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

Martin Garcia-Vazquez *

Feburary 2026

Abstract

Mandatory minimum staff-to-child ratios are a pervasive childcare market regulation in the US, yet little is known about 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 configurations of these 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 these minimum ratios increases the wages of childcare workers by up to 4.33% and wages of lead teachers by up to 2.75%. Despite making paid childcare more expensive, more stringent regulations increase the level of skills at kindergarten entry at virtually each percentile of the skill distribution. Effects on the skill distribution are more pronounced for children born to single mothers than those born to two-parent families. 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. Both family characteristics and the equilibrium adjustment of teacher's wages are key in determining the children who gain and lose.

1 Introduction

Governments extensively regulate the childcare market to ensure minimum quality standards, most notably through mandatory staff-to-child ratios. While these regulations—set in nearly every U.S. state—are intended to foster child development by increasing adult attention, they create a fundamental economic trade-off: they raise the quality of formal care but also its price (Hotz and Xiao (2011)). If higher prices

*Assistant Professor of Economics, Department of Economics, Washington University in St. Louis. Email: garciavazquez@wustl.edu. I thank Mariacristina De Nardi, Jeremy Lise, and Joseph Mullins for their guidance. I also thank Paco Buera, Alessandra Fogli, Bart Hamilton, Fabrizio Perri, Aaron Sojourner, 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.

drive families out of the formal sector and into lower-quality informal arrangements, regulations designed to protect children could unintentionally harm their development. In this paper, I quantify this trade-off and its distributional consequences. In doing so, I provide, to the best of my knowledge, the first piece of evidence on the effects of childcare market regulations on children's skills. To do so, I develop and estimate a structural equilibrium model of the childcare market that accounts for the endogenous equilibrium adjustment of childcare prices and teachers' wages, and how families change their childcare decisions as a result. Using rich data on children and their environments, I estimate how more stringent regulations shift the aggregate skill distribution and identify the characteristics of the children who gain—and lose—the most from these mandates.

Identifying the developmental impacts of these regulations presents a significant empirical challenge. Traditional reduced-form approaches could recover the average or quantile treatment effects of the regulations, but only under data requirements that are not met by the existing survey datasets that contain information on children skills and their families[1]. Moreover identifying the distribution of treatment effects without a structural model typical requires rank-invariance assumptions, which I show are not met in my model since children who are at the top of the skill distribution under stringent regulations could be at the bottom under lenient regulations, and viceversa. This happens because the change in a child's cognitive development depends on the gap between formal quality and their specific informal "outside option," a more stringent mandate can cause significant re-ordering: a child at the top of the skill distribution under lenient regulations may fall toward the bottom if the policy prices the household out of formal care, driving a re-sorting into low-quality informal arrangements. To capture these dynamics, I estimate a structural equilibrium model that captures the price effect of these regulations across the quality ladder, how families respond to these changes in prices by reallocating care across childcare arrangements, how these responses depend on observed and latent family characteristics, and how these family responses change the skills of children.

In the model, two-parent and single-parent families decide in each period how much to consume and save, the leisure and work hours of the adult household members, and crucially, the amount of time the child spends with each of their parents, their relatives, and in paid childcare. Families also choose the type of paid care that they buy

[1]First, typical nationally representative surveys that contain information on children's skills at a young age (such as PSID-CDS, the CNLSY, or the one used in this paper, the ECLS-B) contain at most a couple of hundred observations for the biggest states, such as Texas or California. This makes inference about the effects of regulations on the children's skill distribution infeasible. Second, the ECLS-B, which is arguably the best dataset for the purpose of this paper and the one that I use, follows only one cohort, so one cannot observe two groups of children of the same age in the same state under two different regulatory regimes, yielding a differences-in-differences strategy infeasible. Third, the regulations may be endogenous, as the literature already acknowledges (Currie and Hotz (2004)) and recent anecdotal evidence suggests (https://thefga.org/research/covid19-highlighted-problems-in-childcare/ and https://eu.desmoinesregister.com/story/news/politics/2022/06/16/more-kids-per-class-younger-workers-allowed-iowa-child-care-center-law-regulations/9587915002/

(center and home-based childcare) and its quality. Quality of paid care is produced by competitive childcare providers, who combine the efficiency units of the lead teacher with the number of adults per child in the classroom (which includes the lead teacher, but also childcare workers or teacher aids) in order to maximize profits, taking the wages of teachers and childcare workers as given, as well as the price-quality schedule. In equilibrium, this price-quality schedule is such that higher-quality childcare is more expensive than lower-quality childcare. This continuous price-quality schedule depends on technology parameters, the stringency of the regulations, and the wages of teachers, which are themselves determined in equilibrium to clear the labor market for teachers and childcare workers.

In the model, when regulations become more stringent, the price of childcare goes up. However, it goes up more for lower qualities, since lower-quality producers employ less staff members per child absent the regulation, which implies that a minimum requirement on this staff-to-child ratio distorts their cost more. This causes families to change their demand for quality and quantity of childcare, and some families may find it optimal to exit the childcare market altogether. The effect of these changes in families' decisions on the skills of the children depends on whether their demand for quality of paid care increases or decreases, how families reshuffle the care of their children across childcare arrangements, and the quality of these childcare arrangements. How families decisions change depends on the economic resources of these families, but also on their access to and the quality of their informal childcare arrangements. In the model, all of these are heterogeneous across families. This is not only crucial to predict the direction of the change between childcare arrangements induced by a given change in policy, but also their effect in the cognitive skills of children. For instance, a more stringent regulation may result in higher skills if a family substitutes from paid to maternal care, and maternal care is of higher quality, but in a decrease in skills if the opposite is true. This is particularly important because the quality of informal options has been shown to be very heterogeneous and correlated with economic resources in the United States (Flood et al. (2021)), and it has been hypothesized to be a necessary feature to reconcile the at first seemingly contradictory evidence on the expansion of public childcare programs (Fort et al. (2020),Kottelenberg and Lehrer (2017)).

In the model's partial equilibrium (holding wages of teachers and childcare workers constant), more stringent regulations distort the cost-minimization problem of lower quality childcare providers, which translates into the equilibrium price-quality schedule shifting upwards. Moreover, the distortion is higher for lower qualities, which translates into the equilibrium price-quality schedule becoming flatter. Higher costs may prompt families to reduce paid childcare hours, substituting them with parental or relative care. The quality of these non-paid arrangements influences the resulting impact on children's skills. Conversely, a flatter price-quality schedule may lead to higher-quality paid childcare, benefiting children's skills. These changes in demand for quality and quantity of childcare translate in changes in labor demand for teachers and childcare workers. This in turn changes wages, which can amplify or attenuate the partial equilibrium effects described before.

Even though the model has clear predictions for the partial equilibrium effects on

the price-quality schedule for childcare, the effects of regulations on families' choices and children's skills is a quantitative question, and it requires data to discipline the model's parameters. I show that the model presented in this paper is identified, with the identification proof translating naturally into a multi-step estimation procedure. Combining comprehensive individual-level data on children and their families with state-level data on regulations and labor market outcomes of childcare teachers, I estimate the model parameters. Using the estimated model, I simulate the effects of changing the stringency of mandatory minimum staff-to-child ratios from the set of most lenient regulations observed in my sample of states to the set of the most stringent regulations [2], on the wages of lead teachers and childcare workers and the distribution of cognitive skills of children. I find that going from the most lenient to the most stringent set of regulations increases the wages of lead teachers by up to 2.75% and childcare workers by up to 4.33%. These increases in wages, together with the fact that the regulation directly distorts the cost of providers for given wages makes childcare more expensive for each level of quality. Despite that, this policy increases most percentiles of the skill distribution for both children born to two-parent households and one-parent households, or in other words, the quantile treatment effects (QTE) of more strigent regulations are positive for most percentiles, for both children born to two-parent and one-parent households. This provides support for the existence of these policies if parents underinvest in their children compared to the social optimum, regardless of whether the socially optimal skill distribution features higher skills at the bottom (for example, if the relevant externality is crime (Cunha et al. 2010; Heckman et al. 2017), in the middle, or at the top (for example, if the relevant externality is innovation, Bell et al. 2019). [3]. The magnitude of this increase in skills for each percentile is much larger for children born to one-parent households. This happens because single mothers have lower wages and asset levels, so when regulations are lenient they buy lower-quality childcare. When regulations become more stringent, the price quality schedule both increases and flattens more for lower qualities, which means that both the incentives to exit the childcare market, and to increase qualities conditional on staying, are higher for these levels of quality. However, the incentive to exit the childcare market is counteracted by their higher need for formal childcare, since they have less care provided by relatives available, and their lower wages, lower assets, and absence of spousal income means limits their ability to substitute paid care for maternal care.

The overall effects on the distribution of skills conceal significant heterogeneity. Even though the distribution associated to more stringent regulations is almost uniformly better than the one associated to lower regulations, this near uniform improvement

[2] In my model, a regulation is a vector of 6 numbers

[3] Other reasons for underinvestment are incomplete contracts (parents cannot write investment contracts with their children), or borrowing constraints. The latter is the only source of inefficiency that does not involve the problem of the child when she becomes an adult, and hence is modeled explicitly in this paper. Because the other sources of inefficiency are not explicitly modeled, the model in this paper may be viewed as offering a positive theory of the distribution of children's skills at kindergarten entry, rather than a normative one.

is achieved by re-allocating children across quantiles of the skill distribution (that is, the change in skills associated to a change in regulations is not rank-preserving). For instance, the 10% of children born to single mothers who gain the most from more stringent regulations increase their skills by 41 % of a standard deviation, whereas the bottom 10% of children born to single mothers who lose the most decrease their skills by 5% of a standard deviation. The children who gain and lose the most are different on some key characteristics. The children who gain the most are born to poor families with little access to care provided by relatives. These families buy low-quality paid childcare when regulations are lenient. When regulations are stringent, the qualities that they use to buy are now disproportionally more expensive, and their ability to substitute to informal options is limited. As a consequence, they remain buying paid childcare, only that now they buy it of higher quality. On the other hand, the children who lose the most are born to families with a lot of low-quality care provided by relatives available. When regulations are lenient, these families buy paid childcare, because the alternatives are poor, and paid options are cheap. When regulations become more stringent, paid childcare becomes more expensive and these families use more informal childcare arrangements, trading off in this way child development for consumption.

This paper makes at least two contributions. First, this paper constitutes the first study of the effects of childcare market regulations on the skills of children, both on aggregate and across family types. Of course, childcare market regulations are not the only context in which accounting for equilibrium effects on teachers' wages and accounting for rich household heterogeneity is likely to be key to understand the effects on the skills of children. In principle, that is true of any policy that affects demand for teachers and childcare workers or that induces reallocation of time across different childcare arrangements. For instance, childcare subsidies and the expansion of public early childhood programs are widely studied examples of such contexts. In that sense, and additional contribution of this paper is to provide a quantitative equilibrium model of the US childcare market that allows for rich family heterogeneity and the equilibrium adjustment of the prices of labor inputs used in childcare production.

2 Related Literature

This paper relates to three strands of literature.

First, 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 .[4]

Second, this paper relates the most to a growing literature that uses equilibrium models of the childcare market for policy analysis. This paper contributes to this

[4]For transfer programs see Bernal and Keane (2010, 2011); Mullins (2022), for parent interventions see Sylvia et al. (2021); Gertler et al. (2021); Gomez et al. (2022), and for the introduction, expansion, or universalization of early childhood programs see Heckman et al. (2017); Daruich (2018); Chaparro et al. (2020); Cascio (2023).

literature by integrating a model of the childcare sector in which the price of center and home-based providers depends on the labor market of teachers and regulations, with a household decision model in which children's skills depend on how heterogeneous families decide to allocate their children's care to childcare arrangements of different qualities. In this literature, Moschini (2023) analyzes the impact of childcare subsidies on children's skills in an OLG model with endogenous family formation. In her model, families buy a childcare good of homogeneous quality and the childcare price is a constant fraction of wages. Berlinski et al. (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 et al. (2022) analyze the counterfactual effects of an important childcare market proposal by developing a model that accounts for the cost of childcare through the teacher labor market, acknowledging that increased demand for childcare can elevate costs by raising teacher wages. Relative to their paper, I include skill accumulation. Moreover, relative to both Berlinski et al. (2023) and Borowsky et al. (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[5]. More generally, this paper builds the first model in this literature that yields predictions on the distribution of children's skills in the presence of endogenous childcare prices and wages of teachers.

Finally, this paper also relates to the empirical literature on the effects of childcare market regulations. Papers in this literature use a reduced-form approach to identify the effects of childcare regulations on the provision of quantity and quality of childcare provided[6]. To my knowledge, this paper is the first to examine the effect of childcare regulations on children's skills and to estimate a structural model of these regulations.

3 Model

3.1 Families

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

3.1.1 Choices

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

[5]Another paper that deserves mention is Bodéré (2023), who models dynamic imperfect competition in a local childcare market but does not look at skills of children.

[6]See Chipty (1995); Chipty and Witte (1997); Blau (2003); Currie and Hotz (2004); Blau (2007); Hotz and Xiao (2011); Ali et al. (2024)

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 ($D_t = N$), 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 child's skills 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, because 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 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}). These costs capture that some families may dislike center or home-based care for idyosincratic reasons, like a distaste for those types of arrangements or the cost of driving to them. The cost o_t^D is exponentially distributed, independent of $o_\tau^{D'}$ for all t, τ, and independent of its own leads and lags.

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 G). 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.

[7] I am not dividing c by an equivalence scale. This does not affect results, because family size is exogenous and the equivalence scale shows up as an additive constant in utility.

3.1.4 Time use constraints

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

$$l_t^j + \tau_t^j + n_t^j = \overline{T}^j \,, \tag{TC$_j$}$$

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:

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

where τ_t^r and τ_t^p denote relative and paid care respectively. For single parent-households τ_t^f is 0, so their supervision constraint is:

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

3.1.5 Supply of relative care

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

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

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

3.1.6 Production of cognitive skills

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

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

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. This production function captures that skill gains from reallocating between childcare arrangements depend on the relative qualities of those childcare arrangements. It also does not impose that time spent in each type of care is positive for every family, which is at odds with the data, unlike other specifications in the literature (Caucutt et al. (2020)).

I assume that the qualities 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 I.5)[8]. Finally, quality of center-based care can be purchased on the market at the competitive price schedule $P_t^D(q)$, where D is CB or HB depending on whether the household is buying from a center-based or home-based provider.

For Single-Mother households $\tau^f = 0$, so skill production reduces to:

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

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

3.1.7 Budget constraint

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

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

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 $n_t^f = 0$. Note that the time invariant nature of wages is less restrictive than it may seem, since the time horizon of the model is 5 years.

