^{\alpha_E}\left(\frac{C}{h}\right)^{1-\alpha_E} = q \\ \frac{C}{h} \geqslant \underline{R} \end{aligned}$$

The Karush-Kuhn-Tucker conditions are given by:

$$\begin{aligned} E: w^E - \lambda_q \alpha_E A\left(\frac{E}{h}\right)^{\alpha_E-1}\left(\frac{C}{h}\right)^{1-\alpha_E} = 0 \\ C: w^C - \lambda_q (1-\alpha_E)\left(\frac{E}{h}\right)^{\alpha_E}\left(\frac{C}{h}\right)^{-\alpha_E} - \frac{\mu_{\underline{R}}}{h} = 0 \\ CS: \mu_{\underline{R}}\left[\underline{R} - \frac{C}{h}\right] = 0\,, \end{aligned}$$

where λ_q and $\mu_{\underline{R}}$ are the Lagrange multipliers of the quality constraint and the staff-to-child ratio constraints respectively, and CS stands for Complementary Slackness. The previous set of equations are necessary conditions for a local minimum, so a cost-minimizing choice of E, C has to satisfy the previous KKT conditions.

If $\mu_{\underline{R}} = 0$, a solution to the previous equations has to satisfy:

$$\begin{aligned} E: w^E - \lambda_q \alpha_E A\left(\frac{E}{h}\right)^{\alpha_E-1}\left(\frac{C}{h}\right)^{1-\alpha_E} = 0 \\ C: w^C - \lambda_q (1-\alpha_E)\left(\frac{E}{h}\right)^{\alpha_E}\left(\frac{C}{h}\right)^{-\alpha_E} = 0\,. \end{aligned}$$

After some manipulations, we get:

$$\begin{aligned} E = \left(\frac{w^C}{w^E}\frac{\alpha_E}{1-\alpha_E}\right)^{1-\alpha_E}\frac{qh}{A} \\ C = \left(\frac{w^E}{w^C}\frac{1-\alpha_E}{\alpha_E}\right)^{\alpha_E}\frac{qh}{A}\,. \end{aligned}$$

[17] Note that how h is split between the number of children and hours per child does not matter to the provider, since the price per child per hour is $P(q)$

Note that this choice of C is feasible as long as

$$q \geqslant A\underline{R}\left(\frac{w^C}{w^E}\frac{\alpha_E}{1-\alpha_E}\right)^{\alpha_E} = q^* \; .$$

The previous critical point is the optimum for $q \geqslant q^*$, since it coincides with the only critical point of the relaxed problem in which $\underline{R} = 0$. Let's see what happens when $q < q^*$. By Complementary Slackness it follows that the mandatory minimum staff-to-child ratio binds (otherwise we get the previous expression for C, which violates the minimum ratio constraint). Hence, in this region C is given by:

$$C = \underline{R}h$$

Providing quality q requires:

$$\left(\frac{E}{h}\right)^{\alpha_E} = \left(\frac{h}{C}\right)^{1-\alpha_E}\frac{qh}{A} \; .$$

This together with $\frac{C}{h} = \underline{R}$ implies:

$$E = \left(\frac{1}{\underline{R}}\right)^{\frac{1-\alpha_E}{\alpha_E}}\left(\frac{q}{A}\right)^{\frac{1}{\alpha_E}} h \; .$$

This establishes the result. [18]

A.2 Proof of Lemma 2

The cost function is the value of the cost-minimization problem in the previous subsection. This implies:

$$c(q,h) = w^E E(h,q) + w^C C(h,q) \; ,$$

where $E(h,q)$ and $C(h,q)$ are the conditional factor demands in Lemma 1. This establishes the result

A.3 Proof of Lemma 3

The expression for the price-schedule in Lemma 3 is given by the cost $c(q)$, where $c(q)$ satisfies (in a slight abuse of notation):

$$c(q)h = c(h,q) \; .$$

Hence, proving that $P(q)$ is given by the expression in Lemma 3 is equivalent to arguing $P(q) = c(q)$. Suppose $P(q) > c(q)$. Then the profit function for a generic provider is monotonically increasing in the hours supplied h, which implies that the profit

[18] The fact that the only critical point is indeed an optimum follows from the fact that an optimum exists (From Weierstrass Theorem and a bounding above argument) and that the KKT are necessary conditions.

maximization problem of the provider has no solution, which is inconsistent with equilibrium.

Moreover, if $P(q) < c(q)$ then the profit maximization problem of the provider is solved by offering 0 hours of childcare, which is inconsistent with an equilibrium in which positive hours are offered.

Hence, in an equilibrium in which positive hours are supplied, it must be the case that $P(q) = c(q)$. That establishes the result.

B Value functions are log-additive in skills

We can prove that the value function for the single mother is log-additive in the child's skills. The proof for Two-Parent families is similar and is omitted here.

Proposition 1. *For* $t = 1,2,3$

$$V_t^{SM}(a_t, w^m, q^m, q_t^r, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = \tilde{V}_t^{SM}(a_t, w^m, q^m, q_t^r, \overline{T}^r, o_t^{CB}, o_t^{HB}) + \Gamma_t^\theta \log \theta_t^\theta$$

with

$$\Gamma_4^\theta = \delta_{\theta,4}$$

and

$$\Gamma_t^\theta = \beta\gamma_{\theta,t}\Gamma_{t+1}^\theta + \delta_{\theta,t}$$

Proof. The proof proceeds by backward induction. Hence, we have to establish the result for $t = 3$ first.

Recall that for $t = 3$ the value function of the single mother is given by:

$$\begin{aligned}
V_3^{SM}(a_3, w^m, q^m, q_3^r, \overline{T}^r, \theta_t, o_3^{CB}, o_3^{HB}) = \max_{c_3, a_4, n_3^m, l_3^m, \tau_3^m, q_3^p, \tau_3^p, \tau_3^r, P_3} & U^{SM}(c_3, l_3^m, \tau_3^m) + \delta_{\theta,3} \log \theta_3 \\
-o_3^{CB} 1\{P_3 = CB\} - o_3^{HB} 1\{P_3 = HB\} + \beta(V_a^{SM}(a_4) + \delta_{\theta,4} & \mathbb{E}_3 \log \theta_4 \\
\text{s.t} \quad l_3^m + \tau_3^m + n_3^m = \overline{T}^m & \qquad \text{(TC)} \\
\tau_3^m + \tau_3^r + \tau_3^c = \overline{T} & \qquad \text{(SC)} \\
\tau_3^r \in [0, \overline{T}^r] & \qquad \text{(RC)} \\
1\{D_3 = HB\} P_3^{HB}(q_3^p, \tau_3^p) + 1\{D_3 = CB\} P_3^{CB}(q_3^p, \tau_3^p) + c_3 + a_4 = w^m n_3^m + a_3(1+r) & \qquad \text{(BC)} \\
\log \theta_{t+1} = \log A_t + \gamma_{\theta,t} \log \theta_t + \gamma_{m,t} \frac{\tau_t^m}{\overline{T}} \log q^m + \gamma_{p,t} \frac{\tau^p}{\overline{T}} \log q^p + \gamma_{r,t} \frac{\tau^r}{\overline{T}} \log q_t^r + \eta_{t+1} & \qquad \text{(PF)} \\
a_{t+1} \geqslant \underline{a} & \qquad \text{(AC)} \\
D \in \{CB, HB, N\} & \qquad \text{(DCP)} \\
\text{for } t = 3 &
\end{aligned}$$

If we substitute (PF) in the RHS of the value function equation we get:

$$
\begin{aligned}
V_3^{SM}(a_3, w^m, q^m, q_3^r, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = & \max_{c_3, a_4, n_3^m, l_3^m, \tau_3^m, q_3^p, \tau_3^p, \tau_3^r, D_3} U^{SM}(c, l_3^m, \tau_3^m) + \delta_{\theta,3} \log \theta_3 \\
& - o_3^{CB} 1\{D_3 = CB\} - o_3^{HB} 1\{D_3 = HB\} + \beta \delta_{\theta,4} \gamma_{m,3} \frac{\tau_3^m}{\overline{T}} \log q^m + \\
& \beta \delta_{\theta,4} \gamma_{r,3} \frac{\tau_3^r}{\overline{T}} \log q_3^r + \beta \delta_{\theta,4} \gamma_{p,3} \frac{\tau_3^p}{\overline{T}} \log q_3^p + \\
& + \beta \delta_{\theta,4} \log A_3 + \beta V_a^{SM}(a_4) + \beta \delta_{\theta,4} \gamma_{\theta,3} \log \theta_3 \\
& \text{s.t } (TC), (SC), (RC), (BC), (AC), (DCP)
\end{aligned}
$$

by letting

$$
\Gamma_3^\theta = \beta \delta_{\theta,4} \gamma_{\theta,3} + \delta_{\theta,3}
$$

and

$$
\begin{aligned}
\tilde{V}_3^{SM}(a_3, w^m, q^m, q_3^r, \overline{T}^r, o_3^{CB}, o_3^{HB}) = & \max_{c_3, a_4, n_3^m, l_3^m, \tau_3^m, q_3^p, \tau_3^p, \tau_3^r, D_3} U^{SM}(c_3, l_3^m, \tau_3^m) \\
& - o_3^{CB} 1\{D_3 = CB\} - o_3^{HB} 1\{D_3 = HB\} + \beta \delta_{\theta,4} \gamma_{m,3} \frac{\tau_3^m}{\overline{T}} \log q^m + \\
& \beta \delta_{\theta,4} \gamma_{r,3} \frac{\tau_3^r}{\overline{T}} \log q_3^r + \beta \delta_{\theta,4} \gamma_{p,3} \frac{\tau_3^p}{\overline{T}} \log q_3^p + \\
& \beta \delta_{\theta,4} \log A_3 + \beta V_a^{SM}(a_4) \\
& \text{s.t } (TC), (SC), (RC), (BC), (AC), (DCP)
\end{aligned}
$$

we can establish the result for $t = 3$

What remains to be shown is that the backward induction step works. So suppose that

$$
V_{t+1}^{SM}(a_{t+1}, w^m, q^m, q_{t+1}^r, \overline{T}^r, \theta_{t+1}, o_{t+1}^{CB}, o_{t+1}^{HB}) = \tilde{V}_{t+1}^{SM}(a_{t+1}, w^m, q^m, q_{t+1}^r, \overline{T}^r, o_{t+1}^{CB}, o_{t+1}^{HB}) + \Gamma_{t+1}^\theta \log \theta_{t+1}
$$

We want to show that:

$$
V_t^{SM}(a_t, w^m, q^m, q_t^r, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = \tilde{V}_t^{SM}(a_t, w^m, q^m, q_t^r, \overline{T}^r, o_t^{CB}, o_t^{HB}) + \Gamma_t^\theta \log \theta_t
$$

Note that for $t = 1, 2$ we have:

$$
\begin{aligned}
V_t^{SM}(a_t, w^m, q^m, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = \max_{c_t, a_{t+1}, n_t^m, l_t^m, \tau_t^m, q_t^p, \tau_t^p, \tau_t^r, D_t} & U^{SM}(c_t, l_t^m, \tau_t^m) + \delta_{\theta,t} \log \theta_t \\
& - o_t^{CB} 1\{D_t = CB\} - o_t^{HB} 1\{D_t = HB\} + \\
& \beta \mathbb{E} V_{t+1}^{SM}(a_{t+1}, w^m, q^m, q_t^p, \overline{T}^r, \theta_{t+1}, o_{t+1}^{CB}, o_{t+1}^{HB}) \\
\text{s.t} \quad l_t^m + \tau_t^m + n_t^m = \overline{T}^m & \qquad \text{(TC)} \\
\tau_t^m + \tau_t^r + \tau_t^p = \overline{T} & \qquad \text{(SC)} \\
\tau_t^r \in [0, \overline{T}^r] & \qquad \text{(RC)} \\
1\{D_t = HB\} P_t^{HB}(q_t^p, \tau_t^p) + 1\{D_t = CB\} P_t^{CB}(q_t^p, \tau_t^p) + c_t + a_{t+1} = w^m n_t^m + a_t(1+r) & \qquad \text{(BC)} \\
\log \theta_{t+1} = \log A_t + \gamma_{\theta,t} \log \theta_t + \gamma_{m,t} \frac{\tau_t^m}{\overline{T}} \log q^m + \gamma_{p,t} \frac{\tau^p}{\overline{T}} \log q^p + \gamma_{r,t} \frac{\tau^r}{\overline{T}} \log q^r + \eta_{t+1} & \qquad \text{(PF)} \\
a_{t+1} \geqslant \underline{a} & \qquad \text{(AC)} \\
D_t \in \{CB, HB, N\} & \qquad \text{(DCP)}
\end{aligned}
$$

Using the induction hypothesis we get that:

$$
\begin{aligned}
V_t^{SM}(a_t, w^m, q^m, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = \max_{c_t, a_{t+1}, n_t^m, l_t^m, \tau_t^m, q_t^p, \tau_t^p, \tau_t^r, D_t} & U^{SM}(c_t, l_t^m, \tau_t^m) + \delta_{\theta,t} \log \theta_t \\
- o_t^{CB} 1\{D_t = CB\} - o_t^{HB} 1\{D_t = HB\} + \beta \Big(\mathbb{E} \tilde{V}_{t+1}^{SM}(a_{t+1}, w^m, q^m, q_t^r, \overline{T}^r, o_{t+1}^{CB}, o_{t+1}^{HB}) + \Gamma_{t+1}^{\theta} \mathbb{E}_t \log \theta_{t+1} \Big) & \\
\text{s.t (TC), (SC), (RC), (BC), (AC), (PF), (DCP)} &
\end{aligned}
$$

Substituting (PF) we get:

$$
\begin{aligned}
V_t^{SM}(a_t, w^m, q^m, q_t^r, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = \max_{c_t, a_{t+1}, n_t^m, l_t^m, \tau_t^m, q_t^p, \tau_t^p, \tau_t^r, D_t} & U^{SM}(c_t, l_t^m, \tau_t^m) + \delta_{\theta,t} \log \theta_t \\
- o_t^{CB} 1\{D_t = CB\} - o_t^{HB} 1\{D_t = HB\} + \beta \Gamma_{t+1}^{\theta} \gamma_{m,t} \frac{\tau_t^m}{\overline{T}} \log q^m + & \\
\beta \Gamma_{t+1}^{\theta} \gamma_{r,t} \frac{\tau_t^r}{\overline{T}} \log q_t^r + \beta \Gamma_{t+1}^{\theta} \gamma_{p,t} \frac{\tau_t^p}{\overline{T}} \log q_t^p + & \\
\beta \Gamma_{t+1}^{\theta} \log A_t + \beta \mathbb{E} \tilde{V}_{t+1}^{SM}(a_{t+1}, w^m, q^m, q_{t+1}^r, \overline{T}^r, o_{t+1}^{CB}, o_{t+1}^{HB}) + \beta \Gamma_{t+1}^{\theta} \gamma_{\theta,t} \log \theta_t & \\
\text{s.t (TC), (SC), (RC), (BC), (AC), (DCP)} &
\end{aligned}
$$

By letting

$$\Gamma_t^{\theta} = \beta \Gamma_{t+1}^{\theta} \gamma_{\theta,t} + \delta_{\theta,t}$$

and

$$
\begin{aligned}
V_t^{SM}(a_t, w^m, q^m, q_t^r, \overline{T}^r, \theta_t, o_t^{CB}, o_t^{HB}) = \max_{c_t, a_{t+1}, n_t^m, l_t^m, \tau_t^m, q_t^p, \tau_t^p, \tau_t^r, D_t} & U^{SM}(c_t, l_t^m, \tau_t^m) \\
- o_t^{CB} 1\{D_t = CB\} - o_t^{HB} 1\{D_t = HB\} + \beta \Gamma_{t+1}^{\theta} \gamma_{m,t} \frac{\tau_t^m}{\overline{T}} \log q^m + & \\
\beta \Gamma_{t+1}^{\theta} \gamma_{r,t} \frac{\tau_t^r}{\overline{T}} \log q_t^r + \beta \Gamma_{t+1}^{\theta} \gamma_{p,t} \frac{\tau_t^p}{\overline{T}} \log q_t^p + & \\
\beta \Gamma_{t+1}^{\theta} \log A_t + \beta \mathbb{E} \tilde{V}_{t+1}^{SM}(a_{t+1}, w^m, q^m, q_{t+1}^r, \overline{T}^r, o_{t+1}^{CB}, o_{t+1}^{HB}) & \\
\text{s.t (TC), (SC), (RC), (BC), (AC), (DCP)} &
\end{aligned}
$$

we get the desired result. □

C Solution algorithm for the individual problem

We are going to focus on the Two-Parent family case. The Single Mother case is analogous. Denote the time-invariant type of a generic Two-Parent family as H, that is:

$$H = (w^m, w^f, q^m, q^f, \{q_t^r\}_{t=1}^3, \overline{T}^r) \ .$$

Moreover, denote as Y_t all the choices that can be made by a Two-Parent household at time t except for future assets and the paid-care type choice, that is:

$$Y_t = (c_t, n_t^m, n_t^f, l_t^m, l_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p, q_t^p) \ .$$

Define

$$\Psi_t^{i,ft} = \beta \Gamma_{t+1}^{\theta} \gamma_{i,t} \ ,$$

for time investment category $i = m, f, r, p$ and for $t = 1, 2, 3$, with

$$\Gamma_4^{\theta} = \delta_{\theta,4} \ .$$

For given initial assets a_t, time-invariant type H, and utility costs c_t^{CB}, c_t^{HB} the Two-Parent household solves the following problem:

$$\begin{aligned}
\hat{V}_t^{TP}(a_t, H, o_t^{CB}, o_t^{HB}) = \max_{a_{t+1}, Y_t, D_t} \Big\{ &\log c + \delta_l^m \log l^m + \delta_l^f \log l^f + \delta_\tau^m \log \tau^m + \delta_\tau^f \log \tau^f + \\
&\sum_{i=m,f,r,p} \Psi_t^i \frac{\tau_t^i}{\overline{T}} \log q_t^i + \mathbb{E}_t \tilde{V}_{t+1}^{TP}(a_{t+1}, H, o_{t+1}^{CB}, o_{t+1}^{HB}) \\
&- o_t^{CB} 1(D = CB) - o_t^{HB} 1(D = HB) \Big\} \\
c_t + P_t^p(q^p)\tau^p + a_{t+1} &= a_t(1+r) + w^m n^m + w^f n^f \\
\tau_t^j + l_t^j + n_t^j &= \overline{T}^j \\
\tau_t^m + \tau_t^f + \tau_t^r + \tau_t^p &= \overline{T} \\
\tau^r &\leqslant \overline{T}^r \\
a_{t+1} &\geqslant \underline{a} \\
c_t, l_t^m, l_t^f, n_t^m, n_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p &\geqslant 0
\end{aligned}$$

Note that for the sake of simulating from the probability distribution of individuals' optimal choices, we do not need to solve for the optimal choice for each level of o_t^{CB}, o_t^{HB}, but rather it is enough to solve for the (a_{t+1}, Y_t) for each D and for the choice probabilities for $D = N, CB, HB$. Hence, define the D-specific value functions:

$$\tilde{V}_t^{TP,N}(a_t, H) = \max_{a_{t+1}, Y_t} \Big\{ \log c_t + \delta_l^m \log l^m + \delta_l^f \log l^f + \delta_\tau^m \log \tau^m + \delta_\tau^f \log \tau^f + \sum_{i=m,f,r} \Psi_t^i \frac{\tau_t^i}{\overline{T}} \log q^i + \mathbb{E}_t \tilde{V}_{t+1}^{TP}(a_{t+1}, H, c_{t+1,CB}, c_{t+1,HB}) \Big\} \tag{$\tilde{V}^N$}$$

$$c_t + a_{t+1} = a_t(1+r) + w^m n^m + w^f n^f$$
$$\tau_t^j + l_t^j + n_t^j = \overline{T}^j$$
$$\tau_t^m + \tau_t^f + \tau_t^r = \overline{T}$$
$$\tau^r \leqslant \overline{T}^r$$
$$a_{t+1} \geqslant \underline{a}$$
$$c_t, l_t^m, l_t^f, n_t^m, n_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p \geqslant 0$$

$$\tilde{V}_t^{TP,D}(a_t, H) = \max_{a_{t+1}, Y_t} \Big\{ \log c + \delta_l^m \log l^m + \delta_l^f \log l^f + \delta_\tau^m \log \tau^m + \delta_\tau^f \log \tau^f + \sum_{i=m,f,r,p} \Psi_t^i \frac{\tau_t^i}{\overline{T}} \log q_t^i + \mathbb{E}_t \tilde{V}_{t+1}^{TP}(a_{t+1}, H, c_{t+1,CB}, c_{t+1,HB}) \Big\} \tag{$\tilde{V}^P$}$$

$$c_t + a_{t+1} + P_t^P(q^p)\tau^p = a_t(1+r) + w^m n^m + w^f n^f$$
$$\tau_t^j + l_t^j + n_t^j = \overline{T}^j$$
$$\tau_t^m + \tau_t^f + \tau_t^r + \tau_t^p = \overline{T}$$
$$\tau^r \leqslant \overline{T}^r$$
$$a_{t+1} \geqslant \underline{a}$$
$$c_t, l_t^m, l_t^f, n_t^m, n_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p \geqslant 0$$

for $D = CB, HB$.

Let the policy functions from the D-specific problem be given by $g_{\omega,t}^{TP,D}(a_t, H)$ for $\omega \in \Omega = \{c, a', n^m, n^f, l^m, l^f, \tau^m, \tau^f, \tau^r, \tau^p, q^p\}$ and $D = N, CB, HB$.

We can re-write this problem as:

$$\tilde{V}_t^{TP,D}(a_t, H) = \max_{a_{t+1} \geqslant \underline{a}} \Big\{ \overline{V}_t^{TP,D}(a_t, H; a_{t+1}) + \beta \mathbb{E}_{t+1} \tilde{V}_{t+1}^{TP}(a_{t+1}, H, o_{t+1}^{CB}, o_{t+1}^{HB}) \Big\} \tag{Dynamic choice}$$

where for $D = CB, HB$:

$$\overline{V}_t^{TP,D}(a_{t+1}; a_t, H) = \max_{Y_t} \Big\{ \log c + \delta_l^m \log l^m + \delta_l^f \log l^f + \delta_\tau^m \log \tau^m + \delta_\tau^f \log \tau^f +$$

(Static Choice)

$$\sum_{i=m,f,r,p} \Psi_t^i \frac{\tau_t^i}{\overline{T}} \log q_t^i \Big\}$$

$$c_t + P_t^p(q^p)\tau^p = a_t(1+r) + w^m n^m + w^f n^f - a_{t+1}$$

$$\tau_t^j + l_t^j + n_t^j = \overline{T}^j$$

$$\tau_t^m + \tau_t^f + \tau_t^r + \tau_t^p = \overline{T}$$

$$\tau^r \leqslant \overline{T}^r$$

$$a_{t+1} \geqslant \underline{a}$$

$$c_t, l_t^m, l_t^f, n_t^m, n_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p \geqslant 0$$

and an analogous expression applies to $\overline{V}^{TP,N}$.

The previous discussion implies that if we happen to know $\overline{V}_t^{TP,D}(a_{t+1}; a_t, H)$ and $\mathbb{E}_t \tilde{V}_{t+1}(a_{t+1}, H, o_{t+1,CB}, o_{t+1,HB})$ as a function of a_{t+1}, then we can solve for a_{t+1}, Y in two steps:

1. First, we choose a_{t+1} to solve Dynamic choice. This is a one-dimensional optimization problem.

2. Second, we choose Y_t to solve Static Choice given the choice of assets a_{t+1}. This is an optimization problem in 10 dimensions, but given the special structure of the problem and the discussion in C.2 it is still tractable.

Finally, given $\tilde{V}_t^{TP,D}$ for $D = CB, HB, N$ it is straightforward to find choice probabilities in closed-form according to:

$$\mathbb{P}(D = N|a_t, H) = e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{N})} \qquad \text{(Choice Probabilities)}$$

$$\mathbb{P}(D = HB|a_t, H) = \begin{cases} (1 - e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{N})}) e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} + 1 - e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} - \\ \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{H}} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} (1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{N})}) \\ \text{if } \tilde{V}^{HB} \geqslant \tilde{V}^{CB} \\ (1 - e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{N})}) e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} + \\ e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{HB})} - e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} - \\ \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \left[e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{HB})} - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{N})} \right] \\ \text{otherwise} \end{cases}$$

$$\mathbb{P}(D = CB|a_t, H) = 1 - \mathbb{P}(D = HB|a_t, H) - \mathbb{P}(D = N|a_t, H) \,,$$

and where the particular parametric form follows from the fact that $o^{CB} \sim \exp(\lambda^{CB})$ and $o^{HB} \sim \exp(\lambda^{HB})$. Note that I am leaving the dependence of the Value Functions on a_t, H to avoid clutter of notation.

Next, we need to describe how to compute $\mathbb{E}_t \hat{V}^{TP}_{t+1}(a_t, H, o^{CB}_t, o^{HB}_t)$ for $t = 1, 2$ and $\overline{V}^{TP,D}_t(a_{t+1}; a_t, H)$ for $t = 1, 2$ and $D = CB, HB, N$

C.1 Computing flow indirect value and future expected value as a function of assets

C.1.1 Computing flow indirect value

For given a_t, H, D, t, we want to solve for $\overline{V}^{TP,D}(a_t, H; a_{t+1})$ as a function of a_{t+1}, where $\overline{V}^{TP,D}(a_t, H; a_{t+1})$ is defined as before. In order to do that, we define a grid for assets tomorrow $a'_{grid} = \{\underline{a}, a_1(a_t, H), \dots, a_{max}(a_t, H)\}$. Then, for each a' in that grid we solve for $\overline{V}^{TP,D}_t(a_t, H; a')$ by maximizing Static Choice with respect to Y_t. This can be done by finding all the critical points of the asset conditional lagrangian following the procedure in C.2. Once we have done this, a continuous approximation to $\overline{V}^{TP,D}_t(a_t, H; a_{t+1})$ is given by simply using linear interpolation over the values of this function on a'_{grid}.

