ABSTRACT

Title of Document:	LOSS AVERSION AND THE INTERGENERATIONAL CORRELATION OF INCOME
	Liam Case Malloy, Doctor of Philosophy, 2010
Directed By:	Professor John Shea, Economics

Recent estimates of the intergenerational correlation of income in the United States are centered around 0.6. Existing empirical work is only able to explain about half of this correlation. The first chapter of this dissertation provides a behavioral explanation that accounts for almost half of the unexplained correlation. Heterogeneous agents in the model are loss averse and must choose their education level after learning their "earning ability" and inheriting a reference level of consumption and bequest from the previous generation. These agents make education choices in part to avoid losses relative to reference consumption in the first and second periods of their lives. Agents with high inherited reference consumption choose high levels of education in order to avoid losses in the second period and are therefore likely to have high income and consumption themselves. Those with very low reference consumption are likely to get more education than those in the middle of the reference consumption distribution, as they are less likely to experience a loss in the first period. I find support for this U-shaped education decision rule using the

NLSY97 data set. The dissertation also tries to answer the question of why black and white workers display significant differences in their labor market outcomes. Black workers tend to have less education and earn lower income than their white counterparts at each level of education. The second chapter explores three possibilities (wage discrimination, lower earning ability, and low aspirations) for these gaps within the framework of a model with loss aversion and inherited reference consumption. When people have loss-averse preferences, low aspirations lead to lower levels of chosen education. Loss aversion and low aspirations can lead to education outcomes similar to those caused by outright discrimination or lower earnings ability. When combined with wage discrimination the model can also help explain the larger poverty trap and lower affluence net in black families as opposed to white families. Simulation results compare favorably to intergenerational quintile transition rates in the literature. The model takes many generations to reach educational equality after a period of wage discrimination is ended.

LOSS AVERSION AND THE INTERGENERATIONAL CORRELATION OF INCOME

By

Liam Case Malloy

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor John Shea, Chair
Professor Pablo D'Erasmo
Professor Allan Drazen
Professor L. Bart Landry
Professor Carlos Vegh

© Copyright by
Liam Case Malloy
2010

Contents

List of Tables

List of Figures

Chapter 1

Loss Aversion, Education, and Intergenerational Mobility

Recent estimates of the intergenerational correlation of income in the United States are centered around 0.6. Existing empirical work looking at the effects of parental income on IQ, schooling, wealth, race, and personality is only able to explain about half of this correlation. This paper provides a possible behavioral explanation that could account for almost half of the unexplained correlation. The model has sequential generations in which heterogeneous agents are loss averse and must choose their education level after learning their "earning ability" and inheriting a reference level of consumption and bequest from the previous generation. These borrowing-constrained agents make education investment choices in part to avoid losses relative to reference consumption in the first and second periods of their lives rather than to maximize lifetime resources. Agents with high inherited reference consumption choose high levels of education in order to avoid losses in the second period and are therefore likely to have high income and consumption themselves. Those with very low reference consumption are likely to get more education than those in the middle of the reference consumption distribution, as the opportunity cost of forgone earnings during schooling is less likely to cause them to experience a loss in the first period. I find support for this U-shaped education decision rule using the NLSY97 data set. I simulate the model assuming zero correlation of earning ability between generations and show that it can explain approximately one-half of the unexplained intergenerational correlation of income. The model also offers an explanation for why we see in the data a minor poverty trap for white families, in which children from poor families are likely to remain poor, and a significant affluence net, in which those from rich families are more likely to stay at the top of the income distribution.

1.1 Introduction

A free market system can create a large amount of income inequality. The top 1% of earners in the United States earned 12.4% of the total wage income in the United States in 2007 (Piketty and Saez 2009). Acceptance of this system, especially in the United States, is generally predicated on the idea that each person has a good chance of success, no matter his starting point. However, recent research (e.g. Mazumder 2005, Mulligan 1997) shows that the correlation of income between parents and children is around 0.6. As Mazumder (2005) notes, at this level it will take six generations for the descendants of someone at 25% or 200% of the mean to be within five percent of the average.[1] Furthermore, Hertz (2005) shows that there is a significant affluence net, in which children of parents with income in the top decile have a high chance of remaining in at least the top quintile, while those with parents in the bottom decile are very likely to be poor themselves, caught in the well-documented poverty trap. Both Hertz and Mazumder (2008) show that white and black families experience quite different, and asymmetric, transition rates which would not be implied by a linear

[1] Becker and Tomes (1986) review a number of studies that find an intergenerational correlation of less than 0.2. At this level it would take only two generations for the descendants of someone at 200% of the mean to be within 5% of the average. Unfortunately, these studies suffered from a lack of data so that estimates of lifetime income were much noisier than in recent studies. Mazumder (2005) is able to use up to 16 years of father's income, averaging out the noise of transitory income, and finds an intergenerational correlation of 0.6. Other studies use instrumental variable methods and get similar estimates.

model with a correlation of 0.6.

Theoretically, there are a number of reasons why parents' and children's income (and education) may be correlated across generations. Those parents with more education will also tend to have more income, out of which they can pay for more (and higher quality) education for their children. Parents with fewer financial resources (who will also tend to have lower education) will be less able to finance the education of their children. The problem with this explanation is that it assumes imperfect capital markets in financing education. Given the high correlation between education and income, markets should (and do) exist to finance education expenses. According to data from the National Center for Education Statistics (NCES 2008), over 40% of post-secondary students receive student loans, which average over $5,700 yearly for students at private schools and $4,200 for those at public colleges and universities. In addition, between 25% and 35% of students at in-state public universities receive government and/or institution grants which average almost $3,400 for institutional grants and $2,400 and $3,200 for state and federal government grants, respectively. The typical tuition and room and board at a four-year public school was just under $13,000 in 2007-08, so that many students could finance a large part of their educational expense between grants and loans (all data from NCES 2008). Of course, there is the possibility, explored in this paper as well as others, that capital markets for education are not perfect and that lower-income families may face more severe borrowing constraints than higher income families. If children from poorer families find it more difficult to borrow, this may be one transmission mechanism by which parental income affects child income. And because loans are only available for higher education, children from lower-income families may receive an inferior early education if forced into low-quality public schools. This could generate a poverty trap in which those with the desire and ability to pursue higher education would be unable to do so due to this borrowing constraint.

There is a reasonable question as to how important genetic factors are in transmitting income from one generation to the next. Bowles, Gintis, and Osborne Groves (2005) conduct a meta-analysis of the empirical literature in an attempt to identify the causes of this intergenerational correlation. Their research shows that transmission of intelligence, as measured by IQ, actually plays a very minor role, while schooling and, unfortunately, race, play a more substantial role. Yet their analysis is still only able to account for approximately half of the observed intergenerational correlation of income. Bowles et al. show that the genetic transmission of IQ (despite being fairly high) is less important in transmitting income than one would perhaps suppose, because IQ and earnings are not very highly correlated. However, their paper also shows that there are other inheritable factors that contribute to the determination of income, such as good (or poor) work habits, which may be passed on genetically but also can be passed from one generation to the next culturally.

This paper proposes a novel transmission mechanism between parents and children that focuses on the differences in expectations, or habit formation, among children from different income levels. If children form a reference level of consumption during childhood (based on family consumption) and measure utility in reference to this level, then children of identical ability from different backgrounds may make very different education investments. Specifically, if people have loss-averse utility functions, so that they are risk averse in gains but risk loving in losses, then children from higher-income families will choose higher education than an otherwise identical person from a poorer family. Those at the very bottom of the income distribution will invest more in education than those immediately above them as they find it easier to replace reference consumption, generating a U-shaped education investment decision.

As a hypothetical example, imagine identical twins, Joe and Thomas, who are separated at birth. Joe is adopted by a working-class family while Thomas is adopted by a family from the upper-middle class. Suppose both are interested in electrical work, have similar SAT scores and grades, and have an equal cost of financing their education. They face two possible education and career paths.

One possibility for these hypothetical twins, after finishing high school, is becoming an electrician through an apprenticeship that allows one to work and earn a small amount while becoming qualified to take the licensing exam, after which one can earn around $20 an hour. The other option is a four-year college to get a degree in electrical engineering. This requires a large upfront expense which would have to be financed mainly through loans that would be paid back after finishing the program. This would require consuming at a low level while in school and while paying back the loans, but afterwards earning $35 an hour.

The model presented below asserts that Joe is more likely to choose the electrician route, while Thomas is more likely to choose to become an electrical engineer. The key to this result is that preferences are assumed to be loss averse, while reference consumption is determined by family income during childhood. Joe has a lower level of reference consumption and can enjoy a similar or higher consumption early in life by becoming an electrician, whereas going to school would force him to consume below reference as a young adult. Thomas has higher reference consumption. If he becomes an electrician he would consume below reference his entire life. If he goes to college and becomes an electrical engineer, he will at least be able to equal or exceed reference consumption in his later life.

More generally, people with loss-averse preferences and a low level of reference consumption will invest less in costly education in the first, or education, period of their life in order to avoid first-period losses (compared to their reference level of consumption), even though this will mean reducing the gains in the second, or work, period of life. A person with a high level of reference consumption is much more likely to face losses in one or both periods of life. In order to eliminate (or at least reduce) this loss in the second period of life, the higher-reference person will choose a higher level of education so as to maximize resources in the second period. This choice will lead to a correlation of income (and education and consumption) across generations even if earning ability is independent across generations. On the other hand, a model with habit formation but without loss aversion will only be able to generate intergenerational persistence if borrowing constraints are extremely severe or the curvature of the utility function is quite pronounced. In order to generate a positive intergenerational correlation of income, consuming at a loss needs to have a high enough utility cost compared to future gains that people are willing to reduce total lifetime consumption by investing in less education in order to avoid a first-period loss.

Some sociologists have interviewed those who were qualified for higher education but who didn't go on to a four-year college. In a survey, Connor (2001) finds that 40% of qualified UK students who didn't go on to university, and who had a lower socio-economic status than those who did, stated the desire to start working and earn money as a reason for their decision. One said:

> "what's the point slogging your guts out and there's no guarantee you will get a job whereas I believe in these next three years I will have started and worked a lot higher than [he] will have when he come out of university . . ."

Another study looks at students from the Chicago Public Schools (CCSR 2008) who also tended to come from lower socio-economic backgrounds. One student, whom the authors call Javier, had a 3.95 GPA, scored 21 on the ACT, and even had a scholarship to a four-year business school. He decided to go to a technical school offering an 18-month automotive certification instead. He said:

> "I decided to go to UTI because I was more interested in the program, and it's less time. The other colleges would have been three or four years. I just want to get the studies over with and go to work."

In a study of those in Appalachia, one of the poorer rural areas in the United States, Brown et al (2009) interviewed one student who clearly expressed the desire for certain types of consumption that would be unavailable if he were to go to college.

> "So, I mean, we have to work for stuff! If I go off to college though, and I am in college full-time, there's no way I can work. So who's going to pay for my truck? Who's going to pay for all my bills that I have?"

In addition, a study by Hahn and Price (2008) for the Institute of Higher Education Policy found that "within the non-college-going population, almost half of Black and FRPL [low-income] students...stated that the need to work was "extremely" or "very" important, compared with an average of 38 percent for all non-college-goers. This greater need to work did not appear to be the result of greater family obligations for minority and low-income students."

This evidence suggests that the utility cost of foregone earnings and consumption is an important determinant of education choices for some young adults. But is there evidence that reference consumption depends on a person's level of consumption as a child? Some papers in the literature on aspirations assume this to be the case (e.g. de la Croix, 2000) while others view aspirations as coming from those around them or above them in the income distribution. Genicot and Ray (2009) use an S-shaped utility function (à la loss aversion) but assume that parents choose their children's education investment and that aspirations are a function of the future income distribution rather than the past. However, Waldkirch et al (2004) use the PSID to measure intergenerational consumption patterns and find that, controlling for parent and child income, consumption reference levels are correlated with the consumption of the previous generation. Additionally, McGrath et al (2001) look at the educational aspirations of rural youth in Iowa. Among middle class families, education aspirations and outcomes are similar for the children of the more educated professional middle class and for the children of the less educated farming middle class. These education aspirations and outcomes are both higher than for the group of children whose parents have lower socio-economic status. In a world in which the path to middle-class income goes increasingly through higher education, these children of less educated middle class farmers appear to choose college as a way to maintain their higher place in the income distribution. While there is no direct evidence that students have loss averse preferences when making their education investment decision, the examples cited above are suggestive. It is certainly possible that loss aversion plays a role in this important investment choice.

Formally, I develop a model with sequential generations in which heterogeneous agents are loss averse and inherit their reference level of consumption from their parents. There are two main periods in life: an education period and a work period. There is no uncertainty for agents in the model so that wages are completely determined by observable earning ability and education (which act as complements). The model also allows for the parent generation to give a "warm glow" bequest to the child generation which may be used for education or consumption. A number of different types of borrowing constraints are explored as well. The baseline model allows agents to borrow the complete (direct) cost of education, although agents cannot borrow to finance consumption. I also present results for cases in which there are no borrowing constraints and in which the bottom of the income distribution does not have access to financial markets.

The baseline model with loss aversion and weak borrowing constraints is able to generate an intergenerational correlation of wages of 0.14 even when earning ability is completely independent between generations, suggesting that loss aversion is one possible explanation for the unexplained income correlation between generations. In contrast, a model with standard concave utility and habit formation generates basically no correlation of income under weak borrowing constraints. Only when borrowing constraints are much more severe can the model with habit formation and standard concave preferences generate a similar level of intergenerational correlation of earnings. On the other hand, when there are no borrowing constraints, none of my models generate an intergenerational correlation of wages. Intuitively, if agents can borrow to finance consumption, they can avoid losses in the first period and should always invest in the level of education that maximizes lifetime discounted earnings.

The model with loss aversion generates a possibly counter-intuitive, but testable, implication for education investment decisions. Because those at the bottom of the parental income distribution, with very low levels of reference consumption, will find it easier to consume without a loss in the first period of life, even if they go to school, they will tend to invest in more education than those who are slightly above them in the parental income distribution. As described above, those with high levels of reference consumption will tend to get more education to avoid a loss in the second period of life. This generates a U-shaped decision rule in education with family consumption as the explanatory variable. In contrast, the model with habit formation but without loss aversion will generate either a flat or upward sloping education decision, depending on the severity of borrowing constraints. I test this implication using the NLSY97 data set and find support for a quadratic, U-shaped decision rule in education. I use average family income before the respondent's 18th birthday as the explanatory variable and control for own ability and parental education.

The next section describes the model with particular attention to the assumption of loss-averse preferences. In section three I discuss relevant data moments from the intergenerational persistence literature, which I will use to test my model. These moments include the correlations of income, education, and consumption across generations as well as non-linearities in the transition matrix between income quantiles, especially the poverty trap and the affluence net. The calibration is presented in section four along with simulated results for both the model with loss aversion and the model with only habit formation. The former is able to generate a positive correlation of intergenerational income mainly due to an affluence net, in which the children of the rich choose to get more education than the children of the poor and so have higher earnings. This matches the data for white families fairly well, but cannot explain the different outcomes experienced by black families, who face a significant poverty trap. Section five presents the empirical results from the NLSY97 which support the U-shaped education decision rule implied by the theoretical model. Section six offers a conclusion and a number of avenues for further research, including the mechanisms I explore in Malloy (2009) to explain persistent differences in group outcomes, especially between white and black families.

1.2 Model

1.2.1 Basic Setup

The model presented here is a fairly straightforward partial-equilibrium model with sequential generations. Agents live for three periods. The first period, labeled 0, can be thought of as childhood, in which agents do not have to make any decisions. Children learn their earning ability level, ψ, their reference consumption level, c_0, and the bequest, T, that they will receive at the beginning of period one. Both the reference level of consumption and the bequest come directly from their parents.

Earning ability differs across agents and is normally distributed:

$$\begin{aligned}\psi_i &= \rho\psi_{i-1} + (1-\rho)\bar{\psi} + \chi_i \\ \chi_i &\sim N(0, (1-\rho^2)\sigma_\psi^2)\end{aligned}$$

I assume, in the baseline model, that earning ability is independent across generations. That is, a child's earning ability is completely independent of the earning ability of the parent so that $\rho = 0$. Thus each agent draws a random earning ability from a normal distribution with mean $\bar{\psi}$ and standard deviation σ_ψ.

The second period of life, labeled 1, is the education period. Agents, knowing their earning ability, reference consumption, and bequest, invest in education to maximize lifetime utility. Because there is no uncertainty in this model, agents are also able to choose consumption in periods 1 and 2, saving (or borrowing) between periods, and the bequest level they will leave to the next generation. The value function is described by equation 1.

$$V(c_0, T, \psi) = \max_{e, c_1, c_2, T'} U(z_1) + \beta[U(z_2) + B(T')] \tag{1a}$$

$$s.t.$$

$$s \geq \Omega \tag{1b}$$

$$\varepsilon(e) \leq T + \Omega \tag{1c}$$

$$c_1 + s + \varepsilon(e) = (1-e)\omega(e,\psi) + T \tag{1d}$$

$$c_2 + T' = \omega(e,\psi) + s(1+r) \tag{1e}$$

$$c_1, c_2, T' \geq 0 \tag{1f}$$

Where z_t is the percent change from the reference level of consumption: $z_t = \frac{c_t - c_0}{c_0}$.