3.1.8 Borrowing constraint

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

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

That is, there are borrowing constraints, which are important because they could be a source of underinvestment in children, specially in lower income families.

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 $\overline{\tau}^r$, and the quality of maternal,

[8]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.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.

paternal and relative care $q^f, q^m, \{q_t^r\}_{t=1}^3$[9]. 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

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 Two-Parent Family problem is summarized by the functional equation:

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

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

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

The problem of Single-Mother Households is similar and therefore omitted here.

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 policy functions do not depend on 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 $D_t = N, CB, HB$ and the choice-specific policy functions $g_\omega^{FT,D}(a_t, H)$ for

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

That is, 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 choice probabilities $\mathbb{P}^{FT}(D_t = D|a_t, H)$, choice-specific value functions $\tilde{V}_t^{FT,D}(a_t, H)$, and their corresponding policy functions (see Online Appendix).

3.4 Formal childcare providers

Formal childcare providers can be of two types: Center-Based providers (CB) or Home-Based providers (HB). Home-Based care providers include childcare providers operating at their own home, but also nannies operating at the child's home[10]. The market

[9] Quality of care by relatives is time invariant in the sense that its path is determined and know to families at child's birth

[10] A nanny taking care of only one child is a home-based care provider with a staff-to-child ratio of 1.

for formal childcare has free entry and is perfectly competitive. More precisely, upon entering the market, a childcare provider of type j offering quality q to children of age t can sell each hour of childcare at the competitively determined price $P_t^D(q)$, 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[11]:

$$q = F_t^j(\hat{E}, \hat{C}, k) = A_t^j \left(\frac{\hat{E}}{k}\right)^{\alpha_{E,t}^j} \left(\frac{\hat{C}}{k}\right)^{1-\alpha_{E,t}^j}.$$

Note that I am allowing the parameters of the production function to change with the age of the children and type of provider.

This technology captures that the quality of formal childcare is produced by combining the effectiveness 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[12] in a competitve labor market, with factor prices denoted by w^E and w^C respectively. When making these staffing decisions, childcare providers need to comply with mandatory-minimum ratio imposed by the state at age t, denoted by $\underline{R}_t$

3.4.1 Price of paid care

Given the production function for quality of paid care, we can solve in closed form for the equilibrium price-quality schedule for each type of paid care and for each age, given factor prices and regulations:

Lemma 1 (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 and quality q at age t, the price of childcare has to be given by:*

$$P_t^j(q) = \begin{cases} \overline{P}_t^j q \text{ if } q > q_{j,t}^* \\ \left[\underline{P}_t^j q^{\frac{1}{\alpha_{E,t}^j}} + \kappa_t^j\right] \quad \text{if } q \leqslant q_{j,t}^* \end{cases}$$

where

$$\overline{P}_t^j = \left[w^E \left(\frac{\alpha_{E,t}}{1-\alpha_{E,t}} \frac{w^C}{w^E}\right)^{1-\alpha_{E,t}^j} + w^C \left(\frac{1-\alpha_{E,t}^j}{\alpha_{E,t}^j} \frac{w^E}{w^C}\right)^{\alpha_{E,t}^j} \right] \frac{1}{A_t^j}$$

[11] I choose to focus only on labor inputs because labor accounts for 70 % of operating costs in childcare centers (Helburn and Howes (1996),Brown (2018))

[12] I used E and C to denote total efficiency units of the lead teacher and caregiving hours respectively, and their "hat" version to denote quantities hired to produce one hour of care.

$$\underline{P}_t^j = w^E \left(\frac{1}{\underline{R}_t}\right)^{\frac{1-\alpha_{E,t}^j}{\alpha_{E,t}^j}} \left(\frac{1}{A_t^j}\right)^{\frac{1}{\alpha_{E,t}^j}}, \kappa_t^j = w^C \underline{R}_t, q_t^* = A_t \underline{R}_t \left(\frac{\alpha_{E,t}^j}{1-\alpha_{E,t}^j}\frac{w^C}{w^E}\right)^{\alpha_{E,t}^j}$$

Proof. Comes directly out of the zero profit condition for firms and the form of the costs functions (See Lemma 3 in the Appendix A). Full proof in Appendix A. □

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 ratio is binding (due to an increase in κ_t^j). 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 $\underline{R}$. An increase in the price level is likely to push some families out of the market, but 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.

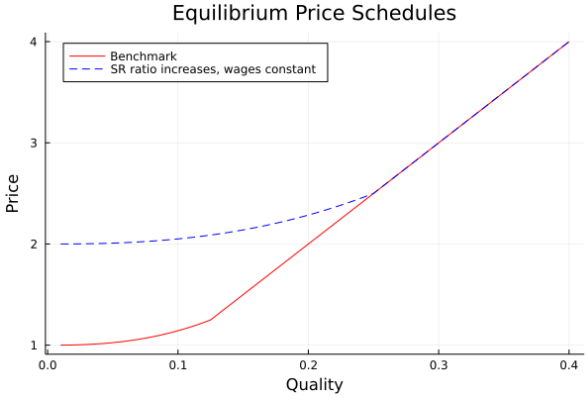


Figure 1: Price schedule for two different regulations and constant factor prices.

While the effect of changing the staff-to-child ratio on the price schedule of quality given factor prices is known (and is given by Lemma 1 and Figure 1), the effect on the demand for lead-teacher efficiency units and total caregiver hours is ambiguous [13]. 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

[13] On 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 extensive margin (families exiting the childcare market altogether) or 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 for which 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 (since the stricter ratio must be met), and a decrease in the demand for lead teacher talent E (since quality is kept constant).

another one in which factor prices decrease. Figure 2 suggests that the redistributive

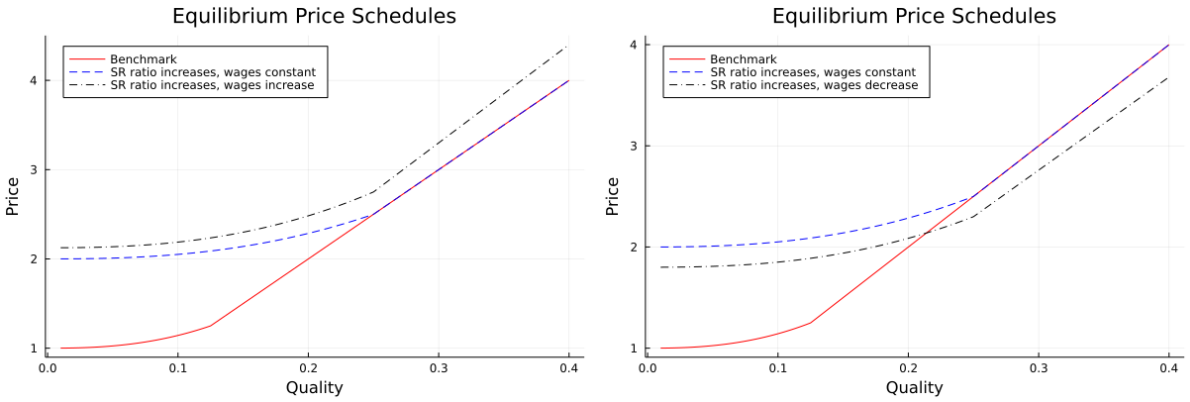


Figure 2: Price schedule for a benchmark level of the mandatory minimum staff-to-child ratio for a benchmark level of regulation and wages (solid line), more stringent regulation but wages as in the benchmark (dashed) and more stringent regulation and different wages (dash-dotted line). The left panel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

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. Keeping factor prices constant, families that buy lower qualities (who are likely poorer) pay a higher price, whereas families that buy high enough qualities (who are likely richer) are not affected. However, if factor prices increase as a consequence of the regulation change, every family pays a higher price for childcare. On the other hand, if factor prices go down as a consequence of the regulation change, higher qualities are cheaper than in the benchmark, which can benefit higher socioeconomic status families.

3.5 Labor supply of teachers and childcare workers

In order to predict how much wages of teachers will change after a change in regulatios, we need to model the labor supply of teachers. The labor supply of lead teachers and childcare workers in this model is very stylized and follows a constant elasticity specification:

$$LT = \overline{LT}(w^{LT})^{\eta_{LT}} ,$$

$$CCW = \overline{CCW}(w^{CCW})^{\eta_{CCW}} .$$

where LT and CCW are the number of lead teachers and childcare workers, w^{LT} and w^{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 they offer, 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 and given by $\overline{H}_{LT}$ and $\overline{H}_{CCW}$ respectively, 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). Hence, the supply of paid childcare production factors is:

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

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

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 (lead-teacher premium and wages of childcare workers), the price-quality schedule, and regulations. Given price-quality schedules, 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' efficiency units and caregiver-hours by childcare providers equates the supply for those factors at the equilibrium wages. A formal equilibrium definition is provided in Appendix C.

4 Data

The main source of data used in this paper is the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). The ECLS-B is a nationally representative survey of children born in the United States in 2001, and contains information on approximately 10700 children. 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. However, ECLS-B only contains limited information on families' assets, so I impute net-worth by using additional information from the Survey of Consumer Finances (SCF) (See Appendix E). Finally, to estimate the model I combine these individual-level data aggregate state-level data on mandatory minimum staff-to-child ratios and wages and labor supply of teachers and childcare workers.

I discuss in more detail now different groups of variables used in the estimation of the model.

4.1 Measures of cognitive skills

ECLS-B contains developmentally-appropriate measures of cognitive development at each wave. To measure cognitive development, I use the Motor and Mental Scores at 9 months and 2 years old, and the Mathematics and Reading scores at 4 and 5 years old.

4.2 Measures of parental quality

ECLS-B contains rich information on parental behavior and attitudes. I use this information to measure parental quality. To have comparable measures of quality for mothers and fathers, I construct one of the measures of maternal and paternal quality from survey questions that were asked to both parents, such as whether the parent plays with the child using toys to build things or takes the child for a walk or to play outside. For the second measure of parental quality, I use responses to survey questions about parenting for the father (such as whether the father reads books to the child, or tell stories, or thinks fatherhood is rewarding). For the second measure of maternal quality I also use similar survey questions, but I combine them with the the Two-Bag Task Score, which is a direct assesment of the quality of the interaction between a parent and their child [14]. See Appendix I.3.1 for details on parental care measurement.

4.3 Non-Parental care

For non-parental childcare arrangements, ECLS-B records the number of caregivers, the number of children being cared for together with the focal child, and the price parents pay for that source of childcare.

For the primary non-parental childcare arrangement, which is the non-parental childcare arrangement the child spends the most time with, ECLS-B contains additional

[14] ECLS-B only measured and recorded Two-Bag Task scores for one of the parents, which was the mother in the overwhelming majority of cases

information. Importantly, it contains the Arnett Caregiver Interaction Scale, (Arnett score from now on), a widely used measure of caregiver behavior (Arnett (1989)) , which is the measure of paid and relative care used in this paper[15]. The Arnett score measures the quality of the interactions of a caregiver with the children she is caring for. Moreover, if the primary non-parental childcare arrangement is a paid provider, we also observe the wage of the primary care provider (e.g, the lead teacher in the case of center-based care), which is used to construct the efficiency units of the lead teacher.

4.4 Socioeconomic and demographic variables

ECLS-B contains rich information on economic and demographic variables of children in the sample. In particular, we can observe earnings and hours worked, which allows us to construct wages (for parents who never work in my sample I impute wages, see Appendix D) for details), and some information on assets, such as house value or broad questions about portfolio composition (I impute net-worth using information common to SCF and ECLS-B. See Appendix E). ECLS-B also contains geographic information (importantly for the analysis here, the state where the family lives), and education of the parents and grandparents.

4.5 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 comes from the Wheelock College in Boston, and the data for 2005 and 2007 comes from the Childcare Licensing Study.

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 to 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. To deal with this issue, the mandatory minimum staff-to-child ratios that I input in the model is the one that applies at the midpoint of the model period for any given state. As an example, the regulation that I use in the model for period 1, which starts when the child is 9 months and ends when she is 2 years old, is the mandatory minimum ratio at 18 months.

Moreover, in the case of Home-Based providers, the minimum required number of adults in the room depends on the age distribution of children (home-based care providers can mix children of different ages in the same room), and on the number of children who are being cared for. Fortunately, the assumptions of the model naturally resolve the ambiguity of what mandatory-minimum ratio to use for home-based

[15]Because the Arnett score of care provided by relatives is not measured when the main non-parental caregiver is not a relative, I assume that relative care quality is a deterministic function of observables (in particular, the education of the maternal grandparents), and I estimate this relationship from cases in which the Arnett score is observed for relatives. See Appendix I.2.

providers. First, note that the model assumes that home-based care providers cannot mix children of different ages in the same room. Second, if home-based care providers of different sizes are subject to different mandatory-minimum ratios, only the size subject to the most lenient ratio survives in equilibrium, given constant returns to scale and perfect competition. Hence, I construct age-specific staff-to-child ratios for home-based care providers at age t in state l by reading the regulation of state l and finding out what would be the ratio that applies if a home-based care providers in state l were to only provide care to age t (as the model assumes) and choose the number of children she provides to minimize the mandatory staff-to-child ratio she is subject to. More details on this in Appendix F.

4.6 Other data

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[16]. 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 [17]

4.7 Sample Selection

I keep all the children in ECLS-B in my sample, except for the ones who are part of a multiple birth (twins, triplets, etc), and the ones whose parental respondent in ECLS-B is not the biological mother.

5 Model Identification

The model is point identified. Here I present here a heuristic identification discussion for some model primitives, but the interested reader can refer to the Appendix for more details and formal identification proofs.

5.1 Measurement parameters for parental and non-parental 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

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

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

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[18]:

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

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). Parametric identification of the measurement system for parental care qualities follows standard identification arguments (Cunha et al. (2010)) given the presence of two measures per factor and one instrument (in this case, I use as instrument the measure of care provided by relatives). The identification of the measurement system for care provided by relatives is a bit less standard. First, using the constant returns to scale assumption for the production of quality in paid care, the loading parameter is identified from the covariation between the noisy measure for care provided by non-parental providers (the Arnett score) and the inputs in production of paid care quality (efficiency units of the lead teacher per child, and number of adults per child). Moreover, the location parameter in the measurement system for the Arnett score is identified from the average skill gain of reallocating an hour of care from the average relative to the average mother. This average skill gain can be obtained as a parameter in the reduced-form equation that arises from combining the restrictions imposed jointly by the measurement systems and the technology of skill formation, in the spirit of Agostinelli and Wiswall (2016b) and Freyberger (2021).

