Getting ahead: The reality of socioeconomic effects on child development

By: Hannah Kosinovsky

Introduction:

In our fast paced and competitive society, parents are doing more and more to make sure that their children are able to perform well. Is it possible that there is no way for parents with less monetary means to compete with the families with a better socioeconomic status? Or is it possible that it just depends how much time the parent has to be with their child: i.e. number of siblings the child has, the number of parents at home, and how well the parents are able to teach their children? These questions and their answers could help drive education policy if it is clear that kids from homes with only one parent, or homes with less money, start school with significantly less tools to learn, educators may be encouraged to spend more time with such children to help them catch up. This may also lead to public university policies that may want to give supposedly more disadvantaged children an opportunity to continue their education even if they started out rough, but show promise.

To evaluate whether or not children from different socioeconomic backgrounds and varying amounts of time devoted to them have the same tools for success, I came up with two questions.

1. If we control for the variables parentscore and momage, which we can interpret as wisdom/ education of the mother, and ability of parents to teach, does higher SES contribute to higher scores?

2. How does the make up of a child's family affect their 2^{nd} BSID score?

Data description:

This data set comes from a national sample of children born in 2001. It includes the characteristics of the children and their families. It also includes two BSID scores. The first score is measured between the ages of 8 and 12 months. And the second score is measured around their second birthday.

"The Bayley Scales of Infant and Toddler Development (Bayley-III is the current version) is a standard series of measurements originally developed by psychologist Nancy Bayley used primarily to assess the motor (fine and gross), language (receptive and expressive), and cognitive development of infants and toddlers, ages 1-42 months… These scores are used to determine the child's performance compared with norms taken from typically developing children of their age (in months)."

Variables in the data set include:

<u>Race:</u> a categorical variable about children's race that includes White, Black, Hispanic(race specified), Hispanic(race unspecified), Asian, Native Hawaiian, American Indian or Alaska Native, and more than one race. Most of the observations are White and Black.

Multibirthstatus: a categorical variable telling whether the infant is a singleton, a twin, or other. Most of the observations are of singletons and there are only 50 children of a higher order than twin.

Parentscore: a score that evaluates how good the parent(s) act as a teacher to the child. From our histogram we see that the score is out of 50 points, but most observations lie at about 35 and no parent scored worse than 12.

SES: a categorical variable that divides families into five groups according to a measure of their socioeconomic status (SES). The five groups are separated into quintiles where the First quintile represents the lowest SES.

Region: location of family. This is a categorical variable that includes Northeast, Midwest, South, and West. The South has the most observations in this dataset.

Momage: This variable represents the mother's age and ranges from 15 to 50 with the most observations at age 30.

Days_premature: infants born before 37 weeks of pregnancy are premature. The variable =0 if born after 37 weeks. It = date of 37 weeks of pregnancy- birth date if born before due date. A majority of the observations are 0 days premature.

num_siblings**:** number of siblings of the child. There are few observations of children with more than 2 siblings, and only 29 observations of children with 6 siblings.

Familystructure**:** a categorical variable describing the structure of the family. It includes categories for both biological parents present, only one biological parent present, one biological and one non-biological parent, and other. Most of the observations have both biological parents present.

Birthweight: a categorical variable of children's birth weight. This is measured in grams. Most of the observations are between 2500 and 3500 grams.

Female: Binary variable that equals 1 if the child is female. The observations are practically evenly split between males and females.

Age_1wave: Age in month when the first test is done. This variable ranges between 8 and 16 months, but a majority of the observations are taken at 9 months.

Age_2wave**:** Age in month when the second test is done. This variable ranges from 23 to 26 months, but most observations are taken at 23 months.

Score_1wave: Standardized values of the first test score

Score_2wave: Standardized values of the second test score

Tables and histograms of these variables will be included in the appendix for clarity.

Methods:

My first question was:

1. If we control for the variables parentscore and momage, which we can interpret as wisdom/ education of the mother, and ability of parents to teach, does higher SES contribute to higher scores?

In order to answer this question, I first regrouped the variable SES into 3 groups so I could use it as an ordinal variable with low, medium, and high values. I generated two new variables that I called "lowSES", the first two quintiles of SES, and "highSES", the last two quintiles of SES. The variable lowSES takes on a value of 1 if the family is in the first two quintiles and 0 otherwise. The variable highSES takes on a value of 1 if the family is in the last two quintiles and 0 otherwise.

When I ran the regression of lowSES highSES momage and parentscore on score 2, I got the model:

$$\widehat{score\_2wave} = \underset{(.0789)}{-.746} - \underset{(.0241)}{.173}lowSES + \underset{(.0239)}{.2037}highSES - \underset{(.0015)}{.0026}momage$$
$$+\underset{(.0019)}{.0202}\,parentscore$$

The coefficients on both lowSES and highSES are significant at the $\alpha = .05$ level. Their p-values are essentially 0 and their 95% confidence intervals do not contain 0. Therefore we can reject the null hypothesis (that the slope on the coefficient is 0).

