

Intergenerational Education Mobility in the United States

Nick Wagner

Spring 2017

1. Introduction

As Americans, we often think of our country as “The Land of Opportunity,” and therefore would expect to see high levels of income mobility. However, when compared with other wealthy countries, particularly in Europe, the United States has relatively low mobility. A system in which children of low earning households are predisposed to also have low earnings, while children from wealthy households enjoy a world overflowing with opportunity does not appear congruent with America’s core values of equality and fairness. Furthermore, it is likely that America’s low income mobility is feeding the growth of the already concerning high level of income inequality in the nation.

One potential explanation for the United States’ low income mobility is that there may be low mobility with respect to education. Since education plays a significant role in determining a person’s income, low education mobility could help explain part of the low mobility with respect to income. On the contrary, if education mobility is high, then it is possible there is some other variable(s) present among children who grow up in high income households which make them more likely to earn a high income. For policy makers whose objectives are to create more equality of opportunity, these implications are important to consider.

This paper will not seek to answer questions regarding the relationship between income mobility, education mobility, and income inequality. Rather, the focus of this paper will be solely on determining the level of intergenerational education mobility in the United States. More

specifically, this paper will use a data set from the National Longitudinal Survey of Youth to estimate the relationship between a person's highest grade completed and their residential father's highest grade completed, as well as explore how this relationship varies between gender, ethnicity, race, and father's educational attainment level. This paper adds to the existing literature on intergenerational mobility by estimating models which control for different variables than models in other papers do. This paper also explores the difference in social mobility for individuals whose father had high educational attainment versus individuals whose father had low educational attainment, a topic rarely discussed in other papers. In addition, this paper will estimate mobility using a data set that is unique to other papers on intergenerational mobility.

2. Literature Review

Extensive research has been done on the intergenerational mobility of socioeconomic status. However, most studies focus on income as the measure of socioeconomic status rather than education. Of the research that has been done on intergenerational mobility in education, essentially all studies conclude there is a positive relationship between the education level of a parent and the education level of that parent's child. The studies differ in the magnitude of this relationship and the variable they use for parent's or parents' education. Many studies use the father's highest grade completed, while others use the average highest grade completed between the mother and father, or the maximum highest grade completed between the mother and father.

One of the most recent papers on intergenerational mobility in education is Ferrare (2016). This paper focuses on how intergenerational mobility in education varied across race and

gender in the United States over the course of the 1900's. Since this paper was only interested in how race and gender affect education mobility, they used the difference in educational attainment across generations as the dependent variable in their multiple regression model. The constant in their model was interpreted as the total mobility of a white woman across all cohorts, which allowed the coefficients of all the independent variables in the model to be interpreted as the difference in estimated mobility between a white woman and the respective indicator variable. However, structuring the multiple regression model in this way does not explicitly show the effect of increasing parents' educational attainment. The simplest model in the paper estimates the mobility of white women, black women, white men, and black men. In this model, the constant and the coefficients on black and male were all positive, showing that being black and male increased estimated mobility in education (Ferrare, 2016).

Hauser and Featherman (1976) studied the trends in intergenerational education mobility seen in men born during the first half of the 20th century. The results in the paper show that for men who were ages 21 to 65 in 1973, after controlling for several background variables including father's occupation, number of siblings, broken family, farm background, southern birth, spanish origin, race, and cohort, a one year increase in their father's education level resulted in an estimated .206 year increase in their education level. The coefficients on the variables Father's education and Father's occupation were both positive, with the rest of the variables containing negative coefficients. The paper shows that all of these variables are statistically significant in the model. One interesting result shown in Hauser and Featherman (1976) is that race had a much smaller effect on educational attainment for men who were ages 21 to 65 in 1962 than for men who were 20 to 64 in 1973. For the 1962 group, being black,

rather than another race, decreased predicted education level by 1.348 years, after accounting for the effects of the various variables outlined previously. In the 1973 group, the associated decrease in predicted educational attainment of being black was only .55 years. The magnitude of this change reflects a significant improvement in racial equality in the education system (Hauser and Featherman, 1976).

Intergenerational education mobility has been estimated using various methods and data sets in India. One of the most in-depth studies is Azam and Bhatt (2015). They use data from the Indian Human Development Survey, 2005, which contains information on 41,554 Indian households. One of the strengths in the data is that they were able to identify the father's years of schooling for 97 percent of the males who took the survey, which helps minimize the potential selection bias in their results. Intergenerational education persistence is measured in two ways in the paper. First, education persistence is estimated in a simple linear regression model which regresses the child's years of schooling on Father's years of schooling. Azam and Bhatt (2015) rationalize the use of years of schooling rather than highest grade completed due to the fact that using highest grade completed creates discontinuities in the distribution. A high Beta on the variable "Father's years of schooling" means a high level of intergenerational education persistence, and hence, low intergenerational education mobility. The second measure of intergenerational education persistence utilized in Azam and Bhatt (2015) is the intergenerational correlation coefficient. For all cohorts and regions, the study reports an intergenerational education elasticity of .634. It should be noted that Azam and Bhatt (2015) uses the term "intergenerational education elasticity" despite using a level-level regression model. The

intergenerational correlation coefficient at the national level ranged from .507 to .535, depending on the birth cohort (Azam and Bhatt, 2015).

