An Analysis of Various Social Factors in Relation to Intergenerational Mobility

Adam Yang

Summer Ventures in Science and Mathematics

The University of North Carolina at Charlotte

Ethical Data Science

Marco Scipioni, Ph.D.

July 25 2023

Abstract

Equality of opportunity is a core concept of American culture. However, many groups of people are disadvantaged, and the disadvantages can be carried over from parent to child. This paper analyzes the effect of two major factors on intergenerational mobility, those being parent income on child household income and parent education level on child education level, and proposes potential ways to begin promoting great equality of opportunity based on the results of the analysis. A Python program was used, through Jupyter Notebook, to organize the data and construct linear regression models for any relevant data. Two-proportion z-tests were also conducted to observe the importance of education on employment status. The regression model for parent income on child household income suggests that AIANs and Blacks are not only the most disadvantaged, but are also the most affected by parent income; therefore, prioritizing parental financial support to these two groups may be an efficient way to promote equality of opportunity. The statistical test for educational attainment proved that education plays a critical role in employment status, and the regression model showed that parent education has a large impact on child education level; it was discovered that an increase in one parent education level could increase the probability of the child achieving a bachelor's degree or higher by roughly 10.41%, and decrease the probability that the child will achieve less than a high school education by roughly 5.05%.

Keywords: intergenerational mobility, equality of opportunity, economic inequality, educational attainment

An Analysis of Various Social Factors in Relation to Intergenerational Mobility

Introduction

America has always been known as the "land of opportunity", a place where all citizens have an equal chance of achieving upwards social mobility. But is that really true today? In 2016, the US Gini coefficient (a measure of income inequality) was 0.435, the highest among other G-7 countries (Kochhar & Cilluffo, 2018). A study conducted by the Pew Research Center found that Blacks and Hispanics earn less than Asians and Whites by a large margin across all income distributions (Kochhar & Cilluffo, 2018). Additionally, the rank mobility curves for Blacks and American Indians and Alaskan Natives are about 13 percentiles lower than the rank mobility curve for Whites (Chetty, Hendren, Jones, & Porter, 2019). The struggles of disadvantaged groups are often passed down from parent to child, making it difficult for these groups to achieve upwards social mobility across generations. The relationship between parent socioeconomic status and child socioeconomic status is known as intergenerational mobility, and a strong intergenerational association indicates children's socioeconomic status is largely determined by their parents' socioeconomic status (Torche, 2015). Evidence suggests that there is indeed a strong intergenerational association in the United States; Chetty, Hendren, Kline, and Saez (2014) found that there was an almost perfectly linear relationship between mean child rank and parent rank. This paper aims to reanalyze and provide an overview of the factors that affect intergenerational mobility to determine the degree to which childrens' future socioeconomic status are affected by their parents' socioeconomic status in the United States, and offer suggestions for how to begin promoting greater equality of opportunity. The specific factors that will be analyzed in this paper include parent income, mother's nationality, and educational attainment.

Data & Methods

This section describes how data was collected and analyzed for each factor of interest, those being parent income, mother's nationality, and educational attainment. All data used in this paper was collected from government datasets, including the United States Census Bureau (USCB) and the United States Bureau of Labor Statistics (BLS).

Parent Income

To observe how parent income affects child household income rank, the dataset "National Statistics by Parent Income Percentile, Gender, and Race" from the USCB was used. The rows of the dataset detail the parent income percentile, while the columns of the data detail the "child information", such as the child household income and child individual income, based on the parent income percentile. The goal was to analyze child household income based on parent income. To do this, a Python program was used, through Jupyter Notebook, to extract the relevant data, those being the columns "par_pctile", "kfr_aian_pooled", "kfr_asian_pooled", "kfr_black_pooled", "kfr_hisp_pooled", and "kfr_white_pooled". The readme for the data can be found here: https://www2.census.gov/ces/opportunity/table7.pdf. The Python program was also used to construct a linear regression model for each racial group, those being Hispanics as a separate group and non-hispanic Whites, Blacks, Asians, and American Indian and Alaskan Natives (AIAN). A scatter plot of the data was also created for visualization purposes.