C.1.2 Computing Expected Continuation value

This part of the solution algorithm proceeds by backward induction. Fix a family time-invariant type H and type-specific grid for assets $a_{grid} = \{\underline{a}, a_1(H), a_2(H), \dots, a_{max}(H)\}$

- For each $a \in a_{grid}$ and for each paid-care type choice P, the Two-Parent family last period problem ($t = 3$) can be solved by finding all the critical points of the last period lagrangian, which are characterized by the FOC in the asset-conditional lagrangian in Appendix L plus the following optimality condition for assets:

$$a_4 = \delta_a c$$

  These critical points can be found by following the procedure in C.2 and letting $a_4 = \delta_a c$ (so no separate numerical optimization is required to find assets). Then, all the critical points found are evaluated, and the one that maximizes the last period flow utility plus the continuation value of assets is selected. Its associated value yields $\tilde{V}^{TP,D}_3(a_{grid}, H)$. From here, we can get the expected value function at time t (and in particular at time $t = 3$) as:

$$\mathbb{E}_t \tilde{V}^{TP}_3(a_t, H, o^{CB}_t, o^{HB}_t) = \mathbb{P}(D = CB | a_t, H)\left(\tilde{V}^{TP,CB}(a_t, H) - \mathbb{E}(o^{CB}_t | D_t = CB)\right) + \mathbb{P}(D = HB | a_t, H)\left(\tilde{V}^{TP,HB}(a_t, H) - \mathbb{E}(o^{HB}_t | D_t = HB)\right) + \mathbb{P}(D_t = N | a_t, H)\tilde{V}^{TP,N}_t(a_t, H)\,,$$

  where the expression for the choice probabilities is available in closed form as a function of $\tilde{V}^{TP,D}$ as in Choice Probabilities and the expression for the conditional expectation of the cost is available in closed form also as a function of $\tilde{V}^{TP,D}$ according to the expression in C.5.3

- At $t = 2, 1$ we want to solve Dynamic choice for each $a_t \in a_{grid}$ and for each $D = CB, HB, N$. Hence, for a fix a_t and choice of paid care P we need to know the static indirect flow payoff $\overline{V}_t^{TP,D}(a_t, H, a_{t+1})$ associated with choosing a_{t+1}. This can be done as described above. Note that we may or may not choose this grid to coincide with a_{grid}. In general, it will make sense not to make them coincide. Once we have done this, a continuous approximation to $\overline{V}_t^{TP,D}(a_t, H; a_{t+1})$ is given by simply using linear interpolation over the values of this function on a'_{grid}. Analogously, we can find a continuous approximation to the expected continuation value by using linear interpolation on a_{grid}. Given these two objects, solving Dynamic choice and getting $\tilde{V}_t^{TP,D}(a_t, H)$ simply amounts to solving an optimization problem with bounds in one dimension. Once we do that for each $a_t \in a_{grid}$ and each $D = CB, HB, N$, we can find $E_1\tilde{V}_2^{TP}(a_t, H)$ for each a_t by using the same closed-form expression for the expected continuation value as for $t = 3$

C.2 Asset-conditional problem at $t = 1, 2$ for the two parent family given paid care type $D = CB, HB$

Here I explain how to solve the asset-conditional problem when the family is choosing paid care for pricing functions given by:

$$P_t^j(q^p, \tau^p) = \begin{cases} \overline{P} q \tau^P & \text{if } q > q_{j,t}^* \\ \left[\underline{P} q^{\frac{1}{\rho_P}} + \kappa_P \right] \tau^P & \text{if } q^P \leqslant q_{j,t}^* \end{cases}$$

Given choices of assets tomorrow a' and private care type P a the Lagrangian of the two-parent family is given by:

$$\begin{aligned} \mathcal{L} = {} & \log c + \delta_l^m \log l^m + \delta_l^f \log l^f + \delta_\tau^m \log \tau^m + \delta_\tau^f \log \tau^f + \Psi^m \frac{\tau^m}{\overline{T}} \log q^m + \Psi^f \frac{\tau^f}{\overline{T}} \log q^f + \\ & \Psi^r \frac{\tau^r}{\overline{T}} \log q^r + \Psi^p \frac{\tau^p}{\overline{T}} \log q^p + \lambda^{SC} [\overline{T} - \tau^m - \tau^f - \tau^r - \tau^p] + \lambda^m [\overline{T}^m - \tau^m - n^m - l^m] + \\ & \lambda^f [\overline{T}^f - \tau^f - n^f - l^f] + \lambda^{BC} [a(1+r) - a' + w^m n^m + w^f n^f - c - P^P(q^p)\tau^p] + \\ & \mu_\tau^r \tau^r + \mu_\tau^p \tau^p + \mu_n^m n^m + \mu_n^f n^f + \omega^r [\overline{T}^r - \tau^r] \end{aligned}$$

Notice that I am omitting time subscripts and family type superscript to avoid clatter of notation. Also note that I am ignoring the terms associated to the non-negativity constraints that are never binding. Taking First Order Conditions we get:

$$\frac{\partial \mathcal{L}}{\partial c} = \frac{1}{c} - \lambda^{BC} = 0$$
$$\frac{\partial \mathcal{L}}{\partial l^m} = \frac{\delta_l^m}{l^m} - \lambda^m = 0$$
$$\frac{\partial \mathcal{L}}{\partial n^m} = -\lambda^m + \lambda^{BC} w^m + \mu_n^m = 0$$
$$\frac{\partial \mathcal{L}}{\partial l^f} = \frac{\delta_l^f}{l^f} - \lambda^f = 0$$
$$\frac{\partial \mathcal{L}}{\partial n^f} = -\lambda^f + \lambda^{BC} w^f + \mu_n^f = 0$$
$$\frac{\partial \mathcal{L}}{\partial \tau^m} = \frac{\delta_\tau^m}{\tau^m} + \frac{\Psi^m}{\overline{\overline{T}}} \log q^m - \lambda^{SC} - \lambda^m = 0$$
$$\frac{\partial \mathcal{L}}{\partial \tau^f} = \frac{\delta_\tau^f}{\tau^f} + \frac{\Psi^f}{\overline{\overline{T}}} \log q^f - \lambda^{SC} - \lambda^f = 0$$
$$\frac{\partial \mathcal{L}}{\partial \tau^r} = \frac{\Psi^r}{\overline{\overline{T}}} \log q^r - \lambda^{SC} + \mu_\tau^r - \omega^r = 0$$
$$\frac{\partial \mathcal{L}}{\partial \tau^p} = \frac{\Psi^p}{\overline{\overline{T}}} \log q^p - \lambda^{SC} - \lambda^{BC} P^D(q^p) + \mu_\tau^p = 0$$
$$\frac{\partial \mathcal{L}}{\partial q^p} = \Psi^p \frac{\tau^p}{\overline{\overline{T}}} \frac{1}{q^p} - \lambda^{BC} \frac{dP^D}{dq^p}(q^p)\tau^P = 0$$

The idea of the solution algorithm for the Static Choice is to find all the critical points of the Lagrangian of the family's problem at time t conditional on choosing childcare type D and then pick the one that yields maximal flow payoff. In order to do that, I split the choice set in multiple regions and show that in most cases finding a critical point of the asset-conditional Lagrangian amounts to solving a one-dimensional root-finding problem. Unfortunately, the function whose root we need to find is not the same in each region, so I adopt a case-by-case approach. The remaining of this sub-appendix details this strategy.

1. **Interior solution** The necessary conditions for an interior solution are given by:

$$\frac{\delta_l^m}{l^m} = \frac{w^m}{c} \tag{TPI1}$$
$$\frac{\delta_l^f}{l^f} = \frac{w^f}{c} \tag{TPI2}$$
$$\frac{\delta_\tau^m}{\tau^m} + \frac{\Psi^m}{\overline{\overline{T}}} \log q^m - \lambda^{SC} = \frac{\delta_l^m}{l^m} \tag{TPI3}$$
$$\frac{\delta_\tau^f}{\tau^f} + \frac{\Psi^f}{\overline{\overline{T}}} \log q^f - \lambda^{SC} = \frac{\delta_l^f}{l^f} \tag{TPI4}$$
$$\frac{\Psi^P}{\overline{\overline{T}}} \log q^P - \lambda^{SC} = \frac{P^D(q^P)}{c} \tag{TPI5}$$
$$\Psi^p \frac{\tau^p}{\overline{\overline{T}}} \frac{1}{q^p} = \frac{1}{c} \frac{dP^D}{dq^p}(q^p)\tau^P \text{ if } q^P \neq q^*_{l,j,t} \tag{TPI6}$$
$$\lambda^{SC} = \frac{\Psi^r}{\overline{\overline{T}}} \log q^r_t \tag{TPI7}$$

These FOC have three possible solutions, one for $q^P < q^*$ (ignoring the subscripts for clarity), another one for $q^P = q^*$ and another one for $q^P > q^*$.

- $q^P < q^*$

  Note that for $q^P < q^*$
  $$\frac{dP^P(q^P)}{dq^P} = \underline{P}^P \frac{1}{\rho_P}(q^P)^{\frac{1-\rho_P}{\rho_P}}$$
  where we are ignoring the subscripts in α to reduce the clutter of notation. Using equation TPI6 then we get:
  $$q^{P,-}(c) = \left(\frac{\rho_P}{\underline{P}^P}\frac{\Psi^P}{\overline{T}}c\right)^{\rho_P} \tag{$q^{P,-}(c)$}$$
  Plugging this in equation TPI5 and using the expression for $P^P(q^P)$ we get that consumption has to satisfy:
  $$\frac{\Psi^P}{\overline{T}}\rho_P \log\left(\frac{\rho_P}{\underline{P}^D}\frac{\Psi^P}{\overline{T}}\right) + \frac{\Psi^P}{\overline{T}}\rho_P \log c - \frac{\kappa_P}{c} - \lambda^{SC} - \rho_P\frac{\Psi^P}{\overline{T}} = 0 \tag{RFc}$$
  where

  Let $g(c)$ be the LHS of the previous expression, and note that $\lim_{c\to 0} g(c) = -\infty$ and $\lim_{c\to\infty} g(c) = \infty$. Hence, by the intermediate value theorem and the fact that g is monotonically increasing, $g(c)$ has only one root in $\mathbb{R}_+$ and a bisection converges linearly to that root. Let that root be c_r. Then q^P is given by $q^{P,-}(c)$ evaluated at c_r

  If $q^P > q^*$ then there cannot be an interior solution with $q^P < q^*$. If $q^P \leqslant q^*$ then the **candidate** to an interior optimum with $q^P < q^*$ is given by:
  $$\begin{aligned}
  c &= c_r \\
  q^P &= \left(\frac{\rho_P}{\underline{P}^D}\frac{\Psi^P}{\underline{T}}c_r\right)^{\rho_P} \\
  l^m &= \frac{\delta_l^m c}{w^m} \\
  l^f &= \frac{\delta_l^f c}{w^f} \\
  \tau^m &= \frac{\delta_\tau^m}{\frac{\delta_l^m}{l^m} - \left(\frac{\Psi^m}{\overline{T}}\log q^m - \frac{\Psi^r}{\overline{T}}\log q_t^r\right)} \\
  \tau^f &= \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f} - \left(\frac{\Psi^f}{\overline{T}}\log q^f - \frac{\Psi^r}{\overline{T}}\log q_t^r\right)} \\
  n^m &= \overline{T}^m - \tau^m - l^m \\
  n^f &= \overline{T}^f - \tau^f - l^f \\
  \tau^P &= \frac{w^m n^m + w^f n^f + a_t(1+r) - c_r - a_{t+1}}{P^D(q^P)} \\
  \tau^r &= \overline{T} - \tau^P - \tau^m - \tau^f
  \end{aligned}$$

- $q^P = q^*$ Using $q^P = q^*$ and TPI5 we get:

$$c^*(q,\lambda^{SC}) = \frac{P^P(q)}{\frac{\Psi^P}{T}\log q^P - \lambda^{SC}} \tag{$c(q^*,\lambda^{SC})$}$$

Evaluating this at $(q = q^*, \lambda^{SC} = \frac{\Psi^r}{T}\log q^r)$ we get c.
From here, we can obtain $l^m, l^f, \tau^m, \tau^f, n^m, n^f, \tau^P, \tau^r$ in the same way as in the candidate to interior optimum with $q^P < q^*$

- $q^P > q^*$ In this case $P^P(q^P) = \overline{P}^P q^P \tau^P$ From TPI5 and TPI6 we get:

$$q^{P,+}(\lambda^{SC}) = \exp\left\{1 + \frac{T}{\Psi^P}\lambda^{SC}\right\} \tag{$q^{P,+}(\lambda^{SC})$}$$

Evaluating this expression at $\lambda^{SC} = \frac{\Psi^r}{T}\log q_t^r$ we get our candidate for q^P. Evaluating $q^{P,+}(\lambda^{SC})$ at the resulting q^P and at $\lambda^{SC} = \frac{\Psi^r}{T}\log q_t^r$ we get consumption as:

$$c = \frac{\overline{P}^P q^P}{\frac{\Psi^P}{T}} \tag{$c^+(q^P)$}$$

From here, we can obtain $l^m, l^f, \tau^m, \tau^f, n^m, n^f, \tau^P, \tau^r$ in the same way as in the candidate to interior optimum with $q^P < q^*$

If in any of the previous sub-cases for q^p any of the variables is negative, that means that no interior optimum can exist with that q^* and we can proceed to check the next sub-case.

2. **Mother does not work ($n^m = 0$) , everything else is interior** $n^f > 0, 0 < \tau^r < \overline{T}^r, \tau^P > 0$. **(At most 3 critical points)** . First, note that since relative care is interior we have:

$$\lambda^{SC} = \frac{\Psi^r}{T}\log q_t^r$$

In this case τ^m and l^m have to satisfy:

$$\frac{\delta_\tau^m}{\tau^m} + \frac{\Psi^m}{T}\log q^m - \lambda^{SC} = \frac{\delta_l^m}{l^m}$$

Letting $\tau^m = \phi_\tau \overline{T}^m$ we can re-write this as:

$$\frac{\delta_\tau^m}{\phi_\tau \overline{T}^m} + \frac{\Psi^m}{T}\log q^m - \lambda^{SC} = \frac{\delta_l^m}{(1-\phi_\tau)\overline{T}^m}$$

If we substract the LHS to the RHS we can notice that by the intermediate value theorem there is a unique solution for ϕ_τ in $(0,1)$ Given that in an optimum $0 < \phi_\tau < 1$ we have that ϕ_τ has to satisfy the following quadratic equation:

$$\underbrace{\left(\Psi^m \frac{\overline{T}^m}{T}\log q^m - \overline{T}^m\lambda^{SC}\right)}_{a}\phi_\tau^2 + \underbrace{\left[\delta_l^m + \delta_\tau^m - \left(\Psi^m \frac{\overline{T}^m}{T}\log q^m - \overline{T}^m\lambda^{SC}\right)\right]}_{b}\phi_\tau + \underbrace{-\delta_\tau^m}_{c} = 0 \tag{QE ϕ_τ^m}$$

The only admissible solution to this quadratic equation is given by:

$$\phi_\tau = \frac{-b+\sqrt{b^2-4ac}}{2a}$$

There are three candidates for q^P, c that can be part of an optimum. These are precisely the same as in 1. For each combination (τ^m, l^m, c, q^P) (there are at most 6) let:

$$\begin{aligned}
l^f &= \frac{\delta_l^f c}{w^f} \\
\tau^f &= \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f} - \left(\frac{\Psi^f}{\overline{T}}\log q^f - \frac{\Psi^r}{\overline{T}}\log q^r\right)} \\
n^f &= \overline{T}^f - \tau^f - l^f \\
\tau^P &= \frac{w^m n^m + w^f n^f + a_t(1+r) - c_r - a_{t+1}}{P^P(q^P)} \\
\tau^r &= \overline{T} - \tau^P - \tau^m - \tau^f
\end{aligned}$$

If for some of these candidates to optimum, some of these variables is negative, we can discard this candidate. If for all of them is negative then no optimum can exist on this region.

3. **Father does not work ($n^f = 0$), everything else is interior $n^m > 0, 0 < \tau^r < \overline{T}^r, \tau^P > 0$ (At most 6 critical points)**

   Symmetric to 2. The corresponding quadratic equation for τ^f is given by:

$$\underbrace{\left(\Psi^f \frac{\overline{T}^f}{\overline{T}} \log q^f - \overline{T}^f \lambda^{SC}\right)}_{a} \phi_\tau^2 + \underbrace{\left[\delta_l^f + \delta_\tau^f - \left(\Psi^f \frac{\overline{T}^f}{\overline{T}} \log q^f - \lambda^{SC}\right)\right]}_{b} \phi_\tau - \underbrace{\delta_\tau^f}_{c} = 0 \quad (\text{QE } \phi_\tau^f)$$

4. **Mother does not work and no relative care ($n^m = 0, \tau^r = 0$), everything else is interior.** In this case TPI2 and TPI6 still hold. Apart from these two conditions a critical point in this region is characterized by the following equations:

$$\lambda^{SC} = \frac{\Psi^m}{\overline{T}} \log q^m + \frac{\delta_\tau^m}{\tau^m} - \frac{\delta_l^m}{l^m}$$

$$\begin{gathered}
\lambda^{SC} = \frac{\Psi^f}{\overline{T}} \log q^f + \frac{\delta_\tau^f}{\tau^f} - \frac{\delta_l^f}{l^f} \\
\frac{\Psi^P}{\overline{T}} \log q^P - \lambda^{SC} = \frac{P^D(q^P)}{c} \\
l^m + \tau^m = \overline{T}^m \\
\tau^P + \tau^f + \tau^m = \overline{T} \\
c + a_{t+1} + P^P(q^P)\tau^P = w^f n^f + a_t(1+r) \\
n^f + l^f + \tau^f = \overline{T}^f
\end{gathered}$$

- $q^P < q^*$

  This case has to be solved numerically. The numerical approach here involves a one-dimensional root-finding algorithm nested within another one-dimensional root-finding algorithm.

  A critical point in this region is characterized by a root of the supervision constraint residual as a function of λ^{SC}, which is defined as:

$$SCR(\lambda^{SC}) = \overline{T} - \tau^m(\lambda^{SC}) - \tau^f(\lambda^{SC}) - \tau^P(\lambda^{SC})$$

  where: $c(\lambda^{SC})$ solves RFc and $q^P(\lambda^{SC})$ is given by evaluating $q^{P,-}(c)$ at $c(\lambda^{SC})$. Moreover

$$l^f(\lambda^{SC}) = \frac{\delta_l^f c(\lambda^{SC})}{w^f}$$

$$\tau^f(\lambda^{SC}) = \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f} - \left(\frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC}\right)}$$

  and

$$n^f(\lambda^{SC}) = \overline{T}^f - \tau^f(\lambda^{SC}) - l^f(\lambda^{SC})$$

  $\tau^m(\lambda^{SC}) = \phi_\tau^m \overline{T}^m$ where ϕ_τ^m solves QE ϕ_τ^m .

  $l^m(\lambda^{SC})$ is given by:

$$l^m(\lambda^{SC}) = \overline{T}^m - \tau^m(\lambda^{SC})$$

$$\tau^P(\lambda^{SC}) = \frac{a_t(1+r) + w^f n^f(\lambda^{SC}) - a_{t+1} - c(\lambda^{SC})}{P^P(q^P(\lambda^{SC}))}$$

- $q^P = q^*$ This case can be solved as a one-dimensional root-finding problem in the budget constraint residual, which is given by:

$$BCR(c) = a_t(1+r) + w^f n^f(c) - P^P(q^*)\tau^P(c) - c - a_{t+1}$$

  with

$$l^f(c) = \frac{\delta_l^f c}{w^f}$$

$$\lambda^{SC}(c) = \frac{\Psi^P}{\overline{T}} \log q^* - \frac{P^P(q^*)}{c}$$

  $\tau^m(c) = \phi_\tau^m \overline{T}^m$, where ϕ_τ^m solves Equation QE ϕ_τ^m .

$$\tau^f(c) = \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(c)} - \left(\frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC}(c)\right)}$$

  and

$$l^m(c) = \overline{T}^m - \tau^m(c)$$

$$\tau^P(c) = \overline{T} - \tau^f(c) - \tau^m(c)$$

$$n^f(c) = \overline{T}^f - \tau^f(c) - l^f(c)$$

- $q^P > q^*$ This case can be solved as a one-dimensional root-finding algorithm on the Supervision Constraint Residual as a function of λ^{SC}:

$$SCR(\lambda^{SC}) := \overline{T} - \tau^m(\lambda^{SC}) - \tau^f(\lambda^{SC}) - \tau^P(\lambda^{SC}) = 0$$

with:

$$\begin{aligned} q^P(\lambda^{SC}) &= \exp\left(1 + \frac{\overline{T}}{\Psi^P}\lambda^{SC}\right) \\ c(\lambda^{SC}) &= \frac{\overline{P}^P q^P(\lambda^{SC})}{\frac{\Psi^P}{\overline{T}}} \\ l^f(\lambda^{SC}) &= \frac{\delta_l^f c(\lambda^{SC})}{w^f} \end{aligned}$$

$\tau^m(\lambda^{SC}) = \phi_\tau^m \overline{T}^m$, where ϕ_τ^m solves Equation QE ϕ_τ^m .

$$\begin{aligned} \tau^f(c) &= \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(\lambda^{SC})} - \left(\frac{\Psi^f}{\overline{T}}\log q^f - \lambda^{SC}\right)} \\ n^f(\lambda^{SC}) &= \overline{T}^f - \tau^f(\lambda^{SC}) - l^f(\lambda^{SC}) \\ l^m(\lambda^{SC}) &= \overline{T}^m - \tau^m(\lambda^{SC}) \\ \tau^P(\lambda^{SC}) &= \frac{w^f n^f(\lambda^{SC}) + a_t(1+r) - c(\lambda^{SC}) - a_{t+1}}{\overline{P}^P q^P(\lambda^{SC})} \end{aligned}$$

5. **Mother does not work and relative care endowment is exhausted ($n^m = 0, \tau^r = \overline{T}$), everything else is interior.**

   Similar to 4, but with $\tau^r = \overline{T}^r$

6. **Father does not work and no relative care ($n^f = 0, \tau^r = 0$), everything else is interior.**

   Symmetric to 4

7. **Father does not work and relative care endowment is exhausted ($n^f = 0, \tau^r = \overline{T}$), everything else is interior.**

   Similar to 6 but with $\tau^r = \overline{T}^r$

8. **Father and mother do not work ($n^m = n^f = 0$), everything else interior**

   In this case TPI3-TPI7 still need to hold. These equations jointly with $n^m = n^f = 0$, the budget constraints, the supervision constraints and the parental time-use constraints determine the critical points.

   Moreover, since τ^r is interior we already know that in a critical point in this region:

$$\lambda^{SC} = \frac{\Psi^r}{\overline{T}}\log q^r$$

   Again, let's analyze the three possible cases for q^P separately:

- $q^P < q^*$ Because TPI5 and TPI6 still need to hold, c has to satisfy RFc. Quality of paid care q^P can be obtained by evaluating $q^{P,-}(c)$ at the resulting c. Since $n^m = n^f = 0$ and TPI3 and TPI4 need to hold, $\tau^m = \phi^m_\tau \overline{T}^m$ and $\tau^f = \phi^f_\tau \overline{T}^f$ where ϕ^m_τ and ϕ^f_τ satisfy QE ϕ^m_τ and QE ϕ^f_τ respectively. From the parental time-use constraints:

$$l^m = \overline{T}^m - \tau^m$$
$$l^f = \overline{T}^f - \tau^f$$

From the budget constraint:

$$\tau^P = \frac{a_t(1+r) - c - a_{t+1}}{P(q^P)}$$

From here we can get τ^r as:

$$\tau^r = \overline{T} - \tau^P - \tau^f - \tau^m$$

- $q^P = q^*$ Evaluating $c(q^*, \lambda^{SC})$ at $(q, \lambda^{SC}) = (q^*, \frac{\Psi^r}{\overline{T}} \log q^r)$. From here, we can get the remaining variables as in the $q^P < q^*$ case
- $q^P > q^*$ Evaluating $q^{P,+}(\lambda^{SC})$ at $\lambda^{SC} = \frac{\Psi^r}{\overline{T}} \log q^r$ we get q^P.
Evaluating $c^+(q^P)$ at $q^P(\lambda^{SC})$ we get $c(\lambda^{SC})$
From here, we can get the rest of the variables as in the previous two cases.

9. **Father and mother do not work and relative care is 0 ($n^f = n^m = \tau^r = 0$)**

In this case TPI3-TPI6 have to hold.

- $q < q^*$ This case can be solved for again as a one-dimensional root-finding problem nested within another one-dimensional root-finding problem.
A critical point in this region is characterized by a root of the supervision constraint residual as a function of λ^{SC}, which is defined as:

$$SCR(\lambda^{SC}) = \overline{T} - \tau^m(\lambda^{SC}) - \tau^f(\lambda^{SC}) - \tau^P(\lambda^{SC})$$

where:
$c(\lambda^{SC})$ solves RFc and $q^P(\lambda^{SC})$ is given by evaluating $q^{P,-}(c)$ at $c(\lambda^{SC})$. Moreover
$\tau^m(\lambda^{SC}) = \phi^m_\tau \overline{T}^m$ where ϕ^m_τ solves QE ϕ^m_τ and $\tau^f(\lambda^{SC}) = \phi^f_\tau \overline{T}^f$ where ϕ^f_τ solves QE ϕ^f_τ.
$l^m(\lambda^{SC})$ is given by:

$$l^m(\lambda^{SC}) = \overline{T}^m - \tau^m(\lambda^{SC})$$

$l^f(\lambda^{SC})$ is given by:

$$l^f(\lambda^{SC}) = \overline{T}^F - \tau^F(\lambda^{SC})$$

and:

$$\tau^P(\lambda^{SC}) = \frac{a_t(1+r) - a_{t+1} - c(\lambda^{SC})}{P^P(q^P(\lambda^{SC}))}$$

- $q^P = q^*$ This case can be solved as a one-dimensional root-finding problem in the budget constraint residual, which is given by:

$$BCR(c) = a_t(1+r) + w^f n^f(c) - P^P(q^*)\tau^P(c) - c - a_{t+1}$$

with

$$\lambda^{SC}(c) = \frac{\Psi^P}{\overline{T}} \log q^* - \frac{P^P(q^*)}{c}$$

$\tau^m(c) = \phi_\tau^m \overline{T}^m$, where ϕ_τ^m solves Equation QE ϕ_τ^m and $\tau^f(c) = \phi_\tau^f \overline{T}^f$, where ϕ_τ^f solves Equation QE ϕ_τ^f .
and

$$\begin{aligned} l^m(c) &= \overline{T}^m - \tau^m(c) \\ l^f(c) &= \overline{T}^f - \tau^f(c) \\ \tau^P(c) &= \overline{T} - \tau^f(c) - \tau^m(c) \\ n^f(c) &= \overline{T}^f - \tau^f(c) - l^f(c) \end{aligned}$$

- $q^P > q^*$ This case can be solved as a one-dimensional root-finding algorithm on the Supervision Constraint Residual as a function of λ^{SC}:

$$SCR(\lambda^{SC}) := \overline{T} - \tau^m(\lambda^{SC}) - \tau^f(\lambda^{SC}) - \tau^P(\lambda^{SC}) = 0$$

with:

$$\begin{aligned} q^P(\lambda^{SC}) &= \exp\left(1 + \frac{\overline{T}}{\Psi^P}\lambda^{SC}\right) \\ c(\lambda^{SC}) &= \frac{\overline{P}^P q^P(\lambda^{SC})}{\frac{\Psi^P}{\overline{T}}} \\ l^f(\lambda^{SC}) &= \frac{\delta_l^f c(\lambda^{SC})}{w^f} \end{aligned}$$

$\tau^m(\lambda^{SC}) = \phi_\tau^m \overline{T}^m$, where ϕ_τ^m solves Equation QE ϕ_τ^m and $\tau^f(\lambda^{SC}) = \phi_\tau^f \overline{T}^f$, where ϕ_τ^f solves Equation QE ϕ_τ^f

$$\begin{aligned} l^f(\lambda^{SC}) &= \overline{T}^f - \tau^f(\lambda^{SC}) \\ l^m(\lambda^{SC}) &= \overline{T}^m - \tau^m(\lambda^{SC}) \\ \tau^P(\lambda^{SC}) &= \frac{a_t(1+r) - c(\lambda^{SC}) - a_{t+1}}{\overline{P}^P q^P(\lambda^{SC})} \end{aligned}$$

10. **Father and mother do not work and relative care is exhausted ($n^f = n^m = 0, \tau^r = \overline{T}^r$)**

    Same as 9 but with $\tau^r = \overline{T}^r$

11. **Relative care is 0, everything else interior ($\tau^r = 0$, everything else interior)**

    In this case TPI1- TPI6 have to hold.