Therefore, from this regression, we can see that families in higher brackets of socioeconomic scores have .2037 higher BSID scores than those with average socioeconomic scores and families with low socioeconomic scores have .1734 lower BSID scores than those with average socioeconomic scores.

The coefficient on momage is insignificant at the $\alpha = .05$ level since its p-value is 0.084, while the coefficient on parentscore is significant at $\alpha = .05$ level since it's p-value is essentially 0. This implies that a mother's age does not always correlate positively with the BSID, or that increased age does not always mean increased education or wisdom, but the ability of a parent to teach is significantly positively correlated with the BSID score.

These findings also imply that with the same mother age and the same parent score, children from families with highSES will still have a higher score than the average. This is really a sad conclusion, because despite parents' willingness and ability to teach, it seems that children from better economic backgrounds will still start school with a head start.

My second question was:

2. How does the make up of a child's family affect their 2nd BSID score?

In order to find this answer, I decided to look at the variables parentscore, familystructure, and num_siblings. Parentscore will be the constant because it represents the parent's ability to teach. The other variables show how much time a parent has to give to their child, since parents with fewer kids and help from their spouse have more time to spend on an individual child.

I want to make a model that allows for score differences between four groups. I chose a base group: a child without both biological parents and with siblings. Then, I generated three new variables.

"bothparents_onlychild" which I defined as both biological parents present from the familystructure variable, and 0 siblings in num_siblings variable. This new variable takes on a value of 1 if the child has both biological parents and has no siblings and 0 otherwise.

"bothparents_siblings" which I defined as both biological parents present from the familystructure variable, and more than 0 siblings in the num_siblings variable. This variable takes on a value of 1 if the child has both biological parents but has siblings.

"not_bothparents_onlychild" which I defined as not having both biological parents present from the familystructure variable, but having 0 siblings in the num_siblings variable. This variable takes on a value of 1 if the child doesn't have both biological parents present and is an only child and takes on a value of 0 otherwise.

The reason I did not count stepparents, or other partners as a second parent is because I think the responsibility of that partner would not be quite the same as that of a biological parent, therefore the one biological parent that is present will still have more responsibility for the child on their shoulders. I also decided to use the second score as the response because the older child has had more time to be affected by the structure of their family.

I ran a regression of bothparents_onlychild, bothparents_siblings, not_bothparents_onlychild, and parentscore on score2 and got the following model:

$$\begin{aligned}\widehat{score\_2wave} = & \underset{(.0686)}{-1.223} + \underset{(.0306)}{.2907}bothparents\_onlychild + \underset{(.0283)}{.1919}bothparents\_siblings \\ & + \underset{(.0373)}{.1224}not\_bothparents\_onlychild + \underset{(.0019)}{.0271}parentscore\end{aligned}$$

Since my base group is defined as children without both biological parents and with siblings, the estimates on the three dummy variables measure the difference in scores relative to my base group. All the coefficients in this regression are significant at $\alpha = .05$ since the p-values of each coefficient are essentially 0, and the 95% confidence intervals on the coefficients do not contain 0. We can see from this model that if a child has both biological parents and is an only child, their second BSID score is .2907 higher than a child with no biological parents and with siblings. A child with both biological parents and siblings has a .1919 higher score than a child without both biological parents and with siblings. A child that doesn't have both biological parents, but is an only child has a .1224 higher 2nd BSID score compared to a child with siblings and not both biological parents present.

We see that the coefficient on bothparents_onlychild is the highest, which makes sense since this child is able to get more attention from his or her parents compared to the others. The coefficient means that a child who is an onlychild and has both biological parents with have a .2907 larger score than my base group.

After running this regression, I thought that it would also be interesting to look at the difference between score 2 and score 1 as the response, and my three new variables and parentscore. I wanted to see how the development of kids from younger to older is affected by their family structure.

To do this, I generated a new variable called "difference" which I defined as the difference between the variables score_2wave and score_1wave. Here I made the assumption that both scores were standardized in the same way. Then, I ran the same regression as above, but with difference as my response. I got the following model:

$$\widehat{difference} = \underset{(.0838)}{-.1758} + \underset{(.0375)}{.1842} bothparents\_onlychild$$
$$+ \underset{(.0345)}{.2145} bothparents\_siblings + \underset{(.0455)}{.0214} not\_bothparents\_onlychild + \underset{(.0023)}{.0017} parentscore$$

In this model, only the coefficients on bothparents_onlychild and both_parents_siblings were significant at $\alpha = .05$ since their p-values were 0 and 0 did not lie in their 95% confidence interval. The p-value for the coefficient on not_bothparents_onlychild was .638 and the p-value for the coefficient on parentscore was .45, therefore they were not significant at $\alpha = .05$. This implies that we can only be certain that the difference in scores is positive for children with both biological parents present. This could be important as a finding for parents who are considering separation.

Conclusion:

From these regressions, I found that the structure of a family and socioeconomic status do make a difference in the BSID scores of children. I cannot be completely certain of my findings, since I cannot feasibly control for all factors that may affect a childs ability, and it is possible that the coefficients on the variables that I thought were significant were actually capturing the effects of other variables that were not present in my model. I used parentscore as a constant in my regressions because it measures a parent's aptitude for teaching their children. The number of family members in a household fluctuates, and the amount of money a family has fluctuates, but parent's ability to be patient and teach is constant.