Educational Attainment

Intergenerational mobility regarding educational attainment was investigated in two steps. First, statistical tests were conducted to determine whether there was significant evidence that educational attainment has a positive impact on employment status. Second, a linear

regression model was constructed to analyze to what degree childrens' educational attainment is affected by their parents' educational attainment. The data used was collected from the USCB and BLS. All conditions for inference for any z-test are satisfied for data obtained from the BLS, as they collect their data by using the Current Population Survey (CPS), which is, according to the BLS, "selected to be representative of the entire population of the United States". Additionally, the sample sizes are much greater than 30 and satisfy the Central Limit Theorem (CLT). The table "Employment status of the civilian noninstitutional population 25 years and over by educational attainment, sex, race, and Hispanic or Latino ethnicity" provided by the BLS was used as the data to test for if educational attainment positively affects employment status; in other words, does a higher education level make it more likely for one to obtain a job? Due to the large sample sizes of this dataset, any differences in employment status can be expected to not have been caused by random chance; however, multiple statistical tests were conducted to verify this. Each statistical test that was conducted was a two-proportions z-test at the significance level of $\alpha = 0.01$, and compared the employment-population ratio between each level of education and the level below it. Four levels of education were observed: less than high school (HS-), high school (HS), some college or associate degree (C/A), and bachelor's degree or higher (B+). The rows of the provided table detail the total civilian population, labor force population, and employment status of different demographic groups, including men, women, and other racial and ethnic groups. The columns of the table detail the level of education that was achieved, ranging from HS- to B+. The education levels of "some college" and "associate degree" are grouped into the section "some college or associate degree", and the education levels of "bachelor's degree only" and "advanced degree" (greater than a bachelor's degree) are grouped into the section "bachelor's degree or higher".

To investigate to what degree childrens' educational attainment is affected by their parents' educational attainment, the data table "Intergenerational Transition Matrices of Educational Attainment by Race and Gender" provided by the USCB was used. The readme for the data can be found here: https://www2.census.gov/ces/opportunity/table7.pdf. A Python program was used, through Jupyter Notebook, to extract the relevant data, including all rows for "pooled" gender across all racial groups, which includes both male and female, and all conditional columns, which show the probability of a child obtaining a certain level of education given their parent's education level. The rows were all averaged together to yield the average probability of a child achieving a certain level of education based on their parent's education, and the data was reorganized into a table comparing the average probability of a child achieving a certain level of education based on their parent's education. This table will be shown later in the analysis.

