THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

FAMILY BACKGROUND, EDUCATION, AND INCOME MOBILITY: A CLOSER LOOK AT THE AMERICAN DREAM

DUNCAN HAYES
SPRING 2015

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Economics
with honors in Economics

Reviewed and approved* by the following:

Dr. Russell Cooper
Professor of Economics
Thesis Supervisor

Dr. Russell Chuderewicz
Senior Lecturer in Economics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

How much of a person's income is determined by their parents? How do the education levels of the parents and children affect this relationship? What other variables are involved in the transmission of education and income across generations? These are the fundamental question that this paper seeks to address. After a literature review that goes over some of the most prominent work in this field, two models are created using data from the Panel Study of Income Dynamics. Model 1 finds the elasticity of children's income with respect to parents' income to be between 0.13 and 0.38, which is at the lower end of what many of the academic studies in this field find. Model 2, which includes only males, finds the elasticity of sons' income with respect to dads' income to be around 0.30, with some regions as high as 0.45. Overall, the data shows clearly that there is a link between the incomes of parents and the incomes of their children, and that education plays a major role in this link as well.

TABLE OF CONTENTS

LIST OF FIGURES iii

LIST OF TABLES iv

Acknowledgements v

Chapter 1 Introduction 1

Chapter 2 Literature Review 4

Chapter 3 Theoretical Model 14

Chapter 4 Data Analysis 17

Chapter 5 Methodology 22

Chapter 6 Results 29

Chapter 7 Conclusion 38

BIBLIOGRAPHY 41

LIST OF FIGURES

Figure 2.1 Intergenerational Mobility by Quintile 12

LIST OF TABLES

Table 2.1 College Selectivity and Socioeconomic Background 6

Table 4.1 Race of 2011 Survey Respondents...18

Table 4.2 Region of 2011 Survey Respondents 20

Table 5.1 Inflation Adjustment by Year 25

Table 6.1-Model 1 Results..29

Table 6.2 Model 1 T-Statistics..29

Table 6.3 Model 2 Results 32

Table 6.4 Model 2 T-Statistics..32

Table 6.5 Income Correlations by Region 35

Table 6.6 Income and Education by Region...36

Table 6.7 Model 2 Results 40 to 50..37

Acknowledgements

I would like to thank Dr. Russell Cooper and Dr. James Tybout for their help in guiding me through the thesis-writing process. Without them, this thesis never would have become a reality. I would also like to thank the Panel Study of Income Dynamics project at the University of Michigan for collecting and providing such detailed and extensive data. Finally, I would like to thank Dr. Gary Solon, as his work provided the basis for much of this thesis.

Chapter 1

Introduction

For most of its existence, the United States of America has been thought of as the land of opportunity. It was and still is a melting pot of people from all over the world, all chasing the common dream of peace and prosperity, both for themselves and for their children. The ability to become prosperous and wealthy through hard work after being born to a family in poverty is what defines the American Dream. In the modern world, the easiest way towards prosperity as a young person is through the pursuit of education. Education opens doors and allows for many more possibilities and potential career paths, and also helps a person to break free of whatever situation they were born into. To provide equal opportunities for all Americans, it is crucial that the education system work as a filter to separate past economic situations from future economic outcomes in the labor market, and enable Americans from all backgrounds to pursue their dreams. The fundamental idea that individuals have some level of ability to move up and down in socioeconomic status is known as economic or income mobility, and the investigation of this idea is the primary purpose of this paper. Income mobility can be thought of as a measure of the equality of opportunities available, while income inequality is a measure of equality of outcomes.

Unfortunately, many Americans are beginning to lose faith in the dream of high mobility. According to the Washington Post, the percentage of Americans who believe in the American Dream (defined by the Washington Post as the belief that if you work hard, you will get ahead) fell 11% in just two years between 2012 and 2014. Interestingly, this particular survey is broken down by education level, and it shows that 38% of people with less than a high school degree believe in the American Dream in 2014, versus 52% of those with a post-graduate degree (Ingraham, 2014). This again suggests that education has an effect on the perception of how easy it is to get ahead by working hard. Traditionally, the U.S. was seen as a global leader in educating its citizens, but this idea too is falling by the wayside as the U.S. slips down the international rankings for education. One statistic that emphasizes this dramatic fall is that currently a greater percentage of young American men have less education than their parents than have more education, 29% versus 20%. This does not bode well for the future of the country, and the statistics keep getting worse. The U.S. now ranks 12th in the world in the percentage of 25-34 year olds with college degrees (Kristof, 2014). The point of these statistics is that if Americans expect their education system to be the great equalizer, they may want to think again.

Ultimately, a person's income depends on many factors, but one of the most important has to do with their parents: how much income they make and how much education they received can have a huge effect on the incomes of their children. Parents who are more highly educated are more likely to talk extensively to their children starting from a young age, and are more likely to have a lot of books around and read to their children. Parents with more disposable income are more likely to send their children to superior private schools, hire expensive private tutors, and pay for test preparation for important standardized tests such as the SATs.

There is also likely some sort of networking effect: on average, people tend to befriend and spend time with other people that are fairly similar to them, so it makes sense that cohorts of similarly educated people with fairly similar levels of income would group themselves together. There is extensive support in the literature for the idea of peer effects, meaning that, for example, children with rich, well-educated parents who spend most of their time around other well-educated, well-off adults and their children are more likely to become rich and well-educated themselves, and vice versa. This effect is particularly prominent during schooling, and since both the school a child attends as well as a child's attitude toward education are largely influenced by the decisions and means of the parents, peer effects are likely to reinforce the transmission of the parent's socioeconomic situation to the child.

When considering these and many other mechanisms in which levels of income and education are transmitted, it would be no surprise if there is some level of positive correlation between a parent's level of income and education and their children's income and education. This paper seeks to investigate and quantify these correlations, first by seeking an answer in the literature surrounding this topic, and then by using empirical analysis with data from the Panel Study of Income Dynamics at the University of Michigan.

Chapter 2

Literature Review

This section of the paper will provide an overview of the existing literature covering the topics of family background, education, and income mobility, and the relationships between them. Several studies have been done on these relationships, although the effect of education is not the specific focus of most of them. A study by Robert Haveman and Timothy Smeeding (2006) provides some statistics on higher education: who attends, who doesn't attend, and how family background affects what kind of universities students attend. This study provides a solid foundation for other academic work, such as Gary Solon's widely cited 1992 study on income mobility, as well as the follow up study which has a larger scope, performed in 2002 by Solon and Laura Chadwick.

Next, a paper by Anders Bjorklund and Marcus Jantti (1997) compares income mobility in the United States and Sweden, which is a good way to establish a relative measure of how the U.S. is doing. Another study that includes international comparisons of income mobility from several different countries is Solon's 2002 paper "Cross-Country Differences in Intergenerational Earnings Mobility." Finally, for a broad, U.S.-focused overview of the topic, the U.S. Departments of Treasury and Education released a paper called "The Economics of Higher Education" in 2012 which, in addition to being very up to date, has a large amount of data on the relationships this thesis seeks to study. This literature provides a solid summary of the work that has been done on education and income mobility, as well as student backgrounds.

Common knowledge suggests that higher education is very important, but is it equally available to everyone? Haveman and Smeeding delve into this topic in their 2006 study "The Role of Higher Education in Social Mobility". Obviously, the ability for anyone to go to college, get a degree, and improve their chances of landing a high income job is critical to social mobility, as the authors acknowledge: “Higher education is expected to promote the goal of social mobility and to make it possible for anyone with ability and motivation to succeed” (Haveman, Smeeding 2006 p.129). However, students from less wealthy backgrounds, who are the ones who could arguably increase their mobility the most with a college degree, attend college at a lower rate than wealthier students. For students who graduated high school between 1980 and 1982, 80% of students from a family in the top income quartile attended college, against only 57% of students from a family in the lowest income quartile (Haveman, Smeeding 2006 p.130). If only 4-year colleges are counted, the numbers look even more unequal: between 1980 and 1992, the percentage of youth from the lowest income bracket enrolled in a 4-year college actually fell from 29% to 28%, while that same percentage for youth from the highest income bracket rose from 55% to 66% (Haveman, Smeeding 2006 p.130). Clearly, these trends are moving in the wrong direction if the goal is a meritocratic education system: it appears that family wealth as opposed to ability is increasingly the determinant of a person’s educational achievement. This is in direct opposition to the idea of a high mobility society.

