**INCOME INEQUALITY EXPLAINED BY EDUCATION INEQUALITY OF COMMUNITY**

By Adil Ryskulov and Zhanna Sarsenova

ABSTRACT: *Contemporary U.S. society faces issues of income inequality between the two genders and the main three racial communities. This income inequality is caused by education inequality between three communities. The higher education level positively correlated with income levels. Thus, the difference in education levels distributions caused differences in income levels.*

1. **INTRODUCTION**

As it is generally known the higher education associates with better wages. The difference in education level represents a difference in skills each individual can offer firms who hire them. Thus, in general, a better individual’s education gets the potential wage rises as well as carrier opportunity. However, this principle is not universally true among all individuals.

The tendency of concentration of individuals within their ethnic, race, religious, or cultural groups is not something uncommon. Many people live within the community in which they have some kind of relationship. The example of such concentration is China Town and Little Italy in Manhattan which have a high concentration of Chinese and Italians within those neighborhoods based on their ethnic background, and Borough Park in Brooklyn which have a concentration of orthodox Jews based on their religious believes. Such concentration creates externalities in wages of members of these communities as close relationships with individuals within the community allow them to build a connection that will allow them to succeed in education and future job placement.

The community effect, where individuals connect, occurs not only based on the geographical residency. The community effect can occur between individuals who share similarities such as the same race or religious beliefs. Thus, the community effect has a broader effect on individuals and members of the community will more likely to follow and support the members of their community over the member's other communities.

In this study, we investigate how education distribution within community will effect on individuals wages. The community effect has key role in education and job placement as more educated community will more likely to influence their younger generation to obtain better education. Also, availability of connection with advanced degrees will make life easier as these individuals give advice with carrier opportunities and directly support by recommending what pass to take to achieve goals. Thus, this study will focus on average education level and education inequality within three main racial groups in State of New York.

1. **LITERATURE REVIEW**

Chun-Hung A. Lin in his study “Education Expansion, Education Inequality, and Income Inequality: Evidence from Taiwan, 1976 – 2003” investigates the relationship between education inequality and income inequality. In this work, Lin states that the decline in education inequality will stimulate the reduction of income inequality by government expansionary policy in education with the goal of increasing average schooling years. The evidence in Taiwan’s time-series study shows the rise in the average schooling year will reduce income inequality within society. Educational Gini coefficient and average schooling years allowing to measure income inequality represented by the income Gini coefficient.

Noman Arshed, Awais Anwar, Nabeela Kousar, and Samra Bukhari in their study "Education Enrollment Level and Income Inequality: A Case of SAARC Economies" proposed the idea that education forms a quadratic relationship with income inequality. In their study, the authors examine the law of diminishing returns of education level in terms of income. The analysis reveals a negative sign in quadratic for the quadratic term. This concludes that higher education level will increase the total income, however, diminishes marginal return from achieving this level in education.

1. **DATA**

The data for this study was provided by The Census Bureau's American Community Survey (acs2017\_ny) for the state of New York for the year 2017. This survey consists of cross-sectional data to perform a comparative analysis between education and income levels among different races and genders within New York State boundaries. The acs2017\_ny data filtered to the level of where it consists only individuals within the labor force, working full time within usual weakly work hours of 37-43 hours, and those who only recognize themselves as one of white, Asian, and African American races.