Results and Analysis

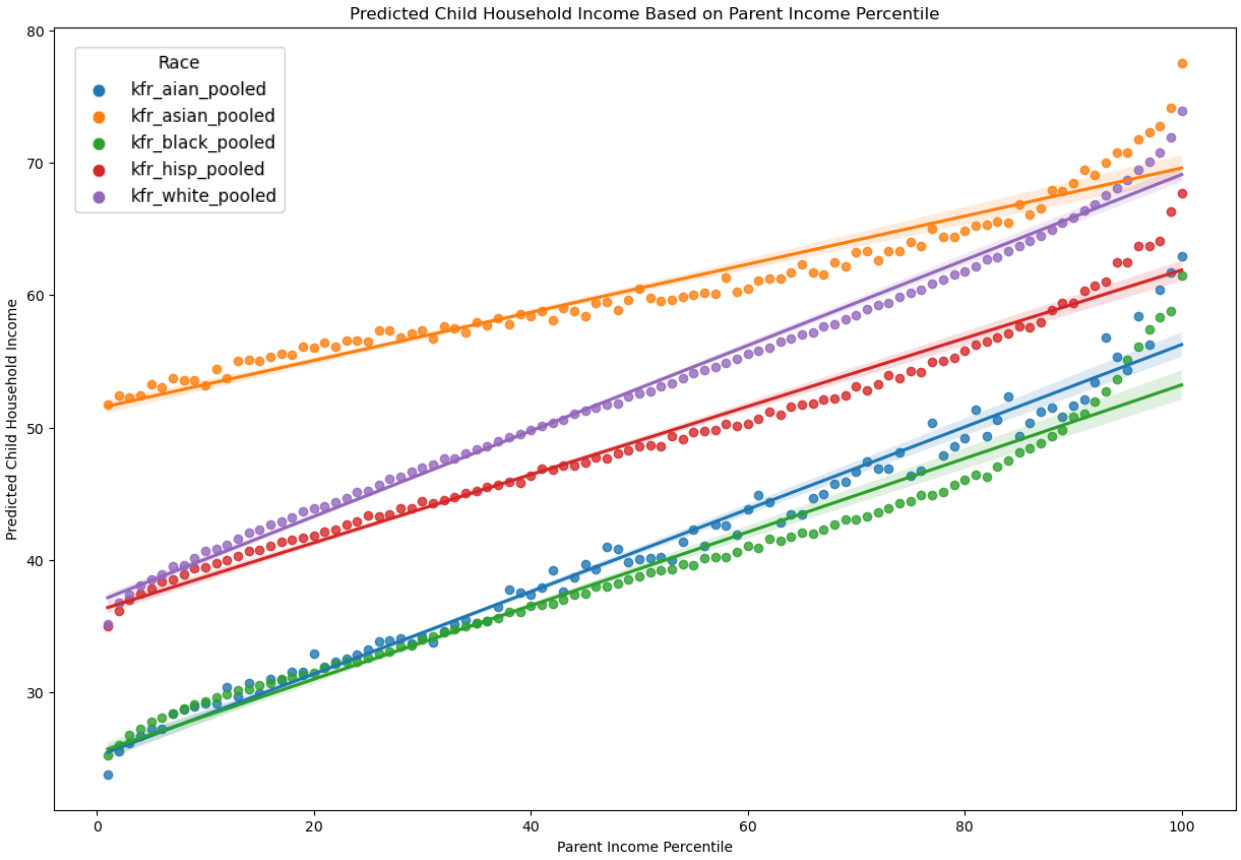
The results for each of the statistical tests and regression models are described in this section.

Parent Income

A linear regression model was constructed, which calculated the line of best fit for each racial group. The r-squared score for the entire model, which includes all racial groups, was 0.964159245630059, indicating that 96.42% of the variability observed in child household income is explained by the regression model. The r-squared score yields a correlation coefficient of 0.98191610926, indicating that there is a strong, positive linear relationship between parent income and child household income.

Figure #1

Scatterplot of the data with line of best fit for each racial group



Note: This graph was constructed using the seaborn library in Jupyter Notebook. Blacks and AIANs noticeably have the lowest projected child household income.

The regression equations for each racial group are provided below:

AIAN: y = 25.15804848 + 0.31108221x

Asian: y = 51.43184848 + 0.1817931x

Black: y = 25.42393939 + 0.2779913x

Hispanic: y = 36.1377697 + 0.25761248x

White: y = 36.82031515 + 0.32299772x

The coefficients for the equations for whites and AIANs are the greatest among all other equations, while the coefficient for the equation for Asians is much lower than all other equations, which suggests that parent income most affects people of white race, with AIANs close behind, while Asians are the least affected by parent income. From this, it can be observed that more parent income is translated into child household income for people of white or AIAN race, meaning that children from these groups receive the most benefits from a higher earning parent. Note that this does not mean that AIANs are an advantaged group, as will be evidenced shortly. Whites start with the second highest predicted household income, while Asians start with the greatest predicted household by a substantial margin; however, because whites are more affected by parent income, whites with parents with income at the top percentile have nearly the same projected income as Asians with parents with income at the top percentile. For children with parents with income at the top percentile, Asians and whites have the greatest projected household income, with Hispanics close behind. AIANs and Blacks have a much lower projected household income, with the latter having the least projected household income for children with parents at the top percentile. Most importantly, not only do AIANs and Blacks have the least projected household income, but they also have the second highest and third highest coefficients respectively, meaning that they are highly affected by parent income and experience a greater increase in child household income for an increase in parent income than all other equations besides the equation for whites. This is a critical factor that will be discussed in the discussion.