Note that this gives a natural minimum to z_t of -1, which is important when thinking about loss aversion, as discussed below. The level of education is e, and the function $B(\cdot)$ measures the warm glow utility derived from giving a bequest, T', to the next generation. The wage is determined by $\omega(\cdot,\cdot)$ which I assume to have the form:

$$\ln[\omega(e,\psi)] = \alpha_1 e + \alpha_2 e^2 + \zeta e\psi + \kappa\psi \tag{2}$$

Where I assume that $\alpha_1, \alpha_2, \zeta, \kappa \geq 0$. This is a fairly standard semi-log wage function that is quadratic in education. The major difference between equation (2) and a more standard wage function from the labor literature is that earning ability, ψ, is observable. The parameter ζ represents the idea captured in quantile wage regressions (e.g. Lemieux 2006) that the return to education is higher for those with higher earning ability. The cost function for education is given by:

$$\varepsilon(e) = \phi_1 e + \phi_2 e^2 \tag{3}$$

With $\phi_1, \phi_2 \geq 0$ so that education costs are at least weakly convex. The wage function and the education function assure that there is some level of education, e^*, which may be at a corner, that uniquely maximizes lifetime resources for each level of earning ability, ψ.

The first two constraints in equation (1) reflect potential borrowing constraints, where borrowing is less than some level Ω (discussed below) and where the agent may only use her bequest and her borrowing in order to finance the education expense, $\varepsilon(e)$. The two constraints (1d) and (1e) represent the per period budget constraints, where agents are not allowed to die in debt and the final constraint ensures that consumption and bequests are non-negative. Savings are assumed to grow at a risk-free rate r. Agents face both a direct cost of education, $\varepsilon(e)$, and an opportunity cost of education so that they work only $(1-e)$ in the first period. The wage per unit of time worked is the same in both periods and depends on the chosen level of education, e. Because the wage is increasing in the chosen level of education, this opportunity cost is also convex.

The first order conditions for each agent are straightforward and are given in equation (4).

$$e : -[\omega(e,\psi) + \varepsilon'(e)]U'(z_1) + \omega_e(e,\psi)U'(z_2) + \gamma_2\varepsilon'(e)[\varepsilon(e) - T - \Omega] \geq 0 \tag{4a}$$

$$s : -U'(z_1) + \beta(1+r)U'(z_2) + \gamma_1 = 0 \tag{4b}$$

$$T' : \beta[-U'(z_2) + B'(T')] = 0 \tag{4c}$$

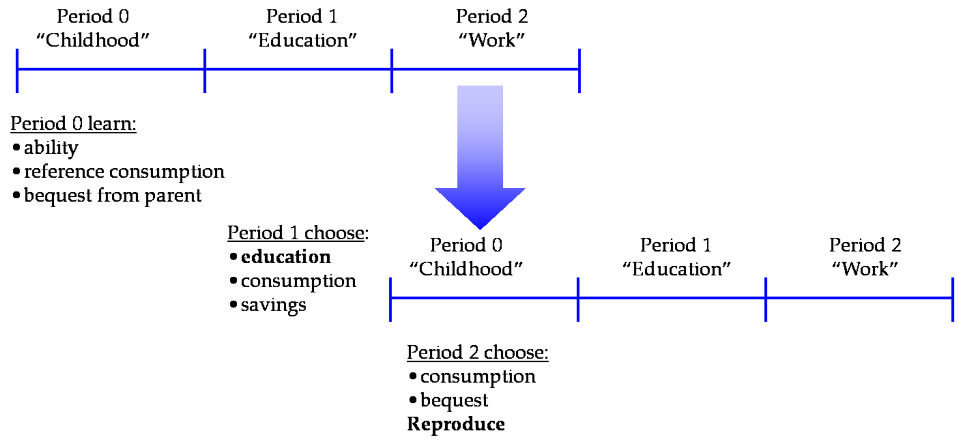
$$\gamma_1 : s + \Omega \geq 0 \tag{4d}$$

$$\gamma_2 : T + \Omega \leq \varepsilon(e) \tag{4e}$$

Equation (4a) characterizes the utility-maximizing level of education, where γ_2 is the multiplier on the education-financing constraint, so long as $U(\cdot)$ is concave. Equation (4b) gives the standard intertemporal Euler equation where γ_1 is the multiplier on the borrowing constraint (1b), (4c) gives the intratemporal Euler equation between consumption and bequests and (4d) and (4e) are the borrowing constraints. However, as will be discussed in more detail below, if agents have loss averse preferences, then $U(z)$ is convex in losses. In this case, the second order condition for a maximum fails and the FOCs cannot be used to solve the system.

The connection between the generations is straightforward. In period 2, agents reproduce so that the next generation's period 0 will match up with period 2 of the previous generation. The consumption level chosen by the parent generation in period 2 becomes the reference level of consumption for the child generation. Further, the bequest level chosen by the parent is given to the child at the beginning of period 1. Figure 1.1 summarizes these intergenerational connections. The intergener-

Figure 1.1: Connections Between Generations



ational persistence in the model comes from the fact that children inherit the reference consumption directly from their parents. As described below, this has important implications in a model with loss aversion as children will try to avoid consuming below this reference level due to the much higher marginal disutility of doing so. In contrast, this is not the case in the model with habit formation as the disutility of consuming below the reference consumption is only marginally higher as utility is everywhere concave.

1.2.2 Loss Aversion

The utility function, $U(z_t)$, will be allowed to take two main forms. In the version of the model with habit formation (but no loss aversion), the per period utility will be given by equation (5).

$$U(z_t) = \frac{1 - e^{-\mu z_t}}{\mu} \tag{5}$$

Equation (5) is a standard constant absolute risk aversion (CARA) utility function which is increasing and concave for all levels of z_t. In this case, agents will try to smooth consumption across the two periods. The utility function is represented in

Figure 1.2: Utility and Marginal Utility with Habit Formation Only

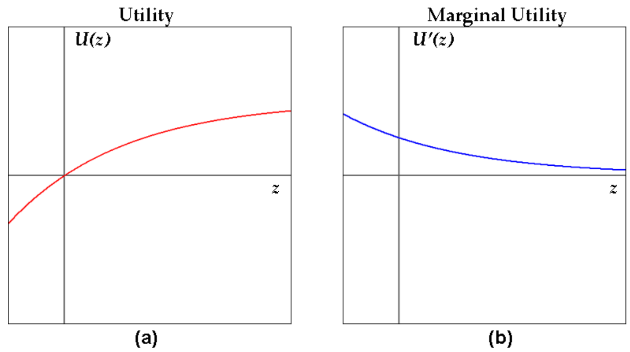


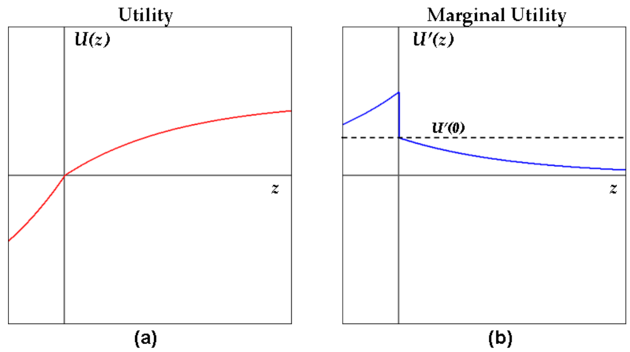
figure 1.2(a) and the marginal utility is in figure 1.2(b).

The more interesting case (at least in this model) is when the utility function represents loss averse preferences. Loss aversion was first introduced by Kahneman and Tversky (1979) as part of their prospect theory. The key insight represented by loss aversion is that agents are risk averse when it comes to gains (as in traditional expected utility theory) but are risk loving when it comes to losses, so that utility is concave in gains but convex in losses. In addition, the marginal utility of a loss is around twice as large as the marginal utility of a comparably sized gain. In this model there is no risk per se (as there is no uncertainty for the agent), so a loss averse utility function means that agents will avoid losses as much as possible. We can represent loss averse preferences by equation (6) as suggested by Kobberling and Wakker (2004).

$$U(z_t) = \begin{cases} \frac{1-e^{-\mu z_t}}{\mu} & \text{for} \quad z_t \geq 0 \\ \lambda \frac{e^{\nu z_t}-1}{\nu} & \text{for} \quad z_t < 0 \end{cases} \tag{6}$$

The loss aversion parameter λ was estimated by Kahneman and Tversky to be around 2.25. A value larger than one means that marginal utility is larger for losses than for comparably sized gains. Note also that utility is concave in gains but convex

Figure 1.3: Utility and Marginal Utility for Strongly Loss Averse Preferences

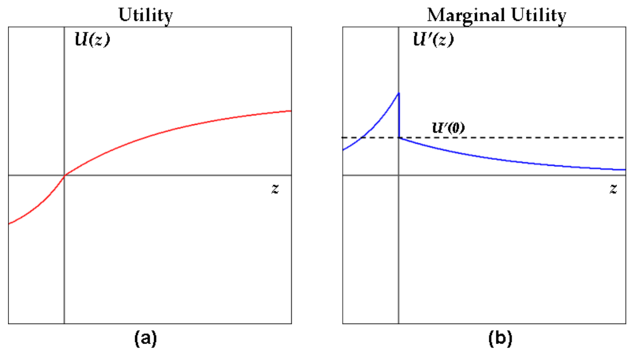


in losses. That is, marginal utility is decreasing away from zero in both directions, so that just as the marginal utility of a gain decreases as the gain increases, so too does the marginal disutility decrease as the loss increases.

Neilson (2002) has suggested a nomenclature for loss aversion which divides potential loss averse utility functions into those exhibiting weak and strong loss aversion. With strong loss aversion, marginal utility is everywhere higher for a loss than for a gain. That is, a loss averse function $U(\cdot)$ is strongly loss averse if $U(0) = 0$ and $U'(y) \leq U'(z) \quad \forall z < 0 < y$. In this case, since z_t is measured in percentage terms (and so has a minimum at -1) we can write $U'(-1) \geq U'(0)^+$ where $U'(0)^+$ is the marginal utility as z_t approaches 0 from the right. A strongly loss averse utility function is represented in figure 1.3(a) and marginal utility in figure 1.3(b).

For a weakly loss averse utility function the average utility of a gain is smaller than the average disutility of a loss. Neilson states that $U(\cdot)$ is weakly loss averse if $U(0) = 0$ and $U(y)/y \leq U(z)/z \quad \forall z < 0 < y$. For a weakly loss averse function it is possible that the marginal utility of a gain close to zero (where the marginal utility of a gain is highest) may be higher than the marginal disutility of a large loss. Note that strong loss aversion implies weak loss aversion, but not the other way around. Weakly loss averse utility and marginal utility are represented in figure 1.4. Note that in this case we have $U'(-1) < U'(0)^+$, represented by the dashed line in figure

Figure 1.4: Utility and Marginal Utility for Weakly Loss Averse Preferences



1.4(b).

Tversky and Kahneman (1992) adopt a constant relative risk aversion (CRRA or power) functional form, and many papers that use loss aversion have used this type of function. However, in a CRRA utility function, marginal utility approaches infinity as the gain or loss approaches zero. Thus it is impossible for a CRRA loss averse utility function to exhibit strong loss aversion. A CARA type loss averse function such as (6) can display either strong or weak loss aversion depending on the relative values of λ and ν. In this paper, I restrict these values so that utility is always strongly loss averse (for all $z_t \geq -1$), as in Bowman et. al (1999). With the CARA-type utility function in equation (6), this requires that $\nu < \ln(\lambda)$.

Because the marginal utility of a loss is always larger than the marginal utility of a gain, the consumption and saving decision in a deterministic two period model is fairly straightforward, as Bowman et. al (1999) show. For a given investment in education, agents will either face a gain in both periods, a loss in both periods, or a loss in one period and a gain in the other period. Because utility is concave in gains, agents will smooth gains across the two periods (as much as possible, given the borrowing constraint). If an agent faces two losses, utility will be maximized by putting as much of the loss into one period as possible and minimizing the loss in the next period. The larger loss will, by necessity, be in period 1, the education period, due to the borrowing constraint. Finally, if the agent faces a loss in period 1 and a gain in period 2, she will choose to minimize the loss as much as possible by reducing the gain in period 2. Eliminating the loss in period 1 may not be possible if the agent faces a borrowing constraint. More details of the solution are given in section 2.5.

Given that loss aversion has both experimental and empirical support it may seem odd that it has not been used more in existing literature. However, loss aversion poses some theoretical problems. Because utility is convex in some areas, in any model with more than two periods there will be multiple equilibria. An agent who groups all her losses into one period is indifferent as to which period that is. Furthermore, if the level of reference consumption is under the control of the agent there can be further perverse results in which the agent will reduce reference consumption in order to increase future utility. This point is discussed further in section 2.4. However, if we are willing to impose further restrictions on the model to rule out such complications, models with loss aversion may shed light on problems that conventional models cannot address.

1.2.3 Borrowing Constraint

Borrowing constraints, represented by Ω in equation (1), interact with loss aversion and habit formation in important ways. Easier credit, or less severe borrowing constraints, will allow agents to avoid losses. Tighter borrowing constraints, on the other hand, make a loss in period one much more likely even if the agent will have a large gain in period two.

I explore four different degrees of borrowing constraint in this paper. First, there may be perfect capital markets. In this case there is no period one borrowing constraint but only a lifetime solvency constraint. In this case we can write:

$$\Omega_{lt} = -[\frac{\omega(e,\psi)}{(1+r)}] \quad \text{(lifetime budget constraint only)} \tag{7a}$$

Second, I assume that agents can borrow up to their cost of education (assuming this is less than the lifetime borrowing constraint). In this case the borrowing constraint becomes:

$$\Omega_{ed} = -\varepsilon(e) \quad \text{(education cost borrowing constraint)} \tag{7b}$$

In this case, students can borrow the entire direct cost of their education, but are unable to borrow in order to finance consumption. In some cases this may be a realistic assumption. Many students do borrow a large portion of the cost of college and graduate school.

Third, I assume that savings between periods one and two must be non-negative. In this case, students cannot borrow at all to finance their education. Students must pay their entire education bill through their bequest, T. Thus we have:

$$\Omega_{no} = 0 \quad \text{(no borrowing allowed between periods)} \tag{7c}$$

In reality the borrowing constraint probably falls somewhere between Ω_{ed} and Ω_{no} for most students. Some education loans require parents to co-sign the loan[2]. In addition, Winter (2009) shows that as many as 18% of households may face borrowing constraints when financing education. The final borrowing constraint I explore in this paper is a function of parental income:

[2]According to the Department of Education (http://federalstudentaid.ed.gov/federalaidfirst/), while Stafford loans only require a student signature, so-called parent PLUS loans require a credit check and co-signer as do most private (i.e. not federally subsidized) loans.

$$\Omega_p = f[\omega(e_{-1}, \psi_{-1})] \quad \text{(borrowing constraint based on parent income)} \tag{7d}$$

Where $\omega(e_{-1}, \psi_{-1})$ is the income of the previous generation. In this case, agents from poorer families will face a more severe borrowing constraint than agents from richer families. As we will see, this may play a significant role in keeping those who are born poor from rising substantially in the income distribution.

1.2.4 Bequest Motive

The function $B(T')$ in equation (1) gives the utility gained from leaving a bequest of size T'. This type of bequest is known as a warm glow bequest because the person giving the bequest does not take the *utility* of the person receiving the bequest directly into consideration. I use the same CARA function for $B(\cdot)$ as is used for gains in utility:

$$B(T') = \frac{1 - e^{-\theta(T'/c_0)}}{\theta} \tag{8}$$

Because z_t is measured in percentage terms, it is useful to scale bequests by the level of reference consumption. If $\theta = \mu$ then the agent sets $\frac{T'}{c_0} = z_2$ when $z_2 \geq 0$ or $T' = c_2 - c_0$. In many cases, however, agents will optimally have zero gain in the second period, in order to minimize losses in period one. An agent consuming at (or below) the reference level of consumption in period 2 will leave no bequest to the next generation.

Allowing agents to take into account the utility of the next generation in deciding on bequest levels, rather than assuming warm glow bequests, would open up a number of complicating issues. First, it would require an expansion of the state space to account for the next generation's (now known) earning ability. Parents would treat a child with high earning ability differently than they would treat a child with low earning ability. This may in fact be the case for parents with multiple children. It seems less likely (but still possible) in the present model where each parent generation only has one child in the next generation.