- $q^P < q^*$
  This case can be solved for again as a one-dimensional root-finding problem nested within another one-dimensional root-finding problem.
  A critical point in this region is characterized by a root of the supervision constraint residual as a function of λ^{SC}, which is defined as:

$$SCR(\lambda^{SC}) = \overline{T} - \tau^m(\lambda^{SC}) - \tau^f(\lambda^{SC}) - \tau^P(\lambda^{SC})$$

  where:
  $c(\lambda^{SC})$ solves RFc and $q^P(\lambda^{SC})$ is given by evaluating $q^{P,-}(c)$ at $c(\lambda^{SC})$. Moreover

$$l^m(\lambda^{SC}) = \frac{\delta_l^m c(\lambda^{SC})}{w^m}$$

$$l^f(\lambda^{SC}) = \frac{\delta_l^f c(\lambda^{SC})}{w^f}$$

$$\tau^m = \frac{\delta_\tau^m}{\frac{\delta_l^m}{l^m(\lambda^{SC})} - \left(\frac{\Psi^m}{\overline{T}} \log q^m - \lambda^{SC}\right)}$$

$$\tau^f = \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(\lambda^{SC})} - \left(\frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC}\right)}$$

  and

$$\tau^P(\lambda^{SC}) = \frac{a_t(1+r) + w^f n^f(\lambda^{SC}) + w^m n^m(\lambda^{SC}) - a_{t+1} - c(\lambda^{SC})}{P^P(q^P(\lambda^{SC}))}$$

- $q^P = q^*$ This case can be solved as a one-dimensional root-finding problem in the budget constraint residual, which is given by:

$$BCR(c) = a_t(1+r) + w^f n^f(c) + w^m n^m(c) - P^P(q^*)\tau^P(c) - c - a_{t+1}$$

  with

$$l^f(c) = \frac{\delta_l^f c}{w^f}$$

$$l^m(c) = \frac{\delta_l^m c}{w^m}$$

$$\lambda^{SC}(c) = \frac{\Psi^P}{\overline{T}} \log q^* - \frac{P^P(q^*)}{c}$$

$$\tau^m(c) = \frac{\delta_\tau^m}{\frac{\delta_l^m}{l^m(c)} - \left(\frac{\Psi^m}{\overline{T}} \log q^m - \lambda^{SC}(c)\right)}$$

$$\tau^f(c) = \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(c)} - \left(\frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC}(c)\right)}$$

$$\tau^P(c) = \overline{T} - \tau^f(c) - \tau^m(c)$$

$$n^f(c) = \overline{T}^f - \tau^f(c) - l^f(c)$$

$$n^m(c) = \overline{T}^m - \tau^m(c) - l^m(c)$$

- $q^P > q^*$ This case can be solved as a one-dimensional root-finding algorithm on the Supervision Constraint Residual as a function of λ^{SC}:

$$SCR(\lambda^{SC}) := \overline{T} - \tau^m(\lambda^{SC}) - \tau^f(\lambda^{SC}) - \tau^P(\lambda^{SC}) = 0$$

with:

$$\begin{aligned}
q^P(\lambda^{SC}) &= \exp\left(1 + \frac{\overline{T}}{\Psi^P}\lambda^{SC}\right) \\
c(\lambda^{SC}) &= \frac{\overline{P}^P q^P(\lambda^{SC})}{\frac{\Psi^P}{\overline{T}}} \\
l^f(\lambda^{SC}) &= \frac{\delta_l^f c(\lambda^{SC})}{w^f} \\
l^m(\lambda^{SC}) &= \frac{\delta_l^m c(\lambda^{SC})}{w^m}
\end{aligned}$$

$$\begin{aligned}
\tau^f(\lambda^{SC}) &= \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(\lambda^{SC})} - \left(\frac{\Psi^f}{\overline{T}}\log q^f - \lambda^{SC}\right)} \\
\tau^m(\lambda^{SC}) &= \frac{\delta_\tau^m}{\frac{\delta_l^m}{l^m(\lambda^{SC})} - \left(\frac{\Psi^m}{\overline{T}}\log q^m - \lambda^{SC}\right)} \\
n^f(\lambda^{SC}) &= \overline{T}^f - \tau^f(\lambda^{SC}) - l^f(\lambda^{SC}) \\
n^m(\lambda^{SC}) &= \overline{T}^m - \tau^m(\lambda^{SC}) - l^m(\lambda^{SC}) \\
\tau^P(\lambda^{SC}) &= \frac{w^f n^f(\lambda^{SC}) + a_t(1+r) - c(\lambda^{SC}) - a_{t+1}}{\overline{P}^P q^P(\lambda^{SC})}
\end{aligned}$$

12. **Relative care is exhausted, everything else interior ($\tau^r = \overline{T}^r$)**

    Same as 11 but with $\tau^r = \overline{T}^r$

13. **Paid care is zero, everything else is interior ($\tau^P = 0$, else interior)** In this case TPI1-TPI5 and TPI7 need to hold.

    The critical points in this region can be solved for as a one-dimensional root-finding problem in the Budget Constraint Residual as a function of c, which is given by:

$$BCR(c) := a_t(1+r) + w^f n^f(c) + w^m n^m(c) - c - a_{t+1}$$

where:

$$
\begin{aligned}
l^m(c) &= \frac{\delta_l^m c}{w^m} \\
l^f(c) &= \frac{\delta_l^f c}{w^f} \\
\lambda^{SC}(c) &= \frac{\Psi^r}{\overline{T}} \log q^r \\
\tau^f(c) &= \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(c)} - \left(\frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC}(c)\right)} \\
\tau^m(c) &= \frac{\delta_\tau^m}{\frac{\delta_l^m}{l^m(c)} - \left(\frac{\Psi^m}{\overline{T}} \log q^m - \lambda^{SC}(c)\right)} \\
\tau^r(c) &= \overline{T} - \tau^m(c) - \tau^f(c) \\
n^m(c) &= \overline{T}^m - l^m(c) - \tau^m(c) \\
n^f(c) &= \overline{T}^f - l^f(c) - \tau^f(c)
\end{aligned}
$$

14. **Paid care and relative care are zero ($\tau^r = \tau^P = 0$) , everything else interior.** In a critical point in this region TPI1-TPI4 have to hold.

    This case can be solved for as a one-dimensional root-finding algorithm on the Budget Constraint Residual as a function of c:

$$
BCR(c) := a_t(1+r) + w^f n^f(c) + w^m n^m(c) - c - a_{t+1}
$$

where:

$$
\begin{aligned}
l^m(c) &= \frac{\delta_l^m c}{w^m} \\
l^f(c) &= \frac{\delta_l^f c}{w^f} \\
\tau^m(c) &= \phi_m^{SC} \overline{T} \\
\tau^f(c) &= (1 - \phi_m^{SC}) \overline{T}
\end{aligned}
$$

where ϕ_m^{SC} is satisfies the following quadratic equation:

$$
a(\phi_m^{SC})^2 + \left(\delta_\tau^m + \delta_\tau^f - a\right) \phi_m^{SC} - \delta_\tau^m = 0 \qquad (\text{QE}\phi_m^{SC}(c))
$$

with a given by:

$$
a = \Psi^m \log q^m - \Psi^f \log q^f + \frac{w^f - w^m}{c} \overline{T}
$$

Note that the only admissible solution [19] to the previous quadratic equation is given by:

$$
\phi_m^{SC} = \frac{a - (\delta_\tau^m + \delta_\tau^f) + \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m}}{2a} \qquad (\phi_m^{SC})
$$

[19] See C.5.2

$$n^m(c) = \overline{T}^m - \tau^m(c) - l^m(c)$$
$$n^f(c) = \overline{T}^f - \tau^f(c) - l^f(c)$$

15. **Relative care is exhausted and paid care is zero ($\tau^r = \overline{T}^r, \tau^P = 0$)**

    Similar to 14 but with $\tau^r = \overline{T}^r$ and with a given by:

$$a = \frac{\overline{T} - \overline{T}^r}{\overline{T}} \Psi^m \log q^m - \frac{\overline{T} - \overline{T}^r}{\overline{T}} \Psi^f \log q^f + \frac{w^f - w^m}{c} \left(\overline{T} - \overline{T}^r \right)$$

16. **Mother does not work, paid care is zero, and everything else is interior ($n^m = \tau^P = 0$, else interior.)**

    In this region TPI2- TPI4 and TPI7 have to hold.

    A critical point in this region solves a root-finding problem in the Budget Constraint Residual as a function of c:

$$BCR(c) := a_t(1+r) + w^f n^f(c) - c - a_{t+1}$$

    Where:

$$l^f(c) = \frac{\delta_l^f c}{w^f}$$
$$\lambda^{SC}(c) = \frac{\Psi^r}{\overline{T}} \log q^r$$
$$\tau^f(c) = \frac{\delta_\tau^f}{\frac{\delta_l^f}{l^f(c)} - \left(\frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC}(c) \right)}$$

    $\tau^m(c) = \phi_\tau^m \overline{T}^m$ and $l^m = (1 - \phi_\tau^m \overline{T}^m)$ where τ^m solves QE ϕ_τ^m

$$n^f(c) = \overline{T}^f - \tau^f(c) - l^f(c)$$
$$\tau^r(c) = \overline{T} - \tau^m(c) - \tau^P(c)$$

17. **Father does not work, paid care is zero, and everything else is interior ($n_f = \tau^P = 0$, else interior.)**

    Similar to 16.

18. **Mother and father do not work, paid care is zero, and everything else is interior ($n^f = n^m = \tau^P = 0$, else interior.)**

    In this case TPI3, TPI4 and TPI7

The critical point in this region (which is unique) can be found in closed form:

$$\lambda^{SC} = \frac{\Psi^r}{\overline{T}} \log q^r$$
$$\tau^m = \phi_\tau^m \overline{T}^m$$
$$\tau^f = \phi_\tau^f \overline{T}^f$$
$$\phi_\tau^m \text{ and } \phi_\tau^f \text{ solve QE } \phi_\tau^m \text{ and QE } \phi_\tau^f \text{ respectively}$$
$$l^m = (1 - \phi_\tau^m)\overline{T}^m$$
$$l^f = (1 - \phi_\tau^f)\overline{T}^f$$
$$c = a_t(1+r) - a_{t+1}$$
$$\tau^r = \overline{T} - \tau^m - \tau^f$$

19. **Mother does not work, paid and relative care are zero, everything else is interior ($n^m = \tau^P = \tau^r = 0$)**

    In this case TPI2- TPI4 have to hold. Critical points in this region can be found by solving a root-finding problem in the supervision constraint residual as a function of c:

    $$\overline{T} - \tau^m(c) - \tau^f(c)$$

    where:

    $$n^f(c) = \frac{c + a_{t+1} - a_t(1+r)}{w^f}$$
    $$l^f(c) = \frac{\delta_l^f c}{w^f}$$
    $$\tau^f(c) = \overline{T}^f - \tau^f(c) - l^f(c)$$
    $$\lambda^{SC}(c) = \frac{\delta_\tau^f}{\tau^f} + \frac{\Psi^f}{\overline{T}} \log q^f - \frac{\delta_l^f}{l^f}$$
    $$\phi_\tau^m \text{solves QE } \phi_\tau^m$$
    $$\tau^m(c) = \phi_\tau^m \overline{T}^m$$
    $$l^m(c) = (1 - \phi_\tau^m)\overline{T}^m$$

20. **Father does not work, paid and relative care are zero, everything else is interior ($n^f = \tau^P = \tau^r = 0$)**

    Similar to 19

21. **Mother does not work, paid care is zero, relative care is exhausted and everything else is interior ($n^m = \tau^P = 0, \tau^r = \overline{T}^r$)**

    Similar to 19 but with the Supervision Constraint Residual given by:

    $$SCR(c) = \overline{T} - \overline{T}^r - \tau^m(c) - \tau^f(c)$$

22. **Father does not work, paid care is zero, relative care is exhausted and everything else is interior ($n^f = \tau^P = 0, \tau^r = \overline{T}^r$)**

    Similar to 21

23. **Mother and father do not work, relative care and paid care are zero ($n^m = n^f = \tau^P = \tau^r = 0$)** From TPI3, TPI4, the parental time-use constraints and the supervision constraint we get that τ^m solves:

$$\frac{\delta^m_\tau}{\tau^m} + \frac{\delta^f_l}{\overline{T}^f - \overline{T} + \tau^m} - \frac{\delta^f_\tau}{\overline{T} - \tau^m} - \frac{\delta^m_l}{\overline{T}^m - \tau^m} + \frac{\Psi^m}{\overline{T}} \log q^m - \frac{\Psi^f}{\overline{T}} \log q^f = 0 \qquad (\text{RF } \tau^m)$$

    Note that since the LHS is strictly decreasing for $\tau^m \in \left[\max\{0, \overline{T} - \overline{T}^f\}, \min\{\overline{T}, \overline{T}^m\}\right]$, the solution in this region is unique and a bisection algorithm with the corresponding bounds converges linearly to the candidate to τ^m.

    From here we get:

$$\begin{aligned} l^m &= \overline{T}^m - \tau^m \\ \tau^f &= \overline{T} - \tau^m \\ l^f &= \overline{T}^f - \tau^f \\ c &= a_t(1+r) - a_{t+1} \end{aligned}$$

24. **Mother and father do not work, relative care is exhausted and paid care is zero ($n^m = n^f = \tau^P = 0, \tau^r = \overline{T}^r$)** This case is similar to 23.

    The root-finding problem is now described by:

$$\frac{\delta^m_\tau}{\tau^m} + \frac{\delta^f_l}{\overline{T}^f - (\overline{T} - \overline{T}^r) + \tau^m} - \frac{\delta^f_\tau}{\overline{T} - \overline{T}^r - \tau^m} - \frac{\delta^m_l}{\overline{T}^m - \tau^m} + \frac{\Psi^m}{\overline{T}} \log q^m - \frac{\Psi^f}{\overline{T}} \log q^f = 0 \qquad (\text{RF}' \ \tau^m)$$

    with bounds given by:

$$\left[\max\{0, \overline{T} - \overline{T}^r - \overline{T}^f\}, \min\{\overline{T} - \overline{T}^r, \overline{T}^m\}\right]$$

    and with:

$$\begin{aligned} l^m &= \overline{T}^m - \tau^m \\ \tau^f &= \overline{T} - \overline{T}^r - \tau^m \\ l^f &= \overline{T}^f - \tau^f \\ c &= a_t(1+r) - a_{t+1} \end{aligned}$$

C.3 Asset-conditional problem at $t = 1, 2$ for the two parent family given paid care type $P = N$

The critical points for this case are coincide with the critical points in regions 13-24 in the previous two cases.

C.4 Last period problem $t = 3$

Same as asset-conditional problem in periods $t = 1, 2$, but now a_4 can be found in closed form given consumption according to:

$$a_4 = \delta_a c$$

C.5 Miscellaneous of useful analytical results

C.5.1 Bounds for root-finding in λ^{SC}

In this sub-appendix we find bounds for the root-search procedures in λ^{SC} that are used in regions 4- 7 and regions 9 - 12.

First, note that from the FOC of the asset-conditional Lagrangian with respect to τ^r and from complementary slackness we get that

$$\lambda^{SC} \geqslant \frac{\Psi^r}{\overline{T}} \log q^r \text{ if } \tau^r = 0$$

and

$$\lambda^{SC} \leqslant \frac{\Psi^r}{\overline{T}} \log q^r \text{ if } \tau^r = \overline{T}$$

Moreover, for the case $q^P < q^*$ we know that λ^{SC} has to be consistent with c not being greater than total resources and with $q^P < q^*$ Together with the fact that the c implied by RFc is strictly increasing in λ^{SC} this implies:

$$\lambda^{SC} \leqslant \frac{\Psi^P}{\overline{T}} \rho_P \log\left(\frac{\rho_P}{\underline{P}^P} \frac{\Psi^P}{\overline{T}}\right) + \frac{\Psi^P}{\overline{T}} \rho_P \log c_{max} - \frac{\kappa_P}{c_{max}} - \rho_P \frac{\Psi^P}{\overline{T}}$$
$$\text{where } c_{max} = \min\left\{a_t(1+r) - a_{t+1} + w^m \overline{T}^m + w^f \overline{T}^f, \frac{\underline{P}^P \overline{T}}{\rho_P \Psi^P} (q^*)^{\frac{1}{\rho_P}}\right\}$$

By a similar logic, for the case $q^P > q^*$ we know that λ^{SC} has to be consistent with c being less than total resources and with $q^P > q^*$. This implies:

$$\frac{\Psi^p}{\overline{T}} (\log q^* - 1) \leqslant \lambda^{SC} \leqslant \frac{\Psi^p}{\overline{T}} \left(\log\left(\frac{c_{max}}{\overline{P}^P} \frac{\Psi^p}{\overline{T}}\right) - 1\right)$$

where now c_{max} is given by:

$$c_{max} = a_t(1+r) - a_{t+1} + w^m \overline{T}^m + w^f \overline{T}^f$$

Hence, except for the case in which $\tau^r = \overline{T}$ and $q^P < q^*$ we have finite lower and upper-bounds to conduct the root-search in λ^{SC}.

C.5.2 Existence and uniqueness of solution for $QE\phi^{SC}_m(c)$

In this sub-appendix we show that the there is only one admissible solution for ϕ^{SC}_m in equation $QE\phi^{SC}_m(c)$ In principle, $QE\phi^{SC}_m(c)$ is a quadratic equation and has two possible solutions, which are given by:

$$\phi^+ = \frac{a - (\delta^m_\tau + \delta^f_\tau) + \sqrt{(\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau}}{2a}$$
$$\phi^- = \frac{a - (\delta^m_\tau + \delta^f_\tau) - \sqrt{(\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau}}{2a}$$

First, we can show that for any value of a the discriminant is positive, which implies that the two possible values of ϕ are real. Let the discriminant be given by:

$$\Delta^{m,SC}_\phi = (\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau$$

If $a \geqslant 0$ this number is obviously positive. Suppose that $a < 0$ Then:

$$\Delta^{m,SC}_\phi = (\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau =$$
$$a^2 + (\delta^m_\tau + \delta^f_\tau)^2 + 2(\delta^m_\tau - \delta^f_\tau)a \geqslant a^2 + (\delta^m_\tau - \delta^f_\tau)^2 + 2(\delta^m_\tau - \delta^f_\tau)a = \Big(a + (\delta^m_\tau - \delta^f_\tau)\Big)^2 \geqslant 0$$

This proves the two roots are real.

Now we can show that ϕ^- is not an admissible solution for any value of a : Suppose $a > 0$. Then $\phi^- < 0$ iff:

$$a - (\delta^m_\tau + \delta^f_\tau) - \sqrt{(\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau} < 0 \iff$$
$$\Big(a - (\delta^m_\tau + \delta^f_\tau)\Big)^2 < \Big(a - (\delta^m_\tau + \delta^f_\tau)\Big)^2 + 4a\delta^m_\tau$$

which if $a > 0$ is always the case. Now, we want to show that if $a < 0$ then $\phi^+ > 1$. To see why note that if $a < 0$ then:

$$(\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau = a^2 + (\delta^m_\tau + \delta^m_\tau)^2 - 2(\delta^f_\tau - \delta^m_\tau)a > a^2 + (\delta^m_\tau + \delta^f_\tau)^2 + 2(\delta^m_\tau + \delta^f_\tau)a = (a + \delta^m_\tau + \delta^f_\tau)$$

Hence:

$$\sqrt{(\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau} < -|a + \delta'_\tau + \delta^f_\tau| < a + \delta^m_\tau + \delta^f_\tau\,,$$

which implies:

$$a - (\delta^m_\tau + \delta^f_\tau) - \sqrt{(\delta^m_\tau + \delta^f_\tau - a)^2 + 4a\delta^m_\tau} < 2a\,.$$

This implies the desired result.

Now we want to show that ϕ^+ is admissible for any value of a.

First we want to show that $\phi^+ \geqslant 0$. If $a > 0$ then we have:

$$\phi^+ \iff a - (\delta_\tau^m + \delta_\tau^f) + \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} \geqslant 0$$

Which is true since:

$$a - (\delta_\tau^m + \delta_\tau^f) + \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} \geqslant a - (\delta_\tau^m + \delta_\tau^f) + |a - (\delta_\tau^m + \delta_\tau^f)| \geqslant 0$$

(Where the first inequality follows from $4\delta_\tau^m a > 0$)

Now suppose $a < 0$. Then we have that:

$$|a - (\delta_\tau^m + \delta_\tau^f)| \geqslant \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2} \geqslant \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m}$$

This implies:

$$-a + \delta_\tau^m + \delta_\tau^f = |a| + |\delta_\tau^m + \delta_\tau^f| \geqslant |a - (\delta_\tau^m + \delta_\tau^f)| \geqslant \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2} \geqslant \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m}$$

Which implies:

$$a - (\delta_\tau^m + \delta_\tau^f) + \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} \leqslant 0$$

which if $a < 0$ implies that $\phi^+ > 0$.

Finally, we want to show that $\phi^+ < 1$.

First suppose that $a > 0$ Note that in this case $\phi^+ < 1$ iff:

$$\begin{aligned} a - (\delta_\tau^m + \delta_\tau^f) + \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} < 2a \iff \\ a + \delta_\tau^m + \delta_\tau^f > \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} \end{aligned}$$

Which is true since:

$$a + \delta_\tau^m + \delta_\tau^f = \sqrt{(a + \delta_\tau^f + \delta_\tau^m)^2} > \sqrt{a^2 + (\delta_\tau^m + \delta_{\tau^f})^2 - 2(\delta_\tau^f - \delta_\tau^m)a} = \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m}$$

Now suppose that $a < 0$: Note that:

$$\begin{aligned} \phi^+ < 1 \iff a - (\delta_\tau^m + \delta_\tau^f) + \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} > 2a \iff \\ \sqrt{(\delta_\tau^m + \delta_\tau^f - a)^2 + 4a\delta_\tau^m} > a + \delta_\tau^m + \delta_\tau^f \end{aligned}$$

Now, note that:

$$a^2 + (\delta_\tau^m + \delta_\tau^f)^2 + 2(\delta_\tau^m + \delta_\tau^f)a < a^2 + (\delta_\tau^m + \delta_\tau^f)^2 - 2(\delta_\tau^f - \delta_\tau^m)a$$

Hence:

$$\begin{aligned} \sqrt{a^2 + (\delta_\tau^m + \delta_\tau^f)^2 - 2(\delta_\tau^f - \delta_\tau^m)a} > \sqrt{a^2 + (\delta_\tau^m + \delta_\tau^f)^2 + 2(\delta_\tau^m + \delta_\tau^f)a} \\ = \sqrt{(a + \delta_\tau^m + \delta_\tau^f)^2} = |a + \delta_\tau^m + \delta_\tau^f| \geqslant a + \delta_\tau^m + \delta_\tau^f \end{aligned}$$

Hence, ϕ^+ is the unique admissible solution to $QE\phi_m^{SC}(c)$.

C.5.3 Expected cost of the optimal paid care choice

Let $\mathbb{E}\mathrm{TrExp}(\lambda, a, b)$ be the expectation of an exponential with parameter λ truncated below by a and above by b:

$$\mathbb{E}\mathrm{TrExp}(\lambda, a, b) = \begin{cases} \frac{\frac{1}{\lambda} - e^{-\lambda b}(1/\lambda + b)}{F_{exp}(b;\lambda)} & \text{if } a \leqslant 0 \\ \frac{(\frac{1}{\lambda})(e^{-\lambda a} - e^{-\lambda b}) + a e^{-\lambda a} - b e^{-\lambda b}}{F_{exp}(b;\lambda)) - F_{exp}(a;\lambda)} & \text{if } a \geqslant 0 \end{cases}$$

Let $\tilde{V}_t^{SM.i} > \tilde{V}_t^{SM.j}$

$$\begin{aligned} \mathbb{E}\left[c_i | P = i\right] = \Big(& (1 - F_{exp}(\tilde{V}^j - \tilde{V}^N; \lambda_j))\mathbb{E}\mathrm{TrExp}(\lambda_i, 0, \tilde{V}^i - \tilde{V}^N)F_{exp}(\tilde{V}^i - \tilde{V}^N; \lambda_i) + \\ & \frac{1}{\lambda_i} F_{exp}(\tilde{V}^j - \tilde{V}^N; \lambda_j) - \frac{\lambda_j}{\lambda_i + \lambda_j} e^{-\lambda_i(\tilde{V}^i - \tilde{V}^j)}(\frac{1}{\lambda_i} + \tilde{V}^i - \tilde{V}^j)F_{exp}(\tilde{V}_j - \tilde{V}^N; \lambda_i + \lambda_j) \\ & - \frac{\lambda_j}{\lambda_i + \lambda_j} e^{-\lambda_i(\tilde{V}^i - \tilde{V}^j)}\mathbb{E}\mathrm{TrExp}(\lambda_i + \lambda_j, 0, \tilde{V}^j - \tilde{V}^N)F_{exp}(\tilde{V}^j - \tilde{V}^N; \lambda_i + \lambda_j)\Big) \Big/ \mathbb{P}(P = i) \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[c_j | D = j\right] = \Big(& (1 - F_{exp}(\tilde{V}^i - \tilde{V}^N; \lambda_i))\mathbb{E}\mathrm{TrExp}(\lambda_i, 0, \tilde{V}^j - \tilde{V}^N)F_{exp}(\tilde{V}^j - \tilde{V}^N; \lambda_j) + \\ & \frac{1}{\lambda_j}\left(F_{exp}(\tilde{V}^i - \tilde{V}^N); \lambda_i) - F_{exp}(\tilde{V}^i - \tilde{V}^j); \lambda_i)\right) - \\ & \frac{\lambda_i}{\lambda_j + \lambda_i} e^{-\lambda_j(\tilde{V}^j - \tilde{V}^i)}(\frac{1}{\lambda_j} + \tilde{V}^j - \tilde{V}^i)(F_{exp}(\tilde{V}^i - \tilde{V}^N; \lambda_i + \lambda_j) - F_{exp}(\tilde{V}^i - \tilde{V}^j; \lambda_i + \lambda_j)) - \\ & \frac{\lambda_i}{\lambda_j + \lambda_i} e^{-\lambda_j(\tilde{V}^j - \tilde{V}^i)}\mathbb{E}\mathrm{TrExp}(\lambda_i, \tilde{V}^i - \tilde{V}^j, \tilde{V}^i - \tilde{V}^N)(F_{exp}(\tilde{V}^i - \tilde{V}^N; \lambda_i + \lambda_j) - F_{exp}(\tilde{V}^i - \tilde{V}^j; \lambda_i + \lambda_j))\Big) \Big/ \mathbb{P}(D = j) \end{aligned}$$

D Simulating optimal households' choices

This Appendix describes how to approximate the distribution of households' optimal choices given a price schedule for Center and Home-Based childcare quality. I am going to specialize the exposition in this Appendix to Two-Parent households, but the same logic applies to Single Mothers.