5.2 Production Function of skills

The identification argument is similar to Agostinelli and Wiswall (2016a) and Agostinelli et al. (2025), except for the fact that the intercept in the measurement system for the Arnett score is identified jointly with technology parameters (as explained above). Let $\widehat{Arnett}_t^j$ be the value of the Arnett score of type of care $j = p, r$ at time t. 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) $\log q_t^j$ can be written as:

$$\log q_{i,t}^{j} = (\widehat{Arnett}_{i,t}^{j} - \mu_t^j)/\alpha_t^{Arnett} \ .$$

Substituting the relationship between $\widehat{Arnett}$ and true quality into the production function for child skills we get:

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

[18] Despite the fact that investment is itself a combination of latent qualities, dedicated measures for the quality of each childcare arrangement are available. For identification results on linear factor models where measures load on multiple latent factors see Williams (2020)

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[19].

5.3 Initial Distribution of household characteristics and assets

Let

$$H_i = (w^m, w^f, q^m, q^f, \{q^r_{i,t}\}_{t=1}^3, \overline{T}^r_i)$$

be the vector of characteriatics of household i, and let $G_{a_{i,1},H_i,\theta_{i,1}}(a_{i,1}, H_i, \theta_{i,0})$ denote the distribution of households over initial levels of wealth, time-invariant family characteristics, and initial skills. This distribution is identified.

First, parental wages are observed for parents that work at least in one of the sample periods, and imputed for the rest. Hence, they can be treated as observable. Second, the quality of care provided by relatives is also modeled as a deterministic function of observables, and this relationship is identified from the covariation between the Arnett score with said observables. In terms of the variables measured with error, assets are imputed from the SCF 2001, so the distribution of the measurement error is identified from the SCF, which is enough to recover the joint distribution of assets with any observable variable assuming that the measurement error is independent of those observable variables. The distribution of parental care quality and initial skills is identified non-parametrically given enough measures and the identification of the measurement parameters (Cunha et al. (2010)). Finally, 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 full support assumptions on the quality of parental care and the utility costs of using paid care, 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 with 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 G provides a formal proof of this result, details on the estimation, and estimation results.

Given the previous discussion, we can identify the distribution of the noisy measures skills, parental quality and assets, conditional on quality and quantity of relative care

[19] At $t = 1$ we do not observe Arnett scores. Because of this, we need to assume that at $t = 1$ relative care quality is a deterministic function of relative care quality at $t = 2$, and the identification argument needs to be extended slightly. See Appendix for a rigourous treatment of this issue.

and wages. Given the additivity of measurement error and the fact that the distribution of the measurement errors are known (from the previous arguments), a standard deconvolution arguments identifies the conditional distribution of skills, parental quality, and assets conditional on the variables that are or can be treated as observable, that is, quantity and quality of relative care and wages.

5.4 Preference Parameters

Preference parameters governing the marginal utility of leisure, time with the child, cognitive skills of the child, and the continuation utility of assets and child's skills $(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t}\}_{t=2}^4, \delta_a)$ are identified from family choices. In particular, the FOC of the household problem can be used to invert in closed form for this parameter given observed choices, measures of parental and relative care quality, and wages.

The parameters governing the distribution of utility costs of using paid care are identified from the choice probabilities for each type of paid care, given knowledge of preference parameters, the distribution of family states, and price-quality schedules. Intuitively, the distribution of utility costs is monotonic (in a first order stochastic dominance sense) in the exponential parameter λ^D, so higher levels of λ^D are associated with lower costs and higher choice probabilities for $D = CB, HB$. Formal identification proofs are available in the Online Appendix for both groups of preference parameters.

5.5 Technology of quality production of paid care 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 inputs. The intercept of this projection combines the efficiency of quality production (log-TFP) the intercept of the measurement system of the Arnett score μ_t^{Arnett}, which is identified. 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 1 [20]. See Appendices H and J for formal identification arguments and details on the estimation.

5.6 Labor Supply of Lead Teachers and Childcare Workers

When estimating labor supply elasticities, I allow for the presence of state-specific unobservable labor supply shifters, that is:

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

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

Including these shifters creates 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 join 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. See Appendix M for formal identification arguments, details on the estimation, and estimation results.

The identification strategy (and the resulting estimation strategy) is robust to mandatory staff-to-child ratios being endogenous to labor supply shocks. This is because identification of preference and technology parameters only depends on labor supply shocks to the extent that they affect wages of teachers, which are observed. Moreover, identification of labor supply parameters is robust to regulations being endogenous to the shock to the extent that lagged fertility is exogenous to the shock.

6 Estimation results and model fit

In this section I discuss model estimation, estimation results, and model fit. For the sake of brevity in presentation, I focus on four important groups of model parameters,

[20] 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). However, one can use the fact that the quality of paid care is known 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

and I relegate the discussion for the remaining parameters to the Appendix.

6.1 Production function of childcare quality

The production function of quality in childcare centers and home-based care providers is one of the novel features of the model in this paper, so it is worth discussing the estimated parameters for said technology and how those estimates are obtained. Parameter estimates imply that the efficiency units of the lead teacher are more important in the production of quality for older children than for younger ones, and home-based care providers are less efficient at producing quality than center-based care providers.

As discussed in the identification section, the importance of lead teacher efficiency units in quality production is identified from the bill share of this factor for childcare centers for which the ratio is not binding. Estimates for $\alpha_{E,t}$ are provided in Table 1:

Table 1: $(\alpha_{E,2}, \alpha_{E,3})$ estimated from wage bill shares

$\alpha_{E,2}$	$\alpha_{E,3}$
0.14	0.24
(0.0127)	(0.0065)

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.

Because the ECLS-B did not conduct a direct observation of non-parental childcare at 9 months, I estimate $\alpha_{E,1}$ by extrapolating from $\alpha_{E,2}$ 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 $\alpha_{E,9mo} = 0.093$).

The estimated $\alpha_{E,t}$ is larger for older ages. This is the case because, in the data, childcare centers for which the mandatory-minimum ratio is not binding choose to have more children per staff member at older ages.

There are two sources of variation that identify the TFP of quality production in centers and home-based care providers. One is the relationship between inputs and measured quality, the other is the relationship between measured quality and prices. Table 2 presents estimates of $\log A_t^{CB}$ and $\log A_t^{HB}$ coming from those two sources:[21]

[21] In the "Prices" row $\log A_t^{HB}$ is estimated from the corresponding $\log A_t^{CB}$ in the prices row and the estimated difference $\log A_t^{CB} - \log A_t^{HB}$ in the "Inputs-Outputs" row. That is $\log A_t^{CB} - \log A_t^{HB}$ is only estimated using differences in measured quality between Center and Home-Based providers given inputs, not from price information. The reason for that is that only observations above the mandatory-minimum ratio are used to estimate log-TFP in the "Prices" row, and mandatory-minimum ratios are imputed for the Home-Based sector (see Data section). When performing counterfactuals, I use the estimates in the "Prices" row. A detailed explanation of the identification and the estimation can be found in Appendix J

Table 2: Estimates of quality production TFP

	$\log A_1^{CB}$	$\log A_1^{HB}$	$\log A_2^{CB}$	$\log A_2^{HB}$	$\log A_3^{CB}$	$\log A_3^{HB}$
Prices	1.65	-	3.70	3.37	3.47	3.20
	(12.53)		(2.29)	(2.30)	(5.74)	(10.33)
Inputs-outputs	-	-	3.27	2.80	4.16	3.90
			(2.21)	(2.21)	(4.89)	(6.22)

NOTE: Bootstrap SE in parentheses. Because inputs of the production function of quality are not observed at 9 months, estimates for $\log A_1^{CB}$ and $\log A_1^{HB}$ based on that information are not available.
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.

As it can be seen, when TFP can be estimated using both information on prices and quality, and measured quality and inputs, the estimated TFPs are similar. This can be seen as failing to reject an informal validation test of the assumptions in the production side of the model (such as constant returns to scale, perfect competition, negligible fixed costs, and labor inputs being the only relevant production factors) [22].

6.2 Distribution of parenting quality

The statistical dependence of parenting quality and other family characteristics can amplify or attenuate the inequality in access to quality childcare arrangements experienced by children. I find that richer parents (in terms of wages) are more effective at fostering cognitive development of children, that children with access to better care provided by relatives also tend to have access to better parental care, and that quality of maternal care and paternal care are positively correlated. This positive correlation is not entirely explained by observables (e.g educated women having children with educated men, and education being positively associated with parenting quality).

From the identification section we know that the joint distribution of states is non-parametrically identified, but given the available data a non-parametric estimator of the joint distribution of parental care with other states is likely to yield noisy estimates. Hence, I impose the following parametric assumption on the dependence of parental care on household-level observables:

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

where

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

where $j = m, f$ with:

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

[22] For instance, markups would be rationalized by the perfect competition model of supply as a lower TFP as compared to the one implied by the one estimated from measures of quality and input choices.

And ν_j assumed to be independent of $X_{j,q}$ for $j = m, f$. Table 3 provides OLS estimates of $\beta_{j,q}$ for $j = m, f$.

Table 3: Observable predictors of parenting quality

Variable	$\log q^m$	$\log q^f$
Own wage	0.09	0.01
	(0.01)	(0.09)
Relative care $\widehat{\text{Arnett}}_2$	0.18	0.27
	(0.03)	(0.16)
High School	0.19	0.27
	(0.02)	(0.15)
College	0.33	0.50
	(0.03)	(0.18)
Post Grad	0.30	0.34
	(0.03)	(0.19)

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. Relative care $\widehat{\text{Arnett}}$ is the Arnett score predicted by family observables (see Appendix I.2). Regressions also include a constant and demographic dummies.
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.

The estimates in Table 3 say that more educated, higher wage parents, and parents with higher-quality relatives are more effective at fostering skill development.

Table 4 show estimates of the parameters of the distribution of the unobserved component of parenting quality. For comparison, I show in Table 5 statistics of the

Table 4: Parameter estimates for the joint distribution of parental care unobservables

$\sigma^2_{\nu,m}$	$\sigma^2_{\nu,f}$	$\rho_{\nu,m,f}$
0.05	0.18	0.75
(0.01)	(0.88)	(1.06)

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.

distribution of the observable part of parental care. Taken together, Tables 4 and 5 say that children whose mothers are effective at fostering skill development have fathers who are also effective at fostering skill development. More details on estimation of the parenting quality distribution can be found in Appendix I.3.

6.3 Production Function of Skills

The Production Function of Skills is estimated using 2SLS.The only source of endogeneity is measurement error, which is addressed by using repeated measurements as

Table 5: 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	0.67

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.

instruments[23] (Agostinelli and Wiswall (2016a)). Estimates are shown in Table 6. Pro-

Table 6: Estimates of PF of skills

Parameter	$t = 1$	$t = 2$	$t = 3$
$\log A_t$	0.04	-0.14	-0.02
	(0.03)	(0.04)	(0.24)
$\gamma_{\theta,t}$	0.30	0.32	0.79
	(0.04)	(0.07)	(1.32)
γ_{par}	0.75	1.10	0.26
	(0.37)	(0.46)	(9.82)
γ_{nonpar}	0.40	0.84	0.22
	(0.42)	(0.35)	(1.44)
N	1900	1300	650

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. 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.

duction function estimates imply that higher quality arrangements are more effective at fostering skills (which can be seen as a validation of the way we measure quality), and that reallocating from lower to higher quality arrangements boosts skills.

6.4 Household preferences and utility costs' of paid care

Preference parameters are key to ensure that families' choices generated by the model are consistent with those observed in the data, and to the extent that regulations affect children's skills through families' choices, that the model captures the effects on regulations changes on skills. Preference parameter estimates suggest that households derive higher utility from maternal time with the child than from paternal child, and that families value cognitive skills of the children at all ages, which rationalizes why some families are willing to pay for high quality childcare.

[23] See Appendix I.5 for a detailed discussion of the estimation of the production function.

The estimation procedure for preference parameters and the parameters of the distributions of choice-specific utility costs follows the logic of the identification proof (see Appendix K) and proceeds in 3 steps. In the first step, the First Order Conditions of single mothers and two-parent families are leveraged to estimate $(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3)$ by Minimum Distance. In the second step, the FOC for assets in the last period is used to estimate δ_a by 2SLS[24]. Finally, in the third step I use Indirect Inference to estimate the parameters of the choice-specific utility cost distributions $\{\lambda_{CB,t}, \lambda_{HB,t}\}$ and to update the estimates of $\{\delta_{\theta,t+1}\}_{t=1}^3$. The reason I update the estimates of $\{\delta_{\theta,t+1}\}_{t=1}^3$ is because in order to capture the equilibrium effects of changing regulations on the demand for childcare is important to match the fraction of households who demand childcare for which the regulation is binding and the total hours of childcare demanded. Hence, I target these statistics, together with the choice probabilities for center-based and home-based care, as well as the additional statistics that are needed in the identification proof to identify $\{\delta_{\theta,t+1}\}_{t=1}^3$[25] [26]. Estimates from the first and second step are shown in Table 7, and estimates from the third step are shown in Table 8.

Table 7: Preference parameter estimates

δ_l^m	δ_l^f	δ_τ^m	δ_τ^f	δ_a
0.25	1.37	1.60	0.04	15.01
(0.01)	(0.04)	0.29)	0.02)	(4.30)

NOTE: Minimum distance estimates for preference parameters. Bootstrap SE in parentheses. "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.

Looking now at the fit (Table 9), we can see that the model is able to reproduce the proportion of families buying center and home-based childcare, but there is a tension between matching the proportion of families buying center-based care for which the mandatory-minimum ratio is binding or close to binding (denoted in Table 9 as Bunching(CB_t)) and the average hours of paid care demanded by families. The reason is that $\delta_{\theta,t}$ governs both. Hence, in order to match the average hours of paid care demanded by families in the data, a lower value of $\delta_{\theta,t}$ is needed, which would generate more bunching around the mandatory minimum ratio. This would lead to exaggerating the equilibrium effects of changing regulations on childcare demand, and ultimately on teachers' wages and childcare prices. More details on the estimation of preference parameters and the distribution of utility costs can be found in Appendix L.

[24] We need to use 2SLS and not OLS because consumption (which is measured as a budget constraint residual) inherits measurement error from imputed assets and price of childcare reported by parents

[25] Like for example, the average price paid for childcare for families whose mother is working and who are using relative care but not exhausting it.

[26] Because the identification proof for $\{\lambda_{CB,t}, \lambda_{HB,t}\}$ is conditional on knowing the distribution of assets at each age, I also target moments of the distribution of assets at each age, to ensure that the distribution of assets generated by the model mimics the empirical distribution. More details on Appendix K.