In addition, not all higher education is created equal: there is a big disparity in the quality of the education provided by Harvard compared to an average community college, for example. Focusing in on the numbers for the most elite universities in the U.S. shows a stunning degree of inequality. Haveman and Smeeding reference an analysis performed by Anthony Carnevale and Stephen Rose, who split U.S. colleges into four tiers based on the Barron index of college selectivity, and split the families of students into four quartiles based on parental income, education, and occupation. Carnevale and Rose (2004) find that when considering only the top tier of universities (which accounts for about 10% of all students) an incredible 74% of the entering class is from the highest quartile of socioeconomic status, while only 3% is from the lowest quartile. The disparity in these numbers is far beyond anything that might be explained by natural ability, and lends credence to the idea that the U.S. education system is much less equal than one might think. No one expects equality of educational outcomes, because outcomes are largely based on individual students. Equality of opportunity, however, is something that the education system should strive for, and the analysis by Carnevale and Rose clearly shows that opportunity is distinctly unequal, particularly in higher education.

Table 2.1 College Selectivity and Socioeconomic Background

College Type-Highest selectivity to lowest	% of students in bottom quartile of socioeconomic status	% of students in top quartile of socioeconomic status
Tier 1	3	74
Tier 2	7	46
Tier 3	10	35
Tier 4	16	35
Community College	21	22

Haveman, Smeeding (2006)-Data from National Education Longitudinal Study of 1988

The above evidence shows that students who have parents with high incomes tend to do better in higher education, but how likely are they to rise to the same income level as their parents? Stated differently, how strong is the correlation between parents' income and their children's in the U.S.? Gary Solon studied this question in a 1992 paper called "Intergenerational Income Mobility in the United States." Solon first reviews the previous work on the topic, and concludes that previous studies have been subject to two major biases that make their results somewhat questionable. For example, a 1985 study by Jere Behrman and Paul Taubman found only a 0.2 correlation coefficient between father and son incomes, which would indicate low intergenerational transmission and high mobility. Other studies by William Sewell and Robert Hauser in 1975 and by William Bielby and Hauser in 1977 found correlation coefficients of 0.18 and 0.16, respectively (Solon, 1992 p.394). However, Solon argues that these studies suffer from both unrepresentative samples and error-ridden data. The studies tend to use short-run data, sometimes for only one year, as opposed to long-term average earnings. The study by Bielby and Hauser actually used the sons' recollections of their parents' income, which Solon argues is extremely unreliable. Solon constructs a model which shows that the likely effect of the errors in the data is to lower the intergenerational correlation coefficient below what it truly is. In addition to data errors, the studies did not use a random sample of the population, but instead more homogenous groups: the Taubman and Behrman study used only white male twins born between 1917 and 1927 who had both served in the armed forces (Solon, 1992 p.396). Obviously, this cohort is not representative of the national average, and Solon again argues that the net effect of this bias is downward, resulting in a lower correlation than an unbiased study would find.

After reviewing previous work on the topic, Solon performs his own analysis using data from the Panel Study of Income Dynamics (PSID) at the University of Michigan. The PSID data are perfect for a study of this type, as Solon explains: "The PSID data are especially well suited for reducing the biases of earlier research. First, because the data come from a national probability sample, they avoid the homogeneity of the samples used in some previous studies. Second, the longitudinal nature of the data makes it possible to explore the empirical importance of using short-run versus long-run status measures" (Solon, 1992 p.397). Solon's analysis focuses on the correlation between the earnings of fathers and their sons, as this correlation has generally been found to be stronger than a more general correlation that includes both genders.

He runs several different regressions using data from different years and with slightly different parameters.

Ultimately, the correlations Solon finds are clustered around 0.4, with some as high as 0.5 (Solon, 1992). These estimates are more than double those found in earlier studies, and indicate a society that is significantly less mobile than previously thought. Solon provides some perspective on how to think about the correlation coefficients: with a coefficient of 0.2, such as those found in earlier studies, a son with a father in the bottom quintile of income has a 30% chance of remaining at the bottom, and a 12% chance of reaching the top quintile. If the intergenerational correlation coefficient is 0.4, however, that same son has a 42% chance of remaining in the bottom quintile, and only a 5% chance of reaching the top (Solon, 1992 p.404).

Another way to interpret these correlations is that they represent the percentage of the parent's financial advantage or disadvantage that is passed on to the next generation. For example, if a father earns $20,000 more than the average person, a correlation of 0.20 would mean that 20% of this advantage is passed on to his son, so his son would earn $4,000 more than average, and a correlation of 0.40 would suggest that 40% of the advantage will be passed on, and so the son would earn $8,000 more than average, all other things equal. Solon's study is not completely definitive, but it shows clear evidence that the United States is a less mobile society than previous research had shown.

There is at least one major omission from Solon's study: the mobility of women. In 2002, Laura Chadwick partnered with Solon to remedy this problem with a study called "Intergenerational Income Mobility among Daughters." Chadwick and Solon are also interested in "assortative mating", or the tendency for people with similar characteristics to marry each other. They reference a study by Michael Kremer in 1997 showing that the spouse correlation in years of education in the U.S. is about 0.6, as well as a study by Steven Haider in 1998 showing the spouse correlation in hourly wages to be above 0.3 (Chadwick, Solon 2002 p.336). Due to this correlation, there is a strong connection between a daughter's parents' income and her husband's earnings; indeed, a study by A.B. Atkinson et al. in 1983 in England found that the correlation between a daughter's husband's earnings and her parents' earnings was as strong as the correlation between a son's earnings and his own parents' earnings (Chadwick, Solon 2002 p.337).

Chadwick and Solon perform their analysis in much the same way as Solon's earlier study: they use data from the PSID, and take averages of income over several years. They find a correlation of 0.43 when running a fairly simple regression of daughters' family income compared to their parents' family income, meaning that the income of daughters has a strong correlation with the income of their parents (Chadwick, Solon 2002 p.340). This result is similar to the result found for fathers and sons in Solon's earlier study, indicating that the result is likely reliable. Chadwick and Solon run a series of regressions with various parameters and sample sizes and find correlation results ranging from 0.35 to 0.49 (Chadwick, Solon 2002 p.342). They also perform a parallel analysis on sons to update Solon's work from 1992, and find an intergenerational correlation of 0.54, which is even higher than Solon's initial finding. The authors conclude by suggesting that further research is needed, preferably with a different set of data. The overall takeaway from both Solon's original study and this updated work is that intergenerational income mobility for both sons and daughters may be significantly lower than expected.

Correlation coefficients can be informative, but it is difficult to get a sense of what they truly mean without another country to compare to the United States. Anders Bjorklund and

Markus Jantti have done just this in their 1997 study comparing income mobility in the U.S. and Sweden. The authors first note that the U.S. and Sweden are very different when it comes to inequality: among OECD countries, the U.S. has the most inequality of disposable income, and Sweden has the least (Bjorklund, Jantti 1997 p.1009). Given this fact, the authors are interested to find out how income mobility relates to this large difference in inequality. They set up their analysis of the Swedish numbers in much the same way that Solon did, except that they include capital gains income in their measurements.