Note that given initial assets a_1 and the time-invariant type of a Two-Parent family H, solving the optimization problem of this Two-Parent family amounts to finding the probability distribution over sequences of optimal choices $\{a_{t+1}, D_t, Y_t\}_{t=1}^3$, where, as in Appendix C a_{t+1} denotes future assets, D_t denotes the choice over types of care $D = N, CB, HB$, and Y_t denotes all the other choices:

$$Y_t = (c_t, n_t^m, n_t^f, l_t^m, l_t^f, \tau_t^m, \tau_t^f, \tau_t^r, \tau_t^p, q_t^p) .$$

Let $\mathbb{P}(\{a_{t+1}, D_t, Y_t\}_{t=1}^3 | a_1, H)$ be the probability of a Two-Parent Household choosing sequence $\{a_{t+1}, D_t, Y_t\}_{t=1}^3$ given that its initial level of assets and time-invariant type is (a_1, H). Here I describe how to compute this probability distribution over sequences. Note that in a given period t the optimal choice of assets tomorrow a_{t+1} and the other continuous choices Y_t is deterministic given the discrete choice D_t. Call this conditional choices $a_{t+1}(a_t, H, D_t)$ and $Y_t(a_t, H, D_t)$ (where a_{t+1} is simply $g_{a',t}^{D_t}(a_t, H)$

and $Y_t(a_t, H, D_t)$ is the image of the cartesian product of the D_t -specific policy functions for the choices in Y_t. This means that for each family starting with (a_1, H) there are at most 27 sequences $\{a_{t+1}, D_t, Y_t\}_{t=1}^{3}$ that occur with positive probability. Each of those sequences can be uniquely identified by its associated history of discrete choices D^3, where $D^t = (P^{t-1}, D_t)$ with $D^1 = D_1$. That is, for each (sub) history of discrete choices D^t there is a (sub) history of choices $a^{t+1}(D^t, a_1, H)$ $Y^t(D^t), a_1, H$ given by:

$$a^{t+1} = \left(a^t, a^{t+1}, ((a^t)_{t-1}, H, (D^t)_t)\right),$$

$$Y^t = \left(Y^{t-1}, Y^t((a^t)_{t-1}, H, (D^t)_t)\right),$$

with

$$a^2 = a_2(a_1, H, D^1),$$

$$Y^1 = Y_1(a_1, H, D^1),$$

where $(a, b, c)_t$ denotes the t-th element of vector (a, b, c)

Now I describe how to find these histories and its associated probabilities

- For a given time invariant family type H we can find $\mathbb{E}_t \tilde{V}_{t+1}^{TP}$ for $t = 1, 2$ following the procedure described in C.1.2.

- For a Household with initial assets and time-invariant type (a_1, H) we can find $a_2(D)$ and $\tilde{V}_t^{TP,D}(a_t, H)$ for $D = N, CB, HB$ by solving Dynamic choice (see Appendix C for details on how to do that). Given $a_2(D_1)$, we can solve Static Choice to get $Y_1(a_1, H, a_2, D_1)$. The probability of choosing $D_1 = D$ ($\mathbb{P}(D_1 = D|a_1, H)$) given that the household starts the period with initial assets a_1 and time-invariant family type H are given in closed form by Choice Probabilities using the computed $\{\tilde{V}^{TP,D}\}_{D=N,CB,HB}$. Now that we know how to find $a_2(a_1, H, D_1)$ and $Y_1(a_1, H, D_1)$ for $D_1 = N, CB, HB$ and the probability distribution over discrete choices D, I explain how to compute choices and probabilities for continuation histories starting at $t = 2, 3$

- For a given history D^{t-1} with initial assets $a^t(D^{t-1}) = a^t(D^{t-1})_t$ and time-invariant family type H we can find $a_{t+1}(D)$ and $\tilde{V}_t^{TP,D}(a_t, H)$ for $D = N, CB, HB$ by solving Dynamic choice (see Appendix C for details on how to do that, and note that in $t = 3$ the static and the dynamic choices can be solved for at once). Given $a_{t+1}(D_t)$, we can solve Static Choice to get $Y_t(a_t, H, a_{t+1}, D_t)$. The probability of choosing $D_t = D$ ($\mathbb{P}(D_t = D|a_t, H)$) given that the household starts the period with initial assets a_t and time-invariant family type H are given in closed form by Choice Probabilities using the computed $\{\tilde{V}^{TP,D}\}_{D=N,CB,HB}$ From this we can construct the 3 continuation histories (corresponding to $D_t = N, CB, HB$) as $a^{t+1}(D^t) = (a^t, a_{t+1}(D_t))$, $Y^t(D^t) = (Y^{t-1}, Y(a_t, H, a_{t+1}, D_t))$. The associated probability to each history is given by

$$\mathbb{P}(D^t|a_1, H) = \mathbb{P}(D^{t-1}|a_1, H)\mathbb{P}(D_t = D|a_t, H).$$

Note that we can write relevant aggregates in terms of the histories of choices and its associated probabilities. For instance, the cumulative distribution function of Two-Parent Households at $t=3$ can be written as:

$$
\begin{aligned}
G(a_3,H) = \int_{a_1,\tilde{H}\leqslant H} \sum_{D_2=N,CB,HB} \sum_{D_1=N,CB,HB} \mathbb{P}^{TP,D_2}(a_2,H)\mathbb{P}^{TP,D_1}(a_1,H) \\
1(g^{D_2}_{2,a'}(a_2,H) \leqslant a_3)1(g^{D_1}_{1,a'}(a_1,H) = a_2)dG_1(a_1,H) = \\
\int_{a_1,\tilde{H}\leqslant H} \sum_{D^3\in\mathcal{D}^3} \mathbb{P}(D^3|a_1,H)1\left(a^3(D^3)_3 \leqslant a_3\right)dG_1(a_1,H)\,,
\end{aligned}
$$

where $\mathcal{D}^3$ denotes the set of all possible histories D^3. The second equality follows from the fact that $a^t(D^t)$ is constructed using optimality given the sequence of discrete choices and the definition of $\mathbb{P}(D^t|a_1,H)$

E Solving for equilibrium factor prices

At a high level, the equilibrium solver proceeds as follows:

- Start with a guess for factor prices w^E_0, w^C_0
- Given factor prices, find the price schedules that satisfy profit maximization and free entry for center and home-based childcare providers according to Lemma 3
- Given the equilibrium price schedules, simulate Single-Mother and Two-Parent Households decisions. Among those decisions, the quality, quantity, and type of paid childcare decisions are going to be informative for the aggregate demand of efficiency units of lead teachers E and the aggregate demand of caregiver hours C.
- Find the demands of E and C by childcare providers necessary to satisfy household demands of quality, quantity, and type of paid childcare. More precisely:

$$
\begin{aligned}
C = \int_{a_1,H} \Big(\sum_{P^3\in\mathcal{P}^3} \mathbb{P}(P^3|a_1,H)\big(C_1^{(P^3)_1}((q^P)^3_1)(\tau^P)^3_2 + C_2^{(P^3)_2}((q^P)^3_2)(\tau^P)^3_2 + \\
C_3^{(P^3)_3}((q^P)^3_3)(\tau^P)^3_3\Big)dG_1(a_1,H)\,,
\end{aligned}
$$

$$
\begin{aligned}
E = \int_{a_1,H} \Big(\sum_{P^3\in\mathcal{P}^3} \mathbb{P}(P^3|a_1,H)\big(E_1^{(P^3)_1}((q^P)^3_1)(\tau^P)^3_2 + E_2^{(P^3)_2}((q^P)^3_2)(\tau^P)^3_2 + \\
E_3^{(P^3)_3}((q^P)^3_3)(\tau^P)^3_3\Big)dG_1(a_1,H)\,.
\end{aligned}
$$

where E^P_t, C^P_t are the conditional factor demands in 1, P^3 denotes a history of childcare-type choices as in D, $(q^P)^3, (\tau^P)^3$ are the elements of $Y^t(P^3)$ corresponding to τ^P and q^P, $Y^t(P^t)$ is constructed as in D, and the integral is calculated by drawing Montecarlo draws from the initial distribution $G_1(a_1,H)$

- Update the guess for factor prices to be a convex combination of the old prices and the prices that would make the factor supply satisfy the new demand. That is:

$$\begin{aligned}
w^C_{new} &= \left(\frac{C-E}{\overline{H}_{CCW}\overline{LT}}\right)^{1/\eta_{CCW}}, \\
w^{LT} &= \left(\frac{E}{\overline{H}_{LT}\overline{LT}}\right)^{1/\eta_{LT}}, \\
w^E_{new} &= w^{LT} - w^C_{new}, \\
(w^E_1, w^C_1) &= \gamma^{damping}(w^E_0, w^C_0) + (1-\gamma^{damping})(w^E_{new}, w^C_{new})
\end{aligned}$$

- Iterate until (w^E_1, w^C_1) converge.

F Imputation of parents' wages

In this appendix, I describe how I impute the wages of mothers and fathers when those are missing. In principle, imputation may not be a good idea if most of the missing wages are missing due to endogenous selection on unobservables on the labor force. This is less of a problem in this context because in order for the wage of a parent not to be observable due to non-participation, that parent would need to not participate in all 4 waves (remember that I am assuming that wages are time-invariant). In fact, of the 3200 families for which the wage of the mother is missing, 1450 are missing because the mother does not participate in the labor force in any wave. For fathers, of the 1600 Two-Parent families for which the wage of the father is missing, only 50 are due to the father not working in any wave[20]. In my sample, a household may have missing wages for some parent in a given period because that parent failed to report earnings or hours worked in that period, or because their combination of hours and earnings implied an hourly wage lower than half the federal minimum wage (which I consider miss-reporting).

Moreover, given that the model does not generate a selection rule with a known parametric form, and given that the model is unlikely to have the single-index property needed for many semi-parametric selection-correction procedures, a model-consistent selection correction for the identification of the joint distribution of wages, other individual characteristics, and assets would be challenging.

In order to impute wages, I follow Appendix D in De Nardi, French, Jones and McGee (2021). This imputation strategy is an analogous version of hotdeck imputation with continuous covariates. First, for the observations with non-missing wages for mothers, I regress their log-wage on a vector of observables z and get fitted values $z'\beta$ and residuals ϵ. The variables in z include a quadratic polynomial of the age of the mother in wage 1, dummies for educational achievement, dummies for self-reported health status, dummies for US Census Region, and Dummies for the 2000 Urban and Rural classification[21]. I split the sub-sample of households with an observed mother wage according to the deciles of $z'\beta$, and I keep the distribution of ϵ for each of those deciles. Then, for the households for which the mother has a missing wage, I calculate $z'\beta$, and I sample an ϵ at random from the set of households in the same decile bin of $z'\beta$ with an observed wage.

The procedure to impute the wages of fathers is analogous. The R^2 or the initial mean-fitting step of the imputation procedure is 34% for mothers and 36% for fathers.

G Measuring assets

In this appendix, I describe how I measure assets in the ECLS-B. I call this "measurement" and not "imputation" because unlike in the case of wages, I account for

[20]Number of families rounded to the nearest 50 to comply with ECLS-B disclosure rules

[21]The ECLS-B provides the 2000 US Census Urban and Rural classification of the households in the sample. The US Census divides areas according to their density and other criteria in Urban Areas, Urban Clusters, and Rural Areas

measurement error in assets when estimating parameters. The ECLS-B does not contain a measure of net worth, nor it does contain information on the monetary value of most assets and debts that are commonly used in the calculation of net worth. However, the ECLS-B does contain some information on the portfolio composition of families. In particular, the ECLS-B asks families the following questions about their investment portfolio

- Whether they own assets from a list of low-risk assets[22]
- Whether they any assets from a list of risky assets[23]
- Whether they own a car or a truck
- Whether they own the house they live in
- If they own the house they live in, whether they have a mortgage
- If they own the house they live in, the value of the house.

All of these questions can be mapped to questions in the SCF 2001. In order to exploit this, I regress[24] net worth on the portfolio variables that can also be found in ECLS-B, and in demographic and economic variables that are common to ECLS-B and SCF 2001. These demographic and economic variables are household income, dummies for the educational attainment of the mother and the father, the ages of the mother and the father, and an interaction of their ages with their educational attainment dummies. Call the vector of portfolio, demographic, and other economic variables z, and the vector of coefficiencts estimated in the regression using SCF 2001 data β. Because all the variables in z are available in ECLS-B, I can construct a measure of assets for each family at each wave as:

$$\tilde{a}_{i,t} = z'_{i,t}\beta \ .$$

The R^2 of the linear prediction for assets is 67%, which implies that the amount of measurement error in $\tilde{a}_{i,t}$ is moderate. Note that because net worth is observable in SCF 2001, the marginal distribution of measurement error in $\tilde{a}_{i,t}$ is identified from computing $a_{i,t} - z'_{i,t}\beta$ in SCF 2001.

[22]The precise question is:"Do you, or anyone in your household,have any money in checking or savings accounts, money market funds, certificates of deposit, or government savings bonds, or Treasury bills, including IRAs?"

[23]The precise question is: "Do you or anyone in your household have any shares of stock in publicly held corporations, mutual funds, or investment trusts, including stocks in IRAs?"

[24]For the sake of comparability, when running this regression, I only use families in SCF with children aged 0-6 years old and that correspond to the definition of Two-Parent Household or Single-Mother Household used in this paper.

H Identification of the relative care endowment function

Recall that the relative care endowment function is a deterministic function of a vector of discrete covariates, call it $Z^{T,r}$. Abusing notation slightly we write this as:

$$\overline{T}^r = \overline{T}^r(Z^{T,r}) < \overline{T}\ .$$

Proposition 2. *Suppose that* $\forall \epsilon > 0$ *and*

$$\mathbb{P}(q^m < \epsilon, q^f < \epsilon | a_1, w^m, w^f, q^r, Z^{T,r}) > 0$$

for all

$$(a_1, w^m, w^f, q^r, Z^{T,r}) \in \operatorname{supp}(a_1, w^m, w^f, q^r Z^{T,r})$$

Then

$$\max_{\tau^r \in \operatorname{supp}(\tau^r | Z^{T,r})} \tau^r = \overline{T}^r(Z^{T,r})$$

and

$$\mathbb{P}(\tau^r = \overline{T}^r(Z^{T,r}) | Z^{T,r}) > 0 \textit{ for all } Z^{T,r} \in \operatorname{supp} Z^{T,r}$$

Proof. It suffices to show that for every $Z^{T,r} \in \operatorname{supp} Z^{T,r}$ there exists a set of states with positive probability measure $\mathcal{X}$ such that

$$\text{for all } (a_1, w^m, w^f, q^m, q^f, \{q^r\}_{t=1}^3, \overline{T}^r, c_1^{CB}, c_1^{HB}) = x \in \mathcal{X} \text{ we have that } \tau_1^r(x) = \overline{T}^r(Z^{T,r})$$

Since $\overline{T}^r(Z^{T,r})$ is time-invariant, I am going to make the argument focusing on the first period, and omitting time subscripts when convenient.

Pick sets bounded above for initial assets $\mathcal{A}$, maternal wages $\mathcal{W}^m$ and paternal wages $\mathcal{W}^f$ such that

$$\mathbb{P}((a_1, w^m, w^f) \in \mathcal{A} \times \mathcal{W}^m, \mathcal{W}^f | Z^{T,r}) > 0$$

Let $q^{r,*}$ be such that the set

$$\{q^r : q^r > q^{r,*}\}$$

has positive probability measure given $(a_1, q^m, q^f) \in \mathcal{A} \times \mathcal{W}^m, \mathcal{W}^f$ and $Z^{T,r}$ (Note that there is at least one $q^{r,*}$ regardless of whether q^r is a continuous or a discrete random variable as long as the distribution of q^r given (a_1, q^m, q^f) is non-degenerate).

Moreover, let

$$q^{m,*} = exp\left\{ \frac{\overline{T}}{\Psi^m} \left(\frac{\Psi^r}{\overline{T}} \log q^{r,*} - \frac{\delta_\tau^m}{\overline{T} - \frac{\overline{T}^r(Z^{T,r})}{2}} \right) \right\}$$

and

$$q^{f,*} = exp\left\{ \frac{\overline{T}}{\Psi^f} \left(\frac{\Psi^r}{\overline{T}} \log q^{r,*} - \frac{\delta_\tau^f}{\overline{T} - \frac{\overline{T}^r(Z^{T,r})}{2}} \right) \right\}$$

Note that

$$\mathbb{P}(q^m < q^{m,*}, q^f < q^{f,*} | (a_1, w^m, w^f) \in \mathcal{A} \times \mathcal{W}^m \times \mathcal{W}^f, q^r \geqslant q^{r,*}, Z^{T,r}) > 0$$

Hence

$$\mathbb{P}((a_1, w^m, w^f) \in \mathcal{A} \times \mathcal{W}^m \times \mathcal{W}^f, q^m < q^{m,*}, q^f < q^{f,*}, q^r \geqslant q^{r,*} | Z^{T,r}) > 0$$

Define

$$c^{j,*} = \sup_{x \in \mathcal{X}_{-cj}} \tilde{V}^{TP,j}(x)$$

where

$$\begin{aligned} \mathcal{X}_{-cj} = \{ & (a_1, w^m, w^f, q^m, q^f, \overline{T}^r) : (a_1, w^m, w^f) \in \mathcal{A} \times \mathcal{W}^m \times \mathcal{W}^f \\ & q^m < q^{m,*}, q^f < q^{f,*}, q^r \geqslant q^{r,*}, \overline{T}^r = \overline{T}^r(Z^{T,r}) \} \end{aligned}$$

And note that $c^{j,*}$ is bounded above because $\mathcal{X}_{-cj}$ is bounded above.

Since $c^j \sim \exp(\lambda^j)$ with c^{HB}, c^{CB} independent of each other and c^j independent of $a_1, w^m, w^f, q^m, q^f, Z^{T,r}$ for $j = CB, HB$ we have that

$$\mathbb{P}(c^j > cj, * | a_1, w^m, w^f, q^m, q^f, Z^{T,r}) > 0 \text{ for all } (a_1, w^m, w^f, q^m, q^f, Z^{T,r}) \text{ for } j = CB, HB$$

Hence

$$\mathbb{P}((a_1, w^m, w^f) \in \mathcal{A} \times \mathcal{W}^m \times \mathcal{W}^f, q^m < q^{m,*}, q^f < q^{f,*}, q^r \geqslant q^{r,*}, c_{CB} > c^*_{CB}, c_{HB} > c^*_{HB} | Z^{T,r}) > 0$$

We are going to show now that if $x \in \mathcal{X}$ with

$$\mathcal{X} = \mathcal{X}_{-cj} \times \{c_{CB} : c_{CB} > c^*_{CB}\} \times \{c_{HB} : c_{HB} > c^*_{HB}\}$$

then

$$\tau^r_1(x) = \overline{T}^r(Z^{T,r})$$

Suppose this is not the case for some $x \in \mathcal{X}$. Because the fixed costs of using home-based and center-based childcare are large enough it is not optimal to use any of those childcare arrangements. Hence, it must be true that it is optimal at state x

$$\tau^f_1 + \tau^m_1 > \overline{T} - \overline{T}^r(Z^{T,r})$$

It must be the case then that either $\tau^m_1 > \frac{\overline{T} - \overline{T}^r(Z^{T,r})}{2}$ or $\tau^f_1 > \frac{\overline{T} - \overline{T}^r(Z^{T,r})}{2}$.

Suppose that the first one is the case. Note that the marginal value of decreasing τ^m_1 and increasing τ^r_1 and l^m by the same amount is given by:

$$-\left[\frac{\delta^m_l}{l^m} + \frac{\delta^m_\tau}{\tau^m_1} + \frac{\Psi^{TP,m}_1}{\overline{T}} \log q^m - \frac{\Psi^{TP,r}_1}{\overline{T}} \log q^r\right] > 0$$

given that $q^m < q^{m,*}$, $q^r > q^{r,*}$ and $\tau^m_1 > \frac{\overline{T} - \overline{T}^r(Z^{T,r})}{2}$. A similar contradiction argument applies if $\tau^f_1 > \frac{\overline{T} - \overline{T}^r(Z^{T,r})}{2}$. Hence, it is optimal to exhaust relative care for all $x \in \mathcal{X}$, which establishes the result. □

The intuition of the result is that within each group of households defined by their level of wages and household wealth, it is always possible to find households with maternal and paternal childcare quality low enough and fixed costs of using center and home-based care high enough, that exhausting relative care has to be optimal.

From here, we get that in order to identify $\overline{T}^r(Z^{T,r})$ we just need to look at the maximum of τ^r given $Z^{T,r}$

The argument for Single Mother households is similar.

H.1 Estimation

From the identification argument above we know that $\overline{T}^r(Z^{T,r} = z)$ can be estimated as the maximum of hours of relative care within the demographic cell defined by $Z^{T,r} = z$. The variables I choose to include in $Z^{T,r}$ are the discretized age of the maternal grandmother, the discretized age of the maternal grandfather and the number of adult relatives living in the same household as the child. More precisely:

$$Z^{T,r} = (\text{GM age}, \text{GF age}, \text{N adult HH})$$

where:

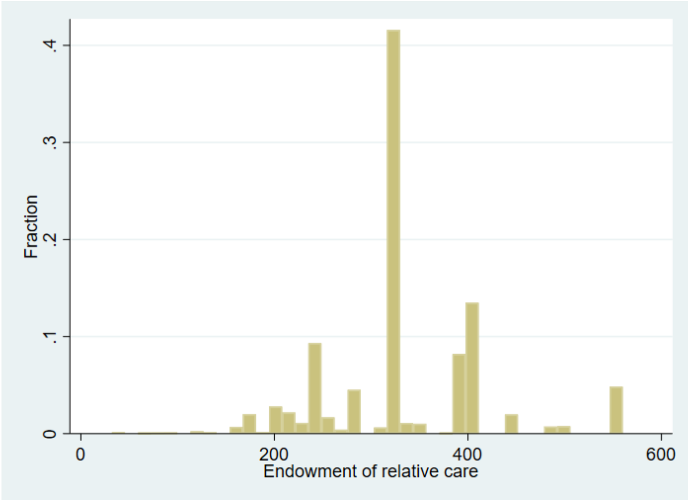
$$\text{GM age} = \begin{cases} 0 & \text{if Maternal Grandmother is dead} \\ 1 & \text{if Maternal Grandmother younger than 50} \\ 2 & \text{if Maternal Grandmother between 50 and 65} \\ 3 & \text{if Maternal Grandmother between 65 and 80} \\ 4 & \text{if Maternal Grandmother older than 80} \end{cases}$$

$$\text{GF age} = \begin{cases} 0 & \text{if Maternal Grandfather is dead} \\ 1 & \text{if Maternal Grandfather younger than 50} \\ 2 & \text{if Maternal Grandfather between 50 and 65} \\ 3 & \text{if Maternal Grandfather between 65 and 80} \\ 4 & \text{if Maternal Grandfather older than 80} \end{cases}$$

$$\text{N adult HH} = \begin{cases} 0 & \text{if no adult relative living in HH} \\ 1 & \text{if 1 adult relative living in HH} \\ 2 & \text{if 2 adult relatives living in HH} \\ 3 & \text{if 3 adult relatives living in HH} \\ 4 & \text{if 4 or more adult relatives living in HH} \end{cases}$$

The resulting estimated distribution of relative care monthly endowments is given by the following histogram:

Figure 8: Histogram of estimated hours of relative care available to families each month



SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

I Share of efficiency units of the lead teacher on the production function of quality (α_E)

Note that the First Order Conditions of the cost minimization problem for a childcare provider producing quality above the kink q^* imply:

$$\alpha_E = \frac{w^E E}{w^C C + w^E E} .$$

That is, childcare providers that offer high enough quality that the mandatory minimum staff-to-child ratio is not binding choose efficiency units of the lead teachers and the hours of childcare workers such that the income share of the efficiency units of the lead teacher is equal to α_E. Hence, α_E is identified from the average income share of the efficiency units of the lead teacher for childcare providers with staff-to-child ratios above the minimum regulated one. Here I describe how I construct the data analog of $\alpha_{E,i,t}$, the income share of the paid childcare provider that child i attends at age t. First, for the subsample of children for which information on paid providers is observed at age t, the wage of the lead teacher $w^{LT}_{i,t}$ can be computed. Note that the hourly wage of the lead teacher includes the remuneration of the care that she provides and her efficiency units:

$$w^{LT} = w^C C + w^E E .$$

Hence, the hourly income going towards paying the efficiency units of the lead teacher is given by:

$$w^E E = w^{LT} - w^C .$$

Given that childcare workers (the staff members of childcare provider different from the lead teacher, CCW from now on) only provide care, their wage is the price of the care factor. Hence, I let w^C in the data to be the average wage of childcare workers at a given year in a given state. I get the wage of CCWs at a given year in a given state from the BLS. Now we have the numerator of $\alpha_{E,i,t}$. In order to calculate the denominator, we need $w^C C$. As discussed before w^C is an aggregate statistic from the BLS. I let C be the average number of adults in the classroom. [25] Now, we can calculate $\alpha_{E,i,t}$ for each child i for which direct observations of paid care providers are conducted at time t. Restricting attention to cases for which the number of adults per children in the classroom is above the minimum regulated one and information on the wage of teachers is available,and cases on a state and year for which the hourly wage of childcare workers is lower than the average wage of lead teachers, we can estimate:

$$\alpha_{E,i,t} = \frac{1}{N}\sum_{i=1}^{N} \alpha_{E,i,t} \ .$$

The results are given in the following table: Because the ECLS-B did not conduct a

Table 5: $(\alpha_{E,2}, \alpha_{E,3})$ estimated from bill shares

$\alpha_{E,2}$	$\alpha_{E,3}$
0.14	0.24
(0.0080)	(0.0056)

NOTE: SE in parenthesis
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

direct observation of non-parental childcare arrangements at 9 months, I estimate $\alpha_{E,1}$ by extrapolating $\alpha_{E,2}$ and $\alpha_{E,3}$ according to:

$$\alpha_{E,t} = \frac{\exp(b_0 + b_1 t)}{1 + \exp(\exp(b_0 + b_1 t))} \ ,$$

where in a slight abuse of notation I am measuring t in months in the formula above. This results in $\alpha_{E,1} = 0.093$ (or $\alpha_{E,9mo} = 0.093$)

J Measurement systems and production functions of cognitive skills

Quality of caregiving is hard to measure. Instead of assuming that quality of caregiving is observable, I assume there is a battery of noisy measures for different types of caregiving quality. For simplicity, I impose parametric restrictions on the relationship between the noisy measures of quality and true latent quality:

$$\widetilde{\log q}^{j,s} = \mu^{j,s} + \alpha^{j,s} \log q^j + \epsilon^{j,s} \ ,$$
$$\text{for } j = m, f, r, p \text{ and } s = 1, \ldots, N_s$$

[25]ECLS-B contains counts at 6 different times of the number of adults and children in the room where the focal child receives care. I average the number of adults across counts to get C and the number of children to get h

The measurement system for children's cognitive skills is analogous:

$$\widetilde{\log\theta}^{s}_{t} = \mu^{s}_{\theta,t} + \alpha^{s}_{\theta,t}\log\theta^{s}_{t} + \epsilon^{s}_{\theta,t} \ .$$

J.1 Identification and estimation of the measurement system for non-parental care

The measure of non-parental care (relative and paid) used in this paper is the Arnett score. The relationship between the Arnett score and true quality of non-parental care is given by the following linear measurement system:

$$\mathrm{Arnett}^{j}_{i,t} = \mu^{\mathrm{Arnett}}_{t} + \alpha^{\mathrm{Arnett}}_{t}\log q^{j}_{t} + \epsilon^{\mathrm{Arnett}}_{t} \text{ for } j = r, p \ .$$

Note that I am allowing the parameters of the measurement system to vary with the age of the child.

