Wenrui Huang

Date: May,12, 2020

**Final Paper Review and Replication**

Paper:

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, *1*(3), 111-34.

Selection bias is a common threat to evidence provided by program evaluation, especially in the area of educational research. Eligible participants in a program might be systematically different from non-participants, mostly in individual and family characteristics. For studies that lack an experiment design, a simple comparison between the eligible and non-eligible groups could bring either upwards or downwards bias to program effect.

To deal with such a concern, Deming (2009) compares siblings who differ in their participation in the program and controlled for a variety of pre-treatment covariates. Using the National Longitudinal Survey of Youth, he assessed the impact of Head Start on kids' academic performance and several important adult outcomes. He found that Head Start has brought a gain of 0.23 standard deviations on a summary index of young adult outcomes for those who participated.

The key assumption in the fixed effect identifying strategy lies in that selection into Head Start among members of the same family is uncorrelated with the unobservable determinants of outcomes. By comparing two or more children within one household, the differences in covariates relevant to the family that could have important influences on children's ability would be differenced out. In the model, the author also included factors of mother's education, permanent family income, and maternal Armed Forced Qualification Test (AFQT) scores that are time-invariant factors. However, even within a family, the selection bias could also happen because parents' decision to enroll their kids in a program might differ given their all eligible kids. To deal with such a non-random selection, the author compared Head Start participants to their unenrolled siblings to identify a pattern that explains differential participation.

One strength of this paper is how the author select sample, determine eligibility and comparison groups. I followed up with the author's steps, and it turns out to be a very complex process and requires thoughtful planning. The original NLSY started in 1979 with 12,686 youths between the ages of 14 and 22. For surveys from1988 it includes the information that a child had ever participated (or were currently enrolled) in Head Start or any other preschool. Firstly, the author restricted the sample to children over four years old by 1990 to maintain the sample as stable since they are no longer eligible. For these kids, they will be 19 by 2004. Then he restricted the sample to families with at least two age-eligible children. Together, these restrictions produce a sample size of 3,698. Finally, the sample was restricted to families where at least one (but not all) children were in Head Start.

The author has also done a great job of dealing with missing values and robustness checks. He produced three sets of variables to define preschool participation. The first simply uses the age variable and excludes those with any missing data. The second substitutes the PPVT age (the age at test) when the original variable is unavailable and substitutes past or future age variables from other survey years (plus or minus 24 months). The third codes inconsistent or missing responses across years as zeros and children for whom mothers report "3 months or less" participation in the program. Finally, he adopted the second strategy in the paper and tested the estimates with the other two robust rules.

When creating the dummy for the "fixed effects" sample, the author carefully integrated the information to structure the comparison sample if the kid attends another preschool and combing the sibling eligibility condition. In doing so, it cannot only estimate the influence of enrolling in Head Start compared to no participation of any preschool program but also allows the comparison between Head Start and other preschools.

However, the author's identification strategy may still fail to exclude the selection bias that happened within the family among two or children. For example, if a household has more than two eligible children, only one was enrolled by their parents. These two children are different in their cognitive development or social-emotional skills. As a result, other parents might likely decide to send a more talented child to join Head Start or a less developed child. Either decision that causes a selection problem might also bias up or downs the program. Another significant concern raised by the author in this identification strategy might lie in that also the estimated effect of Head Start could be moderate if there are treatment spillovers between children, given that Head Start highly focuses on parenting practices. However, the author tested the spillover effect but found very limited and inconsistent evidence. Besides, despite the impressive long-term gains, the author raised another important concern that the identification strategy might attenuate the estimated effect of Head Start if there are treatment spillovers between children.

Despite the impact of 0.23SD gain in adult outcomes are impressive, I would still argue that using a summary index could obscure the actual effect of Head Start on either of the six crucial young adult outcomes: high school graduation, college attendance, "idleness," crime, teen parenthood, and health status. It would be more helpful if the author could estimate the summary index and other important outcomes such as crime and high school graduation individually.

Overall, the fixed effect identification strategy, well-designed sample selection plan, as well as its strict operation steps render this study to relatively reliable evidence in terms of the impact of the Head Start program. The evidence on crucial adult outcomes is vital for answering policy questions. To improve the vanity and reliability of evidence on programs like Head Start, experiment design or combining other causal inference strategies are encouraged.

Results of Table1 replication:

**Table 1—Selected Family and Maternal Characteristics, by Race and Preschool Status**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **White / Hispanic** | | |  |  | **Black** |  |  | Head Start vs. None SD units difference | | |
|  | Head  Start | Preschool | None |  | Head  Start | Preschool | None |  | White/  Hispanic | | Black |
| **Permanent income** | 26,830.54 | 53,482.67 | 35,121.23 |  | 23,327.94 | 32,933.54 | 25,931.49 |  | -0.10 | -0.16 | |
| Fixed effects subsample | 27,417.87 | 42,399.20 | 36,175.59 |  | 25,093.30 | 29,970.77 | 24,030.09 |  | -0.06 | -0.01 | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Mother < high**  **school** | 0.514 | 0.193 | 0.45 |  | 0.36 | 0.21 | 0.4 |  | 0.002 | -0.01 | |
| Fixed effects subsample | 0.54 | 0.25 | 0.40 |  | 0.42 | 0.28 | 0.39 |  | 0.01 | 0.02 | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Mother some**  **college** | 0.223 | 0.401 | 0.212 |  | 0.30 | 0.49 | 0.28 |  | 0.008 | 0.01 | |
| Fixed effects subsample | 0.16 | 0.30 | 0.23 |  | 0.29 | 0.42 | 0.30 |  | -0.05 | 0.00 | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Maternal AFQT** | -0.459 | 0.198 | -0.244 |  | -0.76 | -0.52 | -0.70 |  | -0.131 | -0.10 | |
| Fixed effects subsample | -0.49 | -0.01 | -0.23 |  | -0.78 | -0.62 | -0.76 |  | -0.11 | 0.02 | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Grandmother's**  **education** | 8.51 | 10.7 | 9.31 |  | 9.71 | 10.9 | 9.8 |  | 0.02 | -0.33 | |
| Fixed effects subsample | 8.45 | 10.20 | 9.60 |  | 9.85 | 10.20 | 10.00 |  | 0.12 | 0.09 | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Sample size** | 426 | 838 | 1648 |  | 461 | 276 | 637 |  |  |  | |
| **Sample size — FE** | 265 | 359 | 591 |  | 229 | 163 | 292 |  |  |  | |

]

The original Table 1 from the paper

Table

Description automatically generated