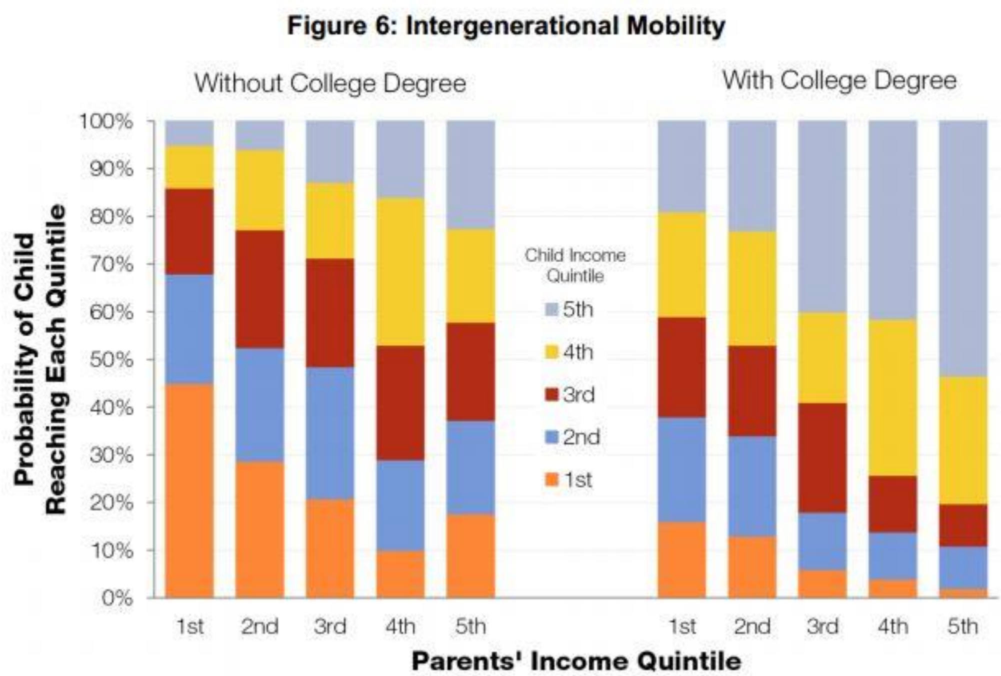
Surprisingly, the authors find significantly lower results for U.S. correlation coefficients than Solon did, but a useful comparison can still be made. Bjorklund and Jantti find that the correlation coefficient for Sweden is between 0.17 and 0.23, while the same coefficient for the U.S. is between 0.23 and 0.33 (Bjorklund, Jantti 1997 p.1014). Also, they find that in the U.S., 40% of sons with poor fathers are poor themselves, and about 40% of sons with rich fathers are also well off. In Sweden, these numbers are closer to 25%. The authors conclude by stating that it is likely that intergenerational income mobility is higher in Sweden than in the U.S., but due to the imprecise nature and limited sample size of their estimates, they cannot be sure.

Another international study in the field of income mobility literature is an analysis by Gary Solon in 2002 comparing intergenerational earnings elasticities in several different countries, and looking at the possible reasons behind the differences. This analysis looks at Canada, Finland, Germany, Malaysia, South Africa, Sweden and the United Kingdom. To find the intergenerational earnings elasticities in each country, multiyear measures of a father's earnings were compared to the same measures of a son's earnings. However, as is true in many of the studies in this field, due to measurement and estimation differences among the data from different countries, some of the observed difference in elasticity is likely due to errors in data collection as opposed to real differences in intergenerational transmission of socioeconomic status. With that said, here are the results that Solon finds: 0.28 in Sweden, 0.23 in Canada, 0.11 in Germany, 0.57 for Britain, 0.44 in South Africa, 0.22 in Finland, and 0.26 in Malaysia (Solon, 2002 p.63).

Several comparisons can be made using this data: intergenerational elasticity is generally quite low in Scandinavian countries, Germany, and Canada, while it is higher in Britain and South Africa. Malaysia has a surprisingly low elasticity; in fact it is less than half of Britain's. Solon suggests that many of the differences in elasticities stem from policy differences in the various countries as well as other economic factors; for example Canada is fairly similar economically to the U.S. in terms of inequality, but it generally has more progressive public policies, potentially leading to its lower intergenerational correlation in income. On the other hand, Solon expected less developed and less economically free countries such as South Africa and Malaysia to have significantly higher intergenerational correlations, but this is not really the case, especially since the South African data is not a very good representative sample of the population according to Solon (Solon, 2002 p.64).

Turning back to the U.S. numbers, the Departments of Treasury and Education published a report called "The Economics of Higher Education" in 2012 that goes into some more detail about family background, education, and potential future earnings. Some of the findings of this report are shown in Figure 2.1. This figure is a lot to take in at first, but basically it splits the population into fifths, with quintile one representing the lowest 20% of income earners and five the highest 20%. The figure shows a stark difference between those with and without college degrees. People without a college degree with parents in the lowest income quintile have almost a 70% chance of remaining in the bottom two income quintiles, with less than a 5% chance of making it to the top quintile.

Figure 2.1 Intergenerational Mobility by Quintile



Source: Brookings analysis of the Panel Study of Income Dynamics (Isaacs, Sawhill, & Haskins, 2011).

Departments of Treasury, Education (2012)

With a college degree, the chances for those people to stay in the bottom two quintiles fall to less than 40%, while the chances of making it to the top quintile increase to almost 20% (Economics of Higher Education, 2012 p.15).

On the other hand, people whose parents resided in the top income quintile, but who do not have college degrees have a fairly even chance of ending up in any one of the quintiles. Interestingly, this group has a higher chance of ending up in the lowest quintile than the group whose parents were in the second highest income quintile. As for those people with all the advantages (a college degree and parents in the top income quintile), the results are not unexpected: they have close to an 80% chance of remaining in the top two income groups, and almost no chance of falling to the lowest quintile. The most encouraging result comes from comparing the probabilities of people with and without college degrees whose parents were in the lowest quintile: getting a degree decreases the chances of remaining in the bottom two quintiles from near 70% to about 40%, and increases the chances of reaching the top two quintiles from under 15% to almost 40% (Economics of Higher Education, 2012 p.16).

These results show definitively that education matters greatly, especially to those who come from poorer families. Getting a degree fundamentally changes the ability of the poorest students to be upwardly mobile, and ultimately, that ability is what the American Dream is all about. The possibility of upward mobility leads people to hope for a better life for themselves and their children, whether it is in the U.S., Europe, or anywhere else around the globe. In the increasingly competitive global economy, education will only become more important, and its effects on inequality and social mobility will only become more profound.

The literature reviewed above covers many of the important issues that arise when considering the relationship between family background, education, and income mobility, but there are of course some unanswered questions. How much of the variation in income mobility is determined simply by different educational choices? How will this relationship change in the future? How does the skyrocketing cost of higher education change the relationship? Are there policies from other countries such as Sweden that the U.S. could adopt to improve its education system, and increase its income mobility? These questions are somewhat beyond the scope of this paper, but further research in these directions could prove both interesting and informative.

Chapter 3

Theoretical Model

The primary component of this thesis is empirical research; however, a brief theory section will shed some more light on the mechanisms by which income and education are transferred between generations. The theory in this section borrows heavily from a 2013 working paper by Gary Solon entitled "Theoretical Models of Inequality Transmission across Multiple Generations." Solon methodically lays out some simple equations that help to explain why the incomes and education levels of parents and children are correlated.

The model assumes that for a given family *i*, there are two generations: the parent's generation *t-1*, and the child's generation *t*. The parent has lifetime earnings $y_{i,t-1}$, which must be allocated between the parent's consumption $C_{i,t-1}$ and investment in the child's human capital $I_{i,t-1}$. The resulting budget constraint for the parent is shown in equation 1.

(1) $$y_{i,t-1} = C_{i,t-1} + I_{i,t-1}$$

The method by which parental investment in the child's human capital $I_{i,t-1}$is translated into actual human capital for the child is shown in equation 2.

(2) $$h_{i,t} = \Theta log I_{i,t-1} + e_{i,t}$$

In this equation, $h_{i,t}$ is the child's human capital, Θ is the return to human capital investment and $e_{i,t}$ is the human capital the child possesses regardless of how much the family invests. The log signifies that investment has diminishing marginal returns. The endowment $e_{i,t}$ is made up of both genetic inheritance and cultural inheritance, or things learned from the parents.

(3) $$e_{i,t} = \delta + \lambda e_{i,t-1} + v_{i,t}$$

In equation 3, δ represents the average level of innate human capital, while $e_{i,t-1}$ represents the innate human capital of the parents and λ is a heritability coefficient between 0 and 1 that indicates how much of the parents innate human capital is passed on to the next generation. Finally, $v_{i,t}$ is a random error term that attempts to account for the randomness of genetics.