J.1.1 Factor loading of the Arnett score

Substituting the expression for the Arnett score measurement system in the production function for quality production in paid childcare providers we get:

$$\begin{aligned}
\mathrm{Arnett}_{i,t} = \mu^{\mathrm{Arnett}}_{t} + \alpha^{\mathrm{Arnett}}_{t}\Big(\log A^{HB}_{t} + 1(D_{i,t} = CB)(\log A^{CB}_{t} - \log A^{HB}_{t}) \\
+\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right)\Big) + \epsilon^{\mathrm{Arnett}}_{t} = \\
\mu^{\mathrm{Arnett}}_{t} + \alpha^{\mathrm{Arnett}}_{t}\Big(\log A^{HB}_{t} + 1(D_{i,t} = CB)(\log A^{CB}_{t} - \log A^{HB}_{t})\Big) + \\
\alpha^{\mathrm{Arnett}}_{t}\Big(\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right)\Big) + \epsilon^{\mathrm{Arnett}}_{t}
\end{aligned}$$

This means that $\alpha^{\mathrm{Arnett}}_{t}$ is identified from a regression of the Arnett score on a constant, a dummy for the paid provider being center-based, and the input composite term

$$\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right) \ .$$

It will be useful later to define:

$$\begin{aligned}
\widehat{\mathrm{Arnett}}_{i,t} = \mu^{\mathrm{Arnett}}_{t} + \alpha^{\mathrm{Arnett}}_{t}\Big(\log A^{HB}_{t} + 1(D_{i,t} = CB)(\log A^{CB}_{t} - \log A^{HB}_{t}) + \\
\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right)\Big) \ .
\end{aligned}$$

Note that we can construct the input composite term because $\alpha_{E,t}$ is identified from the previous section, C and k are the average number of adults and children in the classroom respectively (as in the previous section), and E, the efficiency units of the lead teacher, can be constructed according to:

$$E = \frac{w^{LT} - w^{C}}{w^{E}} \ ,$$

where w^{LT} is the hourly wage of the lead teacher in the classroom, w^{C} is the average wage of childcare workers in which the focal child leaves, and w^{E} is the average salary premium of lead teachers with respect to childcare workers in the state where the average child lives. The intuition of this identification argument for α_t^{Arnett} is the following: The empirical elasticity of the Arnett score with respect to the input composite term is equal to the elasticity of the Arnett score with respect to true quality times the elasticity of true quality with respect to the input composite term. Because the Production Function of quality is constant returns to scale, we know that this last term has to be 1. Hence, the elasticity of the Arnett score with respect to the input composite term has to be equal to the elasticity of the Arnett score with respect to true quality. Estimates of the Arnett loadings are shown in Table 6

Table 6: Estimates of the Arnett score loadings

α_2^{AR}	α_3^{AR}
0.36	0.26
(0.08)	(0.1)

NOTE: SE in parentheses
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

J.1.2 Identification of the Arnett score measurement system additive shifter

See Appendix J.5

J.2 Distribution of relative care

Relative care quality is a function of observables:

$$\log q_t^r = X'_{q,r}\beta_t^{q,r} .$$

This implies the following relationship between the Arnett score and observables in $X_{q,r}$:

$$\mathrm{ARNETT}_{i,t}^r = \mu_t^{\mathrm{ARNETT}} + \alpha_t^{\mathrm{ARNETT}} X'_{q,r}\beta_t^{q,r} + \epsilon_{i,t}^{\mathrm{ARNETT}} .$$

It will be useful to define:

$$\widehat{\mathrm{ARNETT}}_{i,t}^r := \mu_t^{\mathrm{ARNETT}} + \alpha_t^{\mathrm{ARNETT}} X'_{q,r}\beta_t^{q,r} .$$

The estimated dependence between $\widehat{\mathrm{ARNETT}}_{i,t}^r$ and the observables in $X_{q,r}$ is reported in the following table

In the ECLS-B, the Arnett score is not available at 9 months. Because of that, modeling the heterogeneity in relative care quality in the same way as for waves 2 and 3 is not possible. In order to still allow for heterogeneous relative care qualities across

Table 7: Observable predictors of relative care quality

Variable	$\widehat{\text{ARNETT}}_2^r$	$\widehat{\text{ARNETT}}_3^r$
Grandmother High-School	0.30	0.28
	(0.11)	(0.21)
Grandmother College	0.43	0.52
	(0.23)	(0.42)
Grandmother Postgraduate	0.58	0.90
	(0.31)	(0.45)
Grandfather High School	0.23	-0.06
	(0.12)	(0.22)
Grandfather College	0.38	0.51
	(0.21)	(0.36)
Grandfather Postgraduate	0.46	0.29
	(0.25)	(0.59)
Grandfather missing education	-0.17	-0.95
	(0.17)	(0.40)
Constant	-0.42	-0.57
	(0.09)	(0.15)

Table 8: NOTE: SE in parenthesis.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

families without taking a stance on the overall quality of relative care with respect to parental care, I assume that:

$$\log q_2^r = \log q_1^r + g_1^{q,r} .$$

Later I show that $g_1^{q,r}$ is identified. Note that because the location of maternal care is normalized to 0, $g_1^{q,r}$ measures how good or bad at fostering cognitive development become with respect to mothers from period 1 to period 2. For instance, if

$$\mathbb{E}[\log q^m] > \mathbb{E}[\log q_1^r] ,$$

and $g_1^{q,r} > 0$, then the gap in care quality (as measured as how conducive to cognitive development a certain childcare arrangement is) between relatives and mothers becomes narrower between 9 months and 2 years old.

J.3 Parental care

J.3.1 Constructing measures of parental care

I construct two measures of parenting quality for maternal and paternal quality. It is convenient for one of those measures to be constructed from the same items. If both maternal and paternal quality are measured in the same units, then it makes sense to impose $\gamma_{m,t} = \gamma_{f,t}$ in the production function for skills. Therefore, I choose to construct $\widetilde{\log q}^{m,1}$ and $\widetilde{\log q}^{f,1}$ from a subset of questions asked to both the father and the mother of the child in the third wave (when the child is 4 years old). More concretely:

$$\widetilde{\log q}^{j,1} = \sum_{i=1}^{6} 1\left(A_i \neq \text{"Never"}\right) \text{ for j=m,f} .$$

Where each A_i denotes the answer to one of the following questions:

1. How often in the last month did the parent play with child using toys for building things, like Lincoln Logs or Legos.

2. How often in the last month did the parent help the child to bed.

3. How often in the last month did the parent help the child to bathe.

4. How often in the last month did the parent take the child for a walk or to play outside.

5. How often in the last month did the parent help the child to get dressed.

6. How often in the last month did the parent help the child to brush his teeth.

The second measure of maternal care $\widetilde{\log q}^{m,1}$ combines information three different instruments of ECLS-B: The parental interview at 9 months, the parental interview at 2 years old, and the Two-Bags Task at 4 years old. To operationalize that, I compute two sub-scores: An interview score and a Two-Bag Task score [26]. Then, I add the two sub-scores.

- **Interview Score** I transform answers to questions from the parental interviews at 9 months and 2 years old into 0-1 variables, I weigh them by 1 or -1 and I add them up to form a mother interview score

- **Two-Bag Task score** I add up the Emotional Support and Cognitive Development Scales and subtract the Intrusiveness and Detachment scales of the Two-Bag Task measure at 4 years old to form a Two-Bag Task measure

The following table presents the questions from the parent interviews at ages 9 months and 2 years old used to construct the mother interview sub-score, how are they recorded into 0-1 variables, and their weight in the weighted sum:

[26] In the Two Bags Task the parent and the child have to play with two sets of toys. The interaction is then recorded, observed and coded. In this case, different features of the observed behaviour or mood of the parent and child are coded separetaly. I use the Sensitivity, Positive Regard, Intrusiveness, and Stimulation of Cognitive Development at the Two years old assessment, and the Emotional Support, Stimulation of Cognitive Development, Intrusiveness, and Detachment parental scales in the 4 years old assessment. I choose not to use the Negative Regard Scale at either age because it intends to capture the degree negative perception that the child has of the parent, as opposed to a behaviour or mood that the parent presents during the interaction. For a more detailed discussion of the Two Bags Task see Andreassen and Fletcher (2007)

Table 9: Questions used for the mother interview sub-score

Question	Transformation to 0-1
Questions asked when child is 9 months old	
Used book/magazine on parenting	1 if Yes 0 if No (+)
How often play peekaboo with child	0 if Never 1 otherwise (+)
How often tickle child	0 if Never 1 otherwise (+)
How often play/walk with child outside	0 if Never 1 otherwise (+)
Questions asked when child is 2 years old	
How often play chasing games with child	0 if Never 1 if otherwise (+)
How often play games indoors	0 if Never 1 if otherwise (+)
How often play/walk with child outside	0 if Never 1 otherwise (+)
Mother spanks child when child misbehaves	0 if No 1 if Yes (-)
Mother hits back child when child hits her	0 if No 1 if Yes (-)
Mother ignores child when child hits her	0 if No 1 if Yes (-)
Mother makes fun of child when child misbehaves	0 if No 1 if Yes (-)
Mother yells to child when child misbehaves	0 if No 1 if Yes (-)

(+) indicates a weight of +1 and (-) indicates a weight of -1

The ECLS-B does not assess direct interactions of the father and the child, except in the very rare cases in which the father is the parent respondent (the parent responding to the parent interviews), which happens in less than 1% of cases. Therefore, I construct the second measure of father quality only for responses to questions to the father interview when the child is 9 months, 2 years old, and 4 years old. As with the mother interview sub-score, I transform the responses to 0-1 variables and add them up with a weight of 1 or -1 to form a continuous score. The following table summarizes the questions that I use, how I transform them to 0-1 variables, and their weight:

The correlations between the two measures of maternal and paternal care quality is shown in Table 11

J.3.2 Identification and estimation of the measurement system of parental care

So far we have constructed two measures of maternal care quality and two measures of paternal care quality:

$$\widetilde{\log q}^{j,s} = \mu^{j,s} + \alpha^{j,s} \log q^j + \epsilon^{j,s} \,,$$
$$\text{for } j = m, f \text{ and } s = 1, 2 \,,$$

with $\mu^{j,1} = 0$ and $\alpha^{j,1} = 0$ for $j = m, f$ It is going to be useful later to estimate $\mu^{j,2}$ $\alpha^{j,2}$ for $j = m, f$.

Note that if there is an instrument Z that is uncorrelated with measurement error in

Table 10: Questions used for the second measure of father quality

Question	Transformation to 0-1
Questions asked when child is 9 months old	
How often read books to child?	0 if Never 1 otherwise (+)
How often tell stories to child?	0 if Never 1 otherwise (+)
How often sing songs to child	0 if Never 1 otherwise (+)
How often play peekaboo with child	0 if Never 1 otherwise (+)
How often tickle child	0 if Never 1 otherwise (+)
How often play/walk with child outside	0 if Never 1 otherwise (+)
How often talk about child	0 if Never 1 otherwise (+)
How often carry pictures of child	0 if Never 1 otherwise (+)
How often think about child	0 if Never 1 otherwise (+)
How often think that holding/cuddling child is fun	0 if Never 1 otherwise (+)
How often prefer to get things for child instead of himself	0 if Never 1 otherwise (+)
Think father must play with child	1 Agree/Strongly Agree, 0 otherwise (+)
Think father have long term effects on babies	1 Agree/Strongly Agree, 0 otherwise (+)
Think providing is more important than emotional support	0 Agree/Strongly Agree, 1 otherwise (+)
Think fatherhood is rewarding	1 Agree/Strongly Agree, 0 otherwise (+)
Questions asked when child is 2 years old	
How often read books to child?	0 if Never 1 otherwise (+)
How often tell stories to child?	0 if Never 1 otherwise (+)
How often sing songs to child	0 if Never 1 otherwise (+)
Feel like gave up more than expected because of fatherhood	1 Agree/Strongly Agree, 0 otherwise (+)
Expected warmer feelings from fatherhood	1 Agree/Strongly Agree, 0 otherwise (+)
Feel trapped by fatherhood	1 Agree/Strongly Agree, 0 otherwise (+)
Father spanks child when child misbehaves	0 if No 1 if Yes (-)
Father hits back child when child hits her	0 if No 1 if Yes (-)
Father ignores child when child hits her	0 if No 1 if Yes (-)
Father makes fun of child when child misbehaves	0 if No 1 if Yes (-)
Father yells to child when child misbehaves	0 if No 1 if Yes (-)

parental care qualities we have that:

$$cov(\widetilde{\log q}^{j,1}, Z) = cov(\log q^j, Z) \ ,$$

Questions used for the second measure of father quality (continuation)

Question	Transformation to 0-1
Questions asked when child is 2 years old	
Think father must play with child	1 Agree/Strongly Agree, 0 otherwise (+)
Think father have long term effects on child	1 Agree/Strongly Agree, 0 otherwise (+)
Think providing is more important than emotional support	0 Agree/Strongly Agree, 1 otherwise (+)
Think fatherhood is rewarding	1 Agree/Strongly Agree, 0 otherwise (+)
Questions asked when child is 4 years old	
Father spanks child when child misbehaves	0 if No 1 if Yes (-)
Father hits back child when child hits her	0 if No 1 if Yes (-)
Father ignores child when child hits her	0 if No 1 if Yes (-)
Father makes fun of child when child misbehaves	0 if No 1 if Yes (-)
Father yells to child when child misbehaves	0 if No 1 if Yes (-)

(+) indicates a weight of +1 and (-) indicates a weight of -1

Table 11: Correlations between measures of parental care

$\text{corr}(\widetilde{\log q}^{m,1}, \widetilde{\log q}^{m,2})$	$\text{corr}(\widetilde{\log q}^{f,1}, \widetilde{\log q}^{f,2})$
0.16	0.15

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

$$\text{cov}(\widetilde{\log q}^{j,2}, Z) = \alpha^{j,2}\text{cov}(\log q^j, Z) \ ,$$

and hence:

$$\alpha^{2,s} = \frac{\text{cov}(\widetilde{\log q}^{j,2}, Z)}{\text{cov}(\widetilde{\log q}^{j,1}, Z)} \ .$$

We can use the predicted Arnett score for relative care in wave 2 as the instrument Z. Substituting population objects for sample analogs in the previous expression we get the estimates for $\alpha^{m,2}, \alpha^{f,2}$ shown in Table 12

Since $\mathbb{E}[\widetilde{\log q}^{j,1}] = \mathbb{E}[\log q^j]$ it follows that:

$$\mu^{j,2} = \mathbb{E}[\widetilde{\log q}^{j,2}] - \alpha^{j,2}\mathbb{E}[\widetilde{\log q}^{j,1}] \ .$$

Substituting population objects for sample analogs in the previous expression we get the estimates for $\mu^{m,2}, \mu^{f,2}$ shown in Table 13

Table 12: Estimates for $(\alpha^{m,2}, \alpha^{f,2})$

$\alpha_{m,2}$	$\alpha_{f,2}$
5.45	2.10

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 13: Estimates for $(\mu^{m,2}, \mu^{f,2})$

$\mu_{m,2}$	$\mu_{f,2}$
11.51	17.91

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

J.3.3 Distribution of parental care quality

Given that we have two conditionally independent noisy measures of each parental care quality and an instrument, the distribution of parental care quality is non-parametrically identified by Kotlarski's Theorem under mild technical assumptions. By the same token, the same is also true for the conditional distribution of parental care quality conditional on any observable covariate. However, given the available data a fully non-parametric estimator of the joint distribution of non-parental care with other states is likely to be very noisy. Therefore, I assume that the distribution of parental care quality j for $j = m, f$ is given by:

$$\log q^j = \log q^{j,e} + \nu_{j,q} ,$$

where

$$\log q^{j,e} = X'_{j,q}\beta_{j,q} .$$

with:

$$\begin{pmatrix} \nu_m \\ \nu_f \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} \sigma^2_{\nu,m} & \rho_{\nu,m,f}\sigma_{\nu,m}\sigma_{\nu,f} \\ \rho_{\nu,m,f}\sigma_{\nu,m}\sigma_{\nu,f} & \sigma^2_{\nu,f} \end{pmatrix}$$

And ν_j assumed to be independent of $X_{j,q}$ for $j = m, f$. Note that we can write:

$$\widetilde{\log q}^{j,1} = X'_{j,q}\beta_{j,q} + \nu_{j,q} + \epsilon^{j,1} ,$$

$$\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}} = X'_{j,q}\beta_{j,q} + \nu_{j,q} + \frac{\epsilon^{j,2}}{\alpha^{j,2}} .$$

Under the assumption that the measurement error $\epsilon^{j,2}$ is independent of $X_{j,q}$, we can identify $\beta_{j,q}$ from the best linear predictor of $\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}}$ on $X_{j,q}$ (the same applies to the best linear predictor of $\widetilde{\log q}^{j,1}$. The next table presents estimates of $\beta_{j,q}$ for $j = m, f$ for

Table 14: Observable predictors of parenting quality

Variable	$\log q^{m}$	$\log q^{f}$
Own wage	0.1	0.01
	(0.01)	(0.09)
Relative care $\widehat{Arnett}_2$	0.17	0.24
	(0.02)	(0.14)
High School	0.18	0.27
	(0.02)	(0.15)
College	0.31	0.50
	(0.03)	(0.18)
Post Grad	0.27	0.34
	(0.03)	(0.19)
Urban, inside UC	-0.02	-0.15
	(0.02)	(0.11)
Rural	-0.01	-0.01
	(0.02)	(0.11)
Midwest	0.00	-0.14
	(0.02)	(0.13)
South	-0.05	-0.26
	(0.02)	(0.13)
West	0.08	-0.16
	(0.02)	(0.14)
Constant	-0.41	-0.18
	(0.04)	(0.29)

NOTE: SE in parenthesis. Own wage refers to $\log w^{m}$ in the case of mothers and $\log w^{f}$ in the case of fathers.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Two Parent Families and for $j = m$ for Single Mothers obtained from a regression of $\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}}$ on $X_{j,q}$

Note that the variance of $\nu_{j,q}$ is identified from:

$$Var(\nu_{m,q}) = cov(\widetilde{\log q}_{j,1}, \frac{\widetilde{\log q}_{j,2} - \mu_{j,2}}{\alpha_{j,q}}) ,$$

and the covariance of $\nu_{m,q}, \nu_{f,q}$ is identified from:

$$cov(\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}}, \frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}}) .$$

We can estimate $\sigma^2_{j,q}, \rho_{\nu,m,f}$ by substituting population objects by sample analogs in the previous expressions. Results are shown in Table 15 For comparison, I show in Table 16 statistics of the distribution of the observable part of parental care

Table 15: Parameter estimates for the joint distribution of parental care unobservables

$\sigma^2_{\nu,m}$	$\sigma^2_{\nu,f}$	$\rho_{\nu,m,f}$
0.04	0.18	0.74
(0.01)	(0.2)	(0.56)

NOTE: Bootstrap SE in parentheses
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 16: Statistics of the joint distribution of $\log q^{m,e}, \log q^{f,e}$

$Var(\log q^{m,e})$	$Var(\log q^{f,e})$	$corr(\log q^{m,e}, \log q^{f,e})$
0.03	0.06	0.68

NOTE: $corr$ refers to the correlation coefficient.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

In order to get estimates of the variance of skill shocks later on, we are going to need estimates of the variance of the measurement error of the parental quality measures. In order to identify those, note that

$$Var(\log q^j) = cov(\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}}, \widetilde{\log q}^{j,1}) \ ,$$

and

$$Var(\epsilon^{j,1}) = Var(\widetilde{\log q}^{j,1}) - Var(\log q^j)$$

$$Var(\frac{\epsilon^{j,2}}{\alpha^{j,2}}) = Var(\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}}) - Var(\log q^j) \ .$$

Substituting these expressions by sample analogues yields the following estimates: It is

Table 17: Measurement error variances of parental quality measures

$Var(\epsilon^{m,1})$	$Var(\frac{\epsilon^{m,2}}{\alpha^{m,2}})$	$Var(\epsilon^{j,1})$	$Var(\frac{\epsilon^{f,2}}{\alpha^{f,2}})$
0.90	0.18	0.97	1.32

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

also informative to estimate the signal-to-noise ratios:

J.4 Cognitive skills of children

J.4.1 Measures of cognitive skills

A key feature of ECLS-B is that it contains developmentally-appropriate measures of cognitive development at each wave. The measures of cognitive development that I use are the Motor and Mental Scores at 9 months and 2 years old, and the Mathematics and Reading scores at ages 4 and 5 years old.

Table 18: Signal-to-noise ratios for parental qualities

$s^{m,1}$	$s^{m,2}$	$s^{f,1}$	$s^{f,2}$
0.08	0.30	0.17	0.13

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

J.4.2 Identification of the measurement system of cognitive skills

Remember the measurement system for cognitive skills at time t:

$$\widetilde{\log\theta}_t^s = \mu_{\theta,t}^s + \alpha_{\theta,t}^s \log\theta_t + \epsilon_{\theta,t}^s \, .$$

Because cognitive skills are latent, their scale and location are not identified. Hence, I normalize

$$\mu_{\theta,t}^1 = 0$$
$$\alpha_{\theta,t}^1 = 1$$

for all t. I choose this first "normalized" measures to be the mental scale score at times $t = 1,2$ and the math score at $t = 3,4$:

Identification of the measurement parameters for cognitive skills follows similar steps to the identification of the measurement parameters for latent parenting quality. That is, $\{\alpha_{\theta,t}^2\}_{t=1}^4$ are identified from:

$$\alpha_{\theta,t}^2 = \frac{cov(\widetilde{\log\theta}_t^2, Z_t)}{cov(\widetilde{\log\theta}_t^1, Z_t)} \, ,$$

where Z_t is again an instrument. In order to avoid a weak-instrument problem, it makes sense to let Z_t be the same measure as $\widetilde{\log\theta}_t^1$ in a period different from t. That is, I choose Z_1 to be the mental scale score at time 2, I choose Z_2 to be the mental scale score at time 1, Z_3 to be the mathematics score at time 4. and Z_4 to be the math scale score at time 3. Substituting the population covariances by its sample analogues yields the following estimates for $\alpha_{\theta,t}^2$: Once the $\alpha_{\theta,t}^2$ are identified for each t, we can compute the

Table 19: Estimates for the loading of the second measure of cognitive skills

$\alpha_{\theta,1}^2$	$\alpha_{\theta,2}^2$	$\alpha_{\theta,3}^2$	$\alpha_{\theta,4}^2$
0.78	0.74	0.89	0.95

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

variance of latent skill at each t from:

$$Var(\log\theta_t) = \frac{cov(\widetilde{\log\theta}_t^1, cov(\widetilde{\log\theta}_t^2)}{\alpha_{\theta,t}^2} \, ,$$

and the variance of measurement error for the first measure at each t as:

$$Var(\epsilon^1_{\theta,t}) = Var(\widetilde{\log\theta}^1_t) - Var(\log\theta_t) = 1 - Var(\log\theta_t) ,$$

where the second equality follows from the fact that $\widetilde{\log\theta}^1_t$ is standardized, and hence has a variance of 1. Substituting these expressions by their sample analogues yields the following estimates:

Table 20: Estimates for the Variance of $\log\theta_t$

$Var(\log\theta_1)$	$Var(\log\theta_2)$	$Var(\log\theta_3)$	$Var(\log\theta_4)$
0.90	0.63	0.90	0.86

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 21: Estimates for the Variance of $\epsilon^1_{\theta,t}$

$Var(\epsilon^1_{\theta,1})$	$Var(\epsilon^1_{\theta,2})$	$Var(\epsilon^1_{\theta,3})$	$Var(\epsilon^1_{\theta,4})$
0.1	0.37	0.10	0.14

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Given these estimates for the variance of the measurement error and latent skills in each period, we can compute the signal-to-noise ratios for each age as:

$$s^\theta_t = \frac{Var(\log\theta_t)}{Var(\log\theta_t) + Var(\epsilon^1_{\theta,t})} .$$

These signal-to-noise ratios are very high, especially for early ages (see Cunha, Heckman and Schennach (2010)). This speaks to the extraordinary quality of the ECLS-B data.

Table 22: Estimates for the signal-to-noise ratios for the first measure in each wave

s^θ_1	s^θ_2	s^θ_3	s^θ_4
0.87	0.61	0.88	0.86

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

J.4.3 Initial distribution of cognitive skills

In principle, it makes sense to allow initial skills to depend on family characteristics, that is (a_1, H). However, when I estimate a linear regression of $\widetilde{\log\theta}^1_1$ on a_1 and H (using instruments in order to correct for measurement error in initial assets and some components of H), I get that no coefficient is statistically significant, and that all the p-values are above 0.5. Note that this is not because of skills being very imprecisely measured themselves, given that the estimated signal-to-noise ratio in the first wave is very high, and given that if I include birthweight in the same regression its estimated coefficient is statistically very significant. Note that Moschini (2023) also finds that initial child skills are not very correlated with family characteristics.

Because of this, I assume that initial skills are independent of (a_1, H) and normally distributed. The only parameter of the distribution of skills to be estimated is their variance (because the location is normalized to 0), which is estimated to be 0.87 (see the previous discussion on estimating the measurement system for skills).