Second, if a parent takes his child's utility into account when choosing his bequest, it seems only natural that he should also think of his child's utility when choosing consumption in the second period of life. After all, second period consumption in the parent generation becomes the reference level of consumption for the child generation (while not, in this model, directly affecting the child's lifetime utility). A purely altruistic parent would have an incentive to reduce second period consumption (while increasing the bequest) so that the child would have higher lifetime utility. Some parents may indeed reduce consumption on the margin as children are growing up so as to set more reasonable lifetime expectations, but it seems unrealistic that parents would choose less safe cars, more dangerous neighborhoods, or lower quality schools than they can afford.

The assumptions of strong loss aversion, warm glow bequests, and a CARA-type utility function are necessary to find a unique solution to the model but may drive many of the model's results. In order to test these assumptions, I would need to solve the consumption-saving decision for weak loss aversion. Another reasonable extension would be to add an Inada-type utility function to the loss aversion function so that agents never consume zero in the first period. Adding a bequest motive that takes into account the utility of the child may be impossible, as discussed above, but changing the warm glow bequest function to match the data on inter-vivos transfers and end of life bequests may help make the model more accurate. I leave these more complex modeling assumptions for future work.

1.2.5 Solution Methodology

As noted above, with loss averse preferences, the utility function is concave in gains and convex in losses. Given this fact, we cannot use the first order conditions to solve for the utility-maximizing levels of education, consumption, and bequests. However, because the model makes the assumption of strong loss aversion, and given that agents choose consumption for only two periods, we can still solve for how they will divide a certain level of lifetime resources between periods (as in Bowman et al (1999)).

For a given level of education we can calculate both total lifetime resources and the maximum consumption in the first period and minimum consumption in the second period, which will vary depending on the given borrowing constraint. For simplicity, I assume that $\beta = 1 + r = 1$. Letting $R^{lt}(e, \psi)$ denote lifetime resources, $R^1_{max}(e, \psi)$ denote maximum period one resources and $R^2_{min}(e, \psi, T')$ denote minimum period two consumption when period one consumption is maximized we can write:

$$R^{lt}(e, \psi) = (1 - e)\omega(e, \psi) + T + -\varepsilon(e) + \omega(e, \psi) - T' \tag{9a}$$

$$R^1_{max}(e, \psi) = (1 - e)\omega(e, \psi) + T - \varepsilon(e) - \Omega \tag{9b}$$

$$R^2_{min}(e, \psi, T') = \omega(e, \psi) + \Omega(1 + r) - T' \tag{9c}$$

We need to evaluate a number of possible cases. First, when the agent has a non-binding borrowing constraint and enough lifetime resources to consume a gain (or at least avoid a loss) in both periods (i.e. $R^{lt} \geq 2c_0$) then the standard conclusion holds and the agent will smooth consumption across periods so that $c_1 = c_2 = T' + c_0 \geq c_0$. In this case we have $c_2 = \frac{R^{lt}+c_0}{3}$. With the same lifetime resources, but a binding borrowing constraint, the agent will consume as much as possible in period 1 (which may be less than reference consumption) and the remainder in period 2. In this case

$c_1 = R^1_{max} \leq_? c_0 < c_2 = T' + c_0 = \frac{R^2_{min}+c_0}{2}$. Note that this is how the agent will allocate resources for a given education level. Below, I will characterize the level of education that maximizes utility.

If lifetime resources for a given level of education are not large enough to consume at or above the reference level of consumption in both periods but are large enough to consume at reference consumption for one period ($c_0 < R^{lt} < 2c_0$), the agent will put all losses into period one and consume at reference in period two. Consumption will then be: $0 \leq c_1 < c_0 = c_2, T' = 0$. Under loss aversion, the agent will not optimally smooth losses between periods because the marginal utility gained in period one will be lower than the marginal utility lost in period two, given that marginal utility is decreasing in the size of the loss. Assuming that the agent is strongly loss averse, he will not increase the loss in period one in order to consume at a gain in period two because the marginal utility from the loss is always larger than the marginal utility of any sized gain. Note that in some cases, the same utility could be gained by consuming at reference consumption in period one and consuming whatever is left in period two. However, so long as there are borrowing constraints (and $\beta(1 + r) = 1$), it is more likely that consumption in period two will be larger.

If lifetime resources are less than reference consumption ($R^{lt} < c_0$), agents will take as large a loss as possible in period one and consume lifetime resources in period two. That is, $c_1 = 0 < c_2 = R^{lt} < c_0, T' = 0$. The intuition is the same as the previous case. Because marginal utility is decreasing in the size of the loss, the agent wants to minimize one of the losses. Increasing consumption in period one and decreasing period two consumption would reduce utility because the marginal utility gained in period one would be smaller than that lost in period two. These consumption and bequest patterns are summarized in Table 1.

Next, we can write the indirect lifetime utility function associated with each level of education, based on the consumption and bequest pattern solved for above:

Table 1.1: Consumption and Bequest Patterns for a Given Level of Resources

R^{lt}	R^1_{max}	c_1	c_2	T'	Description
$\geq 2c_0$	$\geq c_0$	$\frac{R^{lt}+c_0}{3}$	$\frac{R^{lt}+c_0}{3}$	$\frac{R^{lt}-c_0}{3}$	Spread gains
	$< c_0$	R^1_{max}	$\frac{R^2_{min}+c_0}{2}$	$\frac{R^2_{min}-c_0}{2}$	Borrowing constrained
$c_0 \leq R^{lt} < 2c_0$		$0 \leq c_1 \leq c_0$	c_0	0	Concentrate loss in period 1
$< c_0$		0	R^{lt}	0	Concentrate loss in period 1

$$W(e, c_0, T, \psi) = \max_{c_1, c_2, T'} U(z_1) + \beta[U(z_2) + B(T')]$$

The optimal level of education, $\hat{e}$, maximizes this indirect utility function:

$$\hat{e} = \underset{e}{\text{argmax}}\ W(e, c_0, T, \psi) \tag{10}$$

Subject to the constraints in equation (1).

In my numerical simulations below, I assume the agent has a discrete number of education choices available, designed to mimic a high school dropout, a high school graduate, some college/Associate's degree, Bachelor's degree, Master's degree, and a PhD/professional degree. Given the discrete choice set it is easy to find the utility-maximizing education level by simply comparing total lifetime utility for each education level.

In addition, we can calculate e^*, the level of education that maximizes lifetime resources, in the absence of borrowing constraints:

$$e^* = \underset{e}{\text{argmax}}\,(1-e)\omega(e, \psi) - \varepsilon(e) + \omega(e, \psi) \tag{11}$$

The education level e^*equates the cost (both direct and opportunity) of education with the benefit in terms of wages. No agent will ever choose an education level higher than e^*. Adding an extra unit of education above e^* will reduce lifetime resources and either increase a loss or reduce a gain in one or both periods, and therefore must reduce utility.

However, it is possible that the utility-maximizing level of education is less than the resource-maximizing level ($\hat{e} \leq e^*$). This is possible because of the presence of borrowing constraints and loss aversion. Reducing the level of education from e^* will increase resources available in period one by reducing the direct and opportunity cost of education. These extra resources can then be used to reduce or eliminate a loss in the first period even if it means reducing a gain in the second period. For example, imagine a person who would maximize lifetime resources by receiving a Master's degree. In order to do this, she needs to be in school (and, therefore, not working) for much of period one. If she faces a borrowing constraint, she may well be forced into consuming at a level less than her reference consumption in period one. She may be able to consume well above her reference consumption in period two with her Master's degree, but she is unable to bring this consumption forward due to the borrowing constraint. In this case she may optimally choose some lower level of education, such as a Bachelor's degree, in order to reduce or eliminate the loss in period one, even though it will mean reducing the gain in period two due to her now lower wage in the second period.

For a given borrowing constraint, there are two factors that increase the likelihood that a person will get less education than the amount that would maximize lifetime resources. First, those who have higher earning ability (and, therefore a higher e^*) will be more likely to invest in less education. This is simply because the education choice set is larger for this group. For someone who maximizes lifetime resources with a high school diploma, the only other option is dropping out. Second, those who have a lower reference consumption level will be less likely to invest in the resource-maximizing level of education. This is because they will find it easier to replace reference consumption with a lower level of education than someone who has a high reference level of consumption. Put another way, a person who grows up rich is likely to have a loss in the first period no matter what. In order to minimize or eliminate the loss in the second period, they will need to get the resource maximizing level of education, e^*. The additional loss in period one from getting the extra education will be outweighed by reducing the loss in period two.

The numerical solution methodology is straightforward. Given the optimal consumption/saving/bequest pattern conditioned on education, as shown in Table 1, I calculate the utility-maximizing (as well as the resource-maximizing) level of education for each possible state (where the state consists of reference consumption, earning ability, and bequest). Each education level then implies consumption in each period, saving/borrowing between periods, and a bequest level for the next generation. I simulate an economy with N families, each starting with a normally-distributed randomly assigned earning ability, reference consumption, and bequest. I calculate the education choice, $\hat{e}$, for each person in the initial generation as well as their consumption in each period, savings and bequest. I then assign a random earning ability level to each of the N agents in the next period and carry over their reference consumption and bequest level from the previous generation. I can then calculate the correlation of education, wages, consumption, and bequests between generations. I run the simulation for 100 generations and check to make sure that the distributions of wages, consumption, and bequests have converged[3]. I then throw out the first 10% of generations and use the last 90% to calculate the target moments which I compare to the data. These simulation results are reported in section four, but first I review the data that the model is attempting to match.

1.3 Data

Measuring the correlation of income between generations is conceptually straightforward. However, there are two main problems for the empirical economist to overcome.

[3]For wages, for example, I add up the wages in each decile and then take the norm of the difference between generations. For convergence, I require this to be less than some number ϵ.

First, most data sets do not span a long enough time frame to include income from more than one generation. Second, those data sets that do have a long time span, such as the Panel Study of Income Dynamics (PSID), usually don't have a lot of years of income for both generations. Studies that used only a limited number of years for either or both generations are prone to measurement error because they are likely to be measuring transitory income rather than permanent income (see Bowles et al (2005) for a discussion of the literature).

Early measurements of the correlation of income between generations based on a small number of years of income data for each generation were typically no higher than 0.2. Becker and Tomes (1986), for example, report an average of 0.15. With such a low correlation, any advantage that one family has over another will almost completely disappear after only two generations. Becker and Tomes use these estimates to dismiss the idea that economic success or failure is passed along from one generation to another along with green eyes or curly hair.

More recent studies based on more years of data for at least the parent's generation, however, overturn this conclusion fairly convincingly, finding intergenerational income correlations of at least 0.4. A large data set put together by Mazumder (2005) uses Social Security Administration data in the U.S., matched with the 1984 Survey of Income and Program Participation, which allows him to use up to 16 years of a father's income and four years of the child's income. He finds that the estimated intergenerational income correlation increases as the number of years of income for the father increases. The correlation is less than 0.3 using only two years of father's income but climbs as high as 0.6 when using 16 years of father's income. The latter average is a better proxy for lifetime income as it smooths out the bumps from transitory shocks. A correlation of income of 0.6, as Mazumder notes, has significantly different implications than one of 0.2 or even 0.4, as it implies that the descendants of someone at 200% or 25% of the mean income will not be within 5% of the average until after six generations.

Mulligan (1997) uses the PSID to measure the intergenerational correlation of a number of relevant factors using an instrumental variable approach to overcome measurement error. A regression of child wages on parent's wages yields an estimated coefficient of between 0.32 (OLS) and 0.53 (IV). Family income has a coefficient between generations of between 0.47 (OLS) and 0.71 (IV), while family consumption has a coefficient of between 0.54 (OLS) and 0.77 (IV)[4]. It is interesting to note that parental consumption appears to be more correlated with children's outcomes than is parental income or wages. This fact may support a model such as the one presented in this paper in which agents have (at least) habit formation and possibly loss-averse preferences. Alternatively, it may be capturing consumption smoothing given that measured annual income is fairly noisy so that consumption may be a better measure of permanent income than income itself.

While most studies of intergenerational persistence use linear regressions, a handful take a nonlinear approach. Hertz (2005) constructs transition matrices from parental income deciles to child's deciles using the PSID. Adjusting income for age, he finds that a child with a parent in the bottom income decile has a 31.5% chance of staying in that bottom decile and over a 50 percent chance of being in the bottom quintile. On the other extreme, a child with a parent in the top income decile has a 29.6% chance of staying in that top decile and a 43.3% chance of remaining in the top quintile.

In this paper, I will focus on four key intergenerational income transition rates. The poverty trap measures how likely it is for children born in the bottom quintile of the income distribution to stay there as adults while the affluence net measures how likely children born in the top quintile will stay there. The jump up rate measures

[4] In Mulligan's instrumental variable regressions he uses family income as an instrument for family consumption which is measured with a lot of noise, as well as using instruments such as occupation and school categories as instruments for lifetime parental income.

how often children born in the bottom quintile are able to jump up to the top quintile of the income distribution, while the fall down rate measures how many children from the affluent top quintile fall down to the bottom quintile.

Interestingly, the transition matrices are quite different for white and black families. Mazumder (2008) calculates transition probabilities from each income quintile by race, using the NLSY79 data set. He finds similar transition numbers as Hertz for white families. Using income quintiles, he finds a poverty trap of only 24.9% for white families (where we would expect it to be 20% with perfect mobility) and an affluence net of 38.9%. Comparable numbers in Hertz are 27.2% and 38.5%, respectively. For black families, Mazumder finds a poverty trap of 43.7%, and an affluence net of 21.3%. An affluence net of around 20% is, in fact, no net at all as it is the value we would expect if outcomes were completely random between generations. In both Hertz (2005) and Mazumder (2008), the black and white income quintiles are calculated using the entire population, rather than a subset of only white workers or only black workers. In the simulations reported below I calculate income quintiles in the same manner so as to be better able to compare the model's results with the data.

These matrices also allow us to measure the likelihood of extreme movements in the income distribution between generations. First, how many rags-to-riches stories are there? That is, how many children born in the bottom quintile rise to the top quintile? Mazumder (2008), using the NLSY79, finds that 10.6% of poor white children make the jump to the top compared to only 4.1% of poor black children. Second, how many children fall through the affluence net and go from the top quintile to the bottom? Mazumder finds that 10.4% of white children suffer this reversal of fortune, while 21.6% of black children born into the top quintile fall to the bottom. Because of the significant differences by race, in my numerical calibrations I focus on matching the distribution of white families and leave possible theoretical explanations for why

Table 1.2: Wage Correlation and Quintile Transition Rates in a Linear Model

	Data*		Linear Simulations			
	White	Black	$\rho_\omega = 0$	$\rho_\omega = 0.2$	$\rho_\omega = 0.4$	$\rho_\omega = 0.6$
Correlation of Wages			0.00	0.19	0.41	0.60
Poverty Trap	24.9%	43.7%	20.4%	28.2%	37.7%	49.5%
Affluence Net	38.9%	21.3%	19.8%	27.8%	37.6%	49.2%
Jump Up	10.6%	4.1%	18.7%	13.2%	6.7%	2.2%
Fall Down	10.4%	21.6%	19.8%	14.2%	6.8%	2.4%

* Source for data is Mazumder (2008).

black family transition rates are so different to future work.

Asymmetries between the poverty trap and affluence net and between the jump up rate and the fall down rate cannot be explained in a simple linear model. Imagine that wages are random but that there is some level of correlation, ρ_ω, between the generations:

$$\begin{aligned} \omega_i &= \rho_\omega \omega_{i-1} + (1-\rho_\omega)\bar{\omega} + \epsilon \\ \epsilon &\sim N(0, (1-\rho_\omega^2)\sigma_\epsilon^2) \end{aligned}$$

Where ω_i is the wage of the child and ω_{i-1} is the wage of the parent. It is easy to simulate such a model and generate the implied correlation of income as well as the various transition rates. These are presented in Table 1.2. As you can see, transition rates are symmetric for a given level of ρ_ω. When there is no correlation between the generations, all transition rates (measured in quintiles) are around 20% as we would expect. When the correlation is 0.2, both the poverty trap and the affluence net (measured in quintiles) are around 28% while the jump up and fall down rates are around 13.5%. At a correlation level of 0.6, the poverty trap and affluence net are both almost 50% while the jump up and fall down rates are only around 2%. Compared to these symmetric benchmarks, the data show significant asymmetries such as a larger affluence net for whites and a larger poverty trap for blacks.

Mulligan (1997) also presents data on the intergenerational persistence of education. This is important because education is an important factor in determining wages. He finds that the coefficient on parent's education in a regression of child's education is somewhere between 0.19 and 0.45, with an average estimate of 0.29. With a number around 0.3, Mulligan notes that education will regress towards the mean fairly rapidly. However, the distribution of education has changed significantly over the last few generations. In the last sixty years, we have gone from a world in which only the elite attended college and many never finished high school, to a world in which most people get at least some post-secondary education and more than ten percent get some form of graduate degrees. If the variance of the education distribution has increased as the mean has increased, the measured correlation of education will be depressed.