Next, equation 4 describes the child's lifetime earnings.

(4) $$logy_{i,t} = u + ph_{i,t}$$

In this equation u represents average lifetime income, and p is the earnings return to human capital. At this point, Solon substitutes equation 2 into equation 4 which results in equation 5.

(5) $$logy_{i,t} = u + \gamma logI_{i,t-1} + pe_{i,t} \text{ where } \gamma = \Theta p$$

Equation 5 shows that γ represents the earnings return to human capital investment, and so the child's lifetime earnings are equal to an average value, plus the earnings return to investment times the amount of investment, plus the earnings return to human capital times the amount of innate human capital.

Next, an equation is needed to model the decision that the family makes about how much to invest in the child's human capital. This is shown in equation 6.

(6) $$\mathrm{U_i} = (1-\alpha)\mathrm{logC_{i,t-1}} + \alpha\mathrm{logy_{i,t}}$$

This decision is shown in the form of a Cobb-Douglas utility function, where α is an altruism variable that compares the parent's preferences for their own consumption against their preferences for their children's income. A higher α indicates a more altruistic parent that values their children's incomes higher than their own consumption, and vice versa. Equation 6 can be rewritten to include the choice variable $I_{i,t-1}$, resulting in equation 7.

(7) $$U_i = (1-\alpha)\log\left(y_{i,t-1} - I_{i,t-1}\right) + \alpha u + \alpha\gamma log I_{i,t-1} + \alpha p e_{i,t}$$

Once this equation is rewritten, Solon goes on to give the first-order condition to maximize the utility function, which is shown in equation 8.

(8) $$\frac{\partial U_i}{\partial I_{i,t-1}} = -\frac{1-\alpha}{y_{i,t-1} - I_{i,t-1}} + \frac{\alpha\gamma}{I_{i,t-1}} = 0$$

Taking this first-order condition and solving for the optimal choice of investment results in equation 9.

(9) $$I_{i,t-1} = \{\frac{\alpha\gamma}{[1-\alpha(1-\gamma)]}\} y_{i,t-1}$$

From this result, three major implications can be drawn. First, parents' investment in their children's human capital increases with a higher altruism factor α. Second, parents' investment in their children's human capital also increases with a higher earnings return to human capital investment γ. Finally, parents with higher incomes use part of their additional resources to invest more into their children's human capital. These implications could be considered common sense, but using a model to derive them increases the rigor of the results.

The final step that remains is to derive the regression equation for intergenerational income elasticity. This can be accomplished by substituting equation 9 into equation 5, resulting in equation 10.

(10) $log y_{i,t} = u^* + \gamma log y_{i,t-1} + p e_{i,t}$ where the intercept $u^* = u + \gamma \log\{\frac{\alpha\gamma}{[1-\alpha(1-\gamma)]}\}$

This equation is quite similar to the double-log functional form of the regressions run in the empirical analysis in this thesis. It shows that the main variables which determine a child's lifetime earnings are the coefficient γ on the parent's lifetime earnings and the coefficient *p* on the child's human capital, which are two of the coefficients found in the results section of this paper, particularly in model 2.

Chapter 4

Data Analysis

One of the most important components of this thesis is the data that will be used to run regressions and make conclusions. The primary data source that was used during the course of this research is the Panel Study of Income Dynamics at the University of Michigan. This study started in 1968 with over 18,000 individuals from over 5,000 families in the United States as an attempt to create a nationally representative sample. Since that time, data on these original individuals as well as their descendants has been continuously collected, including such information as income, wealth, education, health, children, philanthropy, and many other variables (PSID online). The National Science Foundation has recognized the PSID as one of the most significant advances funded by the NSF.

For the initial phase of research and data analysis, data was collected from each year of the study, 1968 through 2011. The data available for each year contains over 5,000 variables covering hundreds of different topics. Only a few variables out of the 5,000 will be relevant to the study of intergenerational income mobility, but the PSID website has a search function which makes it much easier to identify relevant variables. This section will identify some of the most important variables, define them, and attempt to explain any quirks or abnormalities. The methodology section goes into more detail about modifications made to the data due to these abnormalities in various years. Since survey questions on the PSID can change from year to year, it is important to make sure to understand exactly what each variable means. The descriptive statistics given for each variable are from model 1, meaning that they cover both genders.

One of the most basic variables that was included is the age of the household head. Age was included because it should have a large effect on income, as well as on amount of education completed below a certain age. Generally, the household head will be at least 18 years of age, but the range for this variable extends from 14 to 120. Another basic descriptive variable is the sex of the head. In this variable, a 1 represents a male while a 2 represents a female. In the most recent survey year, the respondents are heavily skewed towards males, with about 68% of household heads being male and about 32% female. This is important to keep in mind moving forward, as there may be differences in the way that income mobility works for males and females. Several studies in the literature emphasize that gender matters a lot when researching income mobility, so it is an important factor to take into account.

Along with age and sex, another potentially important piece of data is race/ethnicity of the head. The most common choices given in the survey question are 1-White, 2-Black or African American, 3-American Indian or Alaskan Native, 4-Asian, 5-Native Hawaiian or Pacific Islander, 7-Other, 9-N/A or refused. However, this is one of the questions that is worded and coded differently in different years of the PSID. For the people that refuse to answer, the PSID repeatedly asks them the question again later in the survey in what they call additional "mentions". Out of the 8,907 household heads who took the survey, only 87 refused to answer, so these responses can pretty safely be ignored. The racial breakdown of the survey respondents for 2011 is shown in table 4.1.

Table 4.1 Race of 2011 Survey Respondents

Race/Ethnicity	White	Black	American Indian/Alaska Native	Asian	Native Hawaiian or Pacific Islander	Other
Percent of Sample	58.25	35.88	0.75	1.16	0.08	2.91

Interestingly, there is no option for Hispanic, so it must be included in the White and Other categories. Relative to the overall US population, this particular survey seems to over sample

African Americans and under sample Asians, although presumably this problem is solved with the inclusion of the statistical weight variable discussed later. The race data from the PSID is fairly messy: different questions were asked in different years, and the responses are coded differently as well. While race is a very interesting descriptive variable, the inconsistencies in the data make it somewhat unsuitable for inclusion in regressions.

Geographic location could be relevant to the study of income mobility as well. It is quite possible that different areas of the country could have different properties when it comes to heritability of income. In the PSID data, there are variables for both state and region. The state variable includes a number from 1 to 56 that represents which state (or DC) the household head currently lives in, or a 0 if the head lives in a U.S. territory or foreign country. There are 51 codes, but some numbers are skipped which is why the values go up to 56. In addition to this state variable, there is a more general current region variable for 2011. This variable divides up the country into four regions: Northeast, North Central, South, and West. The states included in each region are as follows: Northeast (Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont), North Central (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin), South (Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, Washington DC, West Virginia), and West (Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming). There is then a fifth option for Alaska and Hawaii, and a sixth option for a foreign country. The regional breakdown of the survey respondents is shown in table 4.2. Breaking the data down by state is likely unnecessary, but the differences between regions are much more feasible to study and could yield some interesting results.

Table 4.2 Region of 2011 Survey Respondents

Region	Northeast	North Central	South	West	Alaska and Hawaii	Foreign Country
Percent of Sample	12.92	24.48	43.98	17.91	0.18	0.54

Age, sex, race, and location cover the basic descriptive statistics fairly well. Now it is time to turn to the more interesting data: income and education. The statistic for years of education completed by household heads was the primary measure of education used. A number between 1 and 16 for this variable represents the number of years of education completed, e.g. 12 would be a high school graduate, and 16 a college graduate. A 0 represents no years of education completed, while a 17 represents at least some graduate level work. If the head of household does not know if they attended college or high school, or if the number of years is unknown, a 99 is assigned. In addition, if the survey respondent received a GED, then the number assigned is the last grade completed. The mean value of this variable for 2011 is 13.22, suggesting a fairly well educated sample, as the average person in the sample has completed at least one year of college. This variable seems to be the best measure of education, as it is consistent across years and shows the actual years of education completed, as opposed to some other education variables in the PSID which use brackets of years.