J.5 Production function of skills, Arnett location shifter, and relative care quality growth

Substituting the relationship between the predicted Arnett score and latent quality of non-parental care in the production function for skills we get:

$$\log\theta_{t+1} = \log A_t + \gamma_{\theta,t}\log\theta_t + \gamma_{m,t}\frac{\tau_t^m}{T}\log q^m + \gamma_{f,t}^j\frac{\tau_t^f}{T}\log q^f + \gamma_{p,t}\frac{\tau^p}{T}\frac{\widehat{ARNETT}_t^p - \mu_t^{ARNETT}}{\alpha_t^{ARNETT}} + \gamma_{r,t}\frac{\tau^r}{T}\frac{\widehat{ARNETT}_t^r - \mu_t^{ARNETT}}{\alpha_t^{ARNETT}} + \eta_{t+1}\,.$$

This can be re-arranged as:

$$\log\theta_{t+1} = \log A_t + \gamma_{\theta,t}\log\theta_t + \gamma_{m,t}\frac{\tau_t^m}{T}\log q^m + \gamma_{f,t}^j\frac{\tau_t^f}{T}\log q^f + \frac{\gamma_{p,t}}{\alpha_t^{ARNETT}}\frac{\tau^p}{T}\widehat{ARNETT}_t^p + \frac{\gamma_{r,t}}{\alpha_t^{ARNETT}}\frac{\tau^r}{T}\widehat{ARNETT}_t^r - \frac{\gamma_{p,t}}{\alpha_t^{ARNETT}}\frac{\tau_t^p}{T}\mu_t^{ARNETT} - \frac{\gamma_{r,t}}{\alpha_t^{ARNETT}}\frac{\tau^r}{T}\mu_t^{ARNETT} + \eta_{t+1}\,.$$

Note that the presence of η_{t+1} does not cause an endogeneity issue here. The reason is that because skills are log-additive in the value function (see Appendix B), all the household decisions, and in particular childcare time and paid childcare quality, are independent of η_{t+1}. Note the previous equation cannot be used directly to identify parameters because $\log\theta_t, \log\theta_{t+1}, \log q^m, \log q^f$ are measured with error. However, we have at least two noisy measures of each of those latent factors for each period. Given that α_t^{ARNETT} is identified, $\log A_t$, $\gamma_{j,t}$ for $j = \theta, m, f, p, r$, and μ_t^{ARNETT} are identified from the following moment conditions[27]:

$$\mathbb{E}\Big[Z(\widetilde{\log\theta_{t+1}}^{1} - \log A_t + \gamma_{\theta,t}\log\theta_t + \gamma_{m,t}\frac{\tau_t^m}{T}\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}} + \gamma_{f,t}^j\frac{\tau_t^f}{T}\frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}} + \frac{\gamma_{p,t}}{\alpha_t^{ARNETT}}\frac{\tau^p}{T}\widehat{ARNETT}_t^p + \frac{\gamma_{r,t}}{\alpha_t^{ARNETT}}\frac{\tau^r}{T}\widehat{ARNETT}_t^r - \frac{\gamma_{p,t}}{\alpha_t^{ARNETT}}\frac{\tau_t^p}{T}\mu_t^{ARNETT} - \frac{\gamma_{r,t}}{\alpha_t^{ARNETT}}\frac{\tau^r}{T}\mu_t^{ARNETT})\Big] = 0\,,$$

[27] To be more precise, the only parameters whose identification requires knowledge of α_t^{ARNETT} are $\gamma_{p,t}, \gamma_{r,t}$. μ_t^{ARNETT} is identified from dividing the coefficient of $\frac{\tau_t^r}{T}$ by the coefficient of $\frac{\tau^r}{T}\widehat{ARNETT}_t^r$

with

$$Z = (1, \widetilde{\log\theta_{t+1}}^{2}, \frac{\tau_t^m}{\overline{T}}\widetilde{\log q}^{m,1}, \frac{\tau_t^f}{\overline{T}}\widetilde{\log q}^{f,1}, \frac{\tau^p}{\overline{T}}\widehat{ARNETT}_t^p, \frac{\tau^r}{\overline{T}}\widehat{ARNETT}_t^r, \frac{\tau_t^p}{\overline{T}}, \frac{\tau_t^r}{\overline{T}})' \ .$$

Note that I am using $\frac{\widetilde{\log q}^{j,2}-\mu^{j,2}}{\alpha^{j,2}}$ in the estimation equation instead of $\widetilde{\log q}^{j,1}$ because $\frac{\widetilde{\log q}^{j,2}-\mu^{j,2}}{\alpha^{j,2}}$ has more variation.

The only non-standard part of this identification result is the fact that μ_t^{ARNETT} is identified jointly with the Production Function parameters and using only one measure of non-parental care quality. Note that because μ_1^m and μ_1^f are normalized to 0, μ_t^{ARNETT} measures the extent to which the ARNETT score overstates quality as compared to the measures of parental care. To see why, suppose that $\alpha_t^{ARNETT} = 1$, $\gamma_{m,t} = \gamma_{r,t} = 1$ there is no measurement error, and both the ARNETT score and the measure of parental care are standardized, so their mean is 0 by construction. Then

$$\begin{aligned} \mathbb{E}[\log q^m] - \mathbb{E}[\log q_t^r] = \mathbb{E}[\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}}] - (\mathbb{E}[ARNETT_t] - \mu^r) = \\ \mathbb{E}[\widetilde{\log q}^{m,1}] - (\mathbb{E}[ARNETT_t] - \mu^r) = \mu^r \end{aligned}$$

Hence, allocates one hour of care from a mother of average quality to a relative of average quality the associated change in skills tomorrow is given by:

$$-\frac{1}{\overline{T}}\mu_t^{ARNETT} \ .$$

Note that by the previous discussion, μ_t^{ARNETT} is related to the changes in skills associated to changes in the time children spend in different types of care. Hence, if one wants to analyze the effects on the skills of children of a policy that changes the allocation of time to each childcare arrangement, it is important to estimate μ_t^{ARNETT}. Setting μ_t^{ARNETT} to some value as opposed to estimating it is equivalent to taking a stand on the relative quality of different childcare arrangements.

As noted before, $\widetilde{\log q}^{m,1}$ and $\widetilde{\log q}^{f,1}$ are constructed using the same questions. Hence, to the extent that the way the questions used to construct those measures do not map to different levels of quality for mothers and fathers, it makes sense to make the normalization

$$\gamma_{m,t} = \gamma_{f,t} \ .$$

By the same token, it makes sense to normalize:

$$\gamma_{r,t} = \gamma_{p,t} \ .$$

The estimation equation then simplifies to:

$$\begin{aligned} \widetilde{\log\theta}_{t+1}^{1} = \log A_t + \gamma_{\theta,t}\widetilde{\log\theta}_t^{1} + \gamma_{par,t}(\frac{\tau_t^m}{\overline{T}}\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}} + \frac{\tau_t^f}{\overline{T}}\frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}}) + \\ \frac{\gamma_{nonpar,t}}{\alpha_t^{ARNETT}}(\frac{\tau^p}{\overline{T}}\widehat{ARNETT}_t^p + \frac{\tau^r}{\overline{T}}\widehat{ARNETT}_t^r) - \frac{\gamma_{nonpar,t}}{\alpha_t^{ARNETT}}\frac{\tau_t^r + \tau_t^p}{\overline{T}}\mu_t^{ARNETT} + \\ \eta_{t+1} + \epsilon_{\theta,t+1}^1 + \gamma_{\theta,t}\epsilon_{\theta,t}^1 - \gamma_{par,t}\left(\frac{\tau_t^m}{\overline{T}}\frac{\epsilon^{m,2}}{\alpha^{m,2}} + \frac{\tau_t^f}{\overline{T}}\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right) \ , \end{aligned}$$

where $\epsilon^{m,t}, \epsilon^{f,t}$ are measurement errors in maternal and paternal quality respectively.

The ECLS-B did not perform a direct observation of non-parental childcare arrangements in the first wave, which means that the Arnett score is not available in the first wave for either relative or paid childcare providers. Under the assumption that the change in relative care quality between $t = 1$ and $t = 2$ is the same across families, we can write for families that use no paid care at $t = 1$:

$$\widetilde{\log\theta}_2^1 = \log A_1 + \gamma_{\theta,1}\widetilde{\log\theta}_1^1 + \gamma_{par,1}\left(\frac{\tau_t^m}{T}\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}} + \frac{\tau_t^f}{T}\frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}}\right)$$
$$+\frac{\gamma_{nonpar,1}}{\alpha_2^{ARNETT}}\frac{\tau_1^r}{T}\widehat{ARNETT}_2^r - \frac{\gamma_{nonpar,1}}{\alpha_2^{ARNETT}}\frac{\tau_1^r}{T}(\mu_2^{ARNETT} + \alpha_2^{ARNETT}g_1^{q,r})+$$
$$\eta_2 + \epsilon_{\theta,2}^1 + \gamma_{\theta,1}\epsilon_{\theta,1}^1 + \gamma_{par,1}\left(\frac{\tau_1^m}{T}\frac{\epsilon^{m,2}}{\alpha^{m,2}} + \frac{\tau_1^f}{T}\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right).$$

Note that because μ_2^{ARNETT} is identified, and using similar arguments to the ones before, $g_1^{q,r}$ is identified. Moreover, note that despite the fact that we are conditioning on $\tau^p = 0$ there is no selection bias, because the only unobservables that show up in the previous equation are measurement errors and the idiosyncratic skill shock, which is independent of childcare decisions because of the log-additivity of skills on the value function that was proven in Appendix B. Note that this wouldn't be the case if we let $\tau_1^p > 0$, because then we would have an unobservable factor problem, where the unobservable production factor that would be given by the parental investment composite $\frac{\tau_1^p}{T}\log q_1^p$ (if $\tau_1^p = 0$ that factor is 0).

Table 23: Estimates of PF of skills

Parameter	$t=1$	$t=2$	$t=3$
$\log A_t$	0.08	-0.12	0.01
	(0.02)	(0.04)	(0.3)
$\gamma_{\theta,t}$	0.30	0.31	0.56
	(0.03)	(0.07)	(7.38)
γ_{par}	0.73	11.10	1.38
	(0.34)	(0.39)	(42.43)
γ_{nonpar}	0.34	0.82	0.08
	(0.29)	(0.37)	(1.1)
N	1800	1300	450

NOTE: Two-step estimates of Skill Production Function Parameters. In the first step the loadings of the Arnett score are calculated. In the second step γ_{nonpar} is calculated by multiplying the 2SLS-estimated coefficient of the non-parental time investment composite by the loading of the Arnett score at 2. The rest of the structural parameters are estimated by the 2SLS-estimated coefficients. The estimation equation is given above. Bootstrap SE in parentheses. N refers to the number of observations used in 2SLS stage rounded to the nearest 50 to comply with ECLS-B rules.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 24: Estimates of Arnett measurement system additive shifter and the growth rate of relative care quality between $t = 1$ and $t = 2$

$g_1^{q,r}$	μ_2^{ARNETT}	μ_3^{ARNETT}
1.45	-0.43	-1.17
(11.86)	(0.59)	(4.52)

NOTE: Bootstrap SE in parentheses.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Finally, we need to identify and estimate the variance of the skill shocks η_{t+1}, which matters for making predictions on the distribution of skills in different counterfactual scenarios (although it does not matter for investment behaviour per-se, because of the log-additivity of skills in the value function).

Remember that the residual of the estimation equation is given by:

$$\begin{aligned}
\text{residual}_t = \widetilde{\log\theta}^1_{t+1} - \log A_t - \gamma_{\theta,t}\widetilde{\log\theta}^1_t - \gamma_{par,t}\left(\frac{\tau^m_t}{\overline{T}}\widetilde{\log q}^m - \frac{\tau^f_t}{\overline{T}}\widetilde{\log q}^f\right) \\
-\frac{\gamma_{nonpar,t}}{\alpha^{ARNETT}_t}\left(\frac{\tau^p}{\overline{T}}\widehat{ARNETT}^p_t + \frac{\tau^r}{\overline{T}}\widehat{ARNETT}^r_t\right) - \frac{\gamma_{nonpar,t}}{\alpha^{ARNETT}_t}\frac{\tau^r_t + \tau^p_t}{\overline{T}}\mu^{ARNETT}_t \\
= \eta_{t+1} + \epsilon^1_{\theta,t+1} + \gamma_{\theta,t}\epsilon^1_{\theta,t} - \gamma_{par,t}\left(\frac{\tau^m_t}{\overline{T}}\frac{\epsilon^{m,2}}{\alpha^{m,2}} + \frac{\tau^f_t}{\overline{T}}\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right).
\end{aligned}$$

Taking the expectation of the square of the estimation residual we get:

$$\begin{aligned}
\mathbb{E}[\text{residual}^2_t] = \sigma^2_{\eta,t} + Var(\epsilon^1_{\theta,t+1}) + \gamma^2_{\theta,t}Var(\epsilon_{\theta,t}) + \\
\gamma^2_{par,t}\left(\mathbb{E}\left[\left(\frac{\tau^m_t}{\overline{T}}\right)^2\right]Var\left(\frac{\epsilon^{m,2}}{\alpha^{m,2}}\right) + \mathbb{E}\left[\left(\frac{\tau^f_t}{\overline{T}}\right)^2\right]Var\left(\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right)\right).
\end{aligned}$$

Substituting the population objects in this expression by sample counterparts we get:

Table 25: Estimates of the variance of the skill shocks in each period

$\sigma^2_{\eta,2}$	$\sigma^2_{\eta,3}$	$\sigma^2_{\eta,4}$
0.40	0.55	0.46
0.05	0.05	0.17

NOTE: Bootstrap SE in parentheses.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

K Identification and estimation of the TFP of quality production in centers and homes

TFP in the production function of quality of childcare in Center-Based Providers and Home-Based care providers can be identified from two sources. The first source of identification comes from the fact that once the extent to which the standardized Arnett score overstates or understates the quality of non-parental care (that is, μ^{AR}), and once α_E is known, once can identify the TFP of quality production by comparing the Arnett score observed in paid childcare providers and their inputs of production. More formally, remember from sub-Appendix J.1.1 the following relation:

$$\mathrm{Arnett}_{i,t} = \mu_t^{\mathrm{Arnett}} + \alpha_t^{\mathrm{Arnett}}\Big(\log A_t^{HB} + 1(D_{i,t} = CB)(\log A_t^{CB} - \log A_t^{HB}\Big) + \\ \alpha_t^{\mathrm{Arnett}}\left(\alpha_{E,t}\log\left(\frac{E}{k}\right) + (1-\alpha_{E,t})\log\left(\frac{C}{k}\right)\right) + \epsilon_t^{\mathrm{Arnett}} .$$

Because α_t^{ARNETT} from Appendix J.1.1 and μ_t^{ARNETT} is identified from J.5, $\log A_t^{HB}$ is identified from the constant of a regression of the Arnett score on the imput composite term and a dummy for center-based care, and $\log A_t^{CB}$ is identified from the constant of the constant and the coefficient of the dummy for center-based care in the same regression.

The other source of identification comes from comparing the quality of care purchased by families and the price they pay for it. If families are paying lower prices for higher quality, that is indicative of high efficiency in the production of quality. To be more precise, note that in an interior solution for relative care and paid care, in which the quality of paid care purchased is above the minimum level of quality at which the mandatory minimum staff-to-child ratio binds we have the following relationship between quality purchased and the quality of relative care available:

$$\log q_t^p = 1 + \frac{\gamma_{r,t}}{\gamma_{p,t}}\log q_t^r .$$

Note that $\log q_t^r$ is observed because $\widehat{ARNETT}_2, \widehat{ARNETT}_3$ are observed and $\mu_t^{ARNETT}, \alpha_t^{ARNETT}$ for $t = 2,3$ and $g_1^{r,q}$ are identified. Moreover, the price of paid care paid by families is observed up to classical measurement error:

$$\tilde{P}_{i,t} = P(q_{i,t}^P) + \epsilon_{i,t}^{P,q} .$$

From Lemma 3, we know that the price paid by families whose paid childcare providers' staff-to-child ratios are above the mandated minimum ones is given by:

$$P(q_{i,t}^P) = \left[w^E\left(\frac{\alpha_E}{1-\alpha_E}\frac{w^C}{w^E}\right)^{1-\alpha_E^j} + w^C\left(\frac{1-\alpha_E^j}{\alpha_E}\frac{w^E}{w^C}\right)^{\alpha_E}\right]\frac{1}{A}q_{i,t}^P .$$

The fact that $\epsilon_{i,t}^{P,q}$ is classical and mean zero implies :

$$\mathbb{E}[\tilde{P}_{i,t}|\frac{C}{h} > \underline{R}_l, 0 < \tau_t^r < \overline{T}^r] = \mathbb{E}\left[w^E\left(\frac{\alpha_E}{1-\alpha_E}\frac{w^C}{w^E}\right)^{1-\alpha_E^j} + w^C\left(\frac{1-\alpha_E^j}{\alpha_E}\frac{w^E}{w^C}\right)^{\alpha_E} q_{i,t}^P|\frac{C}{h} > \underline{R}_l, 0 < \tau_t^r < \overline{T}^r\right]\frac{1}{A} .$$

And note that everything but A is a parameter that was identified before or data (given that $q^P_{i,t}$ is known for each family that buys quality high enough with interior relative care as argued before)

The following table presents estimates of $\log A^{CB}_t$ and $\log A^{HB}_t$ coming from those two sources:

Table 26: Estimates of quality production TFP

Variable	Prices	Inputs-outputs
$\log A^{CB}_1$	1.68	-
	(11.70)	-
$\log A^{HB}_1$	-	-
	(-)	(-)
$\log A^{CB}_2$	3.85	3.29
	(2.16)	(2.14)
$\log A^{HB}_2$	3.37	2.80
	(2.17)	(2.17)
$\log A^{CB}_3$	7.1	7.14
	(9.98)	(853.10)
$\log A^{HB}_3$	6.83	6.88
	(10.33)	(852.98)

Table 27: NOTE: Bootstrap SE in parentheses.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

L Identification of preference parameters

In this Appendix, I present identification proofs for the preference parameters in the model. These preference parameters govern the taste for parental leisure, parental time with the child, skills of the child, and continuation assets $((\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a))$, as well as the distribution of fixed utility costs of sending the child to Center-Based and Home-Based care at each age $(\{\lambda_{CB,t}, \lambda_{HB,t}\}_{t=1}^3)$.

Before stating the identification result and discussing the proof, we should introduce some notation and discuss some features of the data. Define the skill gain from reallocating 1 hour of care from childcare arrangement j' to childcare arrangement j as:

$$\Delta_{\theta,i,t}^{j,j'} = \frac{1}{\overline{T}}\left(\gamma_j \log q_i^j - \gamma_{j'} \log q_{i,t}^{j'}\right) .$$

Because $\log q^m$ and $\log q^f$ are measured with mean-independent error, 0-mean error, $\Delta_{\theta,i,t}^{j,j'}$ is measured with mean-independent error, 0-mean error for $j = m, f$ and $j' = m, f, r, p$. That is:

$$\tilde{\Delta}_{\theta,i,t}^{j,j'} = \Delta_{\theta,i,t}^{j,j'} + \frac{1}{\overline{T}}\gamma_{j,t}\epsilon_{i,t}^j := \Delta_{\theta,i,t}^{j,j'} + \epsilon_{i,t}^{\Delta\theta_{j,j'}} \text{ for } j = m, f \text{ and } j' = r, p ,$$

and

$$\tilde{\Delta}_\theta^{j,j'} = \Delta_\theta^{j,j'} + \frac{1}{\overline{T}}\gamma_{j,t}\epsilon^j - \frac{1}{\overline{T}}\gamma_{j',t}\epsilon^{j'} := \Delta_{\theta,i,t}^{j,j'} + \epsilon_{i,t}^{\Delta\theta_{j,j'}} \text{ for } j, j' = m, f \text{ and } j \neq j' ,$$

and note that I am omitting the superindex denoting the measure of parental care for convenience (for each of the two measures of parental care the previous equations still hold).

Assets a_t are imputed (see section G for details), so they are also measured with error, that is:

$$\tilde{a}_{i,t} = a_{i,t} + \epsilon_{i,t}^a .$$

I assume here that the imputation error $\epsilon_{i,t}^a$ is independent of all the household choices and states[28].

ECLS-B asks the parent respondent for each age the child about the price of the main non-parental childcare arrangement. I assume that the answer to this question is a measurement error-ridden measure of the price of paid care, that is:

$$\tilde{P}_{it} = P^{D_{it}}(q_{i,t}^P) + \epsilon_{i,t}^P ,$$

where $D_{i,t} = CB, HB$ is the choice of paid care provider, $q_{i,t}^P$ is the quality of paid care chosen by family i at age t, and $\epsilon_{i,t}^P$ is measurement error in the price of paid care reported by the family, which I assume has mean zero, and is mean independent of all of the choices and states of the family.

$$\tilde{c}_{i,t} = w_i^m n_{i,t}^m + w_i^f n_{i,t}^f + \tilde{a}_{i,t}(1+r) - \tilde{a}_{i,t} - \tilde{P}_{i,t}\tau_t^p .$$

[28]This is indeed an assumption because the imputation procedure only ensures that measurement error in assets is uncorrelated with imputed assets. The results presented in this section go through under mean independence, but other results (such as the identification of the initial distribution of states) require full independence

Hence, we can write the noisy measure of consumption as:

$$\tilde{c}_{i,t} = c_{i,t} + \epsilon^{c}_{i,t} \,,$$

where

$$\epsilon^{c}_{i,t} = (1+r)\epsilon^{a}_{i,t} - \epsilon^{a}_{i,t+1} - \tau^{P}_{i,t}\epsilon^{P}_{i,t} \,.$$

Because $\epsilon^{a}_{i,t}, \epsilon^{a}_{i,t+1}, \epsilon^{P}_{i,t}$ are mean zero and mean independent of all the choices and states of the household, so is $\epsilon^{c}_{i,t}$.

The following proposition establishes sufficient data requirements for the identification of $(\delta^{m}_{l}, \delta^{f}_{l}, \delta^{m}_{\tau}, \delta^{f}_{\tau}, \{\delta_{\theta,t+1}\}^{3}_{t=1}, \delta_{a})$

Proposition 3. *Suppose that the following conditional distributions are observed:*

- *The distribution of* $(\tilde{c}_{i,t}, w^{j}_{i}, l^{j}_{i,t})$ *given* $n^{j}_{i,t} > 0$ *for* $j =, m, f$ *for some* t
- *The distribution of* $(w^{j}_{i}, l^{j}_{i,t}, \Delta^{p,r}, \tilde{P}_{i,t})$ *given* $0 < \tau^{r}_{i,t} < \overline{T}^{r}_{i}$, $\tau^{p}_{i,t} > 0$, *and* $n^{j}_{i,t} > 0$ *for some* $j = m, f$ *for* $t = 2, 3$
- *The distribution of* $\tau^{j}_{i,t}, l^{j}_{i,t}, \Delta^{j,r}_{\theta,i,t}$ *given* $0 < \tau^{r}_{i,t} < \overline{T}^{r}_{i}$ *for* $j = m, f$ *and for* $t = 1, 2, 3$
- *The distribution* $(\tilde{c}_{i,1}, \tilde{c}_{i,3}, \tilde{a}_{i,4})$

Additionally, assume that the measurement error in assets and the price of care $\epsilon^{a}_{i,t}$ *and* $\epsilon^{P}_{i,t}$ *are independent of all the choices and states, are independent of their own lags and leads, and* $\epsilon^{a}_{i,t}$ *is independent of* $\epsilon^{P}_{i,t'}$ *for all* t, t'.

Moreover, suppose that the following conditions hold:

- $\mathbb{E}[\tilde{c}_{i,t} | n^{j}_{i,t} > 0] \neq 0$ *for* $j = m, f$ *for some* t
- $\mathbb{E}[w^{j}_{i} l^{j}_{i,t} \Delta^{p,r}_{\theta,i,t} | 0 < \tau^{r}_{i,t} < \overline{T}^{r}_{i}, \tau^{p}_{i,t} > 0, n^{j}_{i,t} > 0] \neq 0$ *for* $j = m, f$ *and for some* $t = 2, 3$
- $\mathbb{E}[\tilde{c}_{i,1}\tilde{c}_{i,3}] \neq 0$

Then $(\delta^{m}_{l}, \delta^{f}_{l}, \delta^{m}_{\tau}, \delta^{f}_{\tau}, \{\delta_{\theta,t+1}\}^{3}_{t=1}, \delta_{a})$ *are identified.*

Proof. Remember the asset-conditional Lagrangian of the Two-Parent Household:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial c} &= \frac{1}{c} - \lambda^{BC} = 0\\
\frac{\partial \mathcal{L}}{\partial l^m} &= \frac{\delta_l^m}{l^m} - \lambda^m = 0\\
\frac{\partial \mathcal{L}}{\partial n^m} &= -\lambda^m + \lambda^{BC} w^m + \mu_n^m = 0\\
\frac{\partial \mathcal{L}}{\partial l^f} &= \frac{\delta_l^f}{l^f} - \lambda^f = 0\\
\frac{\partial \mathcal{L}}{\partial n^f} &= -\lambda^f + \lambda^{BC} w^f + \mu_n^f = 0\\
\frac{\partial \mathcal{L}}{\partial \tau^m} &= \frac{\delta_\tau^m}{\tau^m} + \frac{\Psi^m}{\overline{T}} \log q^m - \lambda^{SC} - \lambda^m = 0\\
\frac{\partial \mathcal{L}}{\partial \tau^f} &= \frac{\delta_\tau^f}{\tau^f} + \frac{\Psi^f}{\overline{T}} \log q^f - \lambda^{SC} - \lambda^f = 0\\
\frac{\partial \mathcal{L}}{\partial \tau^r} &= \frac{\Psi^r}{\overline{T}} \log q^r - \lambda^{SC} + \mu_\tau^r - \omega^r = 0\\
\frac{\partial \mathcal{L}}{\partial \tau^p} &= \frac{\Psi^p}{\overline{T}} \log q^p - \lambda^{SC} - \lambda^{BC} P^p(q^p) + \mu_\tau^p = 0\\
\frac{\partial \mathcal{L}}{\partial q^p} &= \Psi^p \frac{\tau^p}{\overline{T}} \frac{1}{q^p} - \lambda^{BC} \frac{dP^p}{dq^p}(q^p) \tau^P = 0
\end{aligned}$$

and the FOC for continuation assets:

$$\frac{1}{c} = \frac{\delta_a \beta}{a_4}$$

The proof uses these FOCs to identify the parameters of interest, given the appropriate conditional distributions.

1. **Identification of δ_l^m, δ_l^f**

   Combining the FOCs for c, n^m, and l^m we get:

   $$c = \frac{1}{\delta_l^m} w^m l^m$$

   This implies the following relationship in terms of data objects and measurement error:

   $$\tilde{c}_{i,t} = \frac{1}{\delta_l^m} w_i^m l_{i,t}^m + \epsilon_{i,t}^c$$

   Taking expectations on both sides conditional on $n_{i,t}^m > 0$ we get:

   $$\mathbb{E}[\tilde{c}_{i,t} | n_{i,t}^m > 0] = \frac{1}{\delta_l^m} \mathbb{E}[w_i^m l_{i,t}^m | n_{i,t}^m > 0]$$

   Re-arranging we get:

   $$\delta_l^m = \frac{\mathbb{E}[w_i^m l_{i,t}^m | n_{i,t}^m > 0]}{\mathbb{E}[\tilde{c}_{i,t} | n_{i,t}^m > 0]}\ .$$

   A similar argument using the distribution of $(\tilde{c}_{i,t}, w_{i,t}^f, l_{i,t}^f)$ establishes the identification of δ_l^f.

2. **Identification of** $(\delta_{\theta,4}, \delta_{\theta,4})$

   In an interior solution for relative care ($0 < \tau^r < \overline{T}^r$) the Lagrange multipliers on $\tau^r \geqslant 0$ and $\tau^r_t \leqslant \overline{T}^r$ are 0. Hence, from the FOC for τ^r we get that, in an interior solution for relative care

$$\frac{\Psi^r_t}{\overline{T}} \log q^r_t = \frac{\beta \Gamma^\theta_{t+1} \gamma_{r,t}}{\overline{T}} \log q^r_t = \lambda^{SC}_t .$$

   Combining this with the FOC for paid care τ^P, the FOC for consumption, the FOC for maternal labor supply, and the FOC for leisure, we get that in an interior solution for paid care (so $\mu^p_\tau = 0$) the following relationship has to hold:

$$\beta \Gamma^\theta_{t+1} \Delta^{p,r}_t = \delta^m_l \frac{P_t(q^P_t)}{w^m l^m_t} ,$$

   where I'm using the definition of $\Delta^{p,r}_t$

$$\Delta^{p,r}_t = \frac{1}{\overline{T}} (\gamma_{p,t} \log q^p_t - \gamma_{r,t} \log q^r_t) ,$$

   and the definition of Ψ^p_t, and I am omitting the type of care superscript (D) for convenience.

$$\Psi^p_t = \beta \Gamma^\theta_{t+1} \gamma_{p,t} .$$

   Re-arranging, we get:

$$\beta \Gamma^\theta_{t+1} w^m l^m_t \Delta^{p,r}_t = P^P_t(q^P_t) .$$

   This implies the following equation in terms of data objects and measurement error

$$\tilde{P}_{i,t} = \frac{\beta \Gamma^\theta_{t+1}}{\delta^m_l} \Delta^{p,r}_{i,t} + \epsilon^P_{i,t} .$$

   Taking conditional expectations and re-arranging we get:

$$\frac{\beta \Gamma^\theta_{t+1}}{\delta^m_l} = \frac{\mathbb{E}[\tilde{P}_{i,t} | 0 < \tau^r_{i,t} < \overline{T}^r_i, \tau^p_{i,t} > 0, n^m_{i,t} > 0]}{\mathbb{E}[w^m_i l^m_{i,t} \Delta^{p,r}_{i,t} | 0 < \tau^r_{i,t} < \overline{T}^r_i, \tau^p_{i,t} > 0, n^m_{i,t} > 0]} .$$

   Because δ^m_l is identified from the previous step and β is treated as known, Γ^θ_4 and Γ^θ_3 are identified from the distribution of $(w^m_i, l^m_{i,t}, \Delta^{p,r}, \tilde{P}_{i,t})$ given $0 < \tau^r_{i,t} < \overline{T}^r_i$, $\tau^p_{i,t} > 0$, and $n^m_{i,t} > 0$ for $t = 2, 3$.