The distribution of income by education level is well documented in the United States. Using data from the Current Population Survey, the ratio of overall mean wages to the mean wage of a high school graduate is 1.40. High school dropouts have a median wage that is 71% that of a high school graduate. Those with some college have a median wage 19% higher while those with a BA have annual earnings 60% higher than those with a high school diploma. Wage premia increase as we move up the education ladder. Those with a Master's degree make 180% more at the median than a high school graduate and those with a PhD or professional degree earn over two and a half times as much.

Given the very high return to education, it is perhaps surprising that so few people receive graduate degrees. In the United States only 9.4% of 35 to 44 year olds have received a Master's degree and only 3.5% have a professional degree or PhD. On the other side of the distribution, 8.8% do not have a high school degree (which is surprisingly high given that there is no direct cost for most people to finish high school). In this age range, 27.6% have only a high school degree, about the same have some college or an Associate's degree, and 22.9% have a Bachelor's degree.

Census data also indicate that the variance of annual earnings increases along with

Figure 1.5: Income Distribution by Education

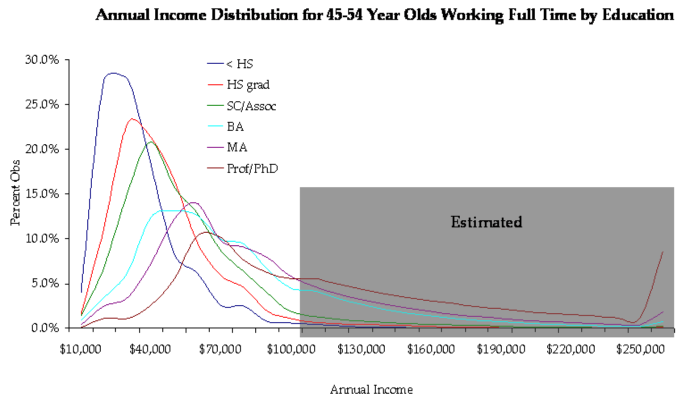


Table 1.3: Median Income and Standard Deviations by Education Level (White Full-Time Workers)

Education Level	Median Wage (HS = 1)	Standard Deviation (est)
HS Dropout	0.71	0.43
High School Grad	1.00	0.52
Some College/Assoc. Degree	1.19	0.61
Bachelor's Degree	1.60	0.87
Master's Degree	1.80	1.01
Prof. Degree/PhD	2.71	1.62

Source: Census, CPS 2008 Annual Social and Economic Supplement

the mean as the amount of education increases. From Figure 1.5, most high school dropouts fall in a fairly narrow band of annual earnings. High school graduates have a wider distribution of earnings while those at the top of the education distribution have a very wide range of annual earnings. For example, almost 80% of high school dropouts earn between \$20,000 and \$50,000 in a year. The 90-10 range widens to \$30,000 to \$110,000 for those with a BA and to approximately \$40,000 to \$210,000 for those with a professional degree or a PhD. Median income and standard deviations by education level are reported in Table 1.3.

Thus, while education is clearly correlated with earnings, it has widely disparate effects on different types of workers. One possible explanation is that the available data does not measure the quality of education. The quality of a bachelor's degree from one institution may be much inferior to that from a different institution. However, studies that measure and correct for quality (e.g. Harmon and Walker 2000) still find that people with the same level of education have very different outcomes, suggesting that luck, effort, and/or inherent differences are also important determinants of wages.

The recent literature using quantile regressions (see, for example, Lemieux 2006) shows that the return to education differs (sometimes substantially) by wage percentile. Those at the 90th percentile of the wage distribution (by education) have much larger returns to education than do those at the 10th percentile. Interestingly, Lemieux shows that the increase in inequality over the last 30 years is mainly due to an increase in the return to education, especially at the higher percentiles. The relevant question for this paper is how much of that economic success at the higher quantiles will be passed along to the next generation.

There is some evidence (Blanden and Gregg, 2004, Corak et al 2004) that holding all else equal, children from families with higher income tend to receive more and higher-quality education. Shea (2000), on the other hand, finds only a minimal relationship between family income and child education. Below, I present my own findings that there is a quadratic relationship between family income and education, perhaps obscuring the relationship for those looking only for a linear coefficient.

1.4 Calibration Results

1.4.1 Parameter Values and Calibration Targets

This section presents numerical results from simulations of the model, comparing the model with loss aversion to that with habit formation but with utility that is concave everywhere. Table 4 presents parameter values that are common to both versions of the model. Most values should not be controversial. I use a CARA parameter of 0.01

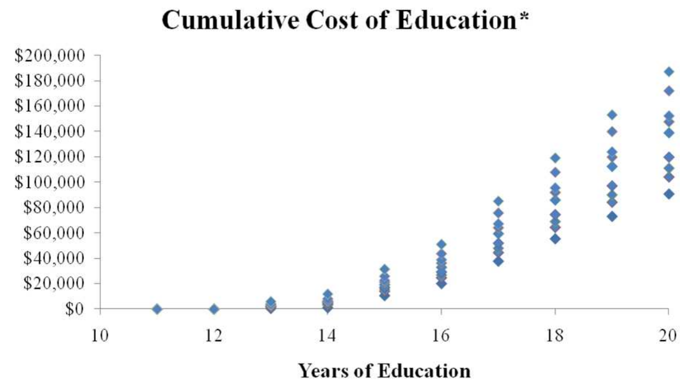
Table 1.4: Parameter Values

Parameter	**Value**	**Description**	**Source/Explanation**
λ	2.25	Loss aversion parameter	Tversky and Kahneman (1992)
$\beta = R = (1+r)$	1	Rate of time preference & interest rate	Simplicity
ϕ_1	0.0	Linear term in cost of education	Estimated from Dept. of Education,
ϕ_2	0.3	Quadratic term in cost of education	National Ctr for Education Statistics
e, ψ	[0,1]	Range for education and earning ability	Normalization
$\bar{\psi}$	0.5	Average ability level	
σ_ψ	0.15	Standard deviation of earning ability, ψ	3 st. dev. above & below the mean

for gains. I use Tversky and Kahneman's (1992) value for the loss aversion parameter, λ, of 2.25. Other estimates of this parameter in the literature generally find it to be between 2 and 3. Given that strong loss aversion requires that the CARA parameter for losses, ν, be less than $\ln(\lambda)$, I let ν be the maximum value of $\ln(\lambda) \approx 0.8$. Because there are only two periods, I allow both β and the gross interest rate R to equal one. This is done without loss of generality so long as we make the standard assumption that $\beta R = 1$. In order to estimate the direct education costs, I use the Department of Education's National Center for Education Statistics, then normalize against the median income of a high school graduate, assuming each period is approximately 20 years. I set the direct cost of the first two levels of education, equivalent to a high school dropout and high school graduate, at zero, as most students in the United States have access to free education through the twelfth grade. The data is graphed in Figure 1.6, showing that education costs tend to be convex. Finally, I choose the minimum, maximum, and standard deviation of earning ability, ψ (which is assumed to be normally distributed), so that the population includes agents with earning ability levels that are up to three standard deviations above and below the mean.

I calibrate the wage function (2) to match features of the education distribution and income distribution by education level for white full-time workers. I focus on white workers in this paper because of the disparate outcomes experienced by white and black families in the intergenerational transmission of income, discussed in sec-

Figure 1.6: Range of Education Costs



* Graphs a range of public education costs assuming a free public high school education, two years of community college, two years at a public in-state college, and four years of various types of professional graduate school. Data from U.S. Department of Education, National Center for Education Statistics, 1987-88 through 2007-08 Integrated Postsecondary Education Data System, "Fall Enrollment Survey" (IPEDS-EF:87-99); "Completions Survey" (IPEDS-C:88-99); "Institutional Characteristics Survey" (IPEDS-IC:87-99).

tion 3. Malloy (2010) discusses possible explanations of these disparate outcomes. I divide education into six possible levels ranging evenly from 0 to 1. I set the wage parameters, $\alpha_1, \alpha_2, \zeta$, and κ from equation (2) so that the model matches the observed median income by education level and the education distribution as closely as possible[5]. Because the minimum of both education, e, and earning ability, ψ, is set at zero, the minimum possible wage in the model, given the semi-log wage function represented by equation (2), is 1. The maximum possible wage occurs when both education and earning ability are at their maximum. Because behavior is significantly different for loss averse agents as opposed to those with habit-formation-only preferences, I calibrate the model separately for these two utility functions. For the baseline calibration, I allow agents to borrow the entire cost of their education so that $\Omega = -\varepsilon(e)$. The observed targets and corresponding moments implied by the calibrated models are presented in Table 1.5.

[5] I use a simple Euclidean norm of the difference between target and calibrated moments of the 12 target moments and search over a grid that varies by values of 0.005.

Table 1.5: Target and Calibrated Moments

Moment	Target	Loss Aversion	Habit Formation Only
Education Distribution			
High School Dropouts	7.9%	8.8%	8.7%
High School Grads	28.5%	23.8%	29.1%
Some College/Assoc. Degree	27.8%	27.5%	24.4%
Bachelor's Degree	23.1%	31.0%	25.3%
Master's Degree	9.2%	7.2%	11.1%
Prof. Degrees/PhD	3.5%	1.9%	1.4%
Median Income by Education Level			
High School Dropout Wage	0.71	0.78	0.77
High School Grad Wage	1.00	1.00	1.00
Some College/Assoc. Degree Wage	1.19	1.57	1.37
Bachelor's Degree Wage	1.60	1.65	1.85
Master's Degree Wage	1.80	2.10	2.75
Prof. Degrees/PhD Wage	2.71	4.04	4.57

Despite the fact that the model is under identified, both models do a fairly good job of matching the targeted moments. They suffer at the extremes as neither is able to generate a large enough percentage of the population investing in the top level of education despite wages that are higher than in the data. This is likely due to the assumption that investing in a professional/PhD in the education period consumes all of the agent's time so that they are unable to earn anything in the first period.

I report the calibrated parameters from equation (2) in Table 1.6. In order to determine whether or not these are reasonable values, I also present the minimum, average, and maximum return to the equivalent of one year of education and the ratio of the maximum wage to the minimum. The return to education averages 8.4% in the loss-averse model and 7.3% in the habit-formation-only model. The labor literature generally finds a value between five and ten percent, so these values seem reasonable.

1.4.2 Education Decision Rules

Based on (11), the resource maximizing level of education, e^*, will depend only on an agent's earning ability. An individual's wage is completely determined by his level of

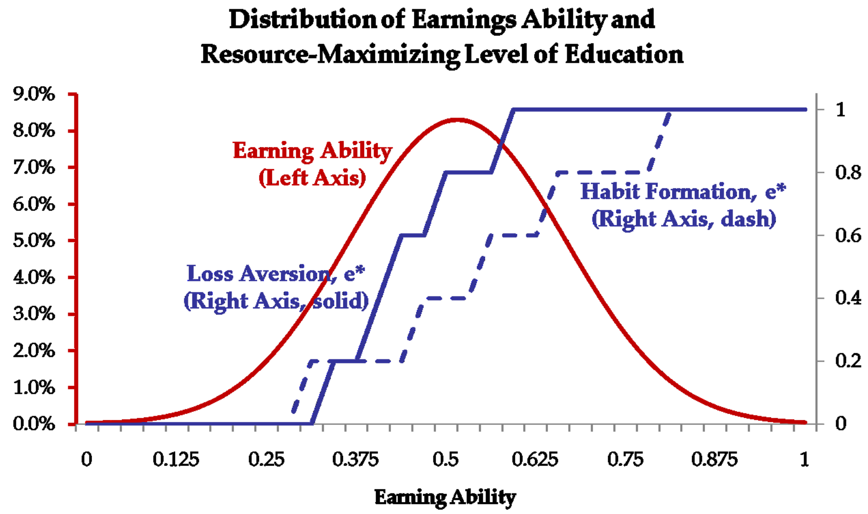
Table 1.6: Calibrated Parameter Values

Parameter	Loss Aversion	Habit Formation Only
α_1: linear term on education	0.285	0.250
α_2: quadratic term on education	0.190	0.000
ζ: interaction term (education*ability)	0.655	0.900
κ: linear term on ability	1.450	1.200
Minimum marginal return to education	3.1%	2.5%
Average marginal return to education	8.4%	7.3%
Maximum marginal return to education	13.9%	12.2%
Max-Min Wage Ratio	13.2	10.5

education and earning ability, while the cost of education (both direct and opportunity) depends only on the level of education chosen. Figure 1.7 graphs the normal distribution of earning ability and the corresponding level of education that would maximize lifetime resources for each model. As expected, e^* is a monotonically increasing function of earning ability. Those with the lowest earning ability maximize resources by choosing the minimum of education because the cost of acquiring more would outweigh the benefit of a higher wage in the second period. Those with the highest earning ability would maximize lifetime resources by investing in the maximum level of education. Note that in this figure the differences between the e^*implied by the loss aversion and habit formation models are due solely to differences in the calibrated parameter values for the wage function shown in table 1.6.

The actual education level chosen to maximize lifetime utility, $\hat{e}$, may be less than or equal to e^* (as discussed earlier, it will never be more than e^*). Figure 1.8 graphs the decision rule for $\hat{e}$, (as a function of reference consumption) for a loss-averse agent with an ability level approximately one standard deviation above the mean, for different levels of bequest. For this individual, the level of education that would maximize lifetime resources is $e^* = 1$, the maximum level of education. However, those with little or no bequest from the previous generation and an intermediate reference consumption level will optimally choose a lower, and sometimes substantially lower, level of education. Intuitively, agents with a moderate reference level of consumption

Figure 1.7: Earnings Ability and Resource Maximizing Education

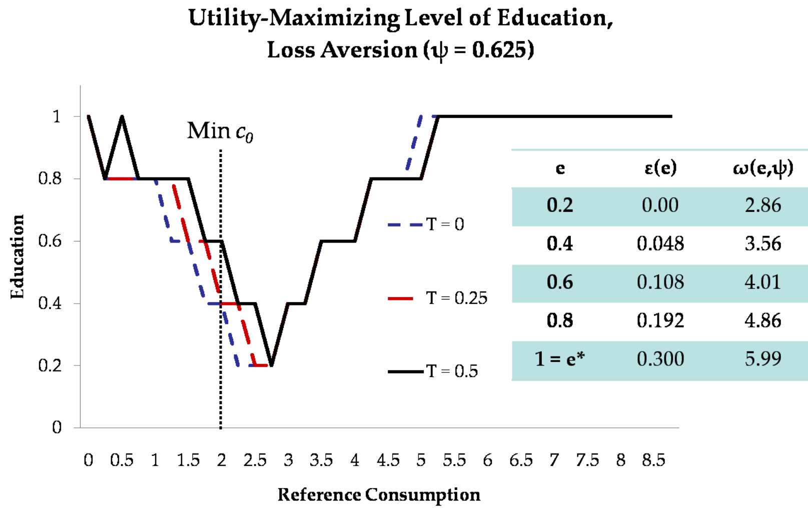


can more easily replace it with a lower level of education, thus avoiding a loss in period 1. Choosing a level of education less than e^* allows them to minimize or eliminate the loss in period one. Overall lifetime utility will be increased, even if choosing $\hat{e} < 1$ means a lower (or non-existent) gain in period two. This shows one mechanism in the model that transmits earnings from one generation to the next: those with lower family consumption choose lower education, and therefore have lower earnings themselves.

As reference consumption increases, the likelihood of a loss in the first period increases (until it is equal to 1). At very high levels of reference consumption, agents cannot avoid losses in period 1 and can only hope to eliminate or reduce losses in period 2. In order to decrease or eliminate the loss in period two, they will maximize second period income by investing in the resource-maximizing level of education, e^*. Thus, those with higher reference (or childhood) consumption will have higher earnings than those individuals with moderate reference consumption, because utility maximization forces them to invest in the highest level of education.

Note that those with the lowest reference consumption will invest in higher levels of education than those directly to their right in the reference consumption distribution.

Figure 1.8: Utility Maximizing Education by Reference Consumption and Bequest Received for Loss-Averse Agent



e	ε(e)	ω(e,ψ)
0.2	0.00	2.86
0.4	0.048	3.56
0.6	0.108	4.01
0.8	0.192	4.86
1 = e*	0.300	5.99

This is because with very low levels of reference consumption it is easier to avoid a loss in the first period, even when getting more education. The dashed line at $c_0 = 2$ is the equilibrium minimum reference consumption for agents in the model simulation. To the left of this line agents would invest in more education, but agents never visit this region of the state space in equilibrium.