As a society, we believe in an American Dream where everyone has the same chance to succeed in life no matter where they come from. This data gives us a reference point for the ability of a child to succeed in life by testing motor skills, speech skills, and cognitive abilities of children. We see from these regressions that socioeconomic status and stability of family structure give children a head start in their development. A child of divorce from a poorer family does not seem to have the same chances for success as a child from a more stable background.

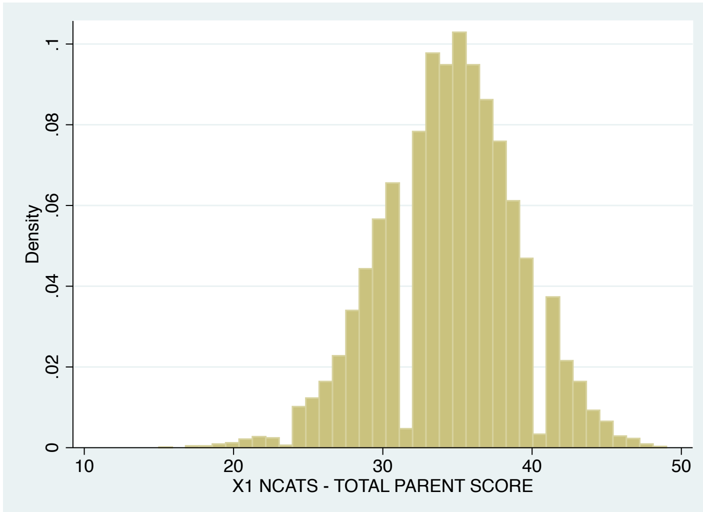
These findings can be applied to education policy. Children who started behind others as a result of their family's condition can be enrolled in special after-school programs, where they can learn skills and get help from elders, from whom they don't receive adequate attention at home. This can also shift how society views success and maybe lead to more "social nets." If we no longer view inability to succeed in society a completely personal fault, access to grants for college could improve, or even on the more extreme side, there could be more housing for the homeless, whose misfortune could be the result of factors in their upbringing that were out of their control.

Appendix

Brief summaries of each variable:

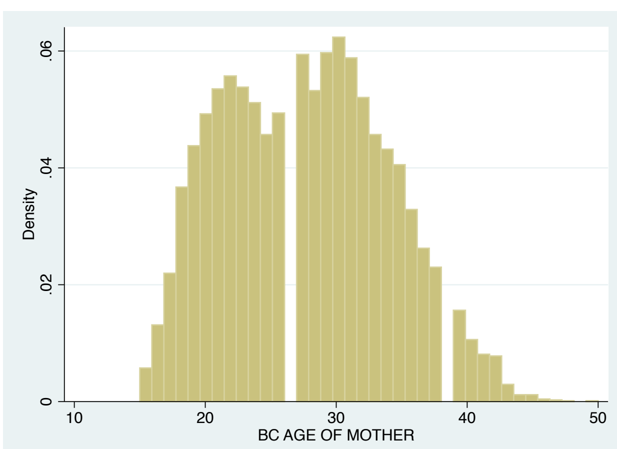
X1 RACE/ETHNICITY - CHILD	Freq.
WHITE, NON-HISPANIC	3,300
BLACK OR AFRICAN AMERICAN,NON-HISPANIC	1,161
HISPANIC, RACE SPECIFIED	980
HISPANIC, NO RACE SPECIFIED	447
ASIAN, NON-HISPANIC	693
NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	28
AMERICAN INDIAN OR ALASKA NATIVE, NON-HI	177
MORE THAN 1 RACE, NON-HISPANIC	577

X1 MULTIPLE BIRTH STATUS INDICATOR	Freq.
SINGLETON	6,148
TWIN	1,165
HIGHER ORDER	50



X1 QUINTILE INDICATOR FOR SOCIOECONOMIC STATUS	Freq.
FIRST QUINTILE (LOWEST)	1,308
SECOND QUINTILE	1,419
THIRD QUINTILE	1,455
FOURTH QUINTILE	1,471
FIFTH QUINTILE (HIGHEST)	1,710

X1 REGION	Freq.
NORTHEAST	1,039
MIDWEST	1,815
SOUTH	2,679
WEST	1,838



days_premature	Freq.
0	5,424
7	443
14	310
21	225

0	2,673
1	2,525
2	1,385
3	529
4	168
5	54
6	29

28	140
35	123
42	121
49	127
56	91
63	102
70	88
77	169

familystructure	Freq.
Biological mother and biological father	5,778
One biological parent present	1,460
One biological parent and one nonbiological	79
Other	46

birthweight	Freq.
<1500 grams	763
1500-2500 grams	1,167
2500-3500 grams	3,401
>3500 grams	2,032

female	Freq.
0	3,754
1	3,609

age_1wave	Freq.
8	1,108
9	2,526
10	1,710
11	857
12	485
13	272
14	160
15	108
16	137

age_2wave	Freq.
23	3,026
24	2,811
25	954
26	572