   Since $\Gamma^\theta_4 = \delta_{\theta,4}$, it follows that $\delta_{\theta,4}$ is identified. Moreover, since $\Gamma^\theta_3 = \beta \gamma_{\theta,3} \Gamma^\theta_4 + \delta_{\theta,3}$, β is treated as known, $\gamma_{\theta,3}$ is identified (see Appendix J.5) and $\Gamma^\theta_3, \Gamma^\theta_4$ are identified, it follows that $\delta_{\theta,3}$ is identified.

   A similar argument applies if the distribution of $(w^f_i, l^f_{i,t}, \Delta^{p,r}, \tilde{P}_{i,t})$ given $0 < \tau^r_{i,t} < \overline{T}^r_i$, $\tau^p_{i,t} > 0$, and $n^f_{i,t} > 0$ is observed.

3. **Identification of** $\delta^m_\tau, \delta^f_\tau$ From the FOCs for τ^m, τ^r, l^m we get that in an interior solution for relative care the following condition needs to hold:

$$\frac{\delta^m_\tau}{\tau^m} + \beta \Gamma^\theta_{t+1} \Delta^{m,r}_\theta = \frac{\delta^m_l}{l^m} ,$$

and note that I am using again the definition of Ψ^m, Ψ^r and $\Delta_\theta^{m,r}$ Re-arranging we get:

$$\frac{\delta_\tau^m}{\delta_l^m} = \frac{\tau^m}{l^m} - \frac{\beta\Gamma_{t+1}^\theta}{\delta_l^m}\Delta_\theta^{m,r}\tau^m \ .$$

This implies that the following equation in terms of observables and measurement error holds:

$$\frac{\delta_\tau^m}{\delta_l^m} = \frac{\tau_{i,t}^m}{l_{i,t}^m} - \frac{\beta\Gamma_{t+1}^\theta}{\delta_l^m}\tilde{\Delta}_{\theta,i,t}^{m,r}\tau_{i,t}^m + \frac{\beta\Gamma_{t+1}^\theta}{\delta_l^m}\tau_{i,t}^m\epsilon_{i,t}^{\Delta\theta_{m,r}} \ .$$

Taking conditional expectations we get:

$$\frac{\delta_\tau^m}{\delta_l^m} = \mathbb{E}[\frac{\tau_{i,t}^m}{l_{i,t}^m}|0 < \tau_{i,t}^r < \overline{T}_i^r] - \frac{\beta\Gamma_{t+1}^\theta}{\delta_l^m}\mathbb{E}[\tilde{\Delta}_{\theta,i,t}^{m,r}\tau_{i,t}^m|0 < \tau_{i,t}^r < \overline{T}_i^r] \ .$$

Since δ_l^m and Γ_{t+1} are identified for $t = 2,3$, δ_τ^m is identified from the distribution of $\tau_{i,t}^m, l_{i,t}^m, \tilde{\Delta}_{\theta,i,t}^{m,r}$ given $0 < \tau_{i,t}^r < \overline{T}_i^r$ for $t = 2,3$ A similar argument establishes the identification of δ_τ^f from the distribution of $\tau_{i,t}^f, l_{i,t}^f, \tilde{\Delta}_{\theta,i,t}^{f,r}$ given $0 < \tau_{i,t}^r < \overline{T}_i^r$ for $t = 2,3$

4. **Identification of** $\delta_{\theta,2}$

   Note that once δ_τ^m is identified, the equation

$$\frac{\delta_\tau^m}{\delta_l^m} = \mathbb{E}[\frac{\tau_{i,1}^m}{l_{i,1}^m}|0 < \tau_{i,1}^r < \overline{T}_i^r] - \frac{\beta\Gamma_2^\theta}{\delta_l^m}\mathbb{E}[\tilde{\Delta}_{\theta,i,1}^{m,r}\tau_{i,1}^m|0 < \tau_{i,1}^r < \overline{T}_i^r] \ .$$

   can be used to identify Γ_2^θ. Using the expression for Γ_2^θ and the fact that β is treated as known, $\gamma_{\theta,1}$ is identified, and Γ_3^θ is identified, we can recover $\delta_{\theta,2}$

5. Identification of δ_a

   From the first order conditions for a_4 we get:

$$\frac{1}{c} = \frac{\beta\delta_a}{a_4} \ .$$

   Re-arranging:

$$a_4 = \beta\delta_a c_3 \ .$$

   This implies that the following equation in terms of observables and measurement error holds:

$$\tilde{a}_{i,4} = \beta\delta_a\tilde{c}_{i,3} + \epsilon_{i,4}^a - \beta\delta_a\epsilon_{i,3}^c \ .$$

   From the expression for the measurement error in consumption and the assumption that measurement error in assets and childcare prices is independent across periods we get that the following conditional moment restriction has to hold:

$$\mathbb{E}[\tilde{c}_{i,1}(\tilde{a}_{i,4} - \beta\delta_a\tilde{c}_{i,3})] = 0 \ .$$

   This conditional moment restriction identifies $\beta\delta_a$ if $\mathbb{E}[\tilde{c}_{i,1}\tilde{c}_{i,3}] \neq 0$, which holds by assumption.

   Because β is known, δ_a is identified.

$\square$

Some comments are in order. First, note that the identification result only requires observing wages for parents that are working. Hence, the fact that I am imputing wages for all parents is immaterial for the identification of the preference parameters discussed so far. Second, the data requirements in Proposition 3 are sufficient but not necessary. One can combine optimality conditions in a different way and arrive at a different set of sufficient data requirements. In fact, $(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a)$ are overidentified. Third, note that the assumptions

$$\mathbb{E}[\tilde{c}_{i,t}|n_{i,t}^j > 0] \neq 0, \mathbb{E}[w_i^j l_{i,t}^j \Delta_{\theta,i,t}^{p,r}|0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^j > 0] \neq 0, \mathbb{E}[\tilde{c}_{i,1}\tilde{c}_{i,3}] \neq 0 \, ,$$

are testable. Moreover, the assumptions

$$\mathbb{E}[\tilde{c}_{i,t}|n_{i,t}^j > 0] \neq 0, \mathbb{E}[w_i^j l_{i,t}^j \Delta_{\theta,i,t}^{p,r}|0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^j > 0] \neq 0 \, ,$$

are always true under the assumption of the model. Under the assumptions of the model, 0 consumption is never optimal provided wages and/or assets are positive (due to the Inada condition). For the same reason, 0 leisure is never optimal. Finally, wages are assumed to be positive and $\Delta_{\theta,i,t}^{p,r}$ is positive because if there is relative care available it is not optimal for families to pay for a childcare arrangement of inferior quality (see First Order Conditions of the asset-conditional problem).

The next proposition provides sufficient conditions for the identification of the paid care utility cost distribution parameters $\{\lambda_t^{CB}, \lambda_t^{HB}\}_{t=1}^3$. The proposition is written for Two-Parent families. The case for Single-Mother Households is similar and is omitted here.

Proposition 4. *Suppose that* $\{\mathbb{P}(D_t = HB), \mathbb{P}(D_t = CB)\}_{t=1}^3$ *are observed.*

Moreover, suppose that the following sets have strictly positive measure

$$\{a_t, H : \tilde{V}_t^D(a_t, H) > \tilde{V}_t^N(a_t, H)\} \textit{ for } D = CB, HB \textit{ and } t = 1, 2, 3$$

Then $\{\lambda_t^{CB}, \lambda_t^{HB}\}_{t=1}^3$ *are identified.*

Proof. First, it is useful to establish that conditional choice probabilities (that is, the probability of using each type of care at t conditional on the state at t being a_t, H) for each type of care are decreasing on the average of their own cost $(1/\lambda)$ and increasing on the average utility cost of the other type of childcare. Taking partial derivatives of the conditional choice probabilities (see expression for conditional choice probabilities in C) we get that if $\tilde{V}^{HB}(a_t, H) \geqslant \tilde{V}^{CB}(a_t, H)$

$$\frac{\partial \mathbb{P}(D = N|a_t, H)}{\partial \lambda_{HB}} = -(\tilde{V}^{HB} - \tilde{V}^N)e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^N)}e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^N)} \leqslant 0 \, ,$$
$$\frac{\partial \mathbb{P}(D = N|a_t, H)}{\partial \lambda_{CB}} = -(\tilde{V}^{CB} - \tilde{V}^N)e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^N)}e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^N)} \leqslant 0 \, ,$$

with strict inequality if $\tilde{V}_t^{HB}(a_t, H), \tilde{V}_t^{CB}(a_t, H) > \tilde{V}^N(a_t, H)$. Also, note that the second inequality follows because by the definition of $\tilde{V}_t^{CB}(a_t, H), \tilde{V}_t^{HB}(a_t, H)$ it must always be the case that:

$$V_t^{CB}(a_t, H) \geqslant V_t^N(a_t, H)\ ,$$
$$V_t^{HB}(a_t, H) \geqslant V_t^N(a_t, H)\ .$$

Taking partial derivatives to the rest of conditional choice probabilities we get:

$$\frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{HB}} = e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^N)} e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^N)} \overbrace{\left[(\tilde{V}^{HB} - \tilde{V}^N) - \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}}(\tilde{V}^{CB} - \tilde{V}^N)\right]}^{\geqslant 0}$$
$$+ \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \left(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^N)}\right) \left[(\tilde{V}^{HB} - \tilde{V}^{CB}) + \frac{1}{\lambda_{CB} + \lambda_{HB}}\right] \geqslant 0$$

again with strict inequality if $\tilde{V}_t^{HB}(a_t, H) > \tilde{V}^N(a_t, H)$.

$$\frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{CB}} = \frac{\lambda_{HB}}{\lambda_{HB} + \lambda_{CB}}(\tilde{V}^{CB} - \tilde{V}^N) e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^N) - \lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^N)}$$
$$- \frac{\lambda_{HB}}{(\lambda_{CB} + \lambda_{HB})^2} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \left(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{HB})}\right) \leqslant 0 \iff$$
$$(\tilde{V}^{CB} - \tilde{V}^N)(\lambda_{HB} + \lambda_{CB}) e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^N)} \leqslant \left(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^N)}\right)$$

which is true by the convexity of $e^{-(\tilde{V}^{CB} - \tilde{V}^N)x}$

Moreover:

$$\frac{\partial \mathbb{P}(D = CB|a_t, H)}{\partial \lambda_{CB}} = - \frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{CB}} - \frac{\partial \mathbb{P}(D = N|a_t, w^m, q^m, q^r, \overline{T}^r)}{\partial \lambda_{CB}} \geqslant 0$$

and

$$\frac{\partial \mathbb{P}(D = CB|a_t, H)}{\partial \lambda_{HB}} = - \frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{HB}} - \frac{\partial \mathbb{P}(D = N|a_t, w^m, q^m, q^r, \overline{T}^r)}{\partial \lambda_{HB}} =$$
$$e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^N)} e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^N)} \frac{\lambda_{CB}}{\lambda_{HB} + \lambda_{CB}}(\tilde{V}^{CB} - \tilde{V}^N)$$
$$- \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \left(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^N)}\right) \left(1 + (\tilde{V}^{HB} - \tilde{V}^{CB})\right) \leqslant 0$$
$$\iff e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^N)} \left(\tilde{V}^{CB} - \tilde{V}^N\right) \left(\lambda_{CB} + \lambda_{HB}\right) \leqslant$$
$$\left(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^N)}\right) (1 + (\lambda_{HB} + \lambda_{CB})(\tilde{V}^{HB} - \tilde{V}^{CB}))$$

which again is true by the convexity of $e^{-(\tilde{V}^{CB} - \tilde{V}^{HB})x}$

By symmetry, if $\tilde{V}^{CB} \geqslant \tilde{V}^{HB}$, $\tilde{V}^{CB}$ is strictly increasing in λ^{CB} and weakly decreasing in λ^{HB}.

Note that when taking derivatives, I am leaving the dependence of $\tilde{V}_t^P$ on a_t, H implicit to avoid clutter of notation. This establishes that $\mathbb{P}(D_t = D|a_t, H)$is weakly increasing in λ^P and weakly decreasing in $\lambda^{P'}$. Moreover, this monotonicity is strict if $\tilde{V}_t^{CB}(a_t, H), \tilde{V}_t^{HB}(a_t, H) > V^N(a_t, H)$ The rest of the proof proceeds by backward induction

- Last period ($t = 3$): Unconditional choice probabilities can be written as:

$$\mathbb{P}(P_3 = CB) = \int_{a_3, H} \mathbb{P}(P_3 = CB|a_t, H) dG_3(a_3, H) := \pi_3^{CB}(\lambda^{CB}, \lambda^{HB})\ ,$$
$$\mathbb{P}(P_3 = HB) = \int_{a_3, H} \mathbb{P}(P_3 = HB|a_t, H) dG_3(a_3, H) := \pi_3^{HB}(\lambda^{HB}, \lambda^{CB})\ ,$$

where we have data objects in the LHS and known functions of $\lambda_3^{CB}, \lambda_3^{HB}$ in the RHS. Moreover, note that $\pi^P(\lambda^{CB}, \lambda^{HB})$ because $\mathbb{P}(D_t = D|a_t, H)$ is strictly increasing in λ_t^P if $\tilde{V}^P > \tilde{V}^N$, it is always weakly increasing in λ^P, weakly decreasing in $\lambda^{P'}$, and the set

$$\{a_t, H : \tilde{V}_t^P(a_t, H) > \tilde{V}_t^N(a_t, H)\} \text{ for } D = CB, HB \text{ and } t = 1, 2, 3$$

has positive measure.

The fact that the RHS is a known function of $\lambda_3^{CB}, \lambda_3^{HB}$ follows from the fact that:

- $\tilde{V}_3^D$ can be computed for $D = N, CB, HB$ because:
  - β is treated as known and $(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a)$ are identified from Proposition 3
  - Price schedules $P_t^P(q)$ can be computed from the observed factor prices w^E, w^C, the observed staff-to-child ratios, and the technology parameters, which are identified from Appendices K, I.
- $G_t(a_t, H)$ is identified (see discussion Section 6.2.2

A simple contradiction argument establishes that $\lambda_3^{CB}, \lambda_3^{HB}$ are identified from $\mathbb{P}(P_3 = CB), \mathbb{P}(P_3 = HB)$.

Suppose that is not the case. Then, there exist two different pairs $(\lambda_3^{CB}, \lambda_3^{HB})$ and $(\tilde{\lambda}_3^{CB}, \tilde{\lambda}_3^{HB})$ that generate the same unconditional choice probabilities. Without loss of generality, assume $\lambda_3^{CB} > \tilde{\lambda}_3^{CB}$. Because π_3^{CB} is strictly increasing in λ_3^{CB}, it must be the case that

$$\tilde{\lambda}_3^{HB} < \lambda_3^{HB} \, ,$$

otherwise we would have

$$\pi_3^{CB}(\lambda_3^{CB}, \lambda_3^{HB}) > \pi_3^{CB}(\tilde{\lambda}_3^{CB}, \tilde{\lambda}_3^{HB}) \, ,$$

which is an immediate contradiction. However, because $\pi_3^{HB}(\lambda_3^{CB}, \lambda_3^{HB})$ is strictly increasing in λ_3^{HB} and weakly decreasing in λ_3^{CB}, we must have

$$\pi_3^{HB}(\tilde{\lambda}_3^{CB}, \tilde{\lambda}_3^{HB}) < \pi_3^{HB}(\lambda_3^{CB}, \lambda_3^{HB}) \, ,$$

which is a contradiction. Similar arguments apply if we start by assuming $\lambda_3^{CB} < \tilde{\lambda}_3^{CB}$, $\lambda_3^{HB} < \tilde{\lambda}_3^{HB}$, or $\lambda_3^{CB} > \tilde{\lambda}_3^{CB}$. This establishes that $(\lambda_3^{CB}, \lambda_3^{HB})$ are identified.

- Induction step: We want to show that if $\lambda_{t+1}^{CB}, \lambda_{t+1}^{HB}$, then $\lambda_t^{CB}, \lambda_t^{HB}$ are identified.

  Again, the unconditional choice probabilities can be written as:

$$\mathbb{P}(D_t = CB) = \int_{a_t, H} \mathbb{P}(D_t = CB|a_t, H) dG_t(a_t, H) := \pi_t^{CB}(\lambda^{CB}, \lambda^{HB}) \, ,$$
$$\mathbb{P}(D_t = HB) = \int_{a_t, H} \mathbb{P}(D_t = HB|a_t, H) dG_t(a_t, H) := \pi_t^{HB}(\lambda^{HB}, \lambda^{CB}) \, ,$$

  where we have data objects in the LHS and known functions of $\lambda_t^{CB}, \lambda_t^{HB}$ in the RHS. The fact that the RHS is a known function of $\lambda_t^{CB}, \lambda_t^{HB}$ follows from the fact that:

- $\tilde{V}_t^P$ can be computed for $D = N, CB, HB$ because:
  * β is treated as known and $(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a)$ are identified from Proposition 3
  * Price schedules $P_t^P(q)$ can be computed from the observed factor prices w^E, w^C, the observed staff-to-child ratios, and the technology parameters, which are identified from Appendices K, I.
  * From the Induction Hypothesis: $\lambda_{t+1}^{CB}, \lambda_{t+1}^{HB}$ are known, which matters for the computation of $\mathbb{E}\hat{V}_{t+1}(a_{t+1}, H, c_{t+1}^{CB}, c_{t+1}^{HB})$
- $G_t(a_t, H)$ is identified (see discussion Section 6.2.2

The contradiction argument goes exactly as before. This establishes the identification of $\{\lambda_t^{CB}, \lambda_t^{HB}\}_{t=1}^3$

□

Some comments are again in order. First, the condition

$$\{a_t, H : \tilde{V}_t^P(a_t, H) > \tilde{V}_t^N(a_t, H)\} \text{ for } D = CB, HB \text{ and } t = 1, 2, 3\,,$$

is testable. That is because if we observe some families using Center-Based care, and some other families using Home-Based care at each age, it must be true that the no-zero measure condition above is satisfied.

Second, note that the identification result assumes that unconditional choice probabilities are observed at each age, instead of conditional choice probabilities. I chose to write the identification result in terms of unconditional choice probabilities for practical reasons. In principle, one could target $\mathbb{P}(D_t = D | a_t, H)$ in estimation. However, (a_t, H) is very high-dimensional, is measured with error, and it's distribution is endogenous and of unknown functional form for $t = 2, 3$, which complicates in practice using $\mathbb{P}(D_t = D | a_t, H)$ in estimation.

M Estimation of Preference Parameters

I estimate preference parameters in three steps. In the first step, I estimate

$$(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3) \, .$$

I translate the joint-restrictions on conditional moments and parameters used in the proof of Proposition 3 into a Minimum-Distance estimator.

More precisely, $\hat{\vartheta} := (\hat{\delta}_l^m, \hat{\delta}_l^f, \hat{\delta}_\tau^m, \hat{\delta}_\tau^f, \{\hat{\delta}_{\theta,t}\}_{t=2}^4)$ solves the following minimization problem:

$$\hat{\vartheta} = \arg\min_{\vartheta} \psi(\vartheta, \pi_N)' W \psi(\vartheta, \pi_N)$$

with

$$\pi_N = \begin{pmatrix} \{\hat{\mathbb{E}}[w_i^m l_{i,t}^m | n_{i,t}^m > 0]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[\tilde{c}_{i,t} | n_{i,t}^m > 0]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[w_i^f l_{i,t}^f | n_{i,t}^f > 0]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[\tilde{c}_{i,t} | n_{i,t}^f > 0]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[\tilde{P}_{i,t} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^m > 0]\}_{t=2}^3 \\ \{\hat{\mathbb{E}}[w_i^m l_{i,t}^m \Delta_{i,t}^{p,r} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^m > 0]\}_{t=2}^3 \\ \{\hat{\mathbb{E}}[\frac{\tau_{i,t}^m}{l_{i,t}^m} | 0 < \tau_{i,t}^r < \overline{T}_i^r]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[w_i^f l_{i,t}^f \Delta_{i,t}^{p,r} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^f > 0]\}_{t=2}^3 \\ \{\hat{\mathbb{E}}[\frac{\tau_{i,t}^m}{l_{i,t}^m} | 0 < \tau_{i,t}^r < \overline{T}_i^r]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{m,r} \tau_{i,t}^m | 0 < \tau_{i,t}^r < \overline{T}_i^r]\}_{t=1}^3 \\ \{\hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{f,r} \tau_{i,t}^f | 0 < \tau_{i,t}^r < \overline{T}_i^r]\}_{t=1}^3 \end{pmatrix}$$

with $\psi(\vartheta, \pi_N)$ given by:

$$\Psi(\vartheta, \pi_N) = \begin{pmatrix} \{\delta_l^m - \frac{\hat{\mathbb{E}}[w_i^m l_{i,t}^m | n_{i,t}^m > 0]}{\hat{\mathbb{E}}[\tilde{c}_{i,t} | n_{i,t}^m > 0]}\}_{t=1}^3 \\ \{\delta_l^f - \frac{\hat{\mathbb{E}}[w_i^f l_{i,t}^f | n_{i,t}^f > 0]}{\hat{\mathbb{E}}[\tilde{c}_{i,t} | n_{i,t}^f > 0]}\}_{t=1}^3 \\ \{\frac{\beta \Gamma_{t+1}^\theta}{\delta_l^m} - \frac{\hat{\mathbb{E}}[\tilde{P}_{i,t} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^m > 0]}{\hat{\mathbb{E}}[w_i^m l_{i,t}^m \Delta_{i,t}^{p,r} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^m > 0]}\}_{t=2}^3 \\ \{\frac{\beta \Gamma_{t+1}^\theta}{\delta_l^f} - \frac{\hat{\mathbb{E}}[\tilde{P}_{i,t} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^f > 0]}{\hat{\mathbb{E}}[w_i^f l_{i,t}^f \Delta_{i,t}^{p,r} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^f > 0]}\}_{t=2}^3 \\ \{\frac{\delta_\tau^m}{\delta_l^m} - \hat{\mathbb{E}}[\frac{\tau_{i,t}^m}{l_{i,t}^m} | 0 < \tau_{i,t}^r < \overline{T}_i^r] - \frac{\beta \Gamma_{t+1}^\theta}{\delta_l^m} \hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{m,r} \tau_{i,t}^m | 0 < \tau_{i,t}^r < \overline{T}_i^r]\}_{t=1}^3 \\ \{\frac{\delta_\tau^f}{\delta_l^f} - \hat{\mathbb{E}}[\frac{\tau_{i,t}^f}{l_{i,t}^f} | 0 < \tau_{i,t}^r < \overline{T}_i^r] - \frac{\beta \Gamma_{t+1}^\theta}{\delta_l^f} \hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{f,r} \tau_{i,t}^f | 0 < \tau_{i,t}^r < \overline{T}_i^r]\}_{t=1}^3 \end{pmatrix}$$

where Γ_{t+1}^θ is given by the expression in B (and note that β is fixed to 0.95 and $\gamma_{\theta,t}$ is estimated on a previous step), and W_n is a diagonal matrix that is not a function of the data.

In the second step I estimate $\beta\delta_a$ by running a 2SLS regression of $\tilde{a}_4$ on $\tilde{c}_3$ using $\tilde{c}_1$ as an instrument. Note that this estimator is the sample analogue of the expression that characterizes δ_a in the proof of Proposition 3.

In the third step, I estimate the parameters of the distribution of fixed utility costs of paid care $\{\lambda_t^{CB}, \lambda_t^{HB}\}$ using indirect inference. From proposition 4 we know that $\{\lambda_t^{CB}, \lambda_t^{HB}\}$

are identified from unconditional choice probabilities. Therefore, the unconditional choice probabilities $\{\{\mathbb{P}(D_t = D)\}_{D=N,CB,HB}\}_{t=1}^{3}$ are the first set of targets that I use in the Indirect Inference step. Because I want to match both the choice probabilities of Two-Parent families and the choice probabilities of Single-Mother Households, I estimate $\{\lambda_t^{CB}, \lambda_t^{HB}\}$ separately for Two-Parent families and Single Mothers (that is, I run two different Indirect Inference estimation procedures).

From 4 it should be noted that identifying $\{\lambda_t^{CB}, \lambda_t^{HB}\}$ from unconditional choice probabilities requires knowing the joint distribution $G_t(a_t, H)$ for each t. While this object is non-parametrically identified, an estimation procedure that leverages this fact is likely to be very data-intensive given the dimensionality of (a_t, H) and the fact that a_t and some elements of H are measured with error. I deal with this issue by targeting statistics of the endogenous joint distribution of a_t and H for $t = 2,3$. By doing that, I reduce the risk that the Indirect Inference estimator yields an estimate of $\{\lambda_t^{CB}, \lambda_t^{HB}\}_{t=1}^{3}$ consistent with the wrong distribution of (a_t, H) [29]. The statistics that I target to deal with this issue are average assets at each age, the variance of assets at each age, and the correlation of assets with $(w^m, w^f, \{\log q^r\}_{t=2}^{3}, \overline{T}^r, \widetilde{\log q}^{m,2}, \widetilde{\log q}^{f,2})$ in the case of Two-Parent families and $(w^m, \{\log q^r\}_{t=2}^{3}, \overline{T}^r, \widetilde{\log q}^{m,2})$ in the case of Single Mothers. Moreover, I also target the amount of bunching around the mandatory minimum staff-to-child ratio at each age in Center-Based providers for Two-Parent families, and the average hours of paid care at each age for Two-Parent families. These statistics are important. Bunching is defined as:

$$\text{Bunching}(CB_t) = \mathbb{P}(\underline{R}_{l,t}^{CB} - \frac{1}{30} < \frac{C}{h} < \underline{R}_{l,t}^{CB} + \frac{1}{30} | D_t = CB) \,.$$

$\frac{C}{h}$ is reported by the parents in the data, and is calculated in model simulations given quality decisions and the observed factor prices according to Lemma 1.