Figure 1.9 gives the education investment decision under habit-formation-only preferences. When agents can borrow to finance the direct cost of education, $\varepsilon(e)$, as in the benchmark model, the decision rule is simply a horizontal line at e^*, which for this agent is equal to 0.6. If there is a more severe borrowing constraint, education will be upward sloping in c_0, as represented by the dashed line, until the constraint no longer binds and the agent can invest in the optimal level of education. The U-shaped education investment decision for loss averse agents and the contrasting flat or upward sloping investment-decision rule for habit-formation-only agents is the main testable implication of the model. In the next section I present evidence which provides support for loss averse preferences.

The final thing to note from Figure 1.8 is that the agent will invest more in education as her bequest increases. In this model, bequests allow agents to avoid or minimize losses in period one. And because utility is standard (i.e. concave and monotonic) in gains, agents will be more likely to invest in the amount of education that will maximize lifetime resources, as in a standard model.

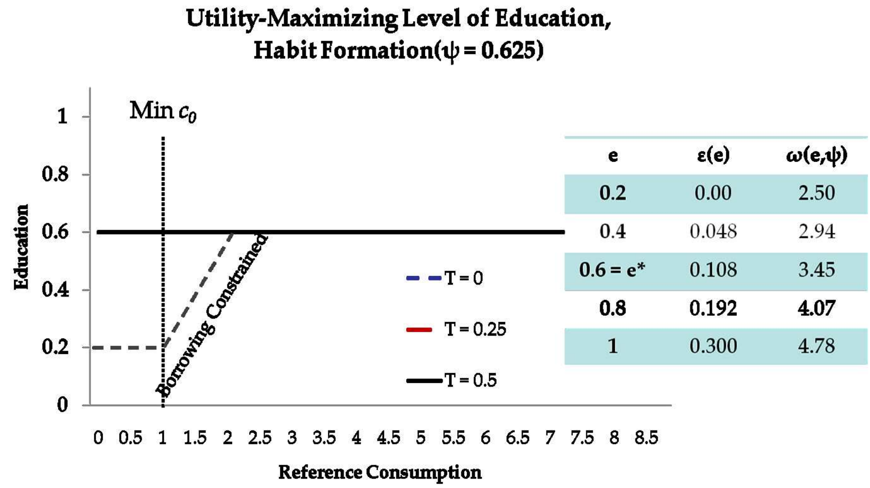
An important implication of Figure 1.8 is that some people will get less education than would maximize lifetime resources. That is, because of their loss averse preferences and the (mild) borrowing constraint, they fail to equate the marginal benefit of an extra year of education with its marginal cost. In the simulation of the loss averse model, about half of the agents invest in less education than would maximize lifetime resources (i.e. $\hat{e} < e^*$). For those who do invest in a lower level of education, the average gap ($e^* - \hat{e}$) is a bit more than 0.4, or about two levels of education. That is, agents who should invest in the maximum level of education, such as a professional degree, on average would invest in only a college degree.

Because this is not a general equilibrium model, it would be inappropriate to measure the possible output and welfare gains that would result from the government solving the imperfect capital markets problem in the model. However, one can make informed speculations. In the model, more education increases productivity. It does this more for some people than for others. If someone with high aptitude becomes an electrical engineer rather than an electrician, or a doctor instead of a nurse, she will be much more productive. Society benefits in because such individuals are producing more output than they would otherwise have done.

1.4.3 Correlations Across Generations

The intergenerational correlation of income depends on a number of mechanisms in the model. As discussed above, loss aversion contributes to a positive correlation for two reasons. First, those with moderately low parental income (and thus reference consumption) have an incentive to invest less in education in order to avoid losses in

Figure 1.9: Utility Maximizing Education by Reference Consumption and Bequest Received for Habit-Formation-Only Agent



the first period of life even though this will reduce gains in the second period due to their lower wage. Second, those with higher reference consumption are likely to be faced with a loss in period one no matter what level of education they choose. In order to avoid or minimize a loss in period two, they will invest in more education in period one and raise their wage and consumption in period two. Thus, the same factors that contribute to a correlation of earnings across generations are also likely to contribute to a positive correlation of education.

Table 1.7 lists intergenerational correlations for wages, education, consumption, and bequests. The data for the correlation of income comes from Mazumder (2005) while the correlation of education and consumption comes from Mulligan (1997). I present results for both types of utility function as well as three different borrowing constraints. When there is no borrowing constraint (columns (1) and (4)), all agents choose the resource-maximizing education level for both models, and because earning ability is independent across generations, there is no correlation of either income or education. With a borrowing constraint in which agents are able to borrow the full direct cost of education (columns (2) and (5)), the loss averse model is able to generate a correlation of income of 0.14 and a correlation of education of 0.24,

Table 1.7: Intergenerational Correlations

	Data	Loss Aversion			Habit Formation Only		
		(1)	(2)	(3)	(4)	(5)	(6)
		No BC	$\Omega = \varepsilon(e)$	$\Omega = f(c_0)$	No BC	$\Omega = \varepsilon(e)$	$\Omega = f(c_0)$
Income	0.60	0.00	0.14	0.17	0.00	-0.00	0.14
Education	0.29	0.00	0.24	0.33	0.00	-0.01	0.26
Consumption (total)	0.68	0.09	0.16	0.11	0.32	0.36	0.36
Consumption (period 2)		0.79	0.76	0.80	0.33	0.32	0.39
Bequests	NA	-0.01	-0.05	-0.02	-0.12	-0.18	-0.10

while the habit-formation-only model still exhibits roughly zero correlation of both income and education. Finally, I present results when the borrowing constraint is a function of reference consumption, $\Omega = f(c_0)$. For simplicity, I assume that those with reference consumption in the bottom twenty percent do not have access to financial markets and so cannot borrow at all while everybody else can borrow the direct cost of education. In this case, habit-formation is able to generate a positive correlation of income.

As discussed in the introduction, most students do have access to both loans and grants that can cover most, if not all, of the tuition at a public, in-state, four-year college. Yet only the model with loss aversion is able to generate a positive and significant correlation of income when agents are able to borrow the full direct cost of education.

It is interesting to note that bequests are negatively correlated across generations. This is true for both models, although less so for loss aversion. Most parents do not leave a bequest in the loss-averse model, while the median parent leaves a small bequest in the habit formation only model. The intuition for this negative correlation is straightforward. A person with a high wage who grew up with a low reference level of consumption (and therefore probably received no bequest) will be able to consume above her reference consumption in period two and leave a bequest for the next generation. The child in the next generation, however, on average would only have average earning ability and would therefore find it difficult to replace his reference consumption in period two. If his consumption in period two is at or below his reference consumption he will leave no bequest for his child. So while bequests do play a role in increasing the intergenerational earnings correlation by making it easier for those with higher reference consumption to invest in the optimal level of education, this factor typically does not operate across more than one generation.

Consumption is correlated across generations in both versions of the model albeit in different ways. Because utility is measured in reference to the consumption level of the individual's parents, both models generate a correlation of consumption. However, in the model with only habit formation, because utility is everywhere concave so that marginal utility is only slightly higher for a loss than for a gain, agents try to smooth consumption as much as possibly between period 1 and period 2. This leads to a correlation of consumption that is approximately the same in period two as in total (period 1 plus period 2). In the model with loss aversion, however, the marginal utility is higher for a loss than for a gain and is decreasing in the size of the loss. An agent facing an unavoidable loss will lump the loss in period 1 so as to reduce or eliminate the loss in period 2. This leads to a much higher correlation of consumption in period 2 than in total (or in period 1).

The model with loss aversion generates a positive intergenerational correlation of income primarily because rich kids tend to stay rich. In other words, the model generates an affluence net. Table 1.8 presents the transition probabilities in the data and in models under various assumptions on preferences and borrowing constraints. When there are no borrowing constraints (columns (1) and (3)), most transition probabilities are close to 20%, indicating no poverty trap or affluence net. The model with loss aversion and the mild borrowing constraint (column 2) has an affluence net, in which the children of those in the top quintile stay in the top quintile, of 46.4% while the affluence net with a more severe borrowing constraint (column 3) is 42.0%. The highest affluence net in the habit formation model is 30.2%, in the model

Table 1.8: Intergenerational Mobility

	Data*	Simulations					
		Loss Aversion			Habit Formation Only		
		(1)	(2)	(3)	(4)	(5)	(6)
	White Families	No BC	$\Omega = \varepsilon(e)$	$\Omega = f(c_0)$	No BC	$\Omega = \varepsilon(e)$	$\Omega = f(c_0)$
Corr. of Income		0.001	0.142	0.171	0.000	-0.001	0.144
Poverty Trap	24.9%	23.4%	22.6%	23.5%	22.8%	23.4%	23.6%
Affluence Net	38.9%	23.4%	46.4%	42.0%	23.2%	23.2%	30.2%
Jump Up	10.6%	23.3%	12.7%	11.8%	23.2%	23.7%	5.8%
Fall Down	10.4%	22.9%	23.6%	23.2%	23.4%	23.7%	22.8%

* Source: Mazumder (2008)

with severe borrowing constraints. None of the versions of the model are generate a significant poverty trap in which the children of the poor stay poor, matching the data of white families in the United States. Both the loss-aversion model, with mild borrowing constraints, and the habit-formation-only model, with more severe borrowing constraints, have a significant affluence net. The loss aversion model with both mild and severe borrowing constraints does a good job of matching the jump up rate (12.7% and 11.8% vs. 10.6% in the data), but neither model can generate the low fall down rate observed in the data.

It is interesting that none of the specifications generate a significant poverty trap. In the model, as in the real world, agents are able to finish high school with no direct cost and can attend community college at little direct cost. For many, a high school degree is enough to pull them out of the bottom quintile so that it is easy, in the model, to escape the poverty trap. In order to generate a significant poverty trap in this model, earning ability has to be correlated across generations. As an exercise, Table 1.9 presents the value of ρ, the correlation between parent and child earning ability, that would be required to generate a correlation of income of 0.45. The intergenerational persistence level found in white families by Hertz (2003) is 0.32 compared to 0.4 for all families. Unfortunately, Mazumder (2005) does not break out the correlation of income for white and black families separately. So while the

Table 1.9: Generating a Higher Correlation of Income Between Generations

	Data	Loss Aversion		Habit Formation Only	
		(1)	(2)	(3)	(4)
	White Families	$\Omega = \varepsilon(e)$	$\Omega = f(c_0)$	$\Omega = \varepsilon(e)$	$\Omega = f(c_0)$
ρ: Persistence of Ability	**?**	**0.34**	**0.32**	**0.46**	**0.38**
Correlation of Income	Target = 0.45	0.45	0.45	0.45	0.45
Correlation of Education	0.23	0.40	0.47	0.43	0.47
Correlation of Consumption	0.68	0.85	0.87	0.63	0.57
Poverty Trap	24.9%	38.5%	37.5%	44.7%	41.0%
Affluence Net	38.9%	61.3%	56.5%	45.0%	40.6%
Jump Up	10.6%	3.5%	5.7%	7.0%	1.7%
Fall Down	10.4%	13.7%	14.1%	7.0%	9.0%

Source for data: Mazumder (2008), Mulligan (1997)

target of 0.45 for white families is somewhat arbitrary, it is consistent with an overall correlation of intergenerational income of approximately 0.6.

All four versions of the model reported in Table 1.9 have trouble matching some key moments. The correlation of education in the data may be low due to the long-run shift from a population that is mainly high school educated to one that is mainly college educated, especially if the variance of education is also increasing. There are no such caveats when looking at the main measures of transition between income quintiles. All four versions of the model generate a poverty trap and an affluence net that are too large when ρ is calibrated to match an overall correlation of 0.45. The model that comes closest to matching the target moments is the one with loss averse preferences and the more severe borrowing constraint. However these values are actually farther from the data than the loss averse model with milder borrowing constraints in which earning ability is completely independent across generations, listed in column 2 in Table 1.8.

Relative to models with no persistence in ability, the models in Table 1.9 have smaller percentages of children who jump up from the bottom quintile to the top and who fall down from the top to the bottom. In fact, the jump up percentage is now much smaller than the 11.0% observed in the data, even when children are able to borrow the entire cost of education. Again, the model with loss aversion and independent earning ability in Table 8 comes much closer to matching the data with a jump up rate of 12.7%, although that model generates a fall down rate that is too high, at around 20%. The models in Table 1.9 come closer to matching the fall down rate in the data of 6.4% for white families. The model with habit formation and mild borrowing constraints has a fall down rate of 7.0%. However, all four models in Table 1.9 have a higher fall down rate than jump up rate, the reverse of what is found in the data.

Taking the results in Table 1.9 as a whole, there is little to recommend the models with persistence in ability over those in Table 1.8 in which earning ability is independent across generations. The models with persistent ability do a better job matching the overall correlation of income, by construction, and the fall down rate. But they generate too little mobility in the bottom and top quintiles, with poverty traps and affluence nets that are much too high. Assuming positive correlation of earning ability between generations means that those who are in the bottom of the income distribution, and who have low earning ability, are more likely to have children who are also of low ability. These children will choose low education and therefore also have a lower income. The same holds for children born at the top of the income distribution only in reverse. Loss aversion actually magnifies this result, especially at the top as is shown by the larger affluence nets in columns 1 and 2 in Table 1.9 as opposed to columns 3 and 4.

As noted in section 3, the variance of income increases with the level of education in the data. There are a number of possible explanations for this, including unobserved heterogeneity in the fields of education (e.g. engineering vs. art history), a magnified effect of luck or effort on those with higher education, or greater heterogeneity in terms of earning ability at higher levels of education than among those who drop out of or only graduate from high school. This last possible explanation can be tested and

Table 1.10: Mean and Standard Deviation of Income by Education Level (Mean HS Grad = 1)

	Data		Loss Aversion		Habit Formation Only	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Drop Out	0.729	0.433	0.617	0.059	0.735	0.049
High School Grad	1.000	0.520	1.000	0.225	1.000	0.058
Some College/Assoc. Degree	1.186	0.606	1.303	0.279	1.346	0.053
Bachelor's Degree	1.758	0.866	1.447	0.315	1.855	0.111
Master's Degree	2.022	1.011	1.793	0.248	2.776	0.222
Professional Degree/PhD	3.105	1.617	3.342	0.724	4.555	0.428

results are presented in the next section. Table 1.10 presents mean income and the standard deviation of income by education level for the models in which agents can borrow the full cost of education. While neither the loss averse nor habit-formation-only model can generate as much variance as the data, given that there is no luck or uncertainty in the model, the loss-averse model generates significantly more variance than the habit-formation-only model. This is because agents with a wide range of ability levels decide to invest in higher levels of education in the loss-averse model, due to the need to avoid losses in the second period.

1.5 Empirical Tests

As noted in section 4, one important difference between the model with loss aversion model and the model with habit formation and concave preferences is the shape of the education decision rule. With loss aversion, the decision rule is U-shaped in reference (or family) consumption. In the model with concave habit formation, education is either flat or upward sloping in reference consumption, depending on the borrowing constraints. I test these implications using the National Longitudinal Survey of Youth (NLSY), specifically the NLSY79, which began in 1979 and initially surveyed young adults between the ages of 17 and 21; and the NLSY97, which began in 1997 and first surveyed the young respondents between the ages of 12 and 17. Both versions collect data on family income, education levels of parents, and the highest level of education completed.

I use the NLSY rather than the PSID because the former includes a (noisy) measure of ability. Controlling for ability is important for estimating the education decision rule. As is shown in figure 6, the maximum level of education, e^*, that an agent will choose is dependent on her inherent earning ability, ψ. Each NLSY79 respondent took the Armed Services Vocational Aptitude Battery (ASVAB) in 1980 when they were between the ages of 18 and 22. The ASVAB is a multiple choice test that measures both reading and quantitative ability as well as vocational information such as mechanical knowledge and shop skills. Each person who takes the test is given an Armed Forces Qualification Test score, which is based only on the arithmetical reasoning, mathematical knowledge, word knowledge, and paragraph comprehension section. This score ranges from 0 to 99, representing percentiles. The NLSY97 cohort, meanwhile, was given the Peabody Individual Achievement Test (PIAT). Most took the test in 1997 when they were between the ages of 12 and 17. The PIAT consists of six subtests, covering such areas as reading recognition and comprehension and mathematics. The test was designed to help diagnose learning disabilities rather than to evaluate new recruits and may do a better job of measuring inherent ability, at least in the areas it tests. Furthermore, the test was given when respondents were, on average, five and a half years younger than the older group taking the ASVAB. It is more likely, therefore, that the NLSY97 test scores measure the inherent ability of the respondents rather than measuring the quantity and quality of education they had already received. PIAT results are also reported in percentile scores, ranging from 0 to 99. Table 1.11 reports summary statistics on test scores and other variables in the NLSY79 and NLSY97.