1. **MODEL**

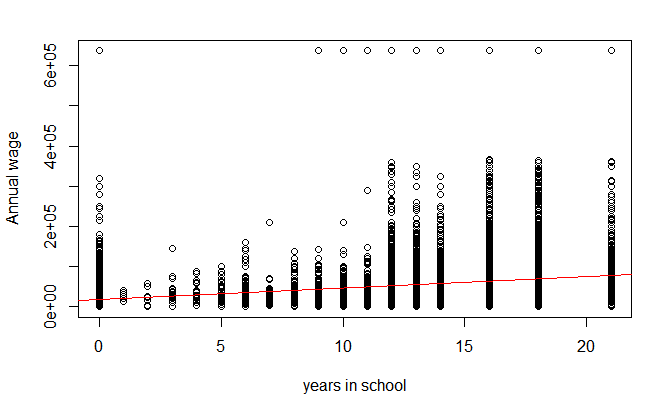
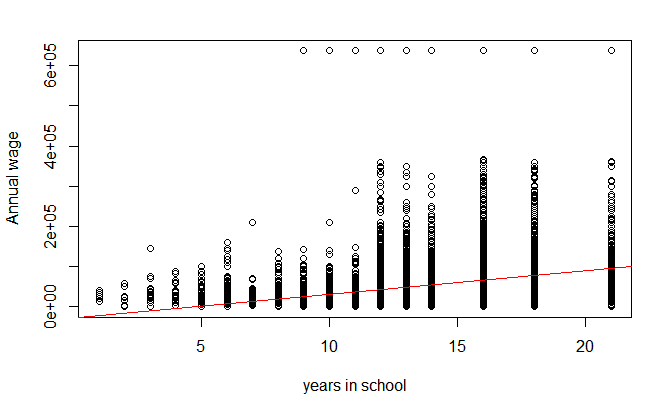
**4.1 Building Model**

The income inequality among different races is one of the main social-political issues in the United States. One of the explanations of differences in income between different races is the education level of individuals. The concentration of individuals within their communities creates externalities for the individual within these communities. The community effect causes the younger generation to follow patterns of the older generation. If the older generation being highly educated creates positive externalities for yonder generation to obtain a good education. The average return from obtaining a particular education level is summarized in Table 1.

**Table 1**: Average annual income from the wage by education level.

|  |  |
| --- | --- |
|  | Mean wage |
| Doctoral degree | $101,824.77 |
| Professional degree | $125,695.37 |
| Master's degree | $78,586.12 |
| Bachelor's degree | $66,873.61 |
| Associate's degree | $49,901.57 |
| GED or alternative credential | $40,629.55 |
| Regular high school diploma | $41,608.54 |

Information from Table 1 reveals a positive correlation between further obtaining a higher education level and benefiting via higher wages. Thus, the income differences among different races can be explained through differences in the proportion of the population by education level. The communities with a small proportion of the population with degrees higher than a high school diploma will have small average wages. In comparison, the community with a large proportion of the population with degrees higher than a high School will have higher average wages. This happens as the younger generation exposed to individuals with higher achievement who can provide advice and recommendation for future carriers serves as a connection for future job placement and allows to go to school as family income will be able to afford it.

As it is already known, the education level and income levels have a positive correlation. Another good indicator is also years of education. The correlation of years obtained in schooling and annual wages revealed a week positive correlation which equal to 0.2157. However, two main outliers exist within this data. First of all, individuals with professional degrees such as nurses and doctors. These individuals obtained professional licensing obtained after a bachelor's degree and each licensing requires a unique additional education. Thus, there are no universal education years for these groups. The second group, are individuals with zero years of schooling. These individuals have higher income in comparison to those who have some school years and could have a higher income as these are highly skilled individuals who obtained their education abroad. Thus, they do not report their international education as education to Census.

Excluding the two groups listed above from the data give more clear correlation. The excluded data also shows a positive correlation between years in school and annual wages which equals to 0.2822, suggesting of week correlation. As correlation between schooling years and income remains week, the correlation in excluded data is relatively stronger.

**Table 2**: Average annual income, average schooling year, and education Gini coefficient among whites, Asians, and African Americans.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean wage | Mean school year | Gini education |
| White | $57,421.85 | 14.4218 | 0.0956 |
| Asian | $59,539.60 | 14.6788 | 0.1095 |
| Black | $47,181.34 | 13.7701 | 0.0900 |

Table 2 above summarizes average annual wages, mean schooling years, and the education Gini coefficient for white, Asian, and African American races. The data reveals that Asians have the highest average annual wages among the three races. Also, Asians have the highest average schooling years and highest education inequality. Meanwhile, African Americans have the lowest average annual wages among the three races. Also, the lowest average schooling years, and lowest education inequality. Education inequality is not a good indicator in determining how well is a community standing. As we can see, African Americans have the lowest educational Gini coefficient they also have the lowest average schooling years. This means that majority of the African American population have equal education levels, however, this education level is the lowest in comparison to whites and Asians.