Educational Attainment

As mentioned in the "Methods" section, the relevant data from the "Employment status of the civilian noninstitutional population 25 years and over by educational attainment, sex, race, and Hispanic or Latino ethnicity" table provided by the BLS was extracted to conduct the

statistical tests. A table of the extracted data is shown below (note: for accuracy and greater consistency, the employment-population ratios shown in the table are my own calculations as opposed to the ratios provided by the BLS). As a refresher, the following abbreviations are being used to represent each level of education: HS-, less than high school; HS, high school; C/A, some college or associate degree; and B+, bachelor's degree or higher.

Table #1

Employment Status of Total US Civilian Noninstitutional Population

Employment status	Less than a high school diploma (HS-)	High school graduates, no college (HS)	Some college or associate degree (C/A)	Bachelor's degree and higher (B+)
Civilian noninstitutional population	19,789	63,707	56,540	85,981
Employed	8,498	34,497	34,523	61,394
Employment-population ratio	42.943	54.149	61.059	71.404

Source: United States Bureau of Labor Statistics, 2022

The first statistical test compares the employment-population ratios of the categories HS- and HS, where $n_1 = 19789$ = the civilian noninstitutional population of people who are in the category HS-, $x_1 = 8498$ = the total number of people in n_1 who are employed, $n_2 = 63707$ = the civilian noninstitutional population of people who are in the category HS, $x_2 = 34497$ = the total number of people in n_2 who are employed, $\hat{p}_1 = x_1/n_1 = 42.943$ = the employment-population ratio of the category HS-, and $\hat{p}_2 = x_2/n_2 = 54.149$ = the employment-population ratio of the category HS. A two-proportions z-test was conducted at the significance level of $\alpha = 0.01$ with the two hypotheses shown below:

H_0: $p_1 = p_2$ (the employment-population ratio of all US citizens in the category HS- is equal to the employment-population ratio of all US citizens in the category HS),

H_a: $p_1 < p_2$ (the employment-population ratio of all US citizens in the category HS- is less than the employment-population ratio of all US citizens in the category HS).

The resultant *z*-statistic was -27.553, yielding a *p*-value of essentially zero; the true result was too small to be displayed on the calculator that was used. Since the *p*-value is less than the significance level of $\alpha = 0.01$, the null hypothesis is rejected; there is convincing evidence that the employment-population ratio of all US citizens in category HS- is less than the employment-population ratio of all US citizens in category HS. This means that the employment-population ratio of all US citizens in the category HS is highly likely to be greater than the employment-population ratio of all US citizens in the category HS-, suggesting that a higher education level makes it more likely for one to obtain a job.

The last two statistical tests compare the categories HS with C/A and C/A with B+. They follow the same exact structure as the statistical test comparing the categories HS- with HS shown above, and have the resultant *z*-statistics -24.185 and -40.727 respectively, which both yielded a *p*-value of essentially zero; the true value small to be displayed on the calculator that was used. The variables and hypotheses for both of these statistical tests are defined below.

For the second statistical test comparing the employment-population ratios of the categories HS and C/A, $n_1 = 63707$ = the civilian noninstitutional population of people who are in the category HS, $x_1 = 34497$ = the total number of people in n_1 who are employed, $n_2 = 56540$ = the civilian noninstitutional population of people who are in the category C/A, $x_2 = 34523$ = the total number of people in n_2 who are employed, $\hat{p}_1 = x_1/n_1 = 54.149$ = the employment-population

ratio of the category HS, and $\hat{p}_2 = x_2/n_2 = 61.059$ = the employment-population ratio of the category C/A. A two-proportions z-test at the significance level of $\alpha = 0.01$ was conducted with the two hypotheses shown below:

H_0: $p_1 = p_2$ (the employment-population ratio of all US citizens in the category HS is equal to the employment-population ratio of all US citizens in the category C/A),

H_a: $p_1 < p_2$ (the employment-population ratio of all US citizens in the category HS is less than the employment-population ratio of all US citizens in the category C/A).