The major drawback of the NLSY data sets, as opposed to the PSID, is that they do not measure consumption. Using the NLSY, I am forced to use total parental

Table 1.11: Summary Statistics for NLSY79 and NLSY97 Data Set

	NLSY79					NLSY97				
Variable	# Obs	Mean	St. Dev.	Min	Max	# Obs	Mean	St. Dev.	Min	Max
R's Highest Grade	7,599	13.31	2.43	7	20	8,853	12.98	2.62	7	20
Family Income (log)	5,422	9.73	0.79	4.64	11.55	7,474	9.89	1.077	1.14	12.47
AFQT	11,878	40.95	28.75	1	99					
PIAT						6,044	70.32	17.39	0	100
Mother's Educ (bio)	11,452	11.21	2.67	4	20	8,255	12.48	2.83	3	20
Father's Educ (bio)	10,531	11.28	3.51	3	20	7,100	12.59	3.17	3	20
Mother's Educ (res)						7,974	12.57	2.88	3	20
Father's Educ (res)						5,691	12.90	3.26	3	20
Black		25.0%					25.99%			
Hispanic		15.8%					21.16%			

income as a proxy for reference consumption. While this does not seem unreasonable, it will bias the estimated decision rule to the extent that family consumption is different than income. I use the average family income for the respondent before his eighteenth birthday. For the early cohort, the NLSY79, this reduces the number of observations significantly, as I have at most two years of family income (age 17 and 18), while I have no data for those who were over 18 at the start of the survey. The newer NLSY97 has at least two years of family income data for all respondents (those who were 17 at the start) and a maximum of seven years for those who were 12. I convert nominal income in the data to real using the CPI deflator from the BLS. The main weakness of the NLSY97 data set is that respondents were only between the ages of 22 and 27 during the last available wave of interviews in 2007 and so may not be done with their educational investment. Respondents to the NLSY79 were in their mid-forties at the last interview and so had most likely completed all of their education.

Regression results are reported in Table 1.12. The dependent variable is the highest grade completed, measured in total years of education. For example a high school graduate would have twelve years while a respondent with a BA would have 16 years. The main explanatory variables are the natural log of family income and its square. Custom sample weights provided by the NLSY were used in all regressions. Columns 1 and 4 provide support for a quadratic education decision rule in both the NLSY79 and NLSY97 data sets when controlling only for race and sex. The linear term in average family income is negative while the quadratic term is positive. Both are significant at 1% for both data sets. The theoretical model implies that one should control for ability as well. The results presented in columns 2 and 5 control for ability, using the AFQT score for the NLSY79 cohort and the PIAT score for the NLSY97 group. As expected, the coefficient on ability is positive (not shown) and significant at 1%. The coefficients on both linear and quadratic family income continue to have the sign implied by the loss aversion model and are still significant at 1%.

While the model does not imply the need to control for parental education levels, there could be a number of reasons to expect that parental education may have a positive correlation with child education, even when controlling for the child's ability level. For example, it could be that the taste for education is passed down from parent to child. It could be that more highly educated parents value education more and so encourage their children to continue their education. Finally, it may be that more highly educated parents pass on skills that allow their children to more easily continue on to higher levels of education themselves. Column 2 shows that when the additional controls of parent education (both mother and father) are introduced for the NLSY79 regression, the coefficients on family income and its square still have the correct signs, but are no longer significant. However, columns 6 and 7, using the NLSY97, still show significant coefficients on both family income and its square, even when controlling for parent education and ability level. The difference between columns 6 and 7 is that column 6 controls for the education of the parents the child lives with (for example a mother and step-father) while column 7 controls for the respondent's biological parents. In both cases we see first a decrease in education as family income increases and then an increase.

Table 1.12: Regression Results for Highest Grade Completed

Variable	NLSY79			
	(1)	(2)	(3)	
Family Income	−4.160***	−1.916***	−0.565	
	(1.166)	(0.729)	(0.734)	
Family Income Squared	0.279***	0.125***	0.039	
	(0.060)	(0.039)	(0.039)	
Controls				
AFQT(79)		✓***	✓***	
Parent's Education			✓***	
Sample Size	2,747	2,719	2,179	
R^2	0.132	0.418	0.437	
	NLSY97			
	(4)	(5)	(6)	(7)
Family Income	−2.274***	−2.003***	−1.896***	−1.272***
	(0.223)	(0.263)	(0.394)	(0.256)
Family Income Squared	0.171***	0.146***	0.115***	0.091***
	(0.012)	(0.014)	(0.021)	(0.014)
Controls				
PIAT(97)		✓***	✓***	✓***
Parent's Education			✓(res)***	✓(bio)**
Sample Size	7,362	5,056	3,013	3,969
R^2	0.165	0.266	0.301	0.311

Significant at: *=10%, **=5%, ***=1%

Robust standard errors in ( )

All regressions control for race and sex and use custom sample weights provided by the NLSY.

The ideal explanatory variable for the regressions in Table 1.12 would be average family consumption for the eighteen years of a person's childhood. Because of data limitations, I am forced to use average family income, measured for between 1 and 7 years. Assuming that yearly family income is a noisy measurement of permanent family income (which should be a good proxy for family consumption), the measurement error on my main explanatory variable will be heteroskedastic, with a larger error variance for observations that average fewer years of family income. As a simple correction for this heteroskedasticity, the regressions presented in Table 1.12 use robust standard errors.

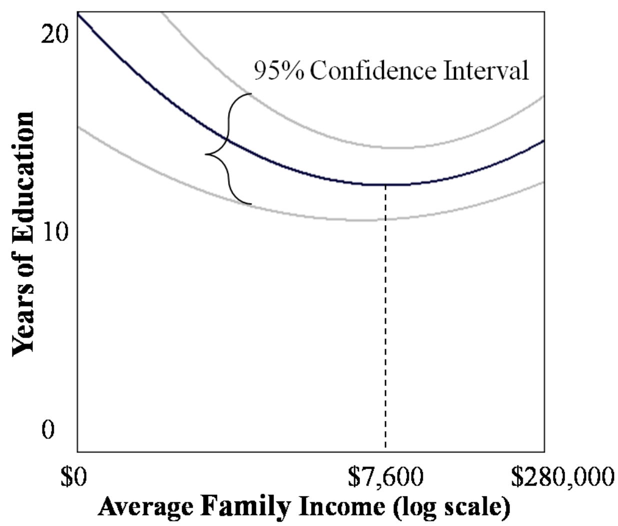
Because I am forced to use family income instead of family consumption as my explanatory variable, the question may arise as to whether or not I would still get a U-shaped eduction decision in the simulated data using the log of the parent's second period income (rather than consumption). In Table 1.13, I present results of the same regressions presented in Table 1.12, using the simulated data discussed in section 4. This simulated data comes from the baseline model in which agents are allowed to borrow the full cost of education. I present results for both the loss aversion model and the habit-formation-only model. Both family income and its square are significant and have the expected sign for all three regressions with the simulated loss aversion data. This is not true for any of the regressions on the simulated habit formation model. In that model, adding ability, column 5, explains almost all the variation in education ($R^2 = 0.922$). Note that adding ability to the loss aversion model only gives an R^2 of 0.34, similar to the results in Table 1.12. Finally, column 7 in Table 1.13 presents the results from a linear regression on the simulated data from the habit-formation-only model. In this case, the coefficient on family income is significant at 5%, but is negative.

Figure 1.10 graphs the decision rule for education implied by column 6 from Table 1.12, along with the 95% confidence interval. The graph is for a white male (who

Table 1.13: Regression Results for Child's Education on Simulated Data with Parent Wage as Explanatory Variable

Variable	Loss Aversion			Habit Formation Only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Family Income	−0.025**	−0.024***	−0.196***	−0.005	0.001	−0.010	−0.009**
	(0.010)	(0.009)	(0.010)	(0.011)	(0.003)	(0.007)	(0.004)
Family Income Squared	0.093***	0.091***	0.114***	0.003	−0.001	0.000	
	(0.005)	(0.004)	(0.004)	(0.005)	(0.001)	(0.002)	
Controls							
Ability (ψ)		✓***	✓***		✓***	✓***	✓***
Parent's Education			✓***			✓*	✓*
Sample Size	50,000	50,000	50,000	50,000	50,000	50,000	50,000
R^2	0.007	0.340	0.357	0.000	0.922	0.922	0.922

Figure 1.10: U-Shaped Education Investment Regression Results with 95% Confidence Interval



scored in the 75th percentile of the PIAT) with parents who each have the average level of parental education. This figure shows significant differences in the level of education received across the family income spectrum. The graph also highlights the main weakness of the results, namely that the minimum of the U-shape is approximately $7,600 (in 2007 dollars), a very low average family income. In fact, the fifth percentile of the average family income in the data set is $5,800 and the tenth percentile is $10,400. So there is not a lot of sample weight behind the left side of the U in Figure 1.10. Only about seven percent of the observations have an average family income less than $7,600.

Another implication of the model with loss aversion that can be tested with the NLSY is that there is a wider distribution of ability at the higher education levels. Table 1.10 shows that the loss aversion model is able to generate more variance at the higher levels of education than the model with only habit formation. This is because agents with a wider range of ability invest in higher levels of education with loss aversion than with only habit formation. Table 1.14 gives the number of observations, the mean ability score (AFQT for the NLSY79 and PIAT for the NLSY97), and standard deviation of the ability score for each level of education (measured in years completed). As we would expect from both types of models, the mean ability level generally increases as the education level increases. Those earning a high school diploma had an average AFQT score of 28.7 and an average PIAT score of 66.5. Those with a BA, or 16 years of education, had an average AFQT score of 64.3 and PIAT score of 80.0.

The data sets show opposite results when it comes to the variance of ability by education. In the older NLSY79, the standard deviation increases almost monotonically until 17 years of education after which it decreases slightly. This seems to support the idea that people choosing more education have a wider range of ability. On the other hand, the NLSY97 shows almost the opposite trend, with standard deviations of ability decreasing as the level of education increases. The likely reason for this discrepancy is that the respondents in the NLSY97 were only between the ages of 22 and 27 in the last round of interviews. Thus only older respondents with very high ability (and determination) have completed the highest levels of education. Only 24 respondents, or 0.5% of the total, have completed more than 18 years of education, and only 7.5% have more than a college degree. In comparison, 3.5% of the total in the NLSY79 cohort have more than 18 years of education and 10.3% have more than a college degree.

Table 1.14: Mean and Variance of Ability Scores by Education in the NLSY

	NLSY79 (AFQT)			NLSY97 (PIAT)		
Highest Grade Completed	Number	Mean	St. Dev.	Number	Mean	St. Dev.
7	37	7.3	8.5	61	59.3	17.7
8	100	8.7	8.7	305	62.1	18.9
9	176	10.9	12.7	389	63.6	15.7
10	172	12.4	13.3	425	63.4	16.0
11	216	13.6	12.8	444	65.0	17.5
12	3,179	28.7	22.4	1,569	66.5	17.2
13	640	37.8	24.2	542	72.1	15.5
14	782	45.4	24.8	585	72.7	15.6
15	367	46.1	25.0	355	75.0	14.7
16	864	64.3	25.1	812	80.0	13.6
17	202	65.4	25.5	327	83.9	12.9
18	286	71.2	22.4	94	87.2	10.5
19	118	76.5	21.6	21	86.6	11.6
20	146	75.0	23.9	3	90.7	4.7

1.6 Conclusion

A simple model with loss-averse preferences and education investment can generate a positive and significant intergenerational income correlation, even when agents are able to borrow the entire (direct) cost of their education and earning ability is completely independent across generations. Previous work has shown that observable factors such as IQ, schooling, and personality, can only explain approximately half of the currently accepted value of the intergenerational correlation of income of 0.6. Loss averse preferences can explain about half the unexplained correlation (0.14). A model with concave preferences and habit formation can only generate similar levels of the intergenerational income correlation when there are severe borrowing constraints that prevent students from borrowing the cost of education. While some students, especially in historically disadvantaged groups, may face difficulty in borrowing for post-secondary schooling, this does not seem to describe the experience of most students in the U.S.

While precise welfare statements are not possible in this partial equilibrium model, it seems likely that loss aversion in this model generates behavior that is socially inefficient, in that a sizable portion of the population invests less in education than is socially optimal. They do this because capital markets are imperfect and they cannot bring earnings from the second period into the first period. Those who can avoid a loss in the first period by investing in less education than what would maximize lifetime resources actually increase their lifetime utility at the cost of total output. If agents could borrow against future earnings they would invest in more education and total output would increase.

The baseline model with loss aversion and weak borrowing constraints generates a positive correlation of income mainly by generating an affluence net. Children born at the top of the income distribution tend to get more education in order to avoid losses in the second period of life. The loss averse model is able to generate an affluence net and a percentage of children who jump up from the bottom quintile to the top that match those rates found by Hertz (2005) and Mazumder (2008) for white families. However, the loss-averse model (with independent earnings ability) is unable to generate a significant poverty trap, again consistent with data for white families. Allowing for the heritability of earning ability does generate a poverty trap, but it also increases the affluence net to what appears to be an unreasonable level, while reducing the jump up rate to an unrealistic level.

I explore alternative explanations for the poverty trap experienced by some black families in related work (Malloy, 2010). Standard explanations for the black poverty trap include labor market discrimination and lower earning ability, perhaps due to lower quality primary education or the health effects related to poverty. In addition to these explanations, I explore the possibility that black children may have lower aspirations than white children. Mazumder (2008) finds that black children born in the bottom half of the income distribution are much less likely to jump up a quintile than are their white counterparts. If these children have generally lower aspirations they will invest in less education. If low aspirations are combined with labor market discrimination, a loss aversion model can generate a significant poverty trap.

If we believe that people are indeed making their education-investment decisions with loss averse preferences and, therefore, that the outcome is not efficient, there is a role for public policy. Many industrialized countries have eliminated the direct cost of post-secondary education for those who qualify. While this would help reduce the inefficiency, additional policies may also improve welfare. The main role of public policy should be improving the market for education loans, and in particular should promote the ability to borrow beyond the direct cost of education so that students are able to live a more normal life while in school. Given that there is a direct individual benefit of education, in the form of higher future wages, it makes sense that the individual should bear the cost. The fact that there are positive social benefits as well, in the form of higher total output, provides ample reason for the government to be involved in improving these capital markets. The United Kingdom actually conducted a pilot experiment in 1999, in which it paid students to stay in school after the age of 16. Dearden et al (2007) found that students did stay in school longer than a control group.

School districts in the United States have also begun experiments in which students are paid either to stay in school, for attendance and good behavior, or for earning higher grades. These experiments are still in their early phases and I do not believe there is any published research on their results. The school district in Tucson, Arizona has started a pilot program of paying students $25 a week for good behavior, good grades, and (implicitly) staying in school. The Education Innovation Lab at Harvard University, run by economist Roland Fryer, has run four experiments in four cities. While it is still too early to measure many outcomes, especially whether or not cash payments prevent students from dropping out, some of the experiments have met with some success.[6]

Is there evidence that loss aversion is important to education-investment decisions? Page, Garboua, and Montmarquette (2007) conduct an experiment in which people are asked to make a costly education-like investment in order to increase future earnings. Their findings suggest that people do have loss-averse preferences when making these types of investments. In this paper, I show that when agents have loss-averse preferences, there is a U-shaped education investment decision in reference consumption, which is not the case if preferences are concave everywhere. Using the NLSY data set I find support for this U-shaped decision rule, with those at the lower end of the parental income distribution investing more in education than those immediately above them.

Loss aversion has been found to help explain phenomena as diverse as trade policy (Tovar, 2009), asset pricing (Yugo 2008), and physician behavior (Rizzo and Zeckhauser, 2003). It has also recently been found to be present in capuchin monkeys (Silberberg et al, 2008). While models with loss aversion do present some serious technical challenges to economic theorists, it is time to make loss aversion a standard part of our economic toolbox. While there may be some cost in analytical tractability of our models, the benefit will be in the ability to better model reality.

[6]Fryer shared some of these results with Time magazine in the article "Should Kids Be Bribed to Do Well in School?" by Amanda Ripley, April 8, 2010, available at http://www.time.com/time/nation/article/0,8599,1978589-1,00.html.

Chapter 2

Low Aspirations, Loss Aversion, and Persistent Group Outcomes

Black and white workers display significant differences in their labor market outcomes. Black workers tend to have less education and earn lower income than their white counterparts at each level of education. This paper explores three possibilities (wage discrimination, lower earning ability, and low aspirations) for these gaps within the framework of a model with loss aversion and inherited reference consumption. When people have loss-averse preferences, low aspirations lead to lower levels of chosen education. Loss aversion and low aspirations can lead to education outcomes similar to those caused by outright discrimination or lower earnings ability. When combined with wage discrimination, the model with low aspirations can also help explain the larger poverty trap and lower affluence net observed for black families as opposed to white families. Simulation results compare favorably to inter-generational quintile transition rates in the literature from both the PSID and the NLSY. The paper also shows that even without lower earning ability or aspirations this type of model offers one explanation for the spike in the relative number of black workers with less than a high school degree that occurred after the Civil Rights legislation of the 1960s. In the model it takes many generations to reach educational equality after a period of wage discrimination is ended.