Table 8: IF estimates for parameters governing taste for skills and utility costs of paid care

Parameter	Estimate (TP)	Estimate (SM)
$\lambda_{CB,1}$	6.58	2.15
	(0.04)	(0.01)
$\lambda_{CB,2}$	1.58	0.82
	(0.00)	(0.00)
$\lambda_{CB,3}$	6.25	1.02
	(0.02)	(0.01)
$\lambda_{HB,1}$	4.06	1.29
	(0.02)	(0.01)
$\lambda_{HB,2}$	0.91	0.27
	(0.00)	(0.00)
$\lambda_{HB,3}$	7.18	0.87
	(0.01)	v(0.04)
$\delta_{\theta,1}$	0.92	0.92
	(0.00)	(0.00)
$\delta_{\theta,2}$	1.01	1.01
	(0.00)	(0.00)
$\delta_{\theta,3}$	0.66	0.66
	(0.00)	(0.00)

NOTES:Standard errors calculated using an analytic asymptotic formula in parenthesis. "SOURCE" below refers to the source of the micro-data used in the indirect inference step. δ_{θ_t} is estimated from data on Two-Parent families only, and assumed to be the same for Single Mothers.
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.

6.5 Elasticity of teachers' labor supply

Labor supply elasticities of lead teachers and childcare workers are key to predict how wages of these workers change when changes in regulations change demand for childcare. Identification of these elasticities is not straightforward when one allows for the presence of state-level unobservables that affect labor supply of teachers and childcare workers, due to the usual simultaneity bias. In order to deal with this problem, I instrument state-level lead teacher and childcare worker wages with state-level lagged fertility. Estimates of η_{LT} and η_{CCW} are shown in Table 10. Both elasticities fall in the range estimated by Blau (1993), who estimates elasticities in the 1.15-1.94 range.

7 Equilibrium effects of Staff-to-Child Ratios

With the estimated parameters in hand, I use the model under the estimated parameters to simulate the wages of teachers and children's skills at kindergarten entry under different regulations.

Table 9: Goodness of fit of the IF step.

Choice Probability	Model (TP)	Data (TP)	Model (SM)	Data (SM)
$\mathbb{P}(P_1 = CB)$	0.08	0.08	0.17	0.14
$\mathbb{P}(P_2 = CB)$	0.16	0.16	0.31	0.23
$\mathbb{P}(P_3 = CB)$	0.49	0.57	0.45	0.35
$\mathbb{P}(P_1 = HB)$	0.12	0.17	0.23	0.18
$\mathbb{P}(P_2 = HB)$	0.20	0.18	0.16	0.15
$\mathbb{P}(P_3 = HB)$	0.08	0.08	0.07	0.08
Bunching(CB_1)	14%	11%	NT	NT
Bunching(CB_2)	5%	4%	NT	NT
Bunching(CB_3)	23%	23%	NT	NT
$\mathbb{E}[\tau_1^p]$	778	360	NT	NT
$\mathbb{E}[\tau_2^p]$	1584	501	NT	NT
$\mathbb{E}[\tau_3^p]$	2238	895	NT	NT

NOTES: "SOURCE" below refers to the "Data" column and "NT" indicates statistitcs that were not targeted for Single Mother Households.
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 10: 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

7.1 Effects on teachers' earnings

First, I use the model to explore the effect of regulations on lead teachers' and childcare workers' wages. In particular, I examone the effect on factor prices 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[27]. Table 11 shows the least, average, and most stringent regulations across states in my data for each age and type of care.

Figures 3 and 4 show the equilibrium effects of changing the stringency of regulations on the lead-teacher premium and the earnings of lead teachers and childcare workers for each US Census Region[28].

[27] I define the least (most) stringent staff-to-child ratio for each age and type of care as the minimum (maximum) staff-to-child ratio across all states at each age and type of care. The average regulation at age t for type of care j is given by the average regulation at age t and type of care j across states.

[28] 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. 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 M). Moreover, I am dropping Alaska, Hawaii, and states for which I do not observe any

Table 11: 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	12	15	18
Average, Centers	5.18	10.40	12.12
Average, Homes	4.40	6.50	6.69
Most stringent, Centers	3	7	8
Most stringent, Homes	2	2	2

To ease interpretation, I report the inverse of the staff-to-child ratios, the number of children per adult.

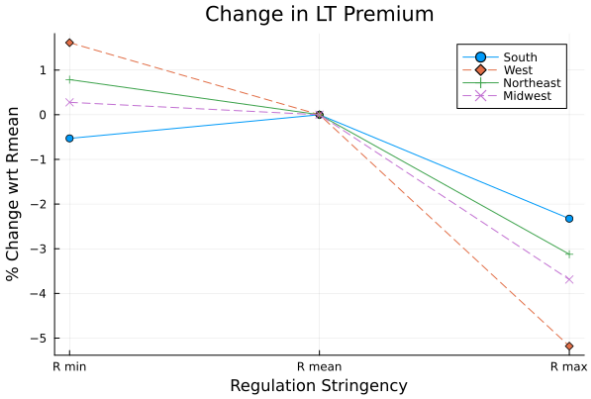


Figure 3: Comparison of average Lead Teacher Premia for different regulation stringencies

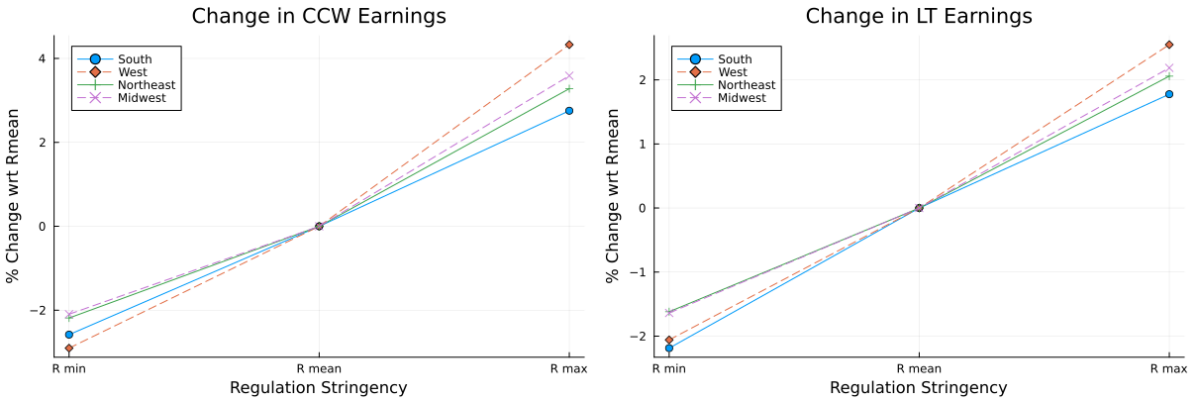


Figure 4: Comparison of Child Care Workers' (left) and Lead Teachers' (right) average earnings for different regulation stringencies

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 non-monotonic in the South, and negative in the other regions. The effect of regulations' stringency on

family with all the necessary observables to simulate their initial state variables.

Given that assumption, states within the same region only differ in their regulations. Because here I am giving all the states the same regulations I can just report the outcomes for each region.

the wage of childcare workers is positive for every region[29]. Going from the average regulation to the most stringent increases childcare worker wages by 2.75 to 4.33%, depending on the region. Moreover, going from the average regulation to the least stringent decreases wages by around 2.1 to 2.9% depending on the region. Despite the fact that more stringent ratios typically decrease the Lead Teacher's *premium*, the effect on lead teachers' *wages* is positive. 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 1.78 %-2.75 % depending on the region, and decreasing the stringency in the same fashion decreases the wages of lead teachers by 1.62 to 2.69 % . The equilibrium effect of regulations on the wages of teachers increases the price of childcare for each level of quality, on top of the partial equilibrium effects that are present for the quality levels for which the regulation is binding after the increase in stringency. Because of this, the price of quality increases for all quality levels, including for quality levels for which the mandatory-minimum ratio does not bind under the most stringent regulations. Despite that, we will see how making the regulation more stringent has mostly positive effects on children's skills.

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 use the model and estimated model parameters to simulate how the skills of children at kindergarten entry would change when going from the most lenient regulation (the lowest mandatory ratio for each age and type of care) to the most stringent one (the highest mandatory ratio for each age and type of care).

Let $F^{Rmin}_{\log\theta,4}(x)$ and $F^{Rmax}_{\log\theta,4}(x)$ be the equilibrium unconditional distribution of cognitive skills at kindergarten entry in the US under the most lenient and the most stringent regulation respectively. Define as $\Delta\log\theta_4(p)$ as the increase in the p-th percentile of the equilibrium unconditional distribution of skills associated with a change in regulations from the most lenient to the most stringent regulation, that is:

$$\Delta\log\theta_4(p) = (F^{Rmax}_{\log\theta,4})^{-1}(p) - (F^{Rmin}_{\log\theta,4})^{-1}(p)$$

Figure 5 plots $\Delta\log\theta_4(p)$ for $p = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99$. From Figure 5 we can see that an increase in the stringency level of regulations is associated with an increase in skills at each percentile of the skill distribution, except for the 1st percentile. In other words, a policymaker who only cares about the unconditional distribution of skills at kindergarten entry prefers very stringent to very lenient mandatory ratios, unless she puts a very high weight on the skills at the 1st percentile.

Given the differences between one-parent and two-parent households, it is to be expected that the skill distribution of children born to single mothers and children born

[29] As previously discussed, this is a quantitative result, and not a theoretical necessity.

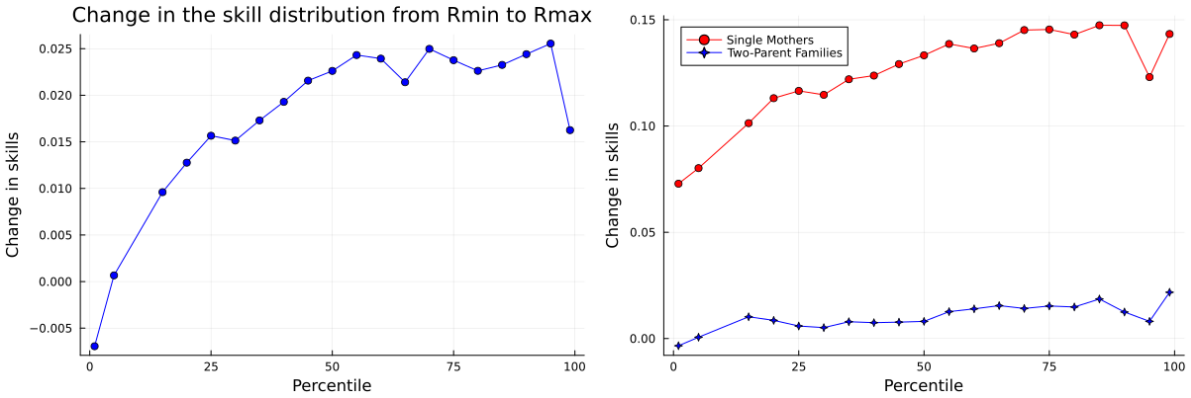


Figure 5: Change associated to going from the least to the most stringent set of regulations in the 1st,10th,20th,30th,40th,50th,60th,70th,80th,90th and 99th percentiles of the overall distribution of skills (left) and for the distribution of skills for children born to Two-Parent and Single-Mother households (right)

to two-parent families will react differently to changes in the stringency of regulations. The right panel of Figure 5 confirms that this is indeed the case. While increases in the stringency of regulations translate into higher skill levels for virtually every percentile[30] of the skill distribution for children born to two-parent families and Single Mothers, the magnitude of the increase is considerably larger for children born to Single Mothers. This is reminiscent of Moschini (2023), who shows that children born to single-parent households are more sensitive to childcare subsidies.

The increase of skills for most skill percentiles seen in Figure 5 can result from increases in skills associated to the regulation change for most children, or from substantial reshuffling of children across the skill distribution. To tell apart the two stories, we need to 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 (measured 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 6 shows that the treatment effects $\Delta \log \theta_{i,4}$ are heterogeneous across children, and more so for children of single mothers than for those of two-parent families, in the sense that the distribution of treatment effects has more mass around 0 for children born to two-parent families. Table 12 illustrates the same point using percentiles of the distribution of $\Delta \log \theta_{i,4}$ for children in two-parent and single-parent households.

Hence, the changes in the overall skill distribution in Figure 5 are the product of some children experiencing large skill gains, and other children experiencing large skill losses. Who are the children who win, and who are the children who lose as a consequence of more stringent regulations? Tables 13 and 14 answer that question.

Tables 13 and 14 detail the pattern of skill redistribution induced by an increase in regulations stringency. First, the children who gain the most are born to the poorest

[30] I qualify this statement by saying "virtually" because there is a very small decrease in skills for the 1st percentile of the skill distribution of children for to two-parent Families, as it can be seen by looking at the right panel of Figure 5.

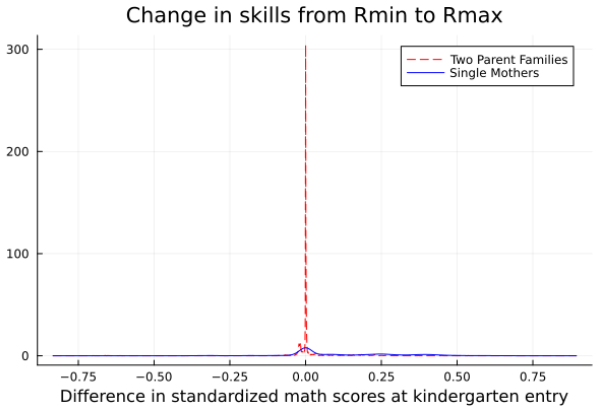


Figure 6: Distribution of GE treatment effects on children skills at pre-school entry. Treatment consists of changing the regulations from the least to the most stringent and letting wages adjust

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

Percentile	Two Parent Families	Single Mothers
1	-62.80 %	-35.89 %
5	-5.20 %	-12.96 %
10	-2.24 %	-5.36 %
25	-0.1 %	0.00 %
50	0.00 %	5.54 %
75	0.18 %	25.20 %
90	4.50 %	41.08 %
95	16.07 %	46.16 %
99	27.88 %	61.38 %

Table 12: 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.

families in terms of wages and assets, and their families also have little access to care provided by relatives. This is true both for the highest gainers among children born to two-parent families and single mothers. Why are these the children who gain the most? As it can be seen in Tables 15 and 16, their families are buying on average lower qualities than the other three groups. This means that after the regulation becomes more stringent the levels of quality that they buy in the benchmark are in a flatter region of the price-quality schedule, which translates into bigger increases in purchased quality of paid care, which is confirmed by Tables 15 and 16. Moreover, these families have on average the lowest endowment of care provided by relatives, so their ability to substitute away from paid care is more limited. In fact, the children of single mothers

Family characteristics of winners and losers for TP families

	$\Delta \log \theta < p_5$	$p_5 \leqslant \Delta \log \theta < 0$	$0 < \Delta \log \theta \leqslant p_{95}$	$\Delta \log \theta > p_{95}$
a_1	270700	400900	157000	130000
w^m	16	18	14	15
w^f	17	22	19	12
$\log q^m$	-0.03	0.05	0.04	-0.02
$\log q^f$	-0.16	-0.04	-0.02	-0.14
$\log q_2^r$	0.57	1.18	1.21	0.46
$\log q_3^r$	0.65	1.27	1.29	0.54
$\overline{T}^r$	3935	4012	4040	3281

Table 13: 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 100 and wages to the nearest dollar.

