**Abstract**

**Key words**

Demographic, Education, Attainment, Scores, Equality.

**Introduction**

Despite the national narrative that America affords its inhabitants an unprecedented land of opportunity, intergenerational social mobility, defined as the likelihood that a child born to parents in the bottom fifth of the income distribution reaches the top fifth, is higher in many other advanced countries. Fewer than eight percent of Americans born in the bottom 20% of the income distribution reach the top 20%, whereas more than 13 percent of Canadians do. However, as Harvard Economist Raj Chetty has demonstrated, significant differences in upward mobility rates exist across the United States. “In this country, of all countries, a person’s zip code shouldn’t decide their destiny,” President Barack Obama said in 2015.

We wish to understand which demographic and geographic factors are associated with social mobility. We will use the CollegeDistance data set that is included in the AER R package. It contains survey data from 1100 USA residents who were high school seniors in 1980 with follow-up data regarding education attainment in 1986. While educational attainment is not a perfect proxy for income, we will use it as our outcome variable, our metric of interest. We will start the construction of the regression model with the 13 other variables, including gender, ethnicity, a composite test score, school and community demographic factors, and family income, among others. After exploratory data analysis that includes checking multicollinearity, we will endeavor to construct a multiple linear or multiple logistic regression model that predicts educational attainment based on the independent variables.

While our data captures educational attainment from more than 30 years ago, we think the construction of a model that predicts such a metric is nonetheless a worthwhile endeavor. An interesting follow-up study may be to repeat such a survey in the present day, construct a model, and then analyze differences.

**Literature Review**

We reviewed three papers for our review that look at social mobility where each study had a slightly different approach at analyzing the different independent variables that affect intergenerational social and economic outcome. Broadly, the first paper reviewed observed economic outcomes children in families that moved from a poorer to a better income neighborhood. The second paper reviewed researched the interplay of a child’s gender from poorer neighborhoods versus higher income neighborhoods and those children’s adult economic outcome. While these two papers studies were focused on communities in the United States, our third paper reviewed looked at differences in Denmark that draw heavily from the theories of sociologist Pierre Bourdieu on the concept of “cultural capital”, which is the theory that individuals can possess a form of “capital” quantified by the amount of knowledge and behavior (culture) that promote social mobility in a socially stratified society (Møllegaard & Jæger 2015).

The first paper we looked at, The Impacts of Neighborhoods on Intergenerational Mobility I:Childhood Exposure Effects found that families that moved from a lower to a better income area had children that had greater social mobility than those that stayed. There was a positive improvement in the child’s adult income, which rose at a rate of about 4% that was proportional the number of years spent growing up in the higher income area (Chetty & Hendren 2016). The dataset composed of federal income tax records from 1996 through 2012 and focused on families with children born between 1980 and 1988 and moved across neighborhoods between 1997 and 2010. This paper reconciles conflicting papers by introducing the number of years growing up in the better neighborhood. Interestingly, a symmetrical finding was found for families moving from higher to lower income neighborhoods in which the children of those families had adult incomes that reduced 4% per years lived in the lower income neighborhood (Chetty & Hendren 2016).

The key variables are actually fairly different from the ones we are using from the dataset in AER. For this paper, those variables are parent’s income, parent location, child’s adult earned income, teenage birth, marriage, educational attainment, and the child’s employment at adulthood.

While the primary independent variable is the child’s adult earned income, ours is educational attainment. Our analysis is also more focused on key variables that are characteristic of a neighborhood, whereas this study is more focused on those associated with the family unit. Our dataset also does not contain any interventional events with each family as this study has. A way to take our study further would be to complete the dataset in AER with family income as well as the child’s adult earned income.

The second paper we reviewed, Childhood Environment and Gender Gaps in Adulthood takes a deeper look into the effects of gender and social mobility. Like the first paper we reviewed, the data was drawn from tax records for families with children born in the 1980s. The major finding of this study found that the traditional gender gap in employment and income earnings are reversed in poor families, particularly in high-poverty neighborhoods. Boys from these families are less likely to work in adulthood and the issue is compounded with single-parent families. The authors write:

Low-income boys who grow up in high-poverty, high-minority areas work less than girls. These areas also have higher rates of crime, consistent with a model in which boys with lower latent earnings potential who grow up in environments of concentrated poverty switch from the formal labor market to crime or other illicit activities (Chetty et al. 2016).