2.1 Introduction

The history of black workers in the United States labor market is marked by inequality of opportunity and of outcomes. In the midst of the civil rights movement in 1960, black workers were paid only about 60% of the wages of their white counterparts according to Card and Krueger (1992). They show that one of the main components of this wage gap was the lower return to schooling for (especially southern) black workers. Given this lower marginal benefit to education and legal barriers to higher education for black students in many parts of the country, it is not surprising that compared to white workers, only about half as many black workers graduated from high school and even less from college compared to white workers. But it has now been over four decades, or two generations, since the civil rights legislation of the 1960s was passed, and there remains a persistent gap in both educational attainment and median income at each education level between black workers and white workers. Controlling only for education, black workers still earn only about 85% (at the median) of their white counterparts and receive only about half as many advanced degrees.

Why have we so far been unable to eradicate this persistent gap in educational and labor market outcomes? The existing literature offers a number of possible explanations. Card and Krueger (1992) show that improved quality of black education contributed to the decline in the black-white wage gap from 1960 to 1980, and Maxwell (1994) concludes that the difference in the quality of education between black and white students during the 1980s can explain as much as two-thirds of the level of the wage gap. Urzua (2008) claims that unobserved cognitive ability is the main driver of the education gap and Neal and Johnson (1996) show that the results of the Armed Forces Qualification Test taken at age 18 (basically at the end of free secondary education) can explain much of the wage gap. These last two beg the question of where these two gaps come from and both leave open the possibility that lower quality primary and secondary education is driving the results.

Another possibility, expored in Grodsky and Pager (2001), is that black workers face a larger wage gap as they move up the career ladder, so that the differences observed in the data are mainly driven by an effective career ceiling faced by black workers. Another possibility is that black workers self select into different types of jobs that pay less. Grodsky and Pager find that black workers face less of a wage gap in the public sector than in the private sector which may push them into that arena.

Underlying each of these papers is the question of whether (and how much) black workers still face outright discrimination in the labor market. That is, how much less are they being paid for being equally productive as white workers? If discrimination is the main driver of the wage gap, then a policy of equal pay for equal work should have helped (and continue to help) eliminate both the wage and education gap. If, on the other hand, the gap is driven by lower quality education, self-selection into different fields or types of jobs, or something else, then the policy solution may be more complex.

Some recent research has considered the idea that black students face a culture that is less accepting of academic success. Fordham and Ogbu (1986) introduced the possibility that successful black students are ostracized by their peers for "acting white" and Fryer (2010) finds additional empirical support for the theory, especially in schools with more interaction between the races. Cook and Ludwig (1996) find little support for this theory, but Kao and Tienda (1998) find that while young students of all races tend to have high educational aspirations, black and Hispanic students find these aspirations harder to maintain as they progress through high school. Both Austen-Smith and Fryer (2005) and Akerlof and Kranton (2000) have built theoretical models in which identity in a group can lead a person to choose what appear to be suboptimal investments in education in order to avoid social costs that outweigh the income benefits.

Another major difference between black and white labor market outcomes is the matrices of intergenerational transition rates between income quintiles. Mazumder (2008) shows that white children are much less likely to be caught in the poverty trap in which those born in the bottom quintile of the parental income distribution remain there as adults. With perfect intergenerational mobility, the transition rates between income quintiles would be a uniform 20%. Only 24.9% of white children born in the bottom quintile stay there as adults whereas 43.7% of black children are caught in this poverty trap. On the other hand, white children enjoy a significant affluence net in which 38.9% of those born in the top quintile stay there as adults. The comparable number for black children is only 21.3%. The difference in these intergenerational transition rates is one of the puzzles that this paper attempts to solve.

The model presented in this paper does not directly punish members in the group for investing more in education, as in Austen-Smith and Fryer (2005), but proposes that members in different groups may measure "success" differently, similar (in spirit) to Akerlof and Kranton (2000). Specifically, this paper extends a behavioral model presented in Malloy (2010) in which heterogeneous agents have loss-averse preferences and measure utility compared to the level of their parents' consumption which they enjoyed as a child. That model showed that agents will often invest in less education than would maximize their lifetime resources in order to avoid a loss in the first (or educational) period of life. This is true even when capital markets allow them to borrow the entire cost of education (but not when there are perfect capital markets that would allow them to move income from the second period into the first).

In this paper, I explore three possible explanations for the black-white gaps in education and median income. The first is that the group of disadvantaged workers face labor market discrimination. While all workers have the same underlying "earning ability", the marginal benefit of investing in each subsequent year of education is lower for disadvantaged workers due to wage discrimination. Second, I allow for the possibility of differences in the distribution of the underlying earning ability, perhaps due to lower quality primary and secondary education, so that the populations of the two groups are not the same once they reach the labor market. Finally, I allow for the possibility that the disadvantaged group has lower aspirations than their counterparts in the advantaged group. That is, reference consumption is lower for this group than for the advantaged group, perhaps because generations of discrimination have led to a culture of lower expectations.

I show that all three of these explanations can, with varying success, match the education distribution of black workers compared to white workers. All three candidate explanations imply a larger poverty trap for the disadvantaged group, in which the children of the poor stay poor, and a lower affluence net implying that those in the disadvantaged group who are lucky enough to be born at the top of the income distribution are not as likely to stay there as those in the advantaged group who were born at the top. However, only wage discrimination is able to explain the black-white wage gap in median income conditioned on education that we observe in the data at all education levels. When there are differences in the distribution of earning ability, but no wage discrimination, then fewer members of the disadvantaged group will invest in higher levels of education, but those that do will be paid the same as those in the advantaged group because their underlying ability is the same. A model with wage discrimination and low aspirations (but the same underlying distribution of earning ability) does the best in matching the observed differences in the education distribution and median income of black and white workers.

Finally, I calculate the transition toward equality that would come in the model after a period of wage discrimination that is finally eliminated. I show that it can take many generations for the education distribution to equalize across groups in a model with loss aversion and inherited reference consumption. If the disadvantaged group develops low aspirations, perhaps due to the generations of outright labor discrimination, the education distributions will never equalize. On the other hand, getting rid of discrimination in the labor market is enough to equalize median income levels (controlling for education level) in the model. If discrimination is completely eliminated this equalization happens immediately. However, due to loss aversion and inherited reference consumption the model exhibits large shifts in the education distribution of the disadvantaged group which actually lead these workers to have a higher median income, especially at the lower levels of education. This is because the earning ability of the median worker of the disadvantaged group is higher than that of the advantaged group at low levels of education. This is not seen in the data. In fact, the education distribution is equalizing faster in the data than the model predicts, showing that there are other forces at work. However, the model does match the observation of a relative spike in high school dropouts that we see in the data after the passage of the civil rights legislation in the 1960s.

The next section provides a brief description of the model and the group level differences. Section 3 describes the data of the black-white education and wage gaps including differences in intergenerational transition rates between income quintiles. I describe the calibration and present simulation results in section 4. Section 5 describes the transition of the education distribution and median income levels after a period of labor market discrimination that is either immediately or slowly eliminated. I present concluding remarks in section 6.

2.2 Model

2.2.1 Basic Setup

The basic model is a sequential generations model in which agents have loss averse preferences as explained in detail in Malloy (2010). Agents live for three periods: period 0, or childhood; period 1, in which agents may invest in education and/or work; and period 2 in which agents work, reproduce, and pass on a bequest to the next generation. In period 0, agents receive their reference consumption, c_0, and a possible bequest, T, from the previous generation. Agents also learn their "earning ability," ψ, which is best thought of as a combination of some sort of inherent ability and the quality of their education up to the mandatory age of enrollment, usually 16 in the United States.

The agent's problem is given by equation (1) below. The per period utility function, $U(z_t)$ is assumed to represent loss averse preferences (discussed below) and each agent receives warm glow utility from leaving a bequest, T', to the next generation, described by $B(T')$. Given the agent's state variables, namely reference consumption, c_0, bequest from the previous generation, T, and earning ability, ψ, she chooses education level, e, consumption in each period, and the bequest she will leave to the next generation to maximize lifetime utility. Utility depends on relative consumption, which is a percent of reference consumption so that $z_t = \frac{c_t - c_0}{c_0}$. The borrowing constraint is represented by Ω in equations (1b) and (1c) and gives a limit to how much may be borrowed (s) in period 1, and also places a limit on how much may be spent on the direct cost of education, $\varepsilon(e)$, in equation (1c). Different values of Ω were discussed in Malloy (2009). In this paper I will generally assume that agents can borrow the full cost of education. Equations (1d) and (1e) represent the per period budget constaints, with the wage, $\omega(e, \psi)$ depending on both education and earning ability, while (1f) assures that consumption and bequests are nonnegative.

$$V(c_0, T, \psi) = \max_{e, c_1, c_2, T'} U(z_1) + \beta[U(z_2) + B(T')] \tag{1a}$$

$$s.t.$$

$$s \geq \Omega \tag{1b}$$

$$\varepsilon(e) \leq T + \Omega \tag{1c}$$

$$c_1 + s + \varepsilon(e) = (1 - e)\omega(e, \psi) + T \tag{1d}$$

$$c_2 + T' = \omega(e, \psi) + s(1 + r) \tag{1e}$$

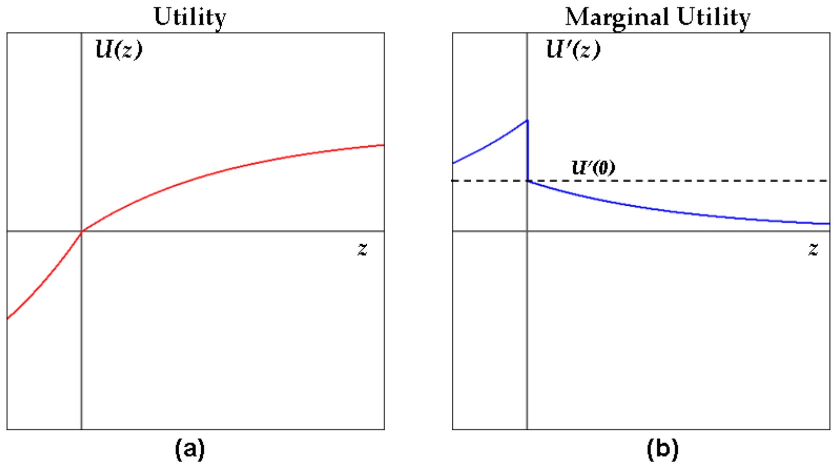
$$c_1, c_2, T' \geq 0 \tag{1f}$$

There are two main features of loss-averse preferences. First, while utility for consumption above the reference level ($z_t \geq 0$) is assumed to be concave, utility for consumption below the reference level ($z_t < 0$) is assumed to be convex, so that there is both a decreasing marginal utility to gains and a decreasing marginal *dis*utility to losses. The main consequence of this is that while an agent will smooth gains across periods whenever possible, he will lump losses into one period in order to avoid (or decrease) a loss in the other period. The second main feature of loss-averse preferences is that the marginal utility around zero is assumed to be steeper for a loss than for a gain. This is captured by the loss aversion parameter, λ, in equation (2) below.

$$U(z_t) = \begin{cases} \frac{1-e^{-\mu z_t}}{\mu} & \text{for} \quad z_t \geq 0 \\ \lambda \frac{e^{\nu z_t}-1}{\nu} & \text{for} \quad z_t < 0 \end{cases} \tag{2}$$

I use constant absolute risk aversion utility, as suggested by Kobberling and Wakker (2004), because it allows for strong loss aversion as discussed in Nielson

Figure 2.1: Strong Loss Aversion Utility and Marginal Utility



(2002). The utility function $U(\cdot)$ is strongly loss averse if $U(0) = 0$ and $U'(y) \leq U'(z) \quad \forall z < 0 < y$. In this case, since z_t is measured in percentage terms (and so has a minimum at -1) we can say that $U(\cdot)$ is strongly loss averse iff $U'(-1) \geq U'(0)^+$ where $U'(0)^+$ is the marginal utility as z_t approaches 0 from the right. A strongly loss averse utility function is represented in figure 2.1(a) and marginal utility in figure 2.1(b). This simplifies the consumption-savings calculation, as in Bowman et al. (1999), and I make the assumption of strong loss aversion throughout the paper.

The direct cost of education, given by $\varepsilon(e)$, is given by equation (3). The function $\varepsilon(e)$ is assumed to be (at least weakly) convex so that the last unit of education costs at least as much as the first unit. Formally, I require $\phi_1, \phi_2 \geq 0$. In fact, the first "unit" of voluntary education, finishing high school, is free for students at public schools in the United States while the last year of school, such as in a medical or law program can be quite expensive, even at a public university.

$$\varepsilon(e) = \phi_1 e + \phi_2 e^2 \tag{3}$$

The function $B(T')$ in equation (1) gives the utility gained from leaving a bequest of size T'. This type of bequest is known as a warm glow bequest because the person giving the bequest does not take the *utility* of the person receiving the bequest directly into consideration. I use the same CARA function for $B(\cdot)$ as is used for gains in utility:

$$B(T') = \frac{1 - e^{-\theta(T'/c_0)}}{\theta} \tag{4}$$

Because z_t is measured in percentage terms, it is useful to scale bequests by the level of reference consumption. If $\theta = \mu$ then the agent sets $\frac{T'}{c_0} = z_2$ when $z_2 \geq 0$ or $T' = c_2 - c_0$. In many cases, however, agents will optimally have zero gain in the second period, in order to minimize losses in period one. An agent consuming at (or below) the reference level of consumption in period 2 will leave no bequest to the next generation.

The nonconcavity of loss-averse preferences presents a number of problems in solving the model. Given the widely replicated experimental evidence of loss aversion in a number of different settings, these problems represent the main reason that loss averse models have not been more widely adapted. In this paper (as in Malloy (2009)), I make a number of assumptions, following Bowman et all (1999), that allow the model to be solved.

First, it should be noted that first order conditions will not allow us to solve the agent's problem, as the second order conditions for a maximization fail due to the convex preferences when consumption is less than the reference level of consumption, c_0. Another potential problem introduced by the nonconcavity of the model is the possibility of multiple equilibria. Because of the decreasing marginal *dis*utility of a loss, an agent who faces a loss will want to lump the loss (as much as possible) into one period. In multi-period models, this gives the very real possibility of multiple equilibria as the agent will not care in which period she takes the loss (so long as the standard assumption that $\beta(1 + r) = 1$ is made).

In order to find a unique equilibrium I first make the assumption of strong loss aversion so that the marginal utility of any sized loss is larger than the marginal utility of any gain, as noted above. This makes the consumption-savings decision for the agent clear cut as he will always (when possible) reduce a gain in one period in order to reduce a loss in another. Also, an agent facing losses in both periods will take as large a loss as possible in one period in order to reduce the loss in the other period. Because the borrowing constraint will bind in the first, or education, period, consumption in that period will never be larger than in the second period. Of course, agents that face gains in both periods will try to smooth them as in a standard model, subject to the borrowing constraint.

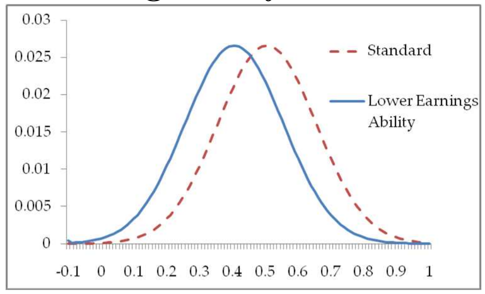
2.2.2 Group Differences

In this version of the model there will be two groups indexed by $j \in \{A, D\}$, where A represents an advantaged majority group and D a (possibly) disadvantaged minority group. There are three possible ways in which the minority group may face disadvantages in the model. First, they may receive lower quality primary education, lowering their average level of earning ability:

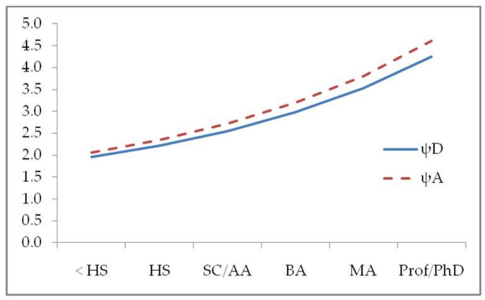
$$
\begin{aligned}
\psi_i^j &= \rho \psi_{i-1}^j + (1-\rho)\bar{\psi}^j + \chi_i \\
\chi_i &\sim N(0, (1-\rho^2)\sigma_\psi^2)
\end{aligned}
$$

The parameter ρ represents the possibility that earning ability is heritable, although in the baseline model I assume that it is independent across generations so that $\rho = 0$. To account for the possibility that the minority group receives lower quality education I allow the mean of the distribution, $\bar{\psi}$, to vary by group so that if education quality is lower for the minority group than for the majority group we have $\bar{\psi}^A > \bar{\psi}^D$. This possibility is illustrated in Figure 2.2, where panel (a) shows

Figure 2.2: Shift in Earning Ability and Resulting Wage Function



(a)



(b)

the shift of the distribution in the disadvantaged group's earning ability to the left and panel (b) shows the resulting wage for someone with earning ability $\bar{\psi}^A$ and $\bar{\psi}^D$ at each level of education. As can be seen by widening of the difference between the wage curves for the advantaged and disadvantaged groups, the marginal return to education is also lower for the disadvantaged group.