**Table 3**: population distribution by education level.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Graduate | Undergrad. | Some Coll. | HS diploma | No HS diploma |
| White | 16.70% | 26.90% | 25.95% | 27.13% | 3.31% |
| Asian | 19.16% | 36.13% | 15.56% | 21.87% | 7.27% |
| Black | 9.95% | 22.22% | 34.54% | 34.54% | 4.97% |
| Male | 12.26% | 25.74% | 24.47% | 32.55% | 4.99% |
| Female | 20.03% | 28.78% | 24.19% | 22.34% | 2.86% |

Table 3 summarizes population distribution by education level. The data from the table above shows that 19.16% of the Asians hold advanced degrees and 36.13% hold a bachelor's degree. While only 9.95% of African Americans hold advanced degrees and 22.22% hold a bachelor's degree. This distribution evidence a low average income from the wages of African Americans comparing to Asians and whites is due to unequal distribution among the population. Thus, the education Gini coefficient with average schooling years among racial groups potentially serves as an alternative and more accurate indicator in estimating an income from wages.

**4.2 Regression Models**

As we discussed above, the community effects influence the younger generation in their education and job placements. The Asian communities have the highest proportion of individuals with advanced and bachelor's degrees. Thus, the younger generation within communities will benefit by having a higher probability to achieve higher degrees and sophisticated career pathways. This happens due to individuals in Asian communities have connections to increase their chances for admissions into the school and job placements.

We start with an econometric model without considering the community effect on income. First of all, control variables are selected based on general factors such as age, education level, and gender. The skills and experience acquired throughout a career, older employees have more working experiences and sophisticated skills and more valuable to firms and in result earn more. The education level demonstrates the skills and abilities of employees to create value for employers. Education level is the most sophisticated indicator of an employee's skill sets and determines an employee's earning. In general, females are discriminated against in hiring and promoting processes. This happed as employers fearing a risks of wasting time on trainings of females as they might get marry which will limit their ability to wok or be flexible with scheduling.

The relationship between age, education level, and gender are listed in the linear function above. The earlier ages, in general, are considered to be the most prolific when yang employees acquire the most skills. In the letter years, employees will acquire specific skills such as how to establish a working environment during a pandemic. Thus, the return from seniority will have a diminishing marginal return from the rise in wages and variable was introduced to model to measure, a negative value is expected. Also, the female who has a larger distribution of advanced and bachelor degree holder are discriminated against in career opportunities and the marital status could be a disadvantage for females in job placement. Thus, variable is introduced to model to measure an effect of marital status to female’s income expectations. As married females considered to bare risks a positive value is expected.

To quantify an effect of the overall community's education level distribution on community member's income from wages average schooling year and educational Gini coefficient of the community introduced to the model. The average education level measures the level of the overall community. As the entire community is relatively more educated all members of the community considered to be better off. Thus, a positive value of expected. Also, the education Gini coefficient measures the community's inequality in education it is important to consider whether the community is dramatically unequal in education. The Gini coefficient alone is not useful as it does indicates where the community overall is standing. Thus, this coefficient only can supplement the average schooling years. The negative value is expected as it is more beneficial to have most people to be equal in education level, so they will be able to utilize their connections.

Finally, we introduced an individual schooling year and the number of children under 5 years old into the model. The community's average schooling year can only estimate an effect of community in an individual's earnings, while to utilize these connections individuals must have an equivalent education level. Which means, positive value of expected. In addition, every following year in education expected to boost earning the marginal growth expected to diminish, which estimates and negative is expected. Children under the age of 5 are not eligible for public schools and require either expensive daycare services or a reduction in income due to restricted scheduling. Which means, children under 5 will reduce earnings and the negative value of expected.