Family characteristics of winners and losers for SM families

	$\Delta \log \theta < p_5$	$p_5 \leqslant \Delta \log \theta < 0$	$0 < \Delta \log \theta \leqslant p_{95}$	$\Delta \log \theta > p_{95}$
a_1	14400	204600	29600	9500
w^m	10	12	10	9
$\log q^m$	-0.21	-0.10	-0.08	-0.11
$\log q_2^r$	0.07	0.76	0.83	0.67
$\log q_3^r$	0.15	0.84	0.91	0.75
$\overline{T}^r$	4377	4600	3850	2417

Table 14: Average characteristics of winners and losers for SM 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 100 and wages to the nearest dollar.

who experience the highest skill gains increase on average their time in paid care[31].

Let's compare now the children that lose the most and the children who experience moderate skill gains. We can see that both for the case of single mothers and two-parent families, these children are born to families with similar economic resources. However, children who experience the largest skill losses are born to families with much worse non-paid childcare options. Because of this, as it can be seen in Tables 15 and 16, they use a lot of paid childcare in the benchmark with lenient regulations. When regulations become more stringent and the price of childcare goes up they respond by lowering

[31]To see why note that in an interior solution for paid care and relative care τ^P has to satisfy:

$$\frac{\psi^P}{\overline{T}} \log q^P - \frac{\psi^r}{\overline{T}} \log q^r = \frac{P^D(q^P)}{c(\tau^P)} .$$

Hence, if purchased quality increases, τ^P may increase if the increase in the price of quality is not too high. Intuitively, the price-flattening effect of regulations pushes families to buy higher quality, which means that, in the margin, developmental gains of an additional hour of paid care are larger, which may push some families to buy higher paid care if the marginal cost in terms of consumption is not too high.

their demand of hours of paid childcare significantly, and they substitute away to lower-quality relative and parental care. On the other hand, the families of children who experience moderate skill gains are less reliant on paid care in the benchmark. When the regulation becomes more stringent, they respond by demanding less paid care and increasing the quality of paid care when they stay on the market. However, the developmental impact of the reduction in paid care is less severe, because paid care is a lower fraction of their total care time, and because their outside options are of higher quality. Finally, let us look at the children who experience moderate skill losses. From Tables 13 and 14 we can see that those children belong to families who are relatively rich in terms of wages and assets, and the quality of their non-paid childcare arrangements is relatively high. Because their non-paid options are good, they buy relatively fewer hours of paid care compared to the other groups. This means that when the price of paid care goes up after the increase in the stringency of regulations the decrease in skills is less severe. There is a key difference between the children who experience moderate skill losses born to two-parent families and those born to single mothers. By looking at Tables 15 and 16 we can see that the only group for which average log-quality of paid care decreases after the regulation becomes more stringent is the children born to two-parent families experiencing moderate losses. The reason for this is that even childcare of very high quality (for which the regulation does not bind) becomes more expensive due to the equilibrium adjustment of factor prices.

Choices of winners and losers for TP families

	$\Delta \log \theta < p_5$	$p_5 \leqslant \Delta \log \theta < 0$	$0 < \Delta \log \theta \leqslant p_{95}$	$\Delta \log \theta > p_{95}$
$\Delta \tau^P$	-1635	-31	-10	-2
$\Delta \tau^r$	1223	19	-14	-30
$\Delta \tau^m$	397	11	23	24
$\Delta \tau^f$	15	1	1	8
$\Delta \log q_1^P$	0.03	-0.04	0.01	0.13
$\Delta \log q_2^P$	0.38	-0.02	0.31	0.55
$\Delta \log q_3^P$	0.62	0.17	0.74	0.89
$\log q_1^P(R_{min})$	0.76	1.61	0.98	0.43
$\log q_2^P(R_{min})$	2.14	2.83	2.04	1.73
$\log q_3^P(R_{min})$	1.1	1.24	0.84	0.62
$\tau^P(R_{min})$	2590	942	369	3388

Table 15: p5 and p95 are the 5th and 95th percentiles of the treatment effect distribution respectively

8 Conclusion

Mandatory minimum staff-to-child ratios are a common licensing requirement for childcare providers across states. While empirical evidence has shown their impact on the market provision of childcare, their effect on children's skills remains unexplored. This paper addresses this gap by constructing and estimating an equilibrium model of

Choices of winners and losers for SM families

	$\Delta \log \theta < p_5$	$p_5 \leqslant \Delta \log \theta < 0$	$0 < \Delta \log \theta \leqslant p_{95}$	$\Delta \log \theta > p_{95}$
$\Delta \tau^P$	-2138	-595	-235	17
$\Delta \tau^r$	1563	219	56	-207
$\Delta \tau^m$	574	375	179	190
$\Delta \log q_1^P$	0.62	0.42	0.89	1.20
$\Delta \log q_2^P$	0.64	0.64	0.86	1.32
$\Delta \log q_3^P$	1.17	0.96	1.11	1.27
$\log q_1^P(R_{min})$	-0.24	0.19	-0.47	-0.62
$\log q_2^P(R_{min})$	1.12	1.15	1.07	0.74
$\log q_3^P(R_{min})$	0.52	0.58	0.52	0.50
$\tau^P(R_{min})$	3042	1484	2186	3755

Table 16: p5 and p95 are the 5th and 95th percentiles of the treatment effect distribution respectively

the childcare market, where paid providers must adhere to mandatory minimum staff-to-child ratios, and teacher wages adjust accordingly to regulatory changes. Stricter mandatory minimum ratios increase the price of childcare, partly because the upward equilibrium adjustment of teachers' wages. Despite the higher costs, more stringent ratios improve the overall distribution of children's skills at kindergarten entry, with the exception of children at the very lower end of the skill distribution. The stringency of mandatory-minimum ratios affects more the children of single mothers than those of two-parent families, and there are children who experience skill losses and skill gains from more stricter ratios in both groups.

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.

References

Agostinelli, F., Saharkhiz, M. and Wiswall, M. (2025). Home and school in the development of children, *International Economic Review* **66**(2): 527–566.

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.

Ali, U., Herbst, C. M. and Makridis, C. A. (2024). Minimum quality regulations and the demand for childcare labor, *Journal of Policy Analysis and Management* **43**(3): 660–695.

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* .

Arnett, J. (1989). Caregivers in day-care centers: Does training matter?, *Journal of applied developmental psychology* **10**(4): 541–552.

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. (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.

Bodéré, P. (2023). Dynamic spatial competition in early education: An equilibrium analysis of the preschool market in pennsylvania, *Technical report*, Working paper.

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.

Brown, J. H. (2018). *Does public pre-k have unintended consequences on the child care market for infants and toddlers?*, Princeton University, Industrial Relations Section.

Cascio, E. U. (2023). Does universal preschool hit the target? program access and preschool impacts, *Journal of Human Resources* **58**(1): 1–42.

Caucutt, E. M., Lochner, L., Mullins, J. and Park, Y. (2020). Child skill production: Accounting for parental and market-based time and goods investments, *Technical report*, National Bureau of Economic Research.

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., 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.

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.

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.

Fort, M., Ichino, A. and Zanella, G. (2020). Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families, *Journal of Political Economy* **128**(1): 158–205.

Freyberger, J. (2021). Normalizations and misspecification in skill formation models, *arXiv preprint arXiv:2104.00473* .

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.

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* .

Helburn, S. and Howes, C. (1996). Child care cost and quality: The future of children, *Financing Child Care* **6**(2): 5.

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.

Kottelenberg, M. J. and Lehrer, S. F. (2017). Targeted or universal coverage? assessing heterogeneity in the effects of universal child care, *Journal of Labor Economics* **35**(3): 609–653.

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.

Williams, B. (2020). Identification of the linear factor model, *Econometric Reviews* **39**(1): 92–109.

Appendices

A Proofs of Lemma 1

Before proving Lemma 1, it is useful to establish some properties of the conditional labor demands and the cost function.Let $h = \tau k$ be the total hours of care in the classroom. Because of constant returns to scale on quality production we can write:

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

Lemma 2 (Factor demands)**.**

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

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

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. This comes out of the unique solution to the KKT necessary conditions □

Lemma 3. *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 2.

Proof. The cost function is the value of the cost-minimization problem. 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 2. □

A.1 Proof of Lemma 1

The expression for the price-schedule in Lemma 1 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 1 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 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 $P(q) = c(q)$.

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 $t = 1,2,3$*

$$V^{SM}_t(a_t, w^m, q^m, q^r_t, \overline{T}^r, \theta_t, o^{CB}_t, o^{HB}_t) = \hat{V}^{SM}_t(a_t, w^m, q^m, q^r_t, \overline{T}^r, o^{CB}_t, o^{HB}_t) + \Gamma^\theta_t \log \theta^\theta_t$$

with

$$\Gamma^\theta_4 = \delta_{\theta,4} \text{ and } \Gamma^\theta_t = \beta \gamma_{\theta,t} \Gamma^\theta_{t+1} + \delta_{\theta,t}$$

Proof. We proceed by backward induction. First, we establish the result for $t = 3$. Substituting (PF) in the RHS of the value function equation:

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

by letting

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

and

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

we prove the result for $t = 3$. The induction step proceeds in a similar fashion with

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

and

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

□

C Definition of Equilibrium

In this appendix, 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* $\overline{C}^S, \overline{E}^S$, *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 1.*
- *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 2*
- $\overline{C}^S, \overline{E}^S$ *are consistent with Supply of* C *and Supply of* E *given wages, that is:*

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

- $g_{\omega,t}^{FT,N}(a_t, H)$ *belong to the argmax of:*

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

  and $g_{\omega,t}^{FT,D}(a_t, H)$, $D = CB, HB$ *belongs to the argmax of:*

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

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

- $\tilde{V}^{FT,D}(a_t,H)$ *solve* $\tilde{V}^N$ *and* $\tilde{V}^P$.
- *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:*

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

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

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

- *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:*

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

- $q \in Q$ *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:*

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

D 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 4450 families for which the wage of the mother is missing, 1100 are missing because the mother does not participate in the labor force in any wave. This means that of the approximately 10300 observations in my sample[32], approximately half of those have missing mother's wages, but approximately only around 10.7 % are missing due to non-participation in the labor force. For fathers, of the 1350 two-parent families for which the wage of the father is missing, only 50 are due to the father not working in any wave[33]. 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 et al. (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[34]. 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 of the regression used in imputation is 34% for mothers and 36% for fathers.

[32]Because of recent changes in the Department of Education, I have no access to ECLS-B data. The approximate total number of observations in my sample has (children who are not part of a multiple birth and whose parent respondent is the biological mother), has been calculated using the DataLab tool of the NCES (see https://nces.ed.gov/datalab/)

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

[34]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

E 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 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[35], whether they any assets from a list of risky assets[36], 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, and 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[37] 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 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 as:

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

The R^2 of the linear prediction for assets is 67%, which implies moderate f measurement error in $\tilde{a}_{i,t}$. 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.

F Measuring minimum ratios for home-based care

A few features institutional features make reducing the mandatory minimum ratio for home-based care providers to a number that can be inputed into the model challenging. First, home-based care providers that provide to a number of children below a certain threshold are not subject to mandatory licensing (and hence not subject to mandatory minimum ratios). Second, home-based care providers with one staff member (from now on Small Homes) sometimes have different mandatory minimum ratios than their

[35]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?"

[36]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?"

[37]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.

counterparts with more than one staff member (from now on Large Homes). Third, because home-based care providers can mix children of different ages in the same room, the mandatory-minimum ratios usually depends on the age distribution of children. As explained in the body of the text, the way that I deal with these complications is by calculating the mandatory-minimum ratio for homes that would apply if mixing children of different ages in the same room was outlowed (as the model imposes) and by picking the configuration of number of staff members and children that minimizes the mandatory minimum ratio (which by the constant returns to scale assumption and perfect competition would be the only one that survives in the equilibrium of my model). Here I explain how to do this.

First, suppose that the threshold number of children for home-based care providers in state l is $\underline{N}_l^{\text{Licensing}}$. That means that the minimum staff-to-child ratio that a home-based care provider operating without a license in state l can have is $1/\underline{N}_l^{\text{Licensing}}$. Second, suppose that the mandatory-minimum ratio for Small Homes providing only to age t in state l is $\overline{R}_{l,t}^{\text{Small Homes}}$ (note that this is just a number and it does not depend on the age distribution of children in the room, because we are assuming that the age distribution is degenerate at age t). Finally, suppose that the the mandatory-minimum ratio for Large Homes providing only to age t in state l is $\overline{R}_{l,t}^{\text{Large Homes}}$. Then, the most lenient mandatory minimum ratio a home-based care provider in state l providing to children of age t can achieve is:

$$\underline{R}_{l,t}^{HB} = \min\left\{1/\underline{N}_l^{\text{Licensing}}, \overline{R}_{l,t}^{\text{Small Homes}}, \overline{R}_{l,t}^{\text{Large Homes}}\right\},$$

which is the mandatory-minimum staff-to-child ratio for home-based providers that I input into the model.

G 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:

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

The following proposition establishes identification of $\overline{T}(Z^{T,r})$:

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

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

$$\textit{for all } (a_1, w^m, w^f, q^r, Z^{T,r}) \in \text{supp}(a_1, w^m, w^f, q^r, Z^{T,r})\,.$$

Then

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

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

$$c_j^* = \sup_{\tilde{x} \in \mathcal{X}_{-c}} \left\{ \tilde{V}^{TP,j}(\tilde{x}) - \tilde{V}^{TP,N}(\tilde{x}) \right\} \text{ for } j = CB, HB$$

where

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

And note that $c^{j,*}$ is finite because $\mathcal{X}_{-c}$ 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

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

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

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

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, at state x the optimal τ^m_1 and τ^f_1 satisfy:

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

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 $\tau^m_1 > \frac{\overline{T} - \overline{T}^r(Z^{T,r})}{2}$, this improves the objective of the Twp-parent family, which contradicts τ^m_1, τ^f_1 being optimal. A similar contradiction argument applies if $\tau^f_1 > \frac{\overline{T} - \overline{T}^r(Z^{T,r})}{2}$, so it's optimal to exhaust relative care for all $x \in \mathcal{X}$ □

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.