Bunching around the ratio is informative of the fraction of children directly affected by changes in the mandatory minimum staff-to-child ratio, and also of the labor demand response to changes in mandatory minimum ratios. Moreover, the average hours of paid care is informative of the importance of the childcare market at each age. In order to target these statistics, I update the estimates of $\delta_{\theta,t+1}$ for $t = 1,2,3$ by re-estimating them in the indirect inference procedure. To make sure that all the parameters that I am estimating by Indirect Inference ($\{\lambda_t^{CB}, \lambda_t^{HB}, \delta_{\theta,t+1}\}_{t=1}^{3}$) are identified, I include

$$\hat{\mathbb{E}}[\frac{\tau_{i,t}^m}{l_{i,t}^m} | 0 < \tau_{i,t}^r < \overline{T}_i^r] \,,$$

$$\hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{m,r} \tau_{i,t}^m | 0 < \tau_{i,t}^r < \overline{T}_i^r] \,,$$

$$\hat{\mathbb{E}}[\frac{\tau_{i,t}^f}{l_{i,t}^f} | 0 < \tau_{i,t}^r < \overline{T}_i^r] \,,$$

[29]Note that the identification proof uses the true $G_t(a_t, H)$ which is non-parametrically identified. However, the proof does not prevent the existence of different $\{\lambda^[CB]_t, \lambda_t^{HB}\}_{t=1}^{3}$ rationalizing the observed choice probabilities for different $G_t()$ for $t = 2,3$

and

$$\hat{\mathbb{E}}[\tilde{\Delta}^{f,r}_{\theta,i,t}\tau^{f}_{i,t}|0<\tau^{r}_{i,t}<\overline{T}^{r}_{i}]$$

as targets (note that these identify $\{\delta_{\theta,t+1}\}_{t=1}^{3}$ given that $\delta^{m}_{\tau},\delta^{f}_{\tau}$ were estimated in the first step). Note that I am letting λ vary across household structures, but not any of the other preference parameters. This choice was made to balance two concerns. First, Single Mothers are different from Two-Parent families as evidenced by the fact that they make very different choices, and capturing these differences is likely to be important when predicting the distributional impacts of childcare market policies. However, because Single Mothers represent only a small fraction of my sample, allowing for total flexibility in preference parameters is likely to be inefficient, given the small sample size for Single Mothers.

Below I detail the Indirect Inference procedure for Two-Parent families. The algorithm for Single-Mothers is similar:

1. Read off the data assets, wages, the endowment of relative care, and quality of relative care, and the observable part of parenting quality, that is:

$$(a_1, w^m, w^f, \overline{T}^r, \log q^r, \log q^{m,e}, \log q^{f,e})\ .$$

   See Appendix G for the imputation of assets, F for the imputation of parental wages, H for the estimation of the relative care endowment for each family, and J.3 for the construction of the observable part of parental care.

2. Simulate the unobservable part of parental $(\nu^{m}_{i,j},\nu^{f}_{i,j})$ care quality from its estimated distribution many times $(j=1,\dots J)$ for each family i and construct $\log q^{m}_{i,j}$ and $\log q^{f}_{i,j}$ according to:

$$\log q^{m}_{i,j} = \log q^{m,e}_{i} + \nu^{m}_{i,j}\ ,$$
$$\log q^{f}_{i,j} = \log q^{f,e}_{i} + \nu^{f}_{i,j}\ .$$

   The resulting simulated dataset:

$$\{\{a_{1,i}, w^{m}_{i}, w^{f}_{i}, \overline{T}^{r}_{i}, \log q^{r}_{i}, \log q^{m}_{i,j}, \log q^{f}_{i,j}\}_{i=1}^{N}\}_{j=1}^{J}\ ,$$

   is an approximation to $G_1(a,H)$.

3. Build pricing schedules for each family i from information on the state the family lives in, the observed factor prices in that state w^E, w^C (remember w^E is simply the average lead-teacher premium in that state), and the observed regulations in that state.

4. Given $\vartheta^{IF}=\{\lambda^{CB}_{t},\lambda^{HB}_{t},\delta_{\theta,t+1}\}_{t=1}^{3}$, solve and simulate choices for each family, aggregate them, and compute the desired statistics.

5. Compute that weighted square distance:

$$\psi^{IF,TP}(\vartheta^{IF})'W\psi^{IF,TP}(\vartheta^{IF})$$

   where W is a diagonal matrix and $\psi^{IF,TP}(\vartheta^{IF})$ is the difference between the simulated and the observed targets.

Table 28: Preference parameter estimates

δ_l^m	0.25
δ_l^f	1.37
δ_τ^m	1.80
δ_τ^f	0.052
δ_a	15.01

NOTE: Minimum distance estimates for preference parameters. "SOURCE" below refers to the data source used to compute the target statistics.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Estimates from the first and second step are shown in Table 28, and estimates from the third step are shown in Tables 29 and 30. Tables 32 and 31 shows how the model fits some important statistics targeted in the Indirect Inference step

Table 29: IF estimates of the distribution of utility costs at each age for Two-Parent families and parental taste for skills at each age.

Parameter	Estimate
$\lambda_{CB,1}$	31.37
$\lambda_{CB,2}$	9.11
$\lambda_{CB,3}$	17.24
$\lambda_{HB,1}$	17.31
$\lambda_{HB,2}$	10.60
$\lambda_{HB,3}$	41.51
$\delta_{\theta,1}$	1.24
$\delta_{\theta,2}$	0.09
$\delta_{\theta,3}$	2.64

NOTE:"SOURCE" below refers to the source of the micro-data used in the indirect inference step. In particular, the statistics targeted in the Indirect Inference step and the initial distribution of states come from "SOURCE".The source of the wages of teachers and childcare workers is the BLS, and the source of regulations is Wheelock College in Boston (see the data section).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 30: IF estimates of the distribution of utility costs at each age for Single Mothers at each age.

Parameter	Estimate
$\lambda_{CB,1}$	3.23
$\lambda_{CB,2}$	1.08
$\lambda_{CB,3}$	1.06
$\lambda_{HB,1}$	1.87
$\lambda_{HB,2}$	0.34
$\lambda_{HB,3}$	0.64

NOTE:"SOURCE" below refers to the source of the micro-data used in the indirect inference step. In particular, the statistics targeted in the Indirect Inference step and the initial distribution of states come from "SOURCE".The source of the wages of teachers and childcare workers is the BLS, and the source of regulations is Wheelock College in Boston (see the data section).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 31: Fit, IF for TP families

Choice Probability	Model	Data
$\mathbb{P}(P_1 = CB)$	0.11	0.08
$\mathbb{P}(P_2 = CB)$	0.25	0.16
$\mathbb{P}(P_3 = CB)$	0.67	0.57
$\mathbb{P}(P_1 = HB)$	0.38	0.17
$\mathbb{P}(P_2 = HB)$	0.28	0.18
$\mathbb{P}(P_3 = HB)$	0.08	0.08
$\text{Bunching}(CB_1)$	15%	11%
$\text{Bunching}(CB_2)$	3%	3%
$\text{Bunching}(CB_3)$	66%	23%
$\mathbb{E}[\tau_1^p]$	1389	360
$\mathbb{E}[\tau_2^p]$	1542	501
$\mathbb{E}[\tau_3^p]$	2920	895

NOTE: "SOURCE" below refers to the "Data" column.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

Table 32: Fit, IF for SM households

Choice Probability	Model	Data
$\mathbb{P}(P_1 = CB)$	0.18	0.15
$\mathbb{P}(P_2 = CB)$	0.3	0.24
$\mathbb{P}(P_3 = CB)$	0.66	0.35
$\mathbb{P}(P_1 = HB)$	0.27	0.19
$\mathbb{P}(P_2 = HB)$	0.16	0.15
$\mathbb{P}(P_3 = HB)$	0.08	0.08

NOTE: "SOURCE" below refers to the "Data" column.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

N Identification and estimation of labor supply elasticities of Lead Teachers and Childcare Workers

In the model, the labor supply of lead teachers and childcare workers is given by:

$$LT_l = \overline{LT}_l (w_l^{LT})^{\eta_{LT}} ,$$

$$CCW_l = \overline{CCW}_l (w_l^{CCW})^{\eta_{CCW}} .$$

Note that in the previous equations, there are no unobservable supply shifters, which I want to allow for in estimation. The labor supply equations including the unobservable supply shifter are given by:

$$LT_l = \overline{LT}_l (w_l^{LT})^{\eta_{LT}} \exp(\xi_{l,t}^{LT}) ,$$

$$CCW_l = \overline{CCW}_l (w_l^{CCW})^{\eta_{CCW}} \exp(\xi_{l,t}^{CCW}) .$$

Taking logs we get

$$\log LT_l = \log \overline{LT} + \eta_{LT} \log(w_l^{LT}) + \xi_{l,t}^{LT} ,$$

$$\log CCW_l = \log \overline{CCW} + \eta_{CCW} \log(w_l^{CCW}) + \xi_{l,t}^{CCW} .$$

Note that a linear regression here is not enough to identify η_{LT}, η_{CCW} due to the familiar simultaneity bias (unobservable supply shifters move the labor supply equation, which affects the observed equilibrium wage, which makes wages mechanically correlated with the unobservable supply shifter). Hence, we need an instrument. The instrument that I use here is fertility lagged two years. Lagged fertility should be a relevant instrument to the extent that the number of children born two years ago increases the demand for childcare services. Moreover, if fertility causes teachers to leave the labor force, fertility **contemporanous** fertility should be endogenous. To the extent that teachers join the labor force again sufficiently fast, lagged fertility shouldn't be endogenous. The estimation equations are given by:

$$\log\left(\frac{LT_{l,t}}{\text{Employment}_{l,t}}\right) = \log\overline{LT} + \eta_{LT}\log w^{LT}_{l,t} + X'_{l,t}\beta_{LT} + \xi^{LT}_{l,t}\ ,$$

$$\log\left(\frac{CCW_{l,t}}{\text{Employment}_{l,t}}\right) = \log\overline{CCW} + \eta_{CCW}\log w^{CCW}_{l,t} + X'_{l,t}\beta_{CCW} + \xi^{CCW}_{l,t}\ .$$

Note that I am dividing the number of Lead Teachers (Childcare workers) in state l at time t by total employment, and that the intercepts of these two equations do not depend on l. Hence, I am only allowing the dependence of $\overline{LT}_{l,t}$ and $\overline{CCW}_{l,t}$ on l to happen through the size of the labor force in state l at time t and $X_{l,t}$. Because current fertility can affect labor supply, and current fertility is correlated with lagged fertility, I include contemporaneous and lagged one period fertility in $X_{l,t}$. The estimates of η_{LT} and η_{CCW} are shown in Table 33. Note that both the estimated elasticity for Lead Teachers and the elasticity for Child Care Workers fall in the range of elasticities estimated by Blau (1993) (he estimates elasticities in the range of 1.15-1.94).

Table 33: Elasticities of labor supply for Lead Teachers and Childcare Workers

η_{LT}	η_{CCW}
1.33	1.58
(1.89)	(2.65)

SE in parenthesis

Note that in order to solve the model and perform counterfactuals it is not the shifter of the extensive margin of labor supply for Lead Teachers and childcare workers that matters, but rather the products $\overline{LTH}_{LT}$ and $\overline{CCWH}_{CCW}$, that is, the shifters of factor supplies. In order to estimate the shifters of factor supplies, one could in principle use estimates for the shifters of the extensive margin of labor supply of Lead Teachers and Childcare workers and combine it with evidence on hours worked by Lead Teachers and Childcare workers. However, there is no guarantee that once the model is solved using those shifters, the wages of childcare workers and lead teachers produced by the model are going to be close to those in the data. Because of that, I use a different strategy here. Note that in equilibrium the supply and demand of factors need to equalize, that is:

$$E^D(w^E, w^C) = \overline{LT}(w^E + w^C)^{\eta_{LT}} \exp^{\xi^{LT}_{l,t}}\ ,$$

and

$$C^D(w^E, w^C) = \overline{LTH}_{LT}(w^E + w^C)^{\eta_{LT}} \exp^{\xi^{LT}_{l,t}} + \overline{CCW}(w^C)^{\eta_{CCW}} \exp^{\xi^{CCW}_{l,t}}\ .$$

Using the previous equation this simplifies to:

$$C^D(w^E, w^C) - E^D(w^E, w^C) = \overline{CCWH}_{CCW}(w^C)^{\eta_{CCW}} \exp^{\xi^{CCW}_{l,t}} .$$

Now, note that given that Household-side and childcare-provider side parameters are identified, factor demands at the observed prices can be calculated using the algorithm in Appendix E. Taking logs, taking expectations, and noting that $\xi^{LT}_{l,t}$ and $\xi^{CCW}_{l,t}$ have mean zero we get:

$$\log\left(\overline{H}_{LT}\overline{LT}\right) = \mathbb{E}[\log E^D(w^E, w^C)] - \eta_{LT}\mathbb{E}[\log(w^E + w^C)] .$$

$$\log\left(\overline{H}_{CCW}\overline{CCW}\right) = \mathbb{E}\left[\log\left(E^D(w^E, w^C) - C^D(w^E, w^C)\right)\right] - \eta_{CCW}\mathbb{E}[\log w^C] .$$

Substituting population objects by its sample analogs we get the estimates in Table 34

Table 34: Estimates of factor supply shifters

$\overline{H}_{LT}\overline{LT}$	$\overline{H}_{CCW}\overline{CCW}$
61.37	26,63

NOTE: In order to estimate $\overline{H}_{CCW}\overline{CCW}$, observations for which $C(w^E, w^C) \leqslant E(w^E, w^C)$ at the observed factor prices are excluded.
NOTE:"SOURCE" below refers to the source of the micro-data used to calculate model-implied factor demands at the observed wages. In particular, the initial distribution of states comes from "SOURCE".The source of the wages of teachers and childcare workers is the BLS, and the source of regulations is Wheelock College in Boston (see the data section).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) of children born in the calendar year 2001.

O Additional results: Teachers' wages

This Appendix examines further the effects of mandatory minimum ratios on the wages of lead teachers and childcare workers Because mandatory minimum ratios are usually expressed as 1 adult per n children, I consider all the combination of ratios in the following list:

$$\underline{R}_1 \in \left\{\frac{1}{8}, \frac{1}{7} \ldots, \frac{1}{3}\right\}$$
$$\underline{R}_2 \in \left\{\frac{1}{12}, \frac{1}{11} \ldots, \frac{1}{5}\right\}$$
$$\underline{R}_3 \in \left\{\frac{1}{14}, \frac{1}{13} \ldots, \frac{1}{7}\right\}$$

For each element in that list, I solve for the equilibrium wages of Lead Teachers and Childcare Workers in each region (remember that all the states within a region are identical, except for the regulations). Table 35 shows the wage-maximizing ratio combination in each region for lead teachers and childcare workers, and 36 shows

Wage maximizing staff-to-child ratios for lead teachers and childcare workers by region

	18 months	3 years old	4 years old
Lead Teachers, South	5	5	7
Lead teachers, West	5	5	7
Lead Teachers, Northeast	4	5	7
Lead Teachers, Midwest	4	5	7
Childcare Workers, South	5	5	7
Childcare Workers, West	6	5	7
Childcare Workers, Northeast	4	5	7
Childcare Workers, Midwest	5	5	7

Table 35: Combination of child-to-staff ratios at each age that maximize wages of teachers and childcare workers in each region. I am reporting children-to-staff ratio instead of staff-to-child ratio for clarity.

Wage minimizing staff-to-child ratios for lead teachers and childcare workers by region

	18 months	3 years old	4 years old
Lead Teachers, South	5	12	14
Lead teachers, West	3	12	13
Lead Teachers, Northeast	8	12	13
Lead Teachers, Midwest	8	12	13
Childcare Workers, South	7	12	10
Childcare Workers, West	3	12	11
Childcare Workers, Northeast	3	12	11
Childcare Workers, Midwest	8	12	12

Table 36: Combination of child-to-staff ratios at each age that minimize wages of teachers and childcare workers in each region. I am reporting children-to-staff ratio instead of staff-to-child ratio for clarity.

wage-minimizing ratios. Again, for ease of interpretation I present children-to-staff ratios, as opposed to staff-to-child ratios.

In conjunction, Tables 35 and 36 show that a regulator seeking to increase the wages of teachers using mandatory-minimum ratios will set stringent regulations (few children per teacher allowed). If the regulator wants to decrease the wages of teachers (for instance, to keep the cost of childcare low), it will choose less stringent ratios. Moreover, note that maximizing wages does not amount to setting the highest possible staff-to-child ratio, nor minimizing ratios amounts to setting the lowest possible ratio. This is because when increasing regulations there are forces that push wages up, and forces that push wages down, as discussed throughout the paper. Finally, Table 37 shows the lowest and the highest wage for lead teachers and childcare workers attainable in each region by manipulating regulations. It shows that regulations are able to affect wages substantially, emphasizing that accounting for general equilibrium effects when examining changes in regulations is important.

Wage range induced by regulations

	Wage range
Lead Teachers, South	9.97-10.47
Lead teachers, West	8.83-9.27
Lead Teachers, Northeast	9.80-10.16
Lead Teachers, Midwest	9.40-9.80
Childcare Workers, South	8.28-8.81
Childcare Workers, West	7.31-7.87
Childcare Workers, Northeast	8.14-8.57
Childcare Workers, Midwest	7.78-8.25

Table 37: Difference between the lowest and highest possible wages achievable by changing regulations. Wages are reported in 2001 dollars

P Optimal ratio by percentile

The previous exercises examine what happens when there is an extreme increase in the stringency of the mandatory minimum ratios, but it is not very informative of what happens for less extreme changes in regulations or offer any prescriptive guidance.

Because of that, I answer the question: What should mandatory minimum staff-to-child ratios be if the policy goal is to maximize the p-th percentile of the distribution of skills at kindergarten entry? In order to answer that question, I solve for the Equilibrium of the model in each US state and simulate the resulting US-wide distribution of skills. For traceability, I impose that mandatory minimum ratios are restricted to be the same for Center-Based and Home-Based providers. Because mandatory minimum ratios are usually expressed as 1 adult per n children, I consider all the combination of ratios in the following list:

$$\underline{R}_1 \in \left\{\frac{1}{8}, \frac{1}{7} \dots, \frac{1}{3}\right\}$$
$$\underline{R}_2 \in \left\{\frac{1}{12}, \frac{1}{11} \dots, \frac{1}{5}\right\}$$
$$\underline{R}_3 \in \left\{\frac{1}{14}, \frac{1}{13} \dots, \frac{1}{7}\right\}$$

From looking at table 38 we see that the optimal regulation features more stringent ratios at younger ages for most objectives (percentiles that we want to maximize). More stringent ratios at younger ages are a feature of reality too. The fact that the stringency of the optimal ratio for most percentiles decreases with age should not be surprising. First, $\alpha_{E,t}$ is lower at younger ages, which implies that the flattening effect of regulations on the price schedule is stronger at younger ages. Second, the TFP of paid care increases with age, which implies that the cognitive development cost associated to families not using paid care increases with the age of the child. Moreover, mandatory minimum ratios that maximize skills at the very bottom of the distribution are more lenient than those that maximize skills at the very top. This makes sense, given that as we have seen before, increasing ratios increases the skills of some poor children (by inducing

Skill-maximizing staff-to-child ratios for each percentile of the distribution of skills.

	18 months	3 years old	4 years old
1st percentile	5	12	14
10th percentile	4	12	12
20th percentile	5	12	14
30th percentile	3	5	14
40th percentile	3	7	14
50th percentile	6	12	11
60th percentile	6	12	11
70th percentile	3	6	14
80th percentile	3	5	13
90th percentile	5	5	9
99th percentile	3	12	11

Table 38: Combination of children-to-staff ratios at each age that maximize skills at each percentile. For simplicity, I impose that the mandatory minimum ratio is the same for Center-Based and Home-Based at each age. Note that to facilitate interpretation, I report the inverse of the staff-to-child ratios, that is, the maximum number of children per adult allowed in the classroom.

Variation in skills induced by ratios at each percentile

	$\Delta \log \theta_4$
1st percentile	4.5%
10th percentile	5.6%
20th percentile	3.2%
30th percentile	1.6%
40th percentile	2.8%
50th percentile	2.0%
60th percentile	2.1%
70th percentile	2.5%
80th percentile	2.2%
90th percentile	2.3%
99th percentile	3.1%

Table 39: Difference in skills at each percentile between the most favorable and the least favorable staff-to-child ratio for skills in each percentile. Skills are measured in Standard deviations of a mathematics test score. For instance, 4.5% at the first percentile means that the minimum and maximum value for the first percentile of the distribution of skills attainable by manipulating ratios is of 4.5% of a standard deviation of the mathematics test score in the data.

their families to buy more quality) while decreasing the skills of other poor children (by inducing their families to buy less paid care).A policy maker that seeks to maximize the top of the distribution only cares about the first effect, while a policy maker that seeks to maximize skills at the top only cares about the second. Table 39 shows the maximum variation that can be induced by mandatory minium ratios at each percentile of the distribution. In order to construct that maximum variation, I minimize and maximize each percentile of the skill distribution with respect to the mandatory-minimum ratios. Table 39 shows that mandatory-minimum ratios are more important for skill growth at the bottom of the distribution. This is to be expected, given that the most affected children (negatively and positively) to mandatory minimum ratios are born to poorer families.

References

Agostinelli, F. and Wiswall, M. (2016a). Estimating the technology of children's skill formation, *Technical report*, National Bureau of Economic Research.

Agostinelli, F. and Wiswall, M. (2016b). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development, *Technical report*, National Bureau of Economic Research.

Andreassen, C. and Fletcher, P. (2007). Early childhood longitudinal study, birth cohort (ecls-b): Psychometric report for the 2-year data collection. methodology report. nces 2007-084., *National Center for Education Statistics* .

Attanasio, O., Bernal, R., Giannola, M. and Nores, M. (2020). Child development in the early years: Parental investment and the changing dynamics of different dimensions, *Technical report*, National Bureau of Economic Research.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C. and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia, *American Economic Review* **110**(1): 48–85.

Attanasio, O., Meghir, C. and Nix, E. (2020). Human capital development and parental investment in india, *The Review of Economic Studies* **87**(6): 2511–2541.

Bell, A., Chetty, R., Jaravel, X., Petkova, N. and Van Reenen, J. (2019). Who becomes an inventor in america? the importance of exposure to innovation, *The Quarterly Journal of Economics* **134**(2): 647–713.

Berlinski, S., Ferreyra, M. M., Flabbi, L. and Martin, J. D. (2020). Child care markets, parental labor supply, and child development.

Berlinski, S., Ferreyra, M. M., Flabbi, L. and Martin, J. D. (2023). Child care markets, parental labor supply, and child development.

Bernal, R. and Keane, M. P. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production, *Journal of Econometrics* **156**(1): 164–189.

Bernal, R. and Keane, M. P. (2011). Child care choices and children's cognitive achievement: The case of single mothers, *Journal of Labor Economics* **29**(3): 459–512.

Blau, D. M. (1993). The supply of child care labor, *Journal of Labor Economics* **11**(2): 324–347.

Blau, D. M. (2003). Do child care regulations affect the child care and labor markets?, *Journal of Policy Analysis and Management* **22**(3): 443–465.

Blau, D. M. (2007). Unintended consequences of child care regulations, *Labour Economics* **14**(3): 513–538.

Borowsky, J., Brown, J. H., Davis, E. E., Gibbs, C., Herbst, C. M., Sojourner, A., Tekin, E. and Wiswall, M. J. (2022). An equilibrium model of the impact of increased public investment in early childhood education, *Technical report*, National Bureau of Economic Research.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. and Pan, Y. (2014). Early childhood investments substantially boost adult health, *Science* **343**(6178): 1478–1485.

Cascio, E. U. (2023). Does universal preschool hit the target? program access and preschool impacts, *Journal of Human Resources* **58**(1): 1–42.

Chaparro, J., Sojourner, A. and Wiswall, M. J. (2020). Early childhood care and cognitive development, *Technical report*, National Bureau of Economic Research.

Chipty, T. (1995). Economic effects of quality regulations in the day-care industry, *The American Economic Review* **85**(2): 419–424.

Chipty, T. and Witte, A. D. (1997). An empirical investigation of firms' responses to minimum standards regulations.

Cunha, F. and Heckman, J. (2007). The technology of skill formation, *American economic review* **97**(2): 31–47.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation, *Journal of human resources* **43**(4): 738–782.

Cunha, F., Heckman, J. J., Lochner, L. and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation, *Handbook of the Economics of Education* **1**: 697–812.

Cunha, F., Heckman, J. J. and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation, *Econometrica* **78**(3): 883–931.

Currie, J. and Hotz, V. J. (2004). Accidents will happen?: Unintentional childhood injuries and the effects of child care regulations, *Journal of health economics* **23**(1): 25–59.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit, *American Economic Review* **102**(5): 1927–1956.

Daruich, D. (2018). The macroeconomic consequences of early childhood development policies, *FRB St. Louis Working Paper* (2018-29).

De Nardi, M., French, E., Jones, J. B. and McGee, R. (2021). Why do couples and singles save during retirement?, *Technical report*, National Bureau of Economic Research.

Elango, S., García, J. L., Heckman, J. J. and Hojman, A. (2015). Early childhood education, *Economics of means-tested transfer programs in the United States, volume 2*, University of Chicago Press, pp. 235–297.

Flood, S., McMurry, J. F., Sojourner, A. and Wiswall, M. J. (2021). Inequality in early care experienced by us children, *Technical report*, National Bureau of Economic Research.

García, J. L., Heckman, J. J. and Ronda, V. (2021). The lasting effects of early childhood education on promoting the skills and social mobility of disadvantaged african americans, *Technical report*, National Bureau of Economic Research.

Gertler, P., Heckman, J. J., Pinto, R., Chang, S. M., Grantham-McGregor, S., Vermeersch, C., Walker, S. and Wright, A. (2021). Effect of the jamaica early childhood stimulation intervention on labor market outcomes at age 31, *Technical report*, National Bureau of Economic Research.

Gomez, M. L., Bernal, R. and Baker-Henningham, H. (2022). Qualitative evaluation of a scalable early childhood parenting programme in rural colombia, *Child: Care, Health and Development* **48**(2): 225–238.

Griffen, A. S. (2019). Evaluating the effects of childcare policies on children's cognitive development and maternal labor supply, *Journal of Human Resources* **54**(3): 604–655.

Heckman, J., Garcia, J., Leaf, D. and Prados, M. (2017). Quantifying the life-cycle benefits of a prototypical early childhood program, *Journal of Political Economy* .

Hotz, V. J. and Xiao, M. (2011). The impact of regulations on the supply and quality of care in child care markets, *American Economic Review* **101**(5): 1775–1805.

McMurry, J. (2021). Child care policy and informal care, *Working paper* pp. 604–655.

Moschini, E. G. (2023). Childcare subsidies and child skill accumulation in one-and two-parent families, *American Economic Journal: Macroeconomics* **15**(1): 475–516.

Mullins, J. (2022). Designing cash transfers in the presence of children's human capital formation, *Working Paper* .

Sylvia, S., Warrinnier, N., Luo, R., Yue, A., Attanasio, O., Medina, A. and Rozelle, S. (2021). From quantity to quality: Delivering a home-based parenting intervention through china's family planning cadres, *The Economic Journal* **131**(635): 1365–1400.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement, *The Economic Journal* **113**(485): F3–F33.

Todd, P. E. and Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps, *Journal of Human capital* **1**(1): 91–136.