The Ordinary Least Square estimated results are shown in Table 4. We estimated three different models described above by introducing new control variables to existing model by creating a new model. The baseline for these models estimation was chosen an education level without high school diploma.

The first model followed all expectations: all variables in this model are statistically significant at a 0.1% significance level. The coefficient estimates have the estimated values either positive or negative. The age squared variable which measures the diminishing marginal return of seniority is negative, however, the estimated value of -27.20 is below expectation relative to the age variable - which equals 3,132.94. As it is expected, the females are discriminated against in hiring and promoting processes as they observe a negative coefficient of -10,573.63, while the variable which determines whether a female is not married has a positive estimated coefficient. This supports an argument where employees consider the risk of females being married will restrict their ability to work and have flexibility in scheduling. Employers are ready to hire females in case if they will be able to discount their wages, also employers are ready to offer a premium for unmarried female employees.

As the first model has all variables statistically significant this model is not accurately estimated wages within New York State residents as independent variables only describe income from wages only in 13.76% of cases of observation. This means the introduction of new variables can improve the model’s estimation.

**Table 4**: OLS estimated results for estimating earnings and components effecting.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Coefficients: |  |  |  |  |  |  |  |  |  |
|  | Estimate | Std. Error |  | Estimate | Std. Error |  | Estimate | Std. Error |  |
|  | -47,253.59 | 2,892.37 | \*\*\* | -173,600.00 | 17,790.00 | \*\*\* | -162,402.32 | 18,987.84 | \*\*\* |
|  | 3,132.94 | 120.71 | \*\*\* | 3,144.00 | 120.70 | \*\*\* | 3,091.00 | 120.52 | \*\*\* |
|  | -27.20 | 1.33 | \*\*\* | -27.35 | 1.33 | \*\*\* | -26.32 | 1.33 | \*\*\* |
|  | 51,270.52 | 1,448.30 | \*\*\* | 50,710.00 | 1,451.00 | \*\*\* | 20,171.39 | 4,161.25 | \*\*\* |
|  | 38,733.50 | 1,385.97 | \*\*\* | 38,340.00 | 1,387.00 | \*\*\* | 19,573.48 | 3,265.47 | \*\*\* |
|  | 18,188.81 | 1,390.45 | \*\*\* | 18,040.00 | 1,394.00 | \*\*\* | 8,533.34 | 2,475.49 | \*\*\* |
|  | 8,035.52 | 1,379.89 | \*\*\* | 8,018.00 | 1,382.00 | \*\*\* | 3,008.64 | 2,029.28 |  |
|  | -10,573.67 | 559.49 | \*\*\* | -10,280.00 | 560.50 | \*\*\* | -10,105.10 | 559.21 | \*\*\* |
|  | 5,165.64 | 554.25 | \*\*\* | 5,320.00 | 554.30 | \*\*\* | 4,956.41 | 553.69 | \*\*\* |
|  |  |  |  | 10,070.00 | 1,576.00 | \*\*\* | 9,554.04 | 1,572.82 | \*\*\* |
|  |  |  |  | -191,500.00 | 80,480.00 | \* | -193,470.62 | 80,325.00 | \* |
|  |  |  |  |  |  |  | -2,427.72 | 1,053.32 | \* |
|  |  |  |  |  |  |  | 208.84 | 37.61 | \*\*\* |
|  |  |  |  |  |  |  | 6,522.51 | 614.84 | \*\*\* |
|  |  |  |  |  |  |  |  |  |  |
|  | 0.1376 |  |  | 0.1388 |  |  | 0.1435 |  |  |
|  | 36,649 |  |  | 36,647 |  |  | 36,644 |  |  |

“\*\*\*” Indicates significance at 0.1% level.

“\*\*” Indicates significance at 1% level.

“\*” Indicates significance at 5% level.

“ . “ Indicates significance at 10% level.