The existing literature on intergenerational education mobility varies greatly in terms of what question each study aims to answer. It is no surprise that studies on different countries produce different estimates for intergenerational education mobility, especially when comparing estimates in countries that are as vastly different as India and the United States. Due to the fact that most studies done on intergenerational education mobility in the United States vary in multiple ways, it is hard to compare results from one study to another.

This paper will take a similar econometric approach as some of the existing literature, but seek to answer different questions using a data set that is unique to the rest of the literature. The model used in this paper will be similar to the more basic econometric models used in the paper by Ferrare (2016), as well as the model used by Hauser and Featherman (1976), which controls for various family background variables. However, the approach we take in estimating education mobility will also have significant differences from the models used in these pieces of literature. First off, the model in this paper will not take into account the same family background variables as the previous models. Furthermore, it will significantly differ from the model used by Ferrare (2016) in that it will use the survey respondent's highest grade completed as the dependent variable rather than the difference in educational attainment between the respondent and their parent's. We set up our model in this way in order to explicitly show the marginal effect of increasing father's highest grade completed by one year. This paper will explore how intergenerational mobility in education differs between children whose father had high educational attainment and those children with fathers who had low educational attainment. We

seek to answer this question in order to shed light on the inequality of opportunity across children from households with different education levels. Mobility can result from underperforming in relation to one's parents or from outperforming one's parents. We would much rather see high mobility for children with low achieving parents and low mobility for high achieving parents than vice versa. Discovering which of these patterns we observe in the data will be one of the main focuses of this paper. Lastly, this paper will build on the existing literature by using a data set that has not been used extensively in other studies done on the intergenerational education mobility of the United States.

3. Data Description

The data used in this paper comes from the 1997 National Longitudinal Survey of Youth, which is a project that follows the lives of 8,984 Americans born between 1980-84 and is funded by the Bureau of Labor Statistics (NLSY97).

The group of participants in the study is made up of two subsamples. The first sample consists of 6,748 respondents and is meant to be representative of all Americans born from 1980-84. The second sample of 2,236 respondents were chosen to over-represent the Hispanic and African American population living in the United States from 1980-84. This over representation of Hispanics and African Americans in the cohort makes controlling for race and ethnicity essential to creating an estimation of intergenerational mobility which is representative of the American population as a whole (NLSY97).

The respondents were ages 12-17 when first interviewed in 1997 and have been surveyed an additional 16 times since then. Through the surveys, respondents provide information on their education, income, family background, and a myriad of other areas of their lives (NLSY97).

It is important to note that all of the individuals from the study are above 30 years old and, therefore, will not likely complete any additional education in the future. This is important because estimating education mobility for a group of individuals who have not completed their educational careers would surely produce inaccurate results. In addition, the study has been able to sustain a high participation rate with approximately 80 percent of the original respondents participating in the most recent survey. Keeping the participation rate high helps to avoid the potential selection bias that could arise from a correlation between dropping out of the study and educational attainment (NLSY97).

A couple of minor changes were made to the data set in order to make the coefficients of the variables easier to interpret. The combined race and ethnicity variable was originally coded as numbers one through four which represented Black, Hispanic, mixed-race, and non-black and non-hispanic respectively. To make the coefficients of the regression easier to interpret, the coding was changed in the data set to let the number one represent being Black or Hispanic and zero represent any other race or ethnicity. Similarly, gender was originally coded as one for male and two for female and this was altered to let zero represent female.

For many of the respondents, the variable for residential father's highest grade completed had a negative value to represent various reasons for unavailable data. Simply excluding these respondents from the data set could cause selection bias in the results. To address this issue, an indicator variable was created which took a value of zero if their father's highest grade

completed was known, and a value of one if there was any type of non-response for father's highest grade completed.

It is also important to consider the fact that for about 1,900 of the 8,984 participants in the study, no data was able to be gathered regarding highest grade completed. These participants could not be included in the regression and this could potentially cause selection bias in the results. The same issue was also present for the Armed Services Vocational Aptitude Battery (ASVAB) test, as the scores on this test were unavailable for 1,329 of the 8,984.