The PSID includes many variables relating to income. The primary variable used in this paper is the labor income of head variable. This seeks to measure all labor income earned by the head of household in the previous year, excluding income earned from farm work and unincorporated business income. This measure is the sum of several other variables measuring individual components of labor income: it adds together wages and salaries, bonuses, overtime, tips, commissions, income from a professional practice or trade, market gardening, income from other jobs, and miscellaneous labor income. Out of all the income variables in the PSID, this variable is the best representation of an average person's gross income, which is the appropriate measure. Using net income instead would introduce an additional level of complexity which isn't really relevant to the study of income mobility. The mean value for labor income in 2010 is $25,210.38, and the median value is just $15,100, however these numbers include many individuals with incomes of zero. Another variable used in the model is total hours of work by the head in the previous year. The mean value for hours worked in 2010 is 1,363. This variable simply represents the total annual work hours of the head from all jobs, including overtime. Including a variable for hours worked will allow comparisons of time spent working versus income earned and other such correlations.

Finally, the models include the most recent statistical weight variable for 2011. This variable, the core-immigrant individual cross-sectional weight, attempts to compensate for the unequal selection probabilities that result from the PSID's combination of two different samples. There are a lot of technicalities involved in how the PSID puts its data together in combination, and this weight variable is an attempt to resolve some of the issues that result from this combination. It also attempts to deal with the issue of non-responsive members of the survey in various years. The intended effect of including this variable is to make the results more representative of the national population.

Chapter 5

Methodology

Model 1

This section will go over the specific methodology used to construct the Model 1 dataset. For this model, data was collected from the PSID for each survey year available, 1968 through 2011. The basic idea of this model is to compare incomes of parents and their children over a prime period in their working life. In general, people earn fairly low incomes at a young age, with incomes rising though middle age and then falling off as the person works less and approaches retirement. Due to this life cycle of income, the period chosen for study was ages 40 to 50, as this should provide a good representation of a person's earning capacity.

The basic structure of the dataset was constructed using a tool provided by the PSID known as the Family Identification Matching System, or FIMS. This tool allows for the creation of unique ID's for each individual, and for their parents as well. In this way, children can be matched with their parents for comparison. The FIMS dataset was specified as an intergenerational file matching individuals to their parents. Adoptive parents were excluded from the data, as well as any family that had a generation missing an ID, and the file format selected was wide, resulting in one observation for each individual. Downloading the FIMS data resulted in over 36,000 observations.

There are two types of variables represented in the FIMS: The 1968 interview number, and the 1968 person number. The first number identifies the family that the individual belongs to, while the second number is unique to the individual. By combining these two numbers together into one variable, a unique ID can be created for each child. This variable, kid_id, was created by multiplying the 1968 interview number by 1000, and then adding the 1968 person number. For example, if an individual had a 1968 interview number of 2, and a person number of 171, their kid_id would become 2171. Along with IDs for the children, FIMS also includes the same variables for their parents, separated into father and mother. Thus, the FIMS includes ER30001_P_F, which is the father's 1968 interview number, and ER30002_P_F, the father's 1968 person number, as well as the same variables for the mother. These variables are then turned into dad_id and mom_id using the same procedure used to generate kid_id. After this process, each child has a unique ID, and their parents who have data available have unique IDs as well.

After the FIMS data was ready, the actual data of study was prepared. Several variables from each year of the survey were chosen. The 1968 interview and person numbers were needed to match this data with the FIMS data, and the individual weight variable from the most recent survey, 2011, will make the data nationally representative when included. Two other variables, the sequence number and the relationship to head, were needed to make sure that only heads of households are included in the dataset. This is necessary because the income variables being used represent the labor income earned by the head of household in the previous year, so including individuals other than household heads would corrupt the data. Other variables included were the individual's sex, age, current state, and current region, as well as the number of hours the head worked last year, and the highest grade of school completed by the head. Finally, the interview number from each year of the survey was included as well. When downloading the dataset, the PSID provides the option to create subset criteria, and this was where the sequence number was set to 1 and the relationship to head was set to 10 to make sure that the only individuals included in the dataset were heads of households that were living in their house in the chosen years.

After the data was downloaded, there was a good amount of modification needed before it was ready to combine with the FIMS and be analyzed. Because the goal of this model is to compare average earnings between ages 40 and 50 for children and their parents, an age variable needed to be created for each survey year that pulled information from the individual age variable if that individual was the head of household in any given year. Doing this resulted in 38 new variables, age1968 through age2011, each containing the ages of all the individuals that were heads of household in that year (surveys were given annually from 1968 to 1997, and then every other year). The same procedure was done for income and hours, so that y1968 through y2011 represent the income of heads of households in the previous year, and hours1968 through hours2011 represent the hours worked by the household head in the previous year.

An adjustment was needed due to a quirk of the PSID in the years 1994 and 1995. During these years, information was collected on a special cohort of Latinos, and was coded differently from the rest of the data. Due to this cohort, income and hours measures for 1994 and 1995 include some data that needs to be excluded: incomes for the Latino cohort were reported as 9,999,999 for some reason, and hours worked the previous year were reported as 9999. This data was removed from the set. Along these lines, there are a few cases where people didn’t know or refused to report their age, and the PSID coded it as 0, 99, 998, or 999. These data points have been removed as well.

After these adjustments were completed, the next step was to adjust the income figures for inflation. All incomes were converted into 2010 dollars by multiplying y1968 through y2011 by the factors shown in table 5.1 (since the variables show the previous year’s income, the factors are for the previous year, for example 6.53 is the factor needed to convert 1967 dollars into 2010 dollars). The variables created through this adjustment are yi1968 through yi2011.

Table 5.1 Inflation Adjustment by Year

Year	**Inflation Multiplier**	**Year**	**Inflation Multiplier**	**Year**	**Inflation Multiplier**	**Year**	**Inflation Multiplier**	**Year**	**Inflation Multiplier**
1968	6.53	1976	4.05	1984	2.19	1992	1.60	2003	1.21
1969	6.27	1977	3.83	1985	2.10	1993	1.55	2005	1.15
1970	5.94	1978	3.60	1986	2.03	1994	1.51	2007	1.08
1971	5.62	1979	3.34	1987	1.99	1995	1.47	2009	1.01
1972	5.38	1980	3.00	1988	1.92	1996	1.43	2011	1.00
1973	5.22	1981	2.65	1989	1.84	1997	1.39		
1974	4.91	1982	2.40	1990	1.76	1999	1.34		
1975	4.42	1983	2.26	1991	1.67	2001	1.27		

Once the incomes are adjusted for inflation, they can all be directly compared against each other. Since this model is only interested in ages 40 through 50, new variables needed to be created that represent incomes at each age, 40 through 50. This was accomplished by generating a variable called income40 that would be equal to yi1968, inflation adjusted income of the head in 1968, if age1968 was equal to 41. Since income is for the previous year, someone who was 41 in 1968 would have been 40 in 1967 when they were earning the income. After the variable was generated, each year's data was added to it if the head's age in that year was 41. Once this command was given for each year of the survey, the income40 variable contained all of the incomes of 40 year old household heads over all the survey years. It was then a simple matter to do the same thing for age 41, age 42, etc., all the way up to age 50, resulting in income40 through income50 as new variables. The same process was also completed for hours worked, giving hours40 through hours50 as new variables.

After income variables were generated for each age 40 through 50, a sum was created, and then an average. The sum variable for each individual is equal to the total of all income earned over ages 40 through 50. The “missing” option was used so that STATA didn’t ignore individuals with one or more missing years. The same command was then used to generate a sum variable for hours worked. To create a variable for average income and hours worked, STATA first needed a count of how many non-missing years of data each individual had between ages 40 and 50, so that it could divide the sum by that many years. This was accomplished using the “rownonmiss” function in STATA. Then, averages for income and hours were created by dividing the sum variables by the rownonmiss variables, resulting in avg_income and avg_hours.

At this point, the data was split into child and parent datasets. The data was saved into two copies, a kids dataset and a parents dataset. In the kids dataset, kid_id was created using the 1968 interview number and 1968 person number in the same way as in the FIMS dataset. In the parents dataset, parent_id was created in the same way, and then separated into mom_id and dad_id. Also, in the parents dataset key variables were renamed so that they would not merge together with the same variables for the kids dataset: avg_income became avg_incomeparent, and avg_hours became avg_hoursparent. Once these changes were made, the final merge process could begin.