“ “ Indicates significance at more than 10% level.

The second model experienced an introduction of two new variables. With the introduction of new variables, existing variables remained statistically significant with a 0.1% significance level. The average schooling age of racial community is also statistically significant with a 0.1% significance level. Meanwhile, an education Gini coefficient also statistically significant the significance level differentiated from other variables and significant at a 5% significance level. This means it fits with the 95% requirement and can be used in estimating income from wages. The r-squared value increased from 0.1376 to 0.1388. This means, an introduction of these two variables describing community effect on earnings expected to increase the accuracy of the model.

The third model experienced an introduction of variables which estimates the effect of an individual’s education level in utilizing connections within the community and the effect of a child under the age of 5 years on earnings. The variable measuring schooling years has a negative coefficient, which was not expected and this variable is statistically significant at a 5% significance level. The squared value of schooling years estimates a diminishing return of additional years of education and is expected to have a negative coefficient, however, the coefficient is positive and the variable is statistically significant at a 0.1% significance level. Lastly, the variable measuring number of children under the age of 5 is positive, against expectation, and statistically significant at a 0.1% significance level. The high school diploma is not statistically significant anymore and estimated to be equally valuable as not having a High School diploma at all.

All introduced variables revealed values against expectations. The explanation is based on the initial assumption of their effect.

The schooling years were expected to measure how efficiently community members will be able to utilize their connections within the community. It was suggested that equivalent years in schooling will be more beneficial to find comparable job opportunities. However, this variable measures the efficiency of education level. The education degrees in the third model significantly different in comparison with the other two, which is due to the introduction of discussing variables – which devaluated them.

Meanwhile, the squared schooling year’s variable was expected to estimate the negative coefficient which would evident for diminishing returns from an additional year of schooling. However, the model shows an opposite effect where an additional year in school accelerates returns in the form of income.

Finally, the main disappointment was an estimated coefficient of the number of children under the age of 5. It was expected to have this coefficient to be negative as each child under the age of 5 requires more attention and as a result leads to a reduction in wages. However, the model reveals a completely opposite effect where the presents of children under the age of 5 will increase an annual wage by about $6,500 per child. Which is contradicts an intuition. The issues of this variable rely on the data we choose. In this work, we limited our data to individuals who are active in labor force participation and usually work between 37 to 43 hours per week. This means we initially eliminated those individuals who have children under the age of 5 and cannot work full time. Thus, the remaining data represents relatively healthy individuals who are able to afford to hire staff who will take care of their children. Also, the data does not exclude male individuals who have kids under 5 and whose spouses not in the labor force. Thus, this variable is misleading and cannot be considered in policy determination as it suggests more children will lead to higher wages.

The third model has the r-square value of 0.1435, which is higher than the previous two models. However, considering the fact that this model has a variable that misleads estimates and a variable which devaluates education level’s values – it cannot be considered to be useful in policy recommendations even it more accurate than previous models.

1. **CONCLUSION**

As it proposed earlier, the community effect is the main driver of income inequality among three racial groups. Asians who have highest income level also the group with highest distribution of higher education levels, highest average schooling years, and highest education inequality. The whites closely follow Asians in this patters, thus, these two groups have similar average wage levels. The data excluded individuals who do not have any schooling and individuals with professional degrees, which will further increase an income gap between Asians and African Americans.

The income gap between Asians, whites, and African Americans caused by education distribution and community effect. This study proposed that wages of particular community member is driven by entire community and average wages between these three races will reveal significant gap, suggesting about racial discrimination. However, closer study reveals that these gap is due to education distribution where African Americans have the lowest percentage of population having advanced degree and bachelor’s degrees and highest distribution of population having high school degree or even not having high school degree.

In addition, community effect where each member if community will prefer to follow or support its community member over other also was empirically evidenced within this study. The average schooling year of entire community and education inequality were selected to measure community effect. The models revealed that these elements are statistically significant to estimate wages. Thus, truly effecting on wages on members.