Though the paper does a comprehensive analysis on the impact of environment based on economic class and the interplay of gender, a deeper analysis that would include race as a more prominent factor would strengthen this paper significantly. This is especially true if we approach the analysis of these factors (race, class, gender, etc) as observations that are not mutually exclusive.

The last two papers we reviewed included clear and concrete independent variables for their analyses, the third paper we reviewed is different in its approach in many ways. The data quantifies three different forms of capital as measured in three major categories: economic, social and cultural. (Møllegaard & Jæger 2015). Economic capital is self-explanatory in that it’s a measurement of income and assets. Cultural capital is measured in less concrete terms such as years of educational attainment of the parents and grandparents, subscriptions to news outlets, and other things. Social capital is interesting in that it is best understood as a kind of network analysis manifesting as professional networks, friendships, familial connections, etc. The paper also addresses the intergenerational effects on educational attainment as far as two-generations, whereas most studies review a single generation prior.

The findings of the paper found that in Denmark, educational attainment or more academically driven educational attainment depended more on cultural capital than of social/economic capital (Møllegaard & Jæger 2015). While these measures are fascinating and the results of the study show that cultural capital measures are certainly valid and useful independent variables, it’s difficult to apply an appropriate weight to some of the cultural capital examples. Generally, applying the appropriate weight of the impact of various aspects of cultural capital are difficult to accomplish and to justify. Further research is needed to properly assess the impact of specific variables associated with cultural capital to apply a stronger model in predicting educational outcomes. For our project, measuring social and cultural capital appropriately will be unfeasible.

Overall, a common measure of all reviewed papers focus on parental income or economic means to assess the probability of social upward mobility of the child’s adulthood income. This “intergenerational mobility” is the common theme as a measure of social mobility. For our project, while we are not focused on income as a measure, our project is focused primarily on educational attainment to serve as a proxy for social mobility. We are also looking to identify the strongest predictors for upward social mobility as measured in educational attainment.

**Methodology**

*Discuss the key aspects of your problem, data set and regression model(s). Given that you are working on real-world data, explain at a high-level your exploratory data analysis, how you prepared thedata for regression modeling, your process for building regression models, and your model selection.*

**Experimentation and Results:**

*Describe the specifics of what you did (data exploration, data preparation, model building, model selection, model evaluation, etc.), and what you found out (statistical analyses, interpretation and discussion of the results, etc.)*

**Discussion and Conclusions:**

*Conclude your findings, limitations, and suggest areas for future work.*

**References**

Chetty, R., & Hendren, N. (2016). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. doi: 10.3386/w23001

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., & Scuderi, B. (2016). Childhood Environment and Gender Gaps in Adulthood. doi: 10.3386/w21936

Møllegaard, S., & Jæger, M. M. (2015). The effect of grandparents’ economic, cultural, and social capital on grandchildrens educational success. Research in Social Stratification and Mobility, 42, 11–19. doi: 10.1016/j.rssm.2015.06.004

References (to be converted to APA format for project):

Relative national intergenerational social mobility: <http://www.ecineq.org/ecineq_nyc17/FILESx2017/CR2/p256.pdf>

Chetty paper on geography of intergenerational mobility: <http://www.equality-of-opportunity.org/assets/documents/mobility_geo.pdf>

Obama quote: <https://talkpoverty.org/2015/12/17/american-dream-zip-codes-affordable-housing/>

CollegeDistance data info: <http://rdocumentation.org/packages/AER/versions/1.2-7/topics/CollegeDistance>

<http://wps.pearsoned.co.uk/wps/media/objects/12401/12699039/empirical/empex_tb/CollegeDistance_DataDescription.pdf>

**Appendices**

*Supplemental tables and/or figures. §♣R statistical programming code.*