After the parents and kids datasets were ready, all that was left to do was merge everything together. First, the FIMS data was combined with the kids data by merging on kid_id. This told STATA to look for instances in which kid_id matched between the FIMS and kids datasets, and combine all the data on that individual together. The same process was then done with the newly merged data and the parents dataset, but this time merging on mom_id and dad_id. The final step was to convert the average income variables into log form, so that when regressions are run, the coefficient on the parent's average income variable is in elasticity form.

Model 2

There were a few important differences in the way that model 2 was constructed. The same general process was used in which the FIMS dataset is combined with one dataset for kids and one for their parents. However, because the literature suggests that the strength of the correlation between sons' and dads' incomes is much stronger than the link between children and their parents more generally, and because the strength of the correlations found in some variants of model 1 were surprisingly low, model 2 focuses exclusively on males. This is accomplished by setting the gender variable equal to 1 when the data is downloaded from the PSID website so that only observations pertaining to males are downloaded. All the same variables from model 1 are included in model 2, but more of them will be used.

Once the data is downloaded, another change from the model 1 methodology is needed. Focusing on ages 40 through 50 for model 1 was an interesting idea, but it ended up limiting the sample size greatly, so model 2 takes a different approach. All ages from 18 through 64 are included for measuring income and hours worked. This provides a much greater range and amount of data, and greatly increases the potential sample size. It is also a more complete picture of someone's lifetime earnings: ages 18 through 64 generally cover the great majority of a person's working life. The process for creating variables income18 through income64 is the same as the process for model 1: variables are created for each survey year and then assigned to the variables for each age if the head of household was that age in that year. Income was then added up and averaged, so that the final variable represents average income over the individual's whole working life. Unlike model 1, the same thing was done for education, and then a maximum value was created that represents each head of household's highest level of education attained, MaxEd.

At this point, the data is ready to be split into kids and parents datasets, just like model 1. Again, each variable that could be relevant is renamed to have either a sons or dads appended onto the end of the variable name. This is done so that when the two datasets are combined together, the variables do not combine due to their different names. In this way, the final dataset resulting from the merge of the FIMS, sons', and dads' datasets has a variable corresponding to sons and dads for each relationship that might be studied, for example, MaxEdSons and MaxEdDads. Once the income variables for sons and dads are converted into log form, the data is ready for analysis.

Chapter 6

Results

Model 1

Table 6.1-Model 1 Results

Coefficient(p-value)	Model 1A	Model1B (weighted)	Model 1C	Model 1D (weighted)
Logavg_incomeParent	0.16175(0.000)	0.35189(0.000)	0.13825(0.013)	0.38469(0.009)
Avg_hoursParent	*	*	0.00006(0.600)	-0.00007(0.778)

Table 6.2 Model 1 T-Statistics

T-statistics	Model 1A	Model 1B	Model 1C	Model 1D
Logavg_incomeParent	5.00	3.80	2.50	2.63
Avg_hoursParent	*	*	0.52	-0.28

The results from model 1 were somewhat limited. Due to the way in which the data was prepared, the sample size of the final regression was only 528 observations, and when the data is weighted only 179 observations. This is a major decrease from the initial sample size of over 30,000 observations, and is largely due to the analytic choices that were made, such as only including income from ages 40 through 50. However, some usable results were obtained from even this small sample size. The p-values and t-statistics in the tables above indicate the statistical significance of each result: a p-value closer to 0 indicates more significant results, while a t-statistic further away from 0 in either direction indicates the same.

The functional forms of models 1C and 1D are similar to equation 10 in the theoretical model: $logy_{i,t} = u^{*} + \gamma logy_{i,t-1} + xH_{i,t-1}$, where $y_{i,t}$ is the child's average annual income, u* is an average value, γ is the coefficient on parent's average annual earnings, $y_{i,t-1}$ is the parents' average annual earnings, x is the coefficient on parents' annual average hours, and $H_{i,t-1}$ is the parents' annual average hours. Models 1A and 1B simply don't include the average hours variable.

In the most basic form (model 1A), a simple regression of the log of children's average income against the log of their parents' average income, the coefficient on parents' income was 0.16175, with a p-value of 0 indicating a statistically significant result. When the data was weighted (model 1B), this same coefficient on parent's income more than doubled to 0.35189, again with a p-value of 0, but with a slightly lower t-statistic. Because both income variables are in log form, these coefficients represent the elasticity of kids' average income with respect to parents' average income. The initial value in model 1A is surprisingly low; the literature suggests that this elasticity may be closer to .25, .30, or even higher. However, when the data is weighted in model 1B the resulting elasticity, 0.35189, is much higher and is closer to what is predicted in the literature. As the weight variable represents an attempt to make the data nationally representative, this disparity in elasticities indicates that income mobility may be much lower in the general population than in the unweighted sample.

There are several potential reasons for the low coefficient in model 1A: probably the most important reason is gender. Many of the studies in the literature on this topic suggest that gender plays a more important role than one might expect in determining the strength of the correlation between the incomes of parents and children. Since model 1 includes both sons and daughters as well as their moms and dads, the strength of the correlation may be diluted somewhat. This explanation is tested in model 2, which focuses only on sons and their dads. Another potential source of error for this model is the method of life cycle analysis that was used. The ages 40 to 50 were chosen due to their place at the peak of an average individual's lifetime earnings curve. This decade is generally late enough in life that people no longer have very young children around and have settled into their career, but early enough that they are unlikely to have retired or otherwise stopped working. However, using only these ages is certainly limiting to the scope of the model, and this issue is addressed in model 2 as well.

An additional variant of model 1 was run that included average hours worked for the parents. Average hours for the children were not included because there would be an extremely strong correlation between income and hours, and that is not what this paper seeks to study. However, it turns out that at least in this particular sample, there is not much correlation between a parent's average hours of work and their child's income. Neither the weighted or un-weighted models found a statistically significant result for parent's hours.

Overall, model 1 was a good starting point, but more analysis is needed to provide a definitive answer to the research questions that this paper seeks to address. Unfortunately, there were some problems with the education variables in this model, so they were not included. Model 2 takes a much closer look at education and how it relates to income. Model 2 represents an attempt to improve upon the methodology of model 1, and find a more comprehensive answer to the questions raised at the beginning of this paper.

Model 2

Table 6.3 Model 2 Results

Coefficient(p-value)	Model 2A	Model 2B (weighted)	Model 2C (weighted)	Model 2D (weighted)
Logavg_incomeDads	0.30112(0.000)	0.28899(0.000)	0.15544(0.000)	0.16848(0.000)
MaxEdSons	*	*	0.12795(0.000)	0.13831(0.000)
MaxEdDads	*	*	*	-0.01714(0.019)

Table 6.4 Model 2 T-Statistics

T-statistics	Model 2A	Model 2B	Model 2C	Model 2D
Logavg_incomeDads	18.55	7.70	4.37	5.32
MaxEdSons	*	*	12.98	14.06
MaxEdDads	*	*	*	-2.36

The results for model 2 were much more robust than those for model 1. The increased age range under measurement allowed a much larger sample size: 3,683 observations as opposed to just 528 in model 1, and 2,373 observations in the weighted portion as opposed to 179 in model 1. In addition, all coefficients in model 2 have a p-value of 0.05 or less, showing that all the variables are at least somewhat significant.

The functional form of model 2D is again similar to the theoretical model, but unlike model 1, education is included: $logy_{i,t} = u^{*} + \gamma logy_{i,t-1} + pe_{i,t} + De_{i,t-1}$ where $y_{i,t}$ is the son's average annual income, u* is an average value, γ is the coefficient on father's average annual earnings, $y_{i,t-1}$is the father's average annual earnings, p is the coefficient on son's education, $e_{i,t}$ is the son's level of education or human capital, D is the coefficient on dad's education, and $e_{i,t-1}$ is the dad's level of education or human capital. Models 2A, 2B, and 2C are simplified versions of this same equation.