4. Econometric Models

Three separate regression models are used in this paper, each aimed at answering a different question about intergenerational mobility in education in the United States. In each model, the coefficient on the variable "Fatherhgc" represents the highest grade completed for the respondent's residential father. The coefficient on Father's highest grade completed represents the amount of intergenerational education persistence. A small coefficient on Father's highest grade completed means there is a high level of mobility because the individual's education level is not greatly associated with their father's educational attainment. On the other hand, a relatively large coefficient on Father's highest grade completed represents a lower level of mobility. The first model simply seeks to estimate the amount of intergenerational mobility in the United States as a whole. This was accomplished through the regression model shown in Equation (1) below.

$$(1) \quad HGC = \beta_0 + \beta_1 \text{Fatherhgc} + \beta_2 \text{FatherhgcMISSING} + \beta_3 \text{ASVAB} + \beta_4 \text{Male} + \beta_5 \text{BlackorHispanic} + u_i$$

The variable “HGC” represents the highest school grade completed by the individual. “FatherhgcMISSING” is the indicator variable discussed earlier which takes a value of one if their was no data available for the educational attainment of the respondent’s residential father, and zero if the data was available. “ASVAB” is the individual’s score on the ASVAB test which the respondents took in 1999. “Male” takes a value of one if Male, and zero if Female. “BlackorHispanic” is one if Black or Hispanic and zero if neither.

The results from this regression model are shown in the table below.

Table 1: Model 1 Regression Results

Variable	Estimate	P-value
Intercept	9.529	$< 2 e^{-16}$
Father’s highest grade completed	.1638	$< 2 e^{-16}$
Father’s highest grade completed is unknown	1.476	$< 2 e^{-16}$
Score on ASVAB test	.05534	$< 2 e^{-16}$
Male	-.6940	$< 2 e^{-16}$
Black or Hispanic	.5866	$< 2 e^{-16}$

All of the variables included in the multiple regression model are very statistically significant with p-values that are essentially zero. The regression results show that for this cohort, after accounting for ASVAB score, gender, whether or not they are Black or Hispanic, and whether or not Father's highest grade completed was known, increasing the individual's residential father's highest grade completed by one year is associated with an estimated increase in highest grade completed of about .16 years for that individual. When compared with other estimates of mobility in the United States and other countries, this model suggests a relatively high level of intergenerational mobility in education.

There are a few important aspects of the model to point out. First, the inclusion of the variable ASVAB diminishes the estimated association between Father's highest grade completed and the individual's highest grade completed. This is due to the fact that individuals whose father had high educational attainment are more likely to score well on the ASVAB and individuals who score well on the ASVAB are more likely to go on to higher levels of education. Usually, we want to include important variables such as this. However, it may be the case that father's educational attainment affects the individual's ASVAB score through the same avenues that it affects the individual's highest grade completed. If this is the case, including ASVAB in the regression only diminishes the true effect of father's educational attainment. This issue is commonly referred to as "overcontrolling" and may be present in this situation.

Also, as discussed earlier, many of the respondents had to be excluded from the regression due to non-responses on some variables. As a result, only 5,717 of the original 8,984 are included in this multiple regression model.

The objective of the second regression model is to explore how intergenerational mobility in education varies across race, ethnicity, and gender. To make this analysis, the interactions (Father's highest grade completed * Black or Hispanic) and (Father's highest grade completed * Male) were added to the regression. This regression model and its results are shown in Equation (2) and Table 2 below.

$$(2) \quad HGC = \beta_0 + \beta_1 \text{Fatherhgc} + \beta_2 \text{FatherhgcMISSING} + \beta_3 \text{ASVAB} + \beta_4 \text{Male} + \beta_5 \text{BlackorHispanic} + \beta_6 (\text{Fatherhgc} * \text{BlackorHispanic}) + \beta_7 (\text{Fatherhgc} * \text{Male}) + u_i$$

Table 2: Model 2 Regression Results

Variable	Estimate	P-value
Intercept	9.529	$< 2 e^{-16}$
Father's highest grade completed	.1638	$< 2 e^{-16}$
Father's highest grade completed unknown	1.448	$< 3.89 e^{-16}$
Score on ASVAB test	.05538	$< 2 e^{-16}$
Male	-.7141	$< 4.67 e^{-13}$
Black or Hispanic	.6403	$< 2.57 e^{-9}$

(Father's highest grade completed) * (Black or Hispanic)	-.0067	.507
(Father's highest grade completed) * (Male)	.0023	.799

The estimates of the coefficients on the interaction variables are small and far from being statistically significant. Based on these results, we have no evidence that intergenerational mobility in education is any different among Blacks and Hispanics than it is for non-blacks and non-hispanics, after accounting for the other variables in the model. In addition, we have no evidence that intergenerational mobility in education varies across gender, after accounting for the other variables in the model.