For the third and final statistical test comparing the employment-population ratios of the categories C/A and B+, $n_1 = 56540$ = the civilian noninstitutional population of people who are in the category C/A, $x_1 = 34523$ = the total number of people in n_1 who are employed, $n_2 = 85981$ = the civilian noninstitutional population of people who are in the category B+, $x_2 = 61394$ = the total number of people in n_2 who are employed, $\hat{p}_1 = x_1/n_1 = 61.059$ = the employment-population ratio of the category C/A, and $\hat{p}_2 = x_2/n_2 = 71.404$ = the employment-population ratio of the category B+. A two-proportions z-test at the significance level of $\alpha = 0.01$ was conducted with the two hypotheses shown below:

H_0: $p_1 = p_2$ (the employment-population ratio of all US citizens in the category C/A is equal to the employment-population ratio of all US citizens in the category B+),

H_a: $p_1 < p_2$ (the employment-population ratio of all US citizens in the category C/A is less than the employment-population ratio of all US citizens in the category B+).

As stated previously, the resultant *p*-value of these two statistical tests is essentially zero; thus, for both statistical tests, the *p*-value is less than the significance level of $\alpha = 0.01$, and the null hypothesis is rejected. All the statistical tests that were performed suggest that a higher education level increases the probability of obtaining a job; thus, it is clear that a higher education level has some sort of positive impact on one's economic status, and educational attainment is an important factor to consider in regards to social mobility. With this in mind, I now present the results of the analysis of how parents' education level affects their childrens' education level. A table of the relevant data that was extracted from the dataset provided by the USCB is shown below.

Table #2

Probability of Children Achieving a Bachelor's Degree or Higher Based on Parent Education Level

Parent Education Level	Child Education Level Probability			
	HS-	HS	C/A	B+
1 (HS-)	0.209516	0.331340	0.28958	0.169586
2 (HS)	0.110456	0.327560	0.33040	0.231618
3 (C/A)	0.080688	0.230100	0.39394	0.295260
4 (B+)	0.051196	0.141942	0.31140	0.495440

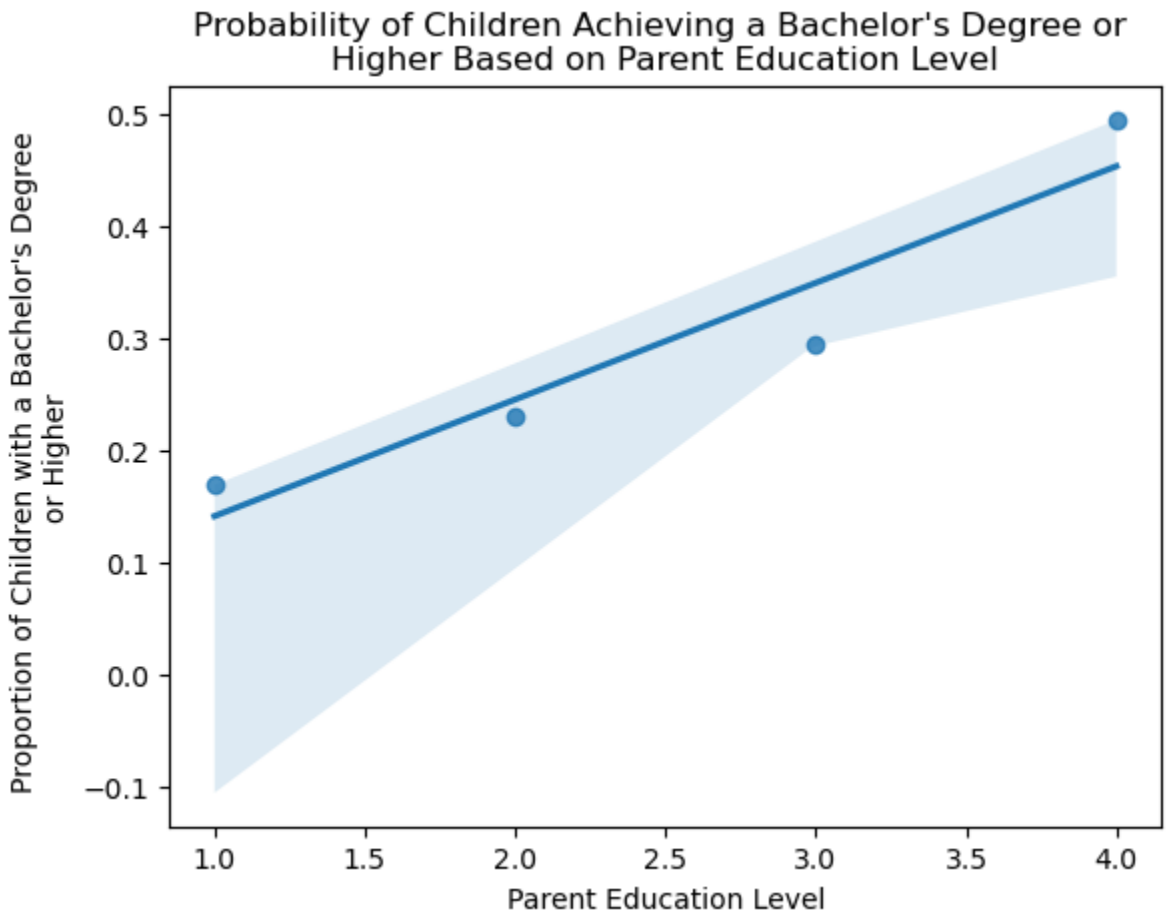
Source: United States Census Bureau, 2018

