Data Construction

for

"Child Care Markets, Parental Labor Supply, and Child Development''

Samuel Berlinski, Maria Marta Ferreyra, Luca Flabbi and Juan David Martin

August 28, 2023

The model is estimated using moments extracted from *The Early Childhood Longitudinal Study, Birth Cohort* (ECLS-B). This study follows from birth through kindergarten a nationally representative sample of children born in 2001. The data were collected over several waves to capture child development. We focus on the second wave, collected between January and December 2003, when the children were two years old.

ECLS-B case-level data are available only to qualified researchers who are granted a restricted-use data license. Information about applying for or amending a restricted-use data license can be found at http://nces.ed.gov/pubsearch/licenses.asp

In order to provide other researchers who have been granted access the possibility to reproduce the moment conditions in the paper and reproduced here in the excel files AggregateMoments.xlsx and Location-SpecificMoments.xlsx, we present a detailed description of how we obtain the relevant moments. This description complements the information already available in the paper's Section 3 and Appendix C.

Household characteristics

We restrict to:

1. Two parents (biological or not) household or a single mother household.
2. One child only who is two years old.
3. Only three people living at home for two parent households or two people for single mother households.

Education

We define two education variables for the parents. First, we create a two-category variable that takes a value of one for those who completed college or more. Second, we create a three-category variable with incomplete high-school, complete high-school (including some college or technical degree) and college degree or more.

Child care types

Parents can choose one of three child care modes in addition to exclusive parental care in our model: child care centers, non-relative care, and relative care. ECLS-B defines these three modes of care in the

following way. They assign to child care providers children whose parents report that the main form of care on a regular basis is child care centers" [..] Such centers include early learning centers, nursery schools, and preschools. "They assign to non-relative care children whose parents report that the main form of care on a regular basis is \someone not related to " the child. [..] \This includes home child care providers, regular sitters. "They assign to relative care children whose parents report that the main form of care on a regular basis is \a relative other than a parent [..], for example, grandparents". In addition to the main form of care used on a regular basis, ECLS-B also collects information on forms of care used on an irregular basis, for example a sitter for the night. We ignore non-parental forms of care used on an irregular basis. We eliminate households with multiple care arrangements as the main form of care or using non-relative or relative caregivers that take care of multiple children at the same time.

Child care cost

Is measured through what the center director reports is the hourly cost for a full-time 2-year-old child attending center based care.

Child care quality

Is measured through the average of the ITERS items. High quality is defined by a center with an average score of more than 4.

Mental score

Bayley Short Form-Research Edition. We use the scale version of this item because it is useful to look at changes in levels. We use the cognitive score at age two as an outcome and at age one for initial conditions.

Pay

Pay is reported before taxes after converting to hourly wages using the pay reporting period.

We drop cases where parents earn above the bottom .5 and top 99.5 of the corresponding educational and gender pay distribution.

Household types

We define full-time work as working at least 40 hours a week and at most 84 hours a week. We define part-time working at least 15 hours a week and less than 40 hours a week.

We keep households where:

1. The father works full-time (in two-parent households) and earns positive wages.
2. Mother doesn't work, works part-time or full-time and if it works (either part-time or full-time) earns positive wages.

We drop households where:

3. Parents' main jobs include shift work.
4. Two-parent households where both parents work full-time but do not use outside child care.
5. Single-mother households where the mother works full- or part-time but does not use outside child care.

Markets

We assign every household to one of four locations. We define locations using 2000 Census county data on population and income. Given our focus on child care markets, we consider as the relevant population the number of children younger than 5 in the county. Given our focus on household decisions, we consider as relevant income the median family income. We build the assignment as follows. Given each household's county of residence, we sort households by county population and calculate their median county population (36,150 children). Households below the median are assigned to Locations 1 and 2 (small locations) and those above are assigned to Locations 3 and 4 (large locations). We then sort households in Locations 1 and 2 by county median family income, the median of which is ($47,150), and assign them to Location 1 or 2 if they are below or above $47,150, respectively. For households in Locations 3 and 4, the corresponding threshold is $53,950; we assign them to Location 3 or 4 if they are below or above the median, respectively.

Initial conditions

As discussed in the paper's Section 4.1, our empirical specification ex-ante household heterogeneity encompasses seven dimensions: parental education, marital status, preferences, wage offers, initial child development, relative care availability, and location. Out of these seven dimensions, four are directly observed in the data: parental education, marital status, initial child development, and location. In addition, wage offers for fathers can be considered observable because all fathers participate in the labor market. To ease replication exercises, we use this information by matching conditional moments of initial child development observations and the proportions of household by parental education, marital status, and location.

Specifically, we build as initial conditions for our simulation a sample with the following characteristics:

1. We create a sample of married and single households that match the proportion of single and married individual in the data by the three levels of education and the four locations we use in the empirical specification of the paper.
2. We assign to fathers' wages extracted from lognormal distributions that fit the means and variances observed in the data, conditioning on education and location.
3. We assign mothers to fathers according to the conditional correlations over education and location observed in the data.
4. We assign to the households the child's cognitive score at age 1 (initial child development value) by extracting them from normal distributions that fit the means and variances observed in the data, conditioning on parents' education, marital status, and location.

These initial condition samples are made available to researchers as: SimData_M0.RData and as SimData_L1.RData, SimData_L2.RData, SimData_L3.RData, SimData_L4.RData. In these publicly available simulated data, sample sizes conditional on observables are rounded to the nearest 50 to comply with ECLS-B confidentiality rules.

Quadratic Form Weights

As discussed in the paper's Section 4.3, the weighting matrix **W** in the quadratic form (26) is diagonal where the weights are the bootstrapped standard deviations of the sample moments. We obtain them running 1,000 extractions with replacement from the original data. The weights are available to

researchers as the SD column in the in the excel files AggregateMoments.xlsx and Location-SpecificMoments.xlsx.