The simple regression of log son's earnings with respect to log dad's earnings yielded an elasticity of 0.30112, almost double the strength of the initial correlation in model 1. This elasticity is still somewhat lower than many of the estimates in the literature for father-son earnings elasticity, but the results of this model show definitively that sons' incomes are heavily influenced by the incomes of their fathers. When the statistical weight is included, the strength of the correlation falls slightly to 0.28899.

It is interesting that in model 1, the weighted correlation was almost double the unweighted correlation, while in model 2 the weighted and unweighted correlations are very similar, and in fact the weighted correlation is slightly lower. One possible explanation for this difference, which is purely conjecture, is that the males in this particular dataset could be more representative of the national population of males than the females in the set are of all American females, and so including females in the model would move the unweighted correlation away from the weighted correlation. In general, as this paper seeks to investigate the American Dream across the whole population, the weighted correlation is probably a more relevant measure to consider.

Model variants 2C and 2D include education variables. Model 2C finds a coefficient of 0.12795 for the education level of the sons and 0.15544 for dads' incomes. This is an interesting result because it suggests that a large part of the correlation between incomes of fathers and sons is related to education level; when education is introduced as a control variable the correlation is cut almost in half. The results from model 2D are somewhat surprising: they actually show a small but moderately significant negative correlation between education of the father and income of the son. This is an unexpected result that doesn't seem to make a lot of sense, and could just be a quirk of the data, but some further investigation is needed.

One possible explanation for the coefficient on father's education is that the effects on children's income are somewhat indirect, and can't be captured in the father's education variable. For example, a regression with son's education as the dependent variable and father's education as the independent variable results in a coefficient of 0.29368 with a p-value of 0 and a t-statistic of 18.20. Clearly, a father's education has a fairly large effect on the education of their son. Likewise, regressing son's education on son's income gives a coefficient of 0.14616 with a p-value of 0 and a t-statistic of 16.52. Therefore, even if the father's education variable does not show a positive correlation in model 2D, these other regressions show that father's education is positively correlated with son's education, and son's education is positively correlated with son's income. In addition to this evidence, running a simple regression of father's education on sons income does result in a small positive coefficient of 0.04442 with a p-value of 0 and a t-statistic of 6.90. This correlation is seemingly lost in the more complex environment of model 2D, but based on these supplemental regressions it seems that there is a fairly small positive correlation between father's education and son's income.

There are a few other ways to break down the data with descriptive statistics. One factor that could lead to important differences in mobility is current geographic location. This is best described using the region variable to break the data into geographical regions and then compare them. The 2011 current region variable is used to group sons into one of 6 regions. The states that are covered by each region are listed in the data analysis section. About a third of the sons' observations do not have a region associated with them, likely because they were no longer in the survey by 2011. The results of this exercise are shown in table 6.5.

Table 6.5 Income Correlations by Region

Region	Coefficient (p-value)	T-statistic	#of Observations
Northeast	0.40386(0.000)	7.54	325
North Central	0.44938(0.000)	8.97	612
South	0.24739(0.000)	8.86	974
West	0.26316(0.000)	4.54	441
Alaska, Hawaii	-1.12457(0.125)	-2.11	5
Foreign Country	0.03152(0.898)	0.13	16
No Region Available	0.27345(0.000)	10.21	1310

The breakdown by region shows a surprising amount of difference between regions. The Alaska and Hawaii and Foreign Country groups can safely be ignored due to their very small sample size. The South, West, and no region available groups are all in a similar range of 0.25 to 0.27. These correlations are similar to the more general numbers found in model 2. However, the Northeast and North Central groups show much higher correlations between 0.40 and 0.45. It was expected that there would be some variation between regions, but these results suggest that a fundamentally lower level of mobility exists in the Northeast and North Central regions relative to the rest of the country.

Obviously, one regression does not prove anything, but it does raise some interesting questions about what might be causing the large differences between regions. More research would be needed to determine if such differences truly exist or are features of this particular dataset. For example, it could be that the 1,310 observations with no region identified for 2011 are primarily from the Northeast and North central regions, causing the correlations for those areas to appear higher than they actually are.

Another possible explanation is differences in the actual levels of education or income between regions. It could be that parents with higher incomes are more likely to transmit that status to their children than parents with lower incomes, and the same could be true for education. The breakdown of income and education levels by region is found in table 6.6.

Table 6.6 Income and Education by Region

Region	Median Sons Income	Average Years of Education of Son
Northeast	$42,531.00	14.23
North Central	$35,270.51	13.77
South	$32,076.00	13.62
West	$37,703.84	14.08
No Region Available	$30,808.89	12.69

Again, there are some significant differences between the regions. The Northeast has a much higher median income than any other region, with the North Central and West occupying a middle ground, and the South and no region groups the lowest. Likewise, for education the Northeast is again on top, with the West close behind, then the North Central, the South, and finally the no region group. Interestingly, both education and income result in the same ordering of regions. It makes some sense that the no region group is last in education, as these are likely people who were out of the survey by 2011 meaning that they were from earlier years, and the average years of education that people complete has increased over time. However, the patterns of education and income by region do not really match the patterns of income correlations seen in table 6.5, suggesting that some other factor is causing income mobility to be much lower in the North and Northeast than the rest of the country. It is possible that differences in policies in the states that make up each region are the primary cause of the varied levels of mobility, but again, more research would be necessary to investigate this idea.

One final variant of model 2 was run in order to better compare it to model 1. To further study the gender differences in income mobility, the age range in model 2 was restricted to 40 to 50 so that the only difference between models 1 and 2 would be the inclusion of either one or both genders. These results are shown in table 6.7.

Table 6.7 Model 2 Results 40 to 50

Dads Income 40 to 50	Unweighted	Weighted
Coefficient(p-value)	0.49312(0.000)	0.38607(0.000)
T-statistic	9.83	3.62

The unweighted elasticity of sons' income with respect to dads' income for ages 40 to 50 is 0.49312, the highest correlation found anywhere in this paper, while the weighted elasticity is closer to those found in models 1 and 2. These results suggest that among fathers and sons, the correlation in incomes is even stronger between ages 40 to 50 than across a whole working life 18 to 64. This makes sense for the same reason that ages 40 to 50 were used for model 1 in the first place: it is a decade that represents some of the prime working years of a person's career. In addition, the difference in mobility between males and females is shown to be quite large in this particular sample, but may be significantly smaller in the general population.

Chapter 7

Conclusion

The fundamental question this paper sought to answer was really all about the American Dream. Can a person, regardless of background, work hard and make it big? Are Americans tied to their parents' socioeconomic status, or can they make their own way in life? The answers are not necessarily straightforward, but based on the evidence uncovered through the empirical research in this paper, U.S. income mobility appears to be something of a mixed bag, with an unfortunate tendency toward the lower end. The correlation between fathers' and sons' incomes is fairly high at around 0.30, and in the Northeast and North Central regions specifically is very high at 0.40-0.45, while the correlation between parents' and children's incomes of both genders across the country more generally is either a far more modest number around 0.15 or a much higher 0.35, depending on whether the data is weighted or unweighted. Either way, these correlations are certainly high enough to be concerned about. As far as the education component, a correlation of about 0.27 was found between fathers' and sons' years of education completed. This indicates that the level of educational attainment, like level of income, is at least partially transmitted from generation to generation.

Obviously, these correlations have major implications for the success of any given child. If that child is born to parents who have high levels of income and education, he or she is much more likely to obtain high levels of income or education as well. Likewise, if a child is born to uneducated, poor parents, he or she has a much higher chance of ending up poor and uneducated as well. This is true almost everywhere in the world, and has been for much of human history. Chances are high that it will always be true to some extent, but in a country where individualism and hard work are emphasized so heavily, it is important to minimize these correlations so as to make the American Dream a reality for as many people as possible.