To allow for linear regression, parent education level was quantified, where each increase of one parent education level represents an increase from one level of education to another; as seen in the table, 1 represents less than high school, 2 represents high school, 3 represents some college or associate degree, and 4 represents bachelor's degree or higher. The two extremes of

the data were analyzed, those being the child education levels of 1 and 4; a linear regression model was created for both. The results are documented below.

Figure #2

Linear regression graph for parent education level vs. child education level 4



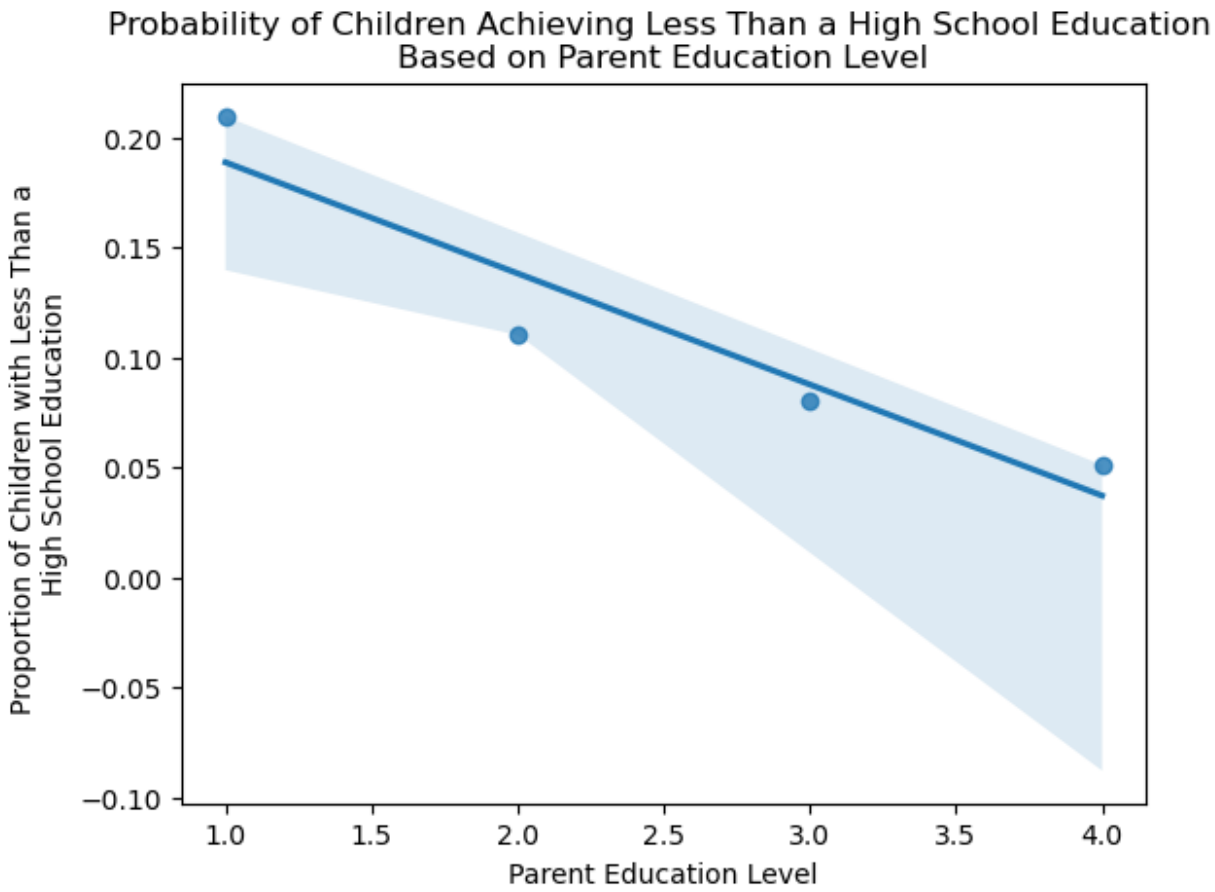
Note: This graph was constructed using the seaborn library in Jupyter Notebook. There is an upwards trend in the probability of a child achieving a bachelor's degree or higher as parent education level increases.