The goal of model three is to determine how educational mobility varies between people whose father had a low level of education and people whose father had a high level of education. To make this comparison, the data set was separated into two parts, one for respondents whose residential father completed less than 12 years of schooling, and one for the respondents whose residential father completed at least 12 years of schooling. Next, a simple linear regression model was estimated for each group of respondents. Variables such as gender, race, and ASVAB score are not included in this model because we are only interested in determining how the association between the father's highest grade completed and the individual's highest grade completed

differs for the group of individuals whose father completed less than 12 years of education versus the group of individuals whose father had at least 12 years of education. This simple model and the results for each group are shown below.

$$(3) \quad HGC = B_0 + B_1 \text{Father}hgc + u_i$$

Table 3: Model 3 Results for Group with High Father’s Education Level

Variable	Estimate	P-value
Intercept	8.21	$< 2 e^{-16}$
Father’s highest grade completed	.45223	$< 2 e^{-16}$

Table 4: Model 3 Results for Group with Low Father’s Education Level

Variable	Estimate	P-value
Intercept	12.711	$< 2 e^{-16}$
Father’s highest grade completed	-.0178	.612

There were 924 people whose residential father had less than 12 years of education and 3,501 people whose residential father had at least 12 years of education. As seen in tables 3 and 4, residential father's educational attainment was highly associated with highest grade completed for the individuals whose father had high educational attainment. On the other hand, there was no apparent association between father's educational attainment and highest grade completed for the group whose residential father completed less than 12 years of education. These results suggest very high mobility in the low Father's education level group and relatively low mobility in the high Father's educational level group. This is an encouraging trend to see. However, even though the mobility is much higher in the group whose residential father had low educational attainment, the average highest grade completed for this group is still lower than it is for people who had a residential father with 12 or more years of education. The average highest grade completed for the group with a residential father who had low educational attainment was 12.56. For people who had a residential father with 12 or more years of education, average highest grade completed was 14.57.

Since important variables such as race, gender, and ASVAB score are not accounted for in the third model, no comparison can be made to the previous models in this paper, or to models in other papers which account for these types of variables. Furthermore, we cannot in any way conclude, based on this model, that the differences in the association between the father's education level and the individual's education level between the two groups is caused by the differences in the father's highest grade completed. The differences in the association could in fact be caused by many other factors which have not been taken into account. The only conclusion we can make from the model is that the association between the father's education

level and the individual's education level was significantly lower for the group with a father who had low educational attainment than it was for the group with a residential father who had high educational attainment.

5. Conclusion

The results in this paper agree with results in other studies which have found there is a significant positive association between a father's education level and the amount of education that father's child will attain. The estimated magnitude of this relationship in this paper is towards the low end when compared to other studies, which suggests a relatively high level of intergenerational mobility in education among the cohort studied in the 1997 National Longitudinal Survey of Youth. This paper provides interesting insights on how educational attainment varies across race which are not explicitly shown in other studies. After accounting for ASVAB score and gender, the difference in mobility in education between African Americans and Hispanics and non-blacks and non-hispanics did not have any statistical significance. This paper also explored the difference in mobility between children whose father had high educational attainment and low educational attainment, a topic rarely investigated in other studies. The results in this paper indicate low mobility for individuals whose residential father had high educational attainment, and very high mobility for individuals whose residential father had low educational attainment.

Despite controlling for several relevant variables in the multiple regression models, it is still inappropriate to conclude a causal relationship between father's highest grade completed and

the individual's highest grade completed. However, this is typically not the objective of studies done on intergenerational mobility as this is a challenging task, and the same is true of this study.

There is still plenty of room for discovery on the subject of intergenerational mobility. This paper did not explore variables such as religion, urban vs. rural upbringing, or number of siblings, all of which could play a role in educational attainment. Nevertheless, this paper provides new insights on intergenerational education mobility in the United States using a group of people not previously studied in the literature. Ideally, the results of this paper will strengthen the overall estimates for intergenerational mobility in education and encourage more research to be done on the relationships between intergenerational mobility in education, intergenerational income mobility, and income inequality.

References

Azam, Mehtabul and Bhatt, Vipul 2015. "Like Father Like Son? Intergenerational Educational Mobility in India." *Demography* 52 (6): 1929-1959.

Ferrare, Joseph, J. 2016. "Intergenerational Education Mobility Trends by Race and Gender in the United States." *Aera Open* 2 (4).

Hauser, Robert M., and Featherman, David L. 1976. "Equality of Schooling: Trends and Prospects." *Sociology of Education* 49 (2): 99-120.

Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1997 cohort, 1997-2013 (rounds 1-16). Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2015.