There are several policy changes that have the potential to reduce the strength of the correlation between the incomes and education of parents and their children. For the most part, education is the key to this process. As mentioned in the introduction, a truly meritocratic education system can help to stand as a barrier between the socioeconomic situation that a student is born into and what they can achieve through hard work. As noted in the literature review section, the American system of education is far from being meritocratic, especially when it comes to higher education. The vast majority of students at the most elite universities are from wealthy backgrounds, which simply perpetuates the existing inequalities in American society. The question of how to get an intelligent, motivated individual who is simply born into a bad situation into an elite university is a complicated one, and there is no simple, easy fix. Indeed, some research has shown that even if students from low-income backgrounds make it into elite universities, they have a lower chance of graduating than students from high-income backgrounds. However, awareness of the transmission of income and education across generations is the first step in crafting policy to address this problem.

There are several areas where this thesis could be expanded upon to continue shedding light on this important topic. The effects of education could be broken down more specifically into high school, undergraduate, and graduate education to gauge the effect on income mobility of each. Additional economic data such as Gini coefficients could be collected for each region in an attempt to explain the differences in mobility by region shown in table 6.5. More regressors could be included to explain a greater percentage of the variation in income between generations. Finally, although the results in this paper show a strong correlation between the incomes of parents and their children, correlation does not necessarily mean causation. More research would be needed to definitively establish a causal link.

The United States is a country with limitless potential. It is almost impossible to imagine what could be accomplished if every child in the U.S. was given a fair shot at success. Obviously, not everyone can be rich, but the opportunity for each individual to work towards middle class status is a goal that should be achievable with the stocks of wealth, resources, and common decency that this country possesses. The U.S. is below potential in a way that has nothing to do with GDP: it is failing the people who need help the most. It would be unfair and unrealistic to promise equality of outcomes, but a level playing field on which all children have an equal opportunity to make something of themselves doesn't seem like too much to ask.

BIBLIOGRAPHY

Atkinson, A., Maynard, A., & Trinder, C. (1983). *Parents and Children: Incomes in Two Generations*. London: Heinemann Educational Books.

Behrman, J., & Taubman, P. (1985). Intergenerational Earnings Mobility in the United States. *The Review of Economics and Statistics, 67*(1), 144-151.

Benabou, R. (1996). Equity and Efficiency in Human Capital Investment: The Local Connection. *The Review of Economic Studies, 63*, 237-264.

Bjorklund, A., & Jantti, M. (1997). Intergenerational Income Mobility in Sweden Compared to the United States. *The American Economic Review 87*(5), 1009-1018.

Blackburn, M., & Neumark, D. (1991). Omitted-Ability Bias and the Increase in the Return to Schooling. *National Bureau of Economic Research Working Paper*.

Card, D., & Krueger, A. (1990). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *National Bureau of Economic Research Working Paper.*

Card, D. (1999). The Causal Effect of Education on Earnings. *Department of Economics, University of California at Berkeley.*

Carnevale, A., & Rose, S. (2004). Socioeconomic Status, Race/Ethnicity, and Selective College Admissions. *America's Untapped Resource: Low-Income Students in Higher Education*. 101-156.

Chadwick, L., & Solon. G. (2002). Intergenerational Income Mobility among Daughters. *The American Economic Review, 92*(1), 335-344.

Chen, Y., Naidu, S., & Yuchtman, N. (2010). Intergenerational Mobility and Institutional Change in 20th Century China. Guanghua School of Management, Peking University, Harvard University.

Clark, D, & Martorell, P. (2014). The Signaling Value of a High School Diploma. *Journal of Political Economy, 122*(2), 282-318.

Cullen, J., Jacob, B., & Levitt, S. (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica, 74*(5), 1191-1230.

Fields, G., & Ok, E. (1996). The Meaning and Measurement of Income Mobility. *Journal of Economic Theory, 71*, 349-377.

Fox, T., & Miller, S. (1965). Economic, Political, and Social Determinants of Mobility: An International Cross-Sectional Analysis. *Sage Publications, Ltd., 9*(1), 76-93.

Garrett, T. (2011). Income Mobility. *National Economic Trends.* Research Division, Federal Reserve Bank of Saint Louis, Saint Louis, MO.

Glomm, G., & Ravikumar, B. (1992) Public versus Private Investment in Human Capital: Endogenous Growth and Income Inequality. *Journal of Political Economy, 100*(4), 818-834.

Haider, S. (1998) Econometric Studies of Long-Run Earnings Inequality. Ph.D. dissertation, University of Michigan, Detroit, MI.

Haveman, R., & Smeeding, T. (2006). "*The Role of Higher Education in Social Mobility.*" *The Future of Children, 16*(2), 125-150.

Ingraham, C. (2014, September 24). Guess Who's Losing Faith in the American Dream? Everyone. *The Washington Post.*

Inkeles, A. (1950). Social Stratification and Mobility in the Soviet Union: 1940-1950*. American Sociological Association, 15*(4), 465-479.

Jaeger, D., & Page, M. (1996) New Evidence on Sheepskin Effects in the Returns to Education. *The Review of Economics and Statistics, 78*(4), (1996): 733-740.

Krebs, T., Krishna, P., & Maloney, W. (2013). Income Mobility and Welfare. IMF Working Paper.

Kremer, M. (1996). How Much Does Sorting Increase Inequality? *The Quarterly Journal of Economics 112*(1) 115-139.

Kristof, N. (2014, October 25). The American Dream is Leaving America. *The New York Times*.

Lee, S., & Brinton, M. (1996). Elite Education and Social Capital: The Case of South Korea. *Sociology of Education, 69*(3), 177-192.

Narayan-Parker, D., and Petesch, P. (2007). *Moving out of Poverty: Cross-Disciplinary Perspectives on Mobility*. World Bank Publications.

O'Brien, M. (2014, October 18). Poor Kids Who Do Everything Right Don't Do Better Than Rich Kids Who Do Everything Wrong. *The Washington Post.*

Personal Consumption Expenditures/Gross Domestic Product. (2014). Economic Data, Federal Reserve Bank of St. Louis.

Rosen, S. (1992). Mincering Labor Economics. *The Journal of Economic Perspectives, 6*(2), 157-170.

Solon, G. (2002). Cross-Country Differences in Intergenerational Earnings Mobility. *The Journal of Economic Perspectives, 16*(3), 59-66.

Solon, G. (1992). Intergenerational Income Mobility in the United States. *The American Economic Review, 82*(3), 393-408.

Solon, G. (2013). Theoretical Models of Inequality Transmission Across Multiple Generations. Working Paper 18790, National Bureau of Economic Research.

Stuhler, J. (2014). Mobility Across Multiple Generations: The Iterated Regression Fallacy. Department of Economics, University College London.

The Economics of Higher Education (2012). Department of the Treasury, Department of Education.

Walpole, M. (2003). Socioeconomic Status and College: How SES Affects College Experiences and Outcomes. *The Review of Higher Education, 27*(1), 45-73.

Woolard, I., and Klasen, S. (2007) "Determinants of Income Mobility and Household Poverty Dynamics in South Africa." *The Journal of Development Studies, 41*(5), 865-897.

ACADEMIC VITA

Duncan Hayes

319 West Hamilton Avenue Apartment C • State College, PA 16801
(814)933-2873 • dwh5213@psu.edu

EDUCATION: **The Pennsylvania State University** — **University Park, PA**

Schreyer Honors College

Bachelor of Science in Economics — May 2015

Minor in Political Science

Paterno Fellows Program

Dean's List: All Semesters

ACTIVITIES: **The State Theatre** — October 2014-January 2015

Development Intern

- Assisted Development Director Chris Rand
- Performed data analytics on ticket sales
- Automated processes to streamline department
- Revised and edited board member handbook

Econ 306-Intro to Econometrics — January 2014-May 2014

Teaching Assistant for Dr. Daniel Goldstein

- Assisted and tutored students
- Graded homework and exams
- Proctored exams and led weekly study groups with more than 50 students

State College Food Bank — November 2013-June 2014

Volunteer

- Pack grocery carts, stock shelves, organize food, handle donations
- Assist clients with delivery and distribution of food

Penn State Economics Association — August 2012-Present

Member

- Discuss and analyze current and past economic events with other students
- Facilitate communication among students, faculty, and alumni
- Attend annual trip to the Federal Reserve

Phi Eta Sigma National Freshman Honor Society — May 2012-Present

Member

- Participate in service opportunities such as THON fundraising
- Network with other high-achieving students