The regression model for child education level 4 represents the probability of a child achieving a bachelor's degree or higher given their parent's education level. The equation for the regression was y = 0.037675 + 0.1041204x, where y represents the probability of a child

achieving a bachelor's degree and x represents the parent's education level. The r-squared score for the regression is 0.9051293902389558, indicating that 90.51% of the variability in the proportion of children who achieve less than a high school education is explained by the regression model. The r-squared score yielded a correlation coefficient of 0.95138288309, indicating a strong, positive linear relationship between parent education level and child education level. The equation suggests that on average, an increase in one level of parent education translates to roughly a 10.41% increase in the probability that the child will achieve a bachelor's degree or higher.

Figure #3

Linear regression graph for parent education level vs. child education level 1



Note: This graph was constructed using the seaborn library in Jupyter Notebook. There is a downwards trend in the probability of a child achieving a bachelor's degree or higher as parent education level increases.

The regression model for child education level 1 represents the probability of a child achieving a less than a high school education given their parent's education level. The equation for the regression was y = 0.239146 + -0.0504728x, where y represents the probability of a child achieving a less than a high school education and x represents the parent's education level. The r-squared score for the regression is 0.8979185524704167, indicating that 89.79% of the variability in the proportion of children who achieve less than a high school education is explained by the regression model. The r-squared score yielded a correlation coefficient of 0.94758564387, indicating a strong, positive linear relationship between parent education level and child education level. The equation suggests that on average, an increase in one level of parent education translates to roughly a 5.05% decrease in the probability that the child will achieve less than a high school education.

Discussion & Conclusion

The results suggest that the socioeconomic status of parents has a significant effect on the socioeconomic status of their children. In order to promote greater equality of opportunity, steps could be taken to balance the effect that parent socioeconomic status has on child socioeconomic status across different demographic groups.

Parent Income

The results from the regression models indicated that AIANs and Blacks have the lowest projected household income out of all other groups. Additionally, AIANs and Blacks are affected by parent income to a higher degree than any other racial group that was analyzed, apart from

Whites. To help promote equality of opportunity, it is reasonable for parental financial support to be prioritized for Blacks and AIANs; parental financial support could also be given to Hispanics, as they are the third most disadvantaged group, though they may need the support less. Since AIANs and Blacks are affected by parental income to a relatively high degree, prioritizing financial support to AIANs and Blacks may also be more effective at narrowing the income gap, as they will experience a greater increase in child household income.

Educational Attainment

The results of the statistical tests reinforce the idea that higher level education has a positive effect on finding a job, highlighting the importance of education in society. The regression model shows how parent education level strongly correlates with child education level; the better the parent education level, the more likely the child is to do better. Interestingly, for parent education levels 1 and 2, children are most likely to achieve an education level of less than high school, followed by achieving an education level of high school. In other words, families that have parents that do not at least have some college degree are likely to produce children that also do not achieve at least some college degree, which creates a repeating cycle. As stated before, higher level education is shown to increase the probability of becoming employed, meaning that families in this cycle are also more likely to be unable to afford a higher level education, which could be a reason why the children are unlikely to achieve some type of college education or degree. This can be mitigated by providing educational support to financially struggling families, such as promoting programs like scholarships and greater financial aid in college.