G.1 Estimation

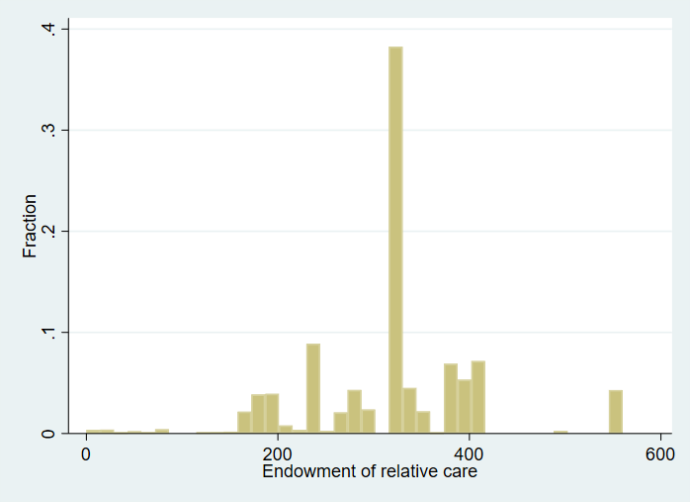
The identification argument above implies that $\overline{T}^r(Z^{T,r} = z)$ can be estimated as the maximum of hours of relative care conditional on $Z^{T,r} = z$. I include in $Z^{T,r}$ the discretized age of the maternal grandmother, the discretized age of the maternal grandfather and the number of adult relatives living in the same household as the child. The estimated distribution of relative care monthly endowments is given by Figure 7.

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

Cost-minimization for a childcare provider with quality above q^* implies:

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

Figure 7: 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.

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 for hours of care and her efficiency units:

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

Hence, the hourly remuneration of 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. [38] Now, we can calculate $\alpha_{E,i,t}$ for each child i for which direct observations of paid care providers are conducted

[38] 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

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:

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

Estimates can be found in the body of the text.

I Measurement systems and production function of 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:

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

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

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

I.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:

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

Note that I am allowing measurement system parameters to vary with child's age.

I.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:

$$Arnett^{p}_{i,t} = \mu^{Arnett}_{t} + \alpha^{Arnett}_{t}\Big(\log A^{HB}_{t} + 1(D_{i,t} = CB)(\log A^{CB}_{t} - \log A^{HB}_{t}) + InpComp_{i,t}\Big) + \epsilon^{Arnett}_{t}$$

with

$$InpComp_{i,t} = \alpha_{E,t} \log\left(\frac{\hat{E}_{i,t}}{k_{i,t}}\right) + (1 - \alpha_{E_{i,t},t}) \log\left(\frac{\hat{C}_{i,t}}{k_{i,t}}\right) \ .$$

This means that α_t^{Arnett} is identified from a regression of the Arnett score on a constant, a dummy for center-based, and the input composite term. It is useful to define:

$$\widehat{Arnett}_{i,t} = \mu_t^{Arnett} + \alpha_t^{Arnett}\left(\log A_t^{HB} + 1(D_{i,t} = CB)(\log A_t^{CB} - \log A_t^{HB}) + InpComp_{i,t}\right) .$$

Note that we can construct the input composite term because $\alpha_{E,t}$ is identified from the previous section, $\hat{C}$ and k are the average number of adults and children in the classroom respectively (as in the previous section), and $\hat{E}$, 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. Estimates of the Arnett loadings are shown in Table 17

Table 17: 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.

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

See Appendix I.5

I.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 $X_{q,r}$:

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

It will be useful to define:

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

Table 18 reports The estimated dependence between $\widehat{Arnett}_{i,t}^r$ and $X_{q,r}$[39].

[39]Because quality of relative care is measured for relatively few families, I use the relation implied by the FOC for families in an interior solution for quality to construct the Arnett score for those families whenever the Arnett score for relative care quality is not observed

Table 18: Observable predictors of relative care quality

Variable	$\widehat{\text{Arnett}}_2^r$	$\widehat{\text{Arnett}}_3^r$
Grandmother High-School	0.23	0.18
	(0.11)	(0.15)
Grandmother College	0.45	0.39
	(0.22)	(0.29)
Grandmother Postgraduate	0.47	0.45
	(0.27)	(0.31)
Grandfather High School	0.24	0.04
	(0.11)	(0.16)
Grandfather College	0.34	0.45
	(0.19)	(0.24)
Grandfather Postgraduate	0.40	0.39
	(0.22)	(0.35)
Grandfather missing education	-0.16	-0.34
	(0.17)	(0.24)
Constant	-0.40	-0.56
	(0.09)	(0.11)

Table 19: 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.

In the ECLS-B, the Arnett score is not available at 9 months. Hence, modeling heterogeneity in relative care quality as in waves 2 and 3 is not possible. To allow for heterogeneous relative care qualities across 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.

I.3 Parental care

I.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:

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

Where each A_i denotes the answer to questions about how often the parent did the following things last month: Play with child using toys for building things, like Lincoln Logs or Legos, help the child to bed, help the child bathe, take the child for a walk or to play outside, help the child to get dressed, 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 and 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 [40]. 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:

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. Table 21 summarizes the questions that I use, how I transform them to 0-1 variables, and their weight. Correlations between the two measures of maternal and paternal care quality are shown in Table 22.

[40] In the Two Bags Task the parent and the child have to play with two sets of toys. The interaction is recorded, observed and coded. Different features of the observed behavior or mood of parent and child are coded separetaly. I use the Sensitivity, Positive Regard, Intrusiveness, and Stimulation of Cognitive Development in the 2 years old assessment, and the Emotional Support, Stimulation of Cognitive Development, Intrusiveness, and Detachment parental scales in the 4 years old assessment. I do not use the Negative Regard Scale at either age because it intends to capture the degree of negative perception that the child has of the parent, as opposed to a behaviour or mood that the parent presents during the interaction. A more detailed discussion of the Two Bags Task is in Andreassen and Fletcher (2007).

Table 20: 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

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

We have two measures for maternal and paternal care qualities:

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

with $\mu^{j,1} = 0$ and $\alpha^{j,1} = 0$ for $j = m, f$. We can estimate $\mu^{j,2}$ $\alpha^{j,2}$ for $j = m, f$. If there is an instrument Z uncorrelated with measurement error in parental care qualities then:

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

and hence:

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

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 23. 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 23

Table 21: 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 (-)

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 22: Correlations between measures of parental care

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

I.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 parental care with other states is likely to be very noisy, so 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

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

Table 23: Estimates for measurement parameters of parental quality

$\alpha_{m,2}$	$\alpha_{f,2}$	$\mu_{m,2}$	$\mu_{f,2}$
5.36	2.10	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.

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 $\widetilde{\log q}^{j,1}$. Estimates of $\beta_{j,q}$ for $j = m, f$ for 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 $X_{j,q}$ are presented in Table 5 in the body of the text. The variance of $\nu_{j,q}$ is identified from:

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

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

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

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 4 in the body of the text.

To obtain estimates of the variance of skill shocks we need estimates of the variance of the measurement error of the parental quality measures. To identify those, note that:

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

Substituting these expressions by sample analogues yields the estimates in Table 24. Table 24 also reports estimates of Signal-to-Noise ratios.

Table 24: Measurement error variances of parental quality measures and Signal-to-Noise Ratios

$Var(\epsilon^{m,1})$	$Var(\frac{\epsilon^{m,2}}{\alpha^{m,2}})$	$Var(\epsilon^{f,1})$	$Var(\frac{\epsilon^{f,2}}{\alpha^{f,2}})$	$s^{m,1}$	$s^{m,2}$	$s^{f,1}$	$s^{f,2}$
0.90	0.18	0.97	1.32	0.08	0.29	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.

I.4 Cognitive skills of children

I.4.1 Identification of the measurement system of cognitive skills

Remember the measurement system for cognitive skills at time t:

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

Because cognitive skills are latent, their scale and location are not identified. I normalize:

$$\mu^{1}_{\theta,t} = 0 \text{ and } \alpha^{1}_{\theta,t} = 1$$

for all t. I choose these "normalized" measures to be the mental scale score at $t = 1, 2$ and the math score at $t = 3, 4$:

Identification of measurement parameters for skills is similar to identification of measurement parameters for parenting quality. Hence, $\{\alpha^{2}_{\theta,t}\}^{4}_{t=1}$ are identified from:

$$\alpha^{2}_{\theta,t} = cov(\widetilde{\log\theta}^{2}_{t}, Z_t)/cov(\widetilde{\log\theta}^{1}_{t}, Z_t) \, ,$$

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 $\widetilde{\log\theta}^{1}_{t}$ 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 $\alpha^{2}_{\theta,t}$: Once the $\alpha^{2}_{\theta,t}$ are identified for each t, we can compute the

Table 25: Estimates for the loading of the second measure of cognitive skills

$\alpha^{2}_{\theta,1}$	$\alpha^{2}_{\theta,2}$	$\alpha^{2}_{\theta,3}$	$\alpha^{2}_{\theta,4}$
0.78	0.74	0.88	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:

$$Var(\log\theta_t) = cov(\widetilde{\log\theta}^{1}_{t}, cov(\widetilde{\log\theta}^{2}_{t})/\alpha^{2}_{\theta,t} \, .$$

Estimates of the variance of $\log\theta_t$ can be found in Table 26. The variance of measurement error for the first measure at each t can be computed 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 $\widetilde{\log\theta}_t^1$ is standardized, and hence has a variance of 1. Substituting these expressions by their sample analogues yields the estimates in Table 27. Given 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:

$$s_t^\theta = \frac{Var(\log\theta_t)}{Var(\log\theta_t) + Var(\epsilon_{\theta,t}^1)} .$$

Estimates of signal-to-noise ratios can be found in Table 27. These signal-to-noise ratios are high, which speaks to the high quality of the ECLS-B data.

Table 26: Estimates for the Variance of $\log\theta_t$

$Var(\log\theta_1)$	$Var(\log\theta_2)$	$Var(\log\theta_3)$	$Var(\log\theta_4)$
0.91	0.64	0.89	0.85

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 27: Estimates for the Variance of $\epsilon_{\theta,t}^1$ and Signal-to-Noise ratios

$Var(\epsilon_{\theta,1}^1)$	$Var(\epsilon_{\theta,2}^1)$	$Var(\epsilon_{\theta,3}^1)$	$Var(\epsilon_{\theta,4}^1)$	s_1^θ	s_2^θ	s_3^θ	s_4^θ
0.09	0.36	0.11	0.15	0.91	0.61	0.88	0.85

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.4.2 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 run 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 all the p-values are above 0.5. This is not because of skills being imprecisely measured, since the estimated signal-to-noise ratio in the first wave is very high, and since I include birthweight in the same regression and 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.

I.5 Skills' PF , 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 and re-arranging we get:

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

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}. 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 latent factor 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:[41] [42]

$$\mathbb{E}\Big[Z(\widetilde{\log\theta_{t+1}}^1 - \log A_t + \gamma_{\theta,t}\log\theta_t + \gamma_{m,t}\frac{\tau_t^m}{\bar{\tau}}\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}} + \gamma_{f,t}^j\frac{\tau_t^f}{\bar{\tau}}\frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}} +$$
$$\frac{\gamma_{p,t}}{\alpha_t^{Arnett}}\frac{\tau^p}{\bar{\tau}}\widehat{Arnett}_t^p + \frac{\gamma_{r,t}}{\alpha_t^{Arnett}}\frac{\tau^r}{\bar{\tau}}\widehat{Arnett}_t^r - \frac{\gamma_{p,t}}{\alpha_t^{Arnett}}\frac{\tau_t^p}{\bar{\tau}}\mu_t^{Arnett} - \frac{\gamma_{r,t}}{\alpha_t^{Arnett}}\frac{\tau^r}{\bar{\tau}}\mu_t^{Arnett})\Big] = 0\ ,$$

with

$$Z = (1, \widetilde{\log\theta_{t+1}}^2, \frac{\tau_t^m}{\bar{\tau}}\widetilde{\log q}^{m,1}, \frac{\tau_t^f}{\bar{\tau}}\widetilde{\log q}^{f,1}, \frac{\tau^p}{\bar{\tau}}\widehat{Arnett}_t^p, \frac{\tau^r}{\bar{\tau}}\widehat{Arnett}_t^r, \frac{\tau_t^p}{\bar{\tau}}, \frac{\tau_t^r}{\bar{\tau}})'\ .$$

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

$$\mathbb{E}[\log q^m] - \mathbb{E}[\log q_t^r] = \mathbb{E}[\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}}] - (\mathbb{E}[Arnett_t] - \mu^r) = \mathbb{E}[\widetilde{\log q}^{m,1}] - (\mathbb{E}[Arnett_t] - \mu^r) = \mu^r$$

Hence, allocating one hour of care from a mother of average quality to a relative of average quality is associated with a change in skills tomorrow of $-\frac{1}{\bar{\tau}}\mu_t^{Arnett}$.

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}.

[41] 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 $\frac{\tau_t^r}{\bar{\tau}}$ by the coefficient of $\frac{\tau^r}{\bar{\tau}}\widehat{Arnett}_t^r$

[42] I am using $\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}}$ in estimation instead of $\widetilde{\log q}^{j,1}$ because $\frac{\widetilde{\log q}^{j,2} - \mu^{j,2}}{\alpha^{j,2}}$ has more variation.