Limitations

Due to time constraints, I was unable to analyze all factors that affect intergenerational mobility. These include the effect of race and gender on educational attainment and the effect of immigration status and parent nationality on child income. The datasets including these pieces of information are available, so further analysis can definitely be done. For immigration status, the average income of immigrants could be compared to the average income of native US citizens by creating a linear regression model for each group, similar to how I compared the different racial groups for parent income and child household income. If possible, a statistical test could be done to see whether the difference in income between the two groups is statistically significant. For parent nationality, I have access to two datasets that include parents from a wide variety of different countries; it could be possible to construct linear regression models of each country and conduct statistical tests comparing the child income of parents from each country to the US national average.

The effect of gender on child educational attainment was actually lightly analyzed in the process of creating this research paper, but was not included due to a lack of further research. A statistical test I conducted on the USCB data table "Intergenerational Transition Matrices of Educational Attainment by Race and Gender" found that while men and women have the same proportion of parents who have a bachelor's degree or higher, more women than men actually achieve a bachelor's degree or higher; another statistical test on the data table "Employment status of the civilian noninstitutional population 25 years and over by educational attainment, sex, race, and Hispanic or Latino ethnicity" from the BLS also reflects this finding. However, despite this, men have a greater employment-population ratio across all education levels compared to women; in fact, the employment-population ratio for men with some college (no

degree) or an associate degree is nearly the same as the employment-population ratio for women with a bachelor's degree or higher. Further analysis could be done to investigate the reason behind this phenomenon, which could help with creating a method to promote greater equality of opportunity among children of different genders, but I did not have enough time to research this in more depth.

A similar strategy to the method used to analyze the effect of gender provided in the previous paragraph could be used to analyze the effect of race on educational attainment, only instead of comparing men to women, different racial groups would be compared with each other and the national average.

References

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2019). Race and Economic Opportunity in the United States: An intergenerational perspective. *The Quarterly Journal of Economics*, *135*(2). https://doi.org/10.1093/qje/qjz042

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623. https://doi.org/10.1093/qje/qju022

Kochhar, R., & Cilluffo, A. (2018, July 12). *Income Inequality in the U.S. Is Rising Most Rapidly Among Asians*. Pew Research Center's Social & Demographic Trends Project; Pew Research Center. https://www.pewresearch.org/social-trends/2018/07/12/income-inequality-in-the-u-s-is-rising-most-rapidly-among-asians/

Torche, F. (2015). Intergenerational Mobility and Equality of Opportunity. *European Journal of Sociology / Archives Européennes de Sociologie / Europäisches Archiv Für Soziologie*, *56*(3), 343–371. https://www.jstor.org/stable/26573215

US Bureau of Labor Statistics. (2019, January 18). *Employment status of the civilian noninstitutional population 25 years and over by educational attainment, sex, race, and Hispanic or Latino ethnicity*. Bls.gov. https://www.bls.gov/cps/cpsaat07.htm

US Census Bureau. (2020, October 26). *Race and Economic Opportunity Data Tables*. The United States Census Bureau. https://www.census.gov/programs-surveys/ces/data/public-use-data/race-and-economic-opportunity-data-tables.html

Additional Files

The Python code that was used to organize the data and construct the linear regression models can be found at the links provided below.

Parent Income:

https://anaconda.cloud/api/nbserve/launch_notebook?nb_url=https%3A%2F%2Fanaconda.cloud%2Fapi%2Fprojects%2F9fc7f3fb-8cc5-4f46-ae2b-2776ff43228d%2Fversions%2F9ecfcac3-dfda-4328-8ab7-69c1c88c7e11%2Ffiles%2FParent%20Income.ipynb

Parent Education Level:

https://anaconda.cloud/api/nbserve/launch_notebook?nb_url=https%3A%2F%2Fanaconda.cloud%2Fapi%2Fprojects%2F594d861a-3db4-4ca3-bd06-7e9ccd935107%2Fversions%2F3d97c0fa-4874-4bc7-93e7-1e2f936df62f%2Ffiles%2FParent%20Education%20Level.ipynb