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. Since $\widetilde{\log q}^{m,1}$ and $\widetilde{\log q}^{f,1}$ are constructed using the same questions, so it makes sense to impose $\gamma_{m,t} = \gamma_{f,t}$. By the same token, it makes sense to impose: $\gamma_{r,t} = \gamma_{p,t}$. The estimation equation then simplifies to:

$$\widetilde{\log\theta}^1_{t+1} = \log A_t + \gamma_{\theta,t}\widetilde{\log\theta}^1_t + \gamma_{par,t}\left(\frac{\tau^m_t}{\overline{\tau}}\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}} + \frac{\tau^f_t}{\overline{\tau}}\frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}}\right) + \frac{\gamma_{nonpar,t}}{\alpha^{Arnett}_t}\left(\frac{\tau^p}{\overline{\tau}}\widehat{Arnett}^p_t + \frac{\tau^r}{\overline{\tau}}\widehat{Arnett}^r_t\right) - \frac{\gamma_{nonpar,t}}{\alpha^{Arnett}_t}\frac{\tau^r_t + \tau^p_t}{\overline{\tau}}\mu^{Arnett}_t + \eta_{t+1} + \epsilon^1_{\theta,t+1} + \gamma_{\theta,t}\epsilon^1_{\theta,t} - \gamma_{par,t}\left(\frac{\tau^m_t}{\overline{\tau}}\frac{\epsilon^{m,2}}{\alpha^{m,2}} + \frac{\tau^f_t}{\overline{\tau}}\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right),$$

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 in the first wave, so the Arnett score for non-parental childcare is not available in that wave. If 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 $t = 1$:

$$\widetilde{\log\theta}^1_2 = \log A_1 + \gamma_{\theta,1}\widetilde{\log\theta}^1_1 + \gamma_{par,1}\left(\frac{\tau^m_t}{\overline{\tau}}\frac{\widetilde{\log q}^{m,2} - \mu^{m,2}}{\alpha^{m,2}} + \frac{\tau^f_t}{\overline{\tau}}\frac{\widetilde{\log q}^{f,2} - \mu^{f,2}}{\alpha^{f,2}}\right) + \frac{\gamma_{nonpar,1}}{\alpha^{Arnett}_2}\frac{\tau^r_1}{\overline{\tau}}\widehat{Arnett}^r_2 - \frac{\gamma_{nonpar,1}}{\alpha^{Arnett}_2}\frac{\tau^r_1}{\overline{\tau}}(\mu^{Arnett}_2 + \alpha^{Arnett}_2 g^{q,r}_1) + \eta_2 + \epsilon^1_{\theta,2} + \gamma_{\theta,1}\epsilon^1_{\theta,1} + \gamma_{par,1}\left(\frac{\tau^m_1}{\overline{\tau}}\frac{\epsilon^{m,2}}{\alpha^{m,2}} + \frac{\tau^f_1}{\overline{\tau}}\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right).$$

Note that because μ^{Arnett}_2 is identified, and using similar arguments to the ones before, $g^{q,r}_1$ 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 $\tau^p_1 > 0$, because then we would have an unobservable factor problem, where the unobservable production factor that would be given by the parental investment composite $\frac{\tau^p_1}{\overline{\tau}}\log q^p_1$ (if $\tau^p_1 = 0$ that factor is 0).

Table 28: Estimates of Arnett measurement system additive shifter and the growth rate of relative care quality between $t = 1$ and $t = 2$

$g^{q,r}_1$	μ^{Arnett}_2	μ^{Arnett}_3
1.34	-0.43	-0.39
(12.61)	(0.73)	(1.04)

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.

We also need variance of η_{t+1}. The estimation equation residual is:

$$\text{residual}_t = \eta_{t+1} + \epsilon^1_{\theta,t+1} + \gamma_{\theta,t}\epsilon^1_{\theta,t} - \gamma_{par,t}\left(\frac{\tau^m_t}{\overline{\tau}}\frac{\epsilon^{m,2}}{\alpha^{m,2}} + \frac{\tau^f_t}{\overline{\tau}}\frac{\epsilon^{f,2}}{\alpha^{f,2}}\right).$$

Taking the expectation of the square of the estimation residual we get:

$$\mathbb{E}[\text{residual}_t^2] = \sigma^2_{\eta,t} + Var(\epsilon^1_{\theta,t+1}) + \gamma^2_{\theta,t} Var(\epsilon_{\theta,t}) + gamma^2_{par,t}\left(\mathbb{E}\left[\left(\frac{\tau^m_t}{\overline{\tau}}\right)^2\right] Var(\frac{\epsilon^{m,2}}{\alpha^{m,2}}) + \mathbb{E}\left[\left(\frac{\tau^f_t}{\overline{\tau}}\right)^2\right] Var(\frac{\epsilon^{f,2}}{\alpha^{f,2}})\right) .$$

Substituting the population objects in this expression by sample counterparts we get the estimates in Table 29.

Table 29: Estimates of the variance of the skill shocks in each period

$\sigma^2_{\eta,2}$	$\sigma^2_{\eta,3}$	$\sigma^2_{\eta,4}$
0.42	0.53	0.23
(0.06)	(0.05)	(0.09)

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.

J Identification and estimation of private providers' TFP

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 I.1.1 the following relation:

$$Arnett^p_{i,t} = \mu^{Arnett}_t + \alpha^{Arnett}_t\left(\log A^{HB}_t + 1(D_{i,t} = CB)(\log A^{CB}_t - \log A^{HB}_t\right) + \quad \text{(A inputs)}$$
$$\alpha^{Arnett}_t InpComp_{i,t} + \epsilon^{Arnett}_t .$$

Because α^{Arnett}_t from Appendix I.1.1 and μ^{Arnett}_t is identified from I.5, $\log A^{HB}_t$ 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 $\log A^{CB}_t$ is identified from 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:

$$\log q^p_t = 1 + \frac{\gamma_{r,t}}{\gamma_{p,t}} \log q^r_t .$$

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 1, 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:

$$P(q_{i,t}^P) = \left[w^E \left(\frac{\alpha_E}{1-\alpha_{E,t}^j} \frac{w^C}{w^E} \right)^{1-\alpha_E^j} + w^C \left(\frac{1-\alpha_E^j}{\alpha_{E,t}^j} \frac{w^E}{w^C} \right)^{\alpha_E} \right] \frac{1}{A} q_{i,t}^P .$$

The fact that $\epsilon_{i,t}^{P,q}$ is classical and mean zero implies :

$$\mathbb{E}[\tilde{P}_{i,t} | \frac{C}{h} > \underline{R}_l, 0 < \tau_t^r < \overline{\tau}^r] = \mathbb{E}\left[w^E \left(\frac{\alpha_E}{1-\alpha_{E,t}^j} \frac{w^C}{w^E} \right)^{1-\alpha_E^j} + w^C \left(\frac{1-\alpha_E^j}{\alpha_E} \frac{w^E}{w^C} \right)^{\alpha_E} q_{i,t}^P | \frac{C}{h} > \underline{R}_l, 0 < \tau_t^r < \overline{\tau}^r \right] \frac{1}{A} .$$

(A prices)

And note that everything but A is a parameter that was identified before or data (given that $q_{i,t}^P$ is known for each family that buys quality high enough with interior relative care as argued before) Estimates for these TFP terms using price variation can be obtained by replacing population expectations by sample expectations in equation A prices. Estimates for the TFP terms using information in inputs and measured quality can be obtained by estimating equation A inputs.

K Identification of preference parameters

Here I present identification proofs for $(\delta_l^m, \delta_l^f, \delta_\tau^m, \delta_\tau^f, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a)$ and $\{\lambda_{CB,t}, \lambda_{HB,t}\}_{t=1}^3$. Before stating and proving the identification result, 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:

$$\Delta_{\theta,i,t}^{j,j'} = \frac{1}{\overline{\tau}} \left(\gamma_j \log q_i^j - \gamma_{j'} \log q_{i,t}^{j'} \right) .$$

Because $\log q^m$ and $\log q^f$ are measured with mean-independent error, 0-mean error, so is $\Delta_{\theta,i,t}^{j,j'}$ for $j = m, f$ and $j' = m, f, r, p$:

$$\tilde{\Delta}_{\theta,i,t}^{j,j'} = \Delta_{\theta,i,t}^{j,j'} + \frac{1}{\overline{\tau}} \gamma_{j,t} \epsilon_{i,t}^j := \Delta_{\theta,i,t}^{j,j'} + \epsilon_{i,t}^{\Delta\theta_{j,j'}} \text{ for } j = m, f \text{ and } j' = r, p ,$$

and

$$\tilde{\Delta}_{\theta}^{j,j'} = \Delta_{\theta}^{j,j'} + \frac{1}{\overline{\tau}} \gamma_{j,t} \epsilon^j - \frac{1}{\overline{\tau}} \gamma_{j',t} \epsilon^{j'} := \Delta_{\theta,i,t}^{j,j'} + \epsilon_{i,t}^{\Delta\theta_{j,j'}} \text{ for } j, j' = m, f \text{ and } j \neq j' ,$$

and note that I am omitting the superindex denoting the measure of parental care for convenience (for each measure of parental care the previous equations still holds). Assets a_t are imputed (see section E), so they are also measured with error, that is:

$$\tilde{a}_{i,t} = a_{i,t} + \epsilon_{i,t}^a .$$

I assume that $\epsilon^a_{i,t}$ is independent of household's choices and states[43].

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:

$$\tilde{P}_{it} = P^{D_{it}}(q^P_{i,t}) + \epsilon^P_{i,t} \, ,$$

where $D_{i,t} = CB, HB$ is the choice of paid care provider, $q^P_{i,t}$ is the quality of paid care chosen by family i at age t, and $\epsilon^P_{i,t}$ 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.

$$\tilde{c}_{i,t} = w^m_i n^m_{i,t} + w^f_i n^f_{i,t} + \tilde{a}_{i,t}(1+r) - \tilde{a}_{i,t} - \tilde{P}_{i,t}\tau^p_t \, .$$

Hence, we can write the noisy measure of consumption as:

$$\tilde{c}_{i,t} = c_{i,t} + \epsilon^c_{i,t} \text{ , where } \epsilon^c_{i,t} = (1+r)\epsilon^a_{i,t} - \epsilon^a_{i,t+1} - \tau^P_{i,t}\epsilon^P_{i,t} \, .$$

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.*

[43] 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 identification of the initial distribution of states requires full independence.

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 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 $n_{i,t}^m > 0$ and re-arranging we get:

   $$\delta_l^m = \frac{\mathbb{E}[w_i^m l_{i,t}^m | n_{i,t}^m > 0]}{\mathbb{E}[\tilde{c}_{i,t} | n_{i,t}^m > 0]} \, .$$

   We can similarly establish identification of δ_l^f using the distribution of $(\tilde{c}_{i,t}, w_{i,t}^f, l_{i,t}^f)$.

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 $\tau_t^r \leqslant \overline{T}^r$ are 0. Hence, from the FOC for τ^r we get that, in an interior solution for relative care

   $$\frac{\Psi_t^r}{\overline{T}} \log q_t^r = \frac{\beta \Gamma_{t+1}^\theta \gamma_{r,t}}{\overline{T}} \log q_t^r = \lambda_t^{SC} \, .$$

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 $\mu_\tau^p = 0$) the following relationship has to hold:

$$\beta \Gamma_{t+1}^\theta \Delta_t^{p,r} = \delta_l^m \frac{P_t(q_t^P)}{w^m l_t^m} ,$$

where I'm using the definition of $\Delta_t^{p,r}$ and the definition of Ψ_t^p, and I am omitting the type of care superscript (D) for convenience. Re-arranging and re-writing in terms of observables and measurement error:

$$\tilde{P}_{i,t} = \frac{\beta \Gamma_{t+1}^\theta}{\delta_l^m} \Delta_{i,t}^{p,r} + \epsilon_{i,t}^P .$$

Taking conditional expectations and re-arranging we get:

$$\frac{\beta \Gamma_{t+1}^\theta}{\delta_l^m} = \frac{\mathbb{E}[\tilde{P}_{i,t} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^m > 0]}{\mathbb{E}[w_i^m l_{i,t}^m \Delta_{i,t}^{p,r} | 0 < \tau_{i,t}^r < \overline{T}_i^r, \tau_{i,t}^p > 0, n_{i,t}^m > 0]} .$$

Because δ_l^m is identified and β is known, Γ_4^θ and Γ_3^θ are identified from the distribution of $(w_i^m, l_{i,t}^m, \Delta^{p,r}, \tilde{P}_{i,t})$ given $0 < \tau_{i,t}^r < \overline{T}_i^r$, $\tau_{i,t}^p > 0$, and $n_{i,t}^m > 0$ for $t = 2, 3$. Since $\Gamma_4^\theta = \delta_{\theta,4}$, it follows that $\delta_{\theta,4}$ is identified. Moreover, since $\Gamma_3^\theta = \beta \gamma_{\theta,3} \Gamma_4^\theta + \delta_{\theta,3}$, β is known, $\gamma_{\theta,3}$ is identified (see Appendix I.5) and $\Gamma_3^\theta, \Gamma_4^\theta$ are identified, it follows that $\delta_{\theta,3}$ is identified. A similar argument applies if the distribution of $(w_i^f, l_{i,t}^f, \Delta^{p,r}, \tilde{P}_{i,t})$ given $0 < \tau_{i,t}^r < \overline{T}_i^r$, $\tau_{i,t}^p > 0$, and $n_{i,t}^f > 0$ is observed.

3. **Identification of** $\delta_\tau^m, \delta_\tau^f$: 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:

$$\frac{\delta_\tau^m}{\tau^m} + \beta \Gamma_{t+1}^\theta \Delta_\theta^{m,r} = \frac{\delta_l^m}{l^m} ,$$

and note that I am using the definition of Ψ^m, Ψ^r and $\Delta_\theta^{m,r}$. Re-arranging and re-writing in terms of observables and measurement error:

$$\frac{\delta_\tau^m}{\delta_l^m} = \frac{\tau_{i,t}^m}{l_{i,t}^m} - \frac{\beta \Gamma_{t+1}^\theta}{\delta_l^m} \tilde{\Delta}_{\theta,i,t}^{m,r} \tau_{i,t}^m + \frac{\beta \Gamma_{t+1}^\theta}{\delta_l^m} \tau_{i,t}^m \epsilon_{i,t}^{\Delta \theta_{m,r}} .$$

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 shows 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 and re-writing in terms of observables and measurement error:

$$\tilde{a}_{i,4} = \beta\delta_a\tilde{c}_{i,3} + \epsilon^a_{i,4} - \beta\delta_a\epsilon^c_{i,3} .$$

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.

□

Some comments are in order. First, the identification result only requires observing wages for parents that are working. Hence, imputing wages for some parents is immaterial for identification of preference parameters discussed so far. Second, 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, $(\delta^m_l, \delta^f_l, \delta^m_\tau, \delta^f_\tau, \{\delta_{\theta,t+1}\}_{t=1}^3, \delta_a)$ are overidentified. Third, the assumptions:

$$\mathbb{E}[\tilde{c}_{i,t}|n^j_{i,t} > 0] \neq 0, \mathbb{E}[w^j_i l^j_{i,t}\Delta^{p,r}_{\theta,i,t}|0 < \tau^r_{i,t} < \overline{T}^r_i, \tau^p_{i,t} > 0, n^j_{i,t} > 0] \neq 0, \mathbb{E}[\tilde{c}_{i,1}\tilde{c}_{i,3}] \neq 0 ,$$

are testable. Moreover, the assumptions

$$\mathbb{E}[\tilde{c}_{i,t}|n^j_{i,t} > 0] \neq 0, \mathbb{E}[w^j_i l^j_{i,t}\Delta^{p,r}_{\theta,i,t}|0 < \tau^r_{i,t} < \overline{T}^r_i, \tau^p_{i,t} > 0, n^j_{i,t} > 0] \neq 0 ,$$

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 $\Delta^{p,r}_{\theta,i,t}$ 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 $\{\lambda^{CB}_t, \lambda^{HB}_t\}_{t=1}^3$. The proposition is written for two-parent families. 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

$$\{a_t, H : \tilde{V}^D_t(a_t, H) > \tilde{V}^N_t(a_t, H)\} \textit{ for } D = CB, HB \textit{ and } t = 1, 2, 3$$

Then $\{\lambda^{CB}_t, \lambda^{HB}_t\}_{t=1}^3$ *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 **??**) 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 $\tilde{V}_t^{CB}(a_t, H), \tilde{V}_t^{HB}(a_t, H)$:

$$V_t^{CB}(a_t, H) \geqslant V_t^{N}(a_t, H) \,,\, V_t^{HB}(a_t, H) \geqslant V_t^{N}(a_t, H) \,.$$

Taking partial derivatives to the rest of conditional choice probabilities we get:

$$\frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{HB}} = e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{N})} e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} \overbrace{\big[(\tilde{V}^{HB} - \tilde{V}^{N}) - \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}}(\tilde{V}^{CB} - \tilde{V}^{N})\big]}^{\geqslant 0}$$
$$+ \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \big(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{N})}\big) \big[(\tilde{V}^{HB} - \tilde{V}^{CB}) + \frac{1}{\lambda_{CB} + \lambda_{HB}}\big] \geqslant 0$$

again with strict inequality if $\tilde{V}_t^{HB}(a_t, H) > \tilde{V}^{N}(a_t, H)$.

$$\frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{CB}} = \frac{\lambda_{HB}}{\lambda_{HB} + \lambda_{CB}} (\tilde{V}^{CB} - \tilde{V}^{N}) e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{N}) - \lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})}$$
$$- \frac{\lambda_{HB}}{(\lambda_{CB} + \lambda_{HB})^2} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \big(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{HB})}\big) \leqslant 0$$

where the last inequality follows from 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

$$\frac{\partial \mathbb{P}(D = CB|a_t, H)}{\partial \lambda_{HB}} = -\frac{\partial \mathbb{P}(D = HB|a_t, H)}{\partial \lambda_{HB}} - \frac{\partial \mathbb{P}(D = N|a_t, w^m, q^m, q^r, \overline{T}^r)}{\partial \lambda_{HB}} =$$
$$e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{N})} e^{-\lambda_{CB}(\tilde{V}^{CB} - \tilde{V}^{N})} \frac{\lambda_{CB}}{\lambda_{HB} + \lambda_{CB}} (\tilde{V}^{CB} - \tilde{V}^{N})$$
$$- \frac{\lambda_{CB}}{\lambda_{CB} + \lambda_{HB}} e^{-\lambda_{HB}(\tilde{V}^{HB} - \tilde{V}^{CB})} \big(1 - e^{-(\lambda_{CB} + \lambda_{HB})(\tilde{V}^{CB} - \tilde{V}^{N})}\big) \big(1 + (\tilde{V}^{HB} - \tilde{V}^{CB})\big) \leqslant 0$$

where the last inequality follows from the convexity of $e^{-(\tilde{V}^{CB} - \tilde{V}^{HB})x}$.

Simmilarly, if $\tilde{V}^{CB} \geqslant \tilde{V}^{HB}$, $\tilde{V}^{CB}$ is strictly increasing in λ^{CB} and weakly decreasing in λ^{HB}.

When taking derivatives, I am leaving the dependence of $\tilde{V}_t^D$ 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^D and weakly decreasing in $\lambda^{D'}$. Moreover, 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:

$$\mathbb{P}(D_3 = CB) = \int_{a_3,H} \mathbb{P}(D_3 = CB|a_t, H) dG_3(a_3, H) := \pi_3^{CB}(\lambda^{CB}, \lambda^{HB}) ,$$
$$\mathbb{P}(D_3 = HB) = \int_{a_3,H} \mathbb{P}(D_3 = HB|a_t, H) dG_3(a_3, H) := \pi_3^{HB}(\lambda^{HB}, \lambda^{CB}) ,$$
$$\mathbb{P}(D_3 = N) = \int_{a_3,H} \mathbb{P}(D_3 = N|a_t, H) dG_3(a_3, H) := \pi_3^{N}(\lambda^{HB}, \lambda^{CB}) ,$$

  where we have data in the LHS and known functions of $\lambda_3^{CB}, \lambda_3^{HB}$ in the RHS. Moreover, $\pi^D(\lambda^{CB}, \lambda^{HB})$ is strictly increasing in λ^D and weakly decreasing in $\lambda^{D'}$ for $D, D' = CB, HB, D' \neq D$, because $\mathbb{P}(D_t = D|a_t, H)$ is weakly increasing in λ^D (strictly if $\tilde{V}^D > \tilde{V}^N$), weakly decreasing in $\lambda^{D'}$, and the following set has positive measure. By a similar token $\pi^N(\lambda^{CB}, \lambda^{HB})$ is strictly decreasing in both arguments.

$$\{a_t, H : \tilde{V}_t^P(a_t, H) > \tilde{V}_t^N(a_t, H)\} \text{ for } D = CB, HB \text{ and } t = 1, 2, 3$$

  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 J, H.
  - $G_t(a_t, H)$ is identified (see discussion Section 5.3

  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 $\pi_3^N(\lambda_3^{CB}, \lambda_3^{HB})$ is strictly increasing decreasing in $(\lambda_3^{CB}, \lambda_3^{HB})$ we must have

$$\pi_3^{N}(\tilde{\lambda}_3^{CB}, \tilde{\lambda}_3^{HB}) > \pi_3^{N}(\lambda_3^{CB}, \lambda_3^{HB}) ,$$

  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 $\lambda^{CB}_{t+1}, \lambda^{HB}_{t+1}$, then $\lambda^{CB}_t, \lambda^{HB}_t$ are identified.

  Again, the unconditional choice probabilities can be written as:

$$\mathbb{P}(D_t = CB) = \int_{a_t,H} \mathbb{P}(D_t = CB|a_t, H) dG_t(a_t, H) := \pi^{CB}_t(\lambda^{CB}, \lambda^{HB}) ,$$
$$\mathbb{P}(D_t = HB) = \int_{a_t,H} \mathbb{P}(D_t = HB|a_t, H) dG_t(a_t, H) := \pi^{HB}_t(\lambda^{HB}, \lambda^{CB}) ,$$

  where we have data in the LHS and known functions of $\lambda^{CB}_t, \lambda^{HB}_t$ in the RHS. The fact that the RHS is a known function of $\lambda^{CB}_t, \lambda^{HB}_t$ follows from the fact that:

  - $\tilde{V}^P_t$ can be computed for $D = N, CB, HB$ because:
    - β is treated as known and $(\delta^m_l, \delta^f_l, \delta^m_\tau, \delta^f_\tau, \{\delta_{\theta,t+1}\}^3_{t=1}, \delta_a)$ are identified from Proposition 3
    - Price schedules $P^P_t(q)$ 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 J, H.
    - From the Induction Hypothesis: $\lambda^{CB}_{t+1}, \lambda^{HB}_{t+1}$ are known, which matters for the computation of $\mathbb{E}\hat{V}_{t+1}(a_{t+1}, H, c^{CB}_{t+1}, c^{HB}_{t+1})$
  - $G_t(a_t, H)$ is identified (see discussion Section 5.3)

  The contradiction argument goes exactly as before. This establishes the result.

□

Some comments are again in order. First, the condition

$$\{a_t, H : \tilde{V}^P_t(a_t, H) > \tilde{V}^N_t(a_t, H)\} \text{ for } D = CB, HB \text{ and } t = 1, 2, 3 ,$$

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, the identification result assumes that unconditional choice probabilities are observed at each age, not conditional choice probabilities. I 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, but (a_t, H) is very high-dimensional, measured with error, and its distribution is endogenous and of unknown functional form for $t = 2, 3$, which complicates using $\mathbb{P}(D_t = D|a_t, H)$ in estimation.

L Estimation of Preference Parameters

I estimate preference parameters in three steps. In the first step, I estimate

$$(\delta^m_l, \delta^f_l, \delta^m_\tau, \delta^f_\tau, \{\delta_{\theta,t+1}\}^3_{t=1}) .$$

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 this minimization problem:

$$\hat{\vartheta} = \arg\min_{\vartheta} \psi(\vartheta, \pi_N)' W \psi(\vartheta, \pi_N)$$

where $\Psi(\theta, \pi_N)$ contains the joint restrictions on parameters and observables from Proposition 3 and W_n is a diagonal matrix that is not a function of 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 single-mother households, I estimate $\{\lambda_t^{CB}, \lambda_t^{HB}\}$ separately for two-parent families and Single Mothers.

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) [44]. 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 2.

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 of the parameters that I

[44] 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 $t = 2, 3$

am estimating by Indirect Inference ($\{\lambda_t^{CB}, \lambda_t^{HB}, \delta_{\theta,t+1}\}_{t=1}^3$) are identified, I include

$$\hat{\mathbb{E}}[\frac{\tau_{i,t}^m}{l_{i,t}^m} | 0 < \tau_{i,t}^r < \overline{T}_i^r] \text{ , } \hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{m,r} \tau_{i,t}^m | 0 < \tau_{i,t}^r < \overline{T}_i^r] \text{ ,}$$

$$\hat{\mathbb{E}}[\frac{\tau_{i,t}^f}{l_{i,t}^f} | 0 < \tau_{i,t}^r < \overline{T}_i^r] \text{ , and } \hat{\mathbb{E}}[\tilde{\Delta}_{\theta,i,t}^{f,r} \tau_{i,t}^f | 0 < \tau_{i,t}^r < \overline{T}_i^r]$$

as targets (note that these identify $\{\delta_{\theta,t+1}\}_{t=1}^3$ given that $\delta_\tau^m, \delta_\tau^f$ were estimated in the first step). Note that λ vary across household structures, but not other preference parameters. In principle, single mothers could have different preferences from two-parent families, but allowing for total flexibility in those 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:

   $$(a_1, w^m, w^f, \overline{T}^r, \log q^r, \log q^{m,e}, \log q^{f,e}) \text{ .}$$

   See Appendix E for the imputation of assets, D for the imputation of parental wages, G for the estimation of the relative care endowment for each family, and I.3 for the construction of the observable part of parental care.

2. Simulate $(\nu_{i,j}^m, \nu_{i,j}^f)$ from its estimated distribution many times ($j = 1, \dots J$) for each family i and construct $\log q_{i,j}^m$ and $\log q_{i,j}^f$ according to:

   $$\log q_{i,j}^m = \log q_i^{m,e} + \nu_{i,j}^m \text{ , } \log q_{i,j}^f = \log q_i^{f,e} + \nu_{i,j}^f \text{ .}$$

   The resulting simulated dataset:

   $$\{\{a_{1,i}, w_i^m, w_i^f, \overline{T}_i^r, \log q_i^r, \log q_{i,j}^m, \log q_{i,j}^f\}_{i=1}^N\}_{j=1}^J \text{ ,}$$

   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 $\vartheta^{IF} = \{\lambda_t^{CB}, \lambda_t^{HB}, \delta_{\theta,t+1}\}_{t=1}^3$, 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.

Estimates and fit are shown in the body of the text.

M Identification and estimation of labor supplies

Taking logs in the labor supply equations we get:

$$\log LT_l = \log \overline{LT} + \eta_{LT} \log(w_l^{LT}) + \xi_{l,t}^{LT} ,$$

$$\log CCW_l = \log \overline{CCW} + \eta_{CCW} \log(w_l^{CCW}) + \xi_{l,t}^{CCW} .$$

A linear regression 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, **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:

$$\log\Big(\frac{LT_{l,t}}{\text{Employment}_{l,t}}\Big) = \log \overline{LT} + \eta_{LT} \log w_{l,t}^{LT} + X_{l,t}' \beta_{LT} + \xi_{l,t}^{LT} ,$$

$$\log\Big(\frac{CCW_{l,t}}{\text{Employment}_{l,t}}\Big) = \log \overline{CCW} + \eta_{CCW} \log w_{l,t}^{CCW} + X_{l,t}' \beta_{CCW} + \xi_{l,t}^{CCW} .$$

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 is correlated with lagged fertility, I include contemporaneous and lagged one period fertility in $X_{l,t}$. See body of the paper for estimates of elasticities.

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. Note that in equilibrium the supply and demand of factors need to equalize, that is:

$$E^D(w^E, w^C) = \overline{LT}(w^E + w^C)^{\eta_{LT}} \exp^{\xi_{l,t}^{LT}} ,$$

and

$$C^D(w^E, w^C) = \overline{LTH}_{LT}(w^E + w^C)^{\eta_{LT}} \exp^{\xi_{l,t}^{LT}} + \overline{CCW}(w^C)^{\eta_{CCW}} \exp^{\xi_{l,t}^{CCW}} .$$

Using the previous equation this simplifies to:

$$C^D(w^E, w^C) - E^D(w^E, w^C) = \overline{CCWH}_{CCW}(w^C)^{\eta_{CCW}} \exp^{\xi_{l,t}^{CCW}} .$$

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 the Computational Appendix. Taking logs, expectations, and noting that $\xi_{l,t}^{LT}$, $\xi_{l,t}^{CCW}$ have zero mean:

$$\log\left(\overline{H}_{LT}\overline{LT}\right) = \mathbb{E}[\log E^{D}(w^{E}, w^{C})] - \eta_{LT}\mathbb{E}[\log(w^{E} + w^{C})] \,.$$

$$\log\left(\overline{H}_{CCW}\overline{CCW}\right) = \mathbb{E}\left[\log\left(E^{D}(w^{E}, w^{C}) - C^{D}(w^{E}, w^{C})\right)\right] - \eta_{CCW}\mathbb{E}[\log w^{C}] \,.$$

Substituting population objects by sample analogs we get the estimates[45] in Table 30.

Table 30: Estimates of factor supply shifters

$\overline{H}_{LT}\overline{LT}$	$\overline{H}_{CCW}\overline{CCW}$
34.93	33.89
(0.61)	(0.51)

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. Standard Errors (in parenthesis) are calculated using an analytic formula
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".
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.

[45] 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.