Second, the model allows for the possibility of labor market discrimination. In general, a worker's wage will be equal to her marginal product and will be a function of both the worker's earning ability, ψ, and her chosen level of education, $\hat{e}$. In addition, the disadvantaged group may face some haircut, δ, to its wage. This form of discrimination is not "rational" on the part of firms which, in a competitive environment, should be willing to pay each worker a wage equal to his marginal product. If, however, the firms have some level of monopsony power or there is a level of collusion between firms so that this type of discrimination has become accepted and institutionalized, the workers in group D will have no choice but to accept the wage offered. The wage function is given by equation (5).

$$\ln[\omega^j(e,\psi)] = (1-\delta^j)[\alpha_1 e + \alpha_2 e^2 + \zeta e\psi + \kappa\psi] \tag{5}$$

Here the parameters α_1 and α_2 capture the (log) quadratic return to education, κ is the (log) linear return to earning ability, and ζ is the interaction term between earning ability and education so that those with higher earning ability get a greater return to their investments in education. For the advantaged group, A, there is no haircut in the wage so that $\delta^A = 0$, whereas the disadvantaged group D faces possible discrimination so that $0 \leq \delta^D < 1$.

Finally, there may be a behavioral difference between the two groups. With loss averse preferences, utility is measured relative to some reference point. In this model, reference consumption is equal to the previous generation's period two consumption. That is, agents measure utility by comparing current period consumption to the level of consumption they enjoyed as a child. If the disadvantaged group, perhaps due to generations of discrimination, feels they won't get a fair shake in life, they may lower their expectations and measure utility against a lower level of reference consumption. Formally, I model reference consumption as:

$$c^j_{0,i} = (1+a^j)c^j_{2,i-1}$$

Reference consumption for person i in group j is equal to his parent's consumption in period 2 multiplied by some group specific factor, $(1+a^j)$. If $a^j > 0$ then agents in group j measures utility against a reference consumption greater than that of their parents. While formal expectations do not enter into this model, we can interpret a positive a^j to mean that children expect to do better than their parents. On the other hand, if $a^j < 0$, then children measure utility against some level of consumption less than that of their parents. That is, they expect to do worse than their parents.

For simplicity, and in order to keep the model stationary, I assume that $a^A = 0$ while I allow a^D to be less than zero. In the popular discourse, and in an economy with positive average growth, we usually ask if children expect to do better than their parents in real terms. In this model, with no growth, we can think of this mechanism as asking whether children expect to do better (or the same, or worse) than their parents in terms of their placement in the income distribution.

The main effect of the parameter a^j is that it can change the utility maximizing level of education for two agents from different groups who are otherwise identical. The solution methodology is described in the next section, but the intuition is simply that by shifting the reference point, agents can face a loss at different levels of education investment. For example, imagine a person with earning ability above the mean who would maximize lifetime resources by getting a master's degree, maybe an MBA. Investing this much in education may force the person to realize a loss. For a child from the advantaged group, if c_0 is high enough, a first period loss is unavoidable in any case, so the optimal decision is to choose the resource maximizing level of education. But for a disadvantaged child, if a^D is low enough, the adjusted reference level of consumption may be low enough that the child will be able to avoid the loss in the first period by reducing her investment in education. Therefore, she may be "satisfied" with only getting a bachelor's degree while her counterpart from the advantaged group goes on to get the MBA, accepting the inevitable loss in the first period.

In this way, wages for the disadvantaged group will be lower than those with similar earning ability in the advantaged group through the equilibrium level of education chosen by the members in each group. Even if there is no current discrimination and no differences in the distribution of earning ability across groups, there may still exist wage differentials that are best understood as the legacy of past discrimination which has affected current preferences and expectations. Note, however, that the model implies that these wage differentials will disappear once one controls for education.

Of course, in reality it is quite possible that these three possible explanations for the white-black disparities in education and wages work at the same time. In section 4, I present both simulations that allow only for a single explanation, in order to note differences and similarities in the three, and simulations involving combinations of the three, to see which combination can do the best in matching the data.

2.2.3 Solution Methodology

The solution methodology here is the same as in Malloy (2009). Briefly, the assumption of strong loss aversion allows me to solve the consumption and savings problem for each agent for a given level of education. I can then compare the utility derived by this solution for each education level and choose the level of education that maximizes utility, $\hat{e}$. I can compare this to the level of education that would maximize lifetime resources, e^*, which is the level that would be chosen in a standard model with perfect capital markets.

The numerical solution methodology is straightforward. Given the optimal consumption/saving/bequest pattern conditioned on education, I calculate the utility-maximizing (as well as the resource-maximizing) level of education for each possible state (where the state consists of reference consumption, earning ability, and bequest). This education level then gives consumption in each period, saving/borrowing between periods, and a bequest level for the next generation. I simulate an economy with N families (N^A in the advantaged group and N^D in the disadvantaged group, where $N = N^A + N^D$), each starting with a normally-distributed randomly assigned earning ability, reference consumption, and bequest. I calculate the education choice, $\hat{e}$, for each person in the initial generation as well as their consumption in each period, savings and bequest. I then assign a random earning ability level to each of the N

agents in the next period and carry over their reference consumption and bequest level from the previous generation. I can then calculate the correlation of education, wages, consumption, and bequests between generations. I run the simulation for 100 generations and check to make sure that the distribution of wages, consumption, and bequests has converged[1].

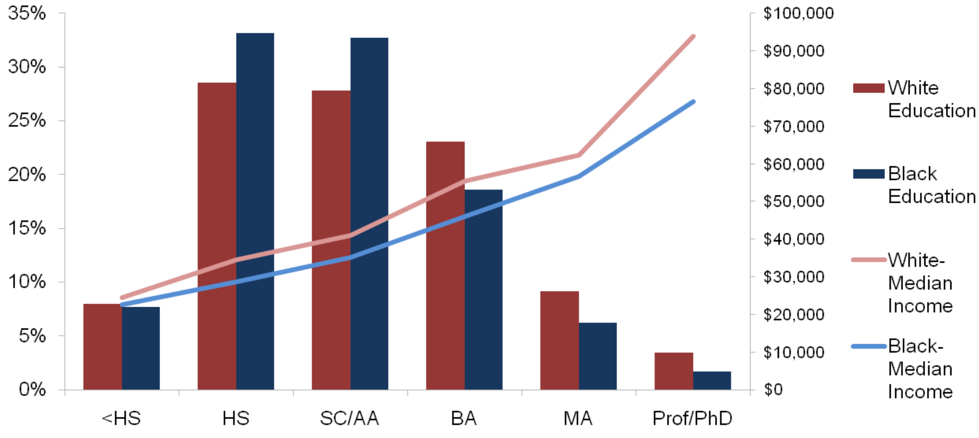
Because each agent's choice depends only on her specific state variables (reference consumption, earning ability, and bequest), I can run the simulation separately for each group and calculate the education distribution and median wage at each education level. I then combine the two populations in order to calculate the wage and education gap of the disadvantaged group as well as the overall intergenerational transition rates between income quintiles for each group.

2.3 Data

Differences in education levels and labor market outcomes between white and black workers in the United States have been well documented, even if explanations for these differences have differed. This section reviews key facts to establish target moments for the calibration and goal posts by which the various calibrations in the next section can be tested. As we will see, while the model offers a number of possible explanations for the lower education levels and wages for black workers, not all explanations will be able to match the other statistics, such as the income quintile transition rates, which differ substantially by race. Other literature reviewed here will add stronger a priori support for some of the model mechanisms than others.

[1] For wages, for example, I add up the wages in each decile and then take the norm of the difference between generations. For convergence, I require this to be less than some number ϵ.

Figure 2.3: Education Distribution and Median Income Levels by Race



2.3.1 Education Levels and Labor Market Outcomes

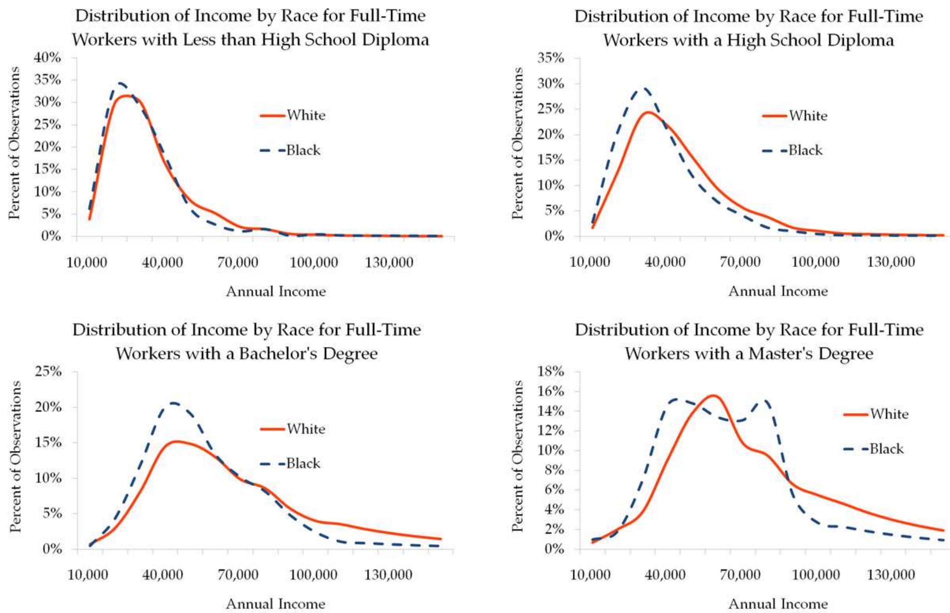
The columns in figure 2.3 plot the education distribution (on the left axis) for white and black full-time workers aged 25-64 in 2007 according to the Current Population Survey. The percentage of high school dropouts is almost identical for both groups: 7.95% for white workers and 7.67% for black workers. The major difference in the two populations is that compared to white workers, a much larger percentage of black workers finish with a high school degree or with something short of a bachelor's degree. At the upper end of the education distribution, only 1.67% of black workers have a PhD or professional degree as opposed to 3.45% of white workers, and those with master's degrees account for only 6.24% of black workers but almost 10% of white workers. So at least for full-time workers, the education distribution for black workers is not a simple shift to the left of the distribution for white workers, but rather a clumping in the high school graduate and some college categories. The same is true if the sample is restricted to 25-34 year olds.

A more inclusive sampling including all people over the age of 25 shows a similar pattern, with the only significant change being at the bottom. The percentage of white people over the age of 25 who have not graduated from high school jumps to almost 9% while the percentage of black people who have not finished high school is 9.8%. These numbers are 8.8% and 9.4%, respectively, when only including ages 25-64. This shows that there is a small but significant number of poorly educated people who do not participate in the labor market and that the percentage of such people is slightly higher for black people than for their white counterparts. It should be noted that the CPS accounts only for the non-institutionalized population. The prison population represents a much larger percentage of the black population, and of these 44% have not finished high school, compared to 27% of white prisoners in 1997 (Harlow, 2003). In this paper, I will focus on those working full time in the labor market and abstract from the issues facing those at the very bottom of the income and education distributions such as chronic unemployment, the non-formal labor market, and the risk of prison.

The solid lines (and right axis) of figure 2.3 show the median income by education level for white and black workers. Black workers consistently earn less than white workers at any given education level. Note, however, that the difference is not constant across education levels. For example, for those who have not graduated from high school, black workers earn 92% of what white workers earn. Black high school graduates and college graduates earn just under 83% of what white workers in the same category earn while those with Master's degrees earn 91% of their white counterparts. At the top level of education, black workers with a PhD or professional degree earn less than 82% as much as their white counterparts.

The income distribution by education level for black workers is substantially different in some cases than the distribution for white workers. Figure 4 gives the income distribution for white and black workers at four different education levels. At the lower education levels, high school dropout and high school graduate, the income distribution for black workers is shifted slightly to the left with a fatter left tail and a skinnier right tail. Note, however, that the mode of each distribution is more or less

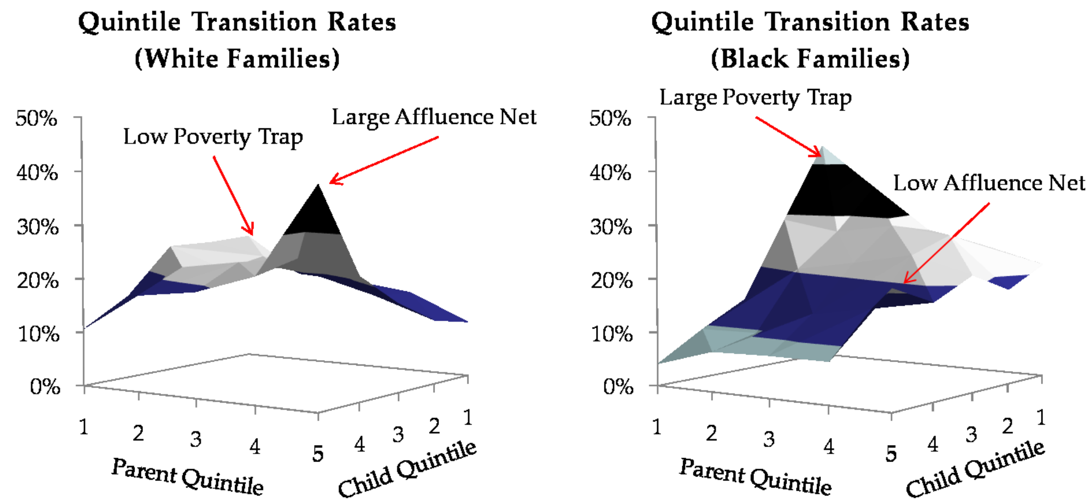
Figure 2.4: Income Distribution by Race at Four Education Levels



the same. The distribution for those with a Bachelor's degree is similar. However, the distribution for black workers with higher levels of education is substantially different than that for white workers. For those with a Master's degree, there appears to be a much more pronounced bimodal distribution for black workers than for white workers. This suggests the possibility that different types of workers achieve a high level of education in the different racial groups.

Intergenerational transition rates between income quintiles also differ by race. Hertz (2005) (in the Panel Study of Income Dynamic, PSID) and Mazumder (2008) (in the National Longitudinal Study of Youth, NLSY) calculate the intergenerational transition rates between income quintiles. Transition rates for the overall population are fairly symmetric. However, significant asymmetries emerge when the data are broken down by race. Figure 2.5 graphs the transition rates from Mazumder (2008) for white and black families. I look at four main transition rates in the data: the well documented poverty trap, defined as the fraction of children born in the bottom quintile who stay in the bottom; the affluence net, defined as the fraction of those

Figure 2.5: Intergenerational Income Quintile Transition Rates



Source: Mazumder (2008)

born into the top quintile who stay at the top; the fall down rate, defined as the fraction of children born in the top quintile who fall down to the bottom quintile, and the jump up rate, defined as the fraction of those born at the bottom who jump up to the top quintile.

If there was perfect intergenerational mobility, all quintile transition rates would be 20%. For all NLSY workers, Mazumder estimates that the poverty trap is 33.5%, the affluence net is 37.8%, the jump up rate is 7.4%, and the fall down rate is 10.9%. A symmetric transition matrix such as this is what we would expect to see if earning ability was highly correlated between generations. Suppose, for instance, that wages of successive generations are determined as follows:

$$
\begin{aligned}
\omega_i &= \rho_\omega \omega_{i-1} + (1-\rho_\omega)\bar{\omega} + \epsilon \\
\epsilon &\sim N(0, (1-\rho_\omega^2)\sigma_\epsilon^2)
\end{aligned}
$$

Where ω_i is the wage of the child and ω_{i-1} is the wage of the parent. When $\rho_\omega = 0.4$, the poverty trap and affluence net are around 37.7% and the jump up and fall down rates are around 6.7%. All of these rates are fairly close to the data.

However, intergenerational mobility becomes highly asymmetric when the data are disaggregated by race. Figure 2.5 shows that white families experience a significant affluence net (38.9%), a fairly low poverty trap (24.9%), and similar jump up and fall rates of around 10.5%. Almost the opposite is true for black families. These families face a large poverty trap (43.7%), a low affluence net (21.3%), a very low jump up rate (4.1%), and a fairly large fall down rate (21.6%). Hertz (2005) finds very similar transition rates in the PSID for white families and even more extreme transition rates, in the same direction as Mazumder (2008), for black families. Given the asymmetric transition rates by race, a successful model will need to explain why poor black children are more likely to be caught in the poverty trap and why rich white children are more likely to be supported by the affluence net.

2.3.2 Possible Explanations for Racial Disparities

There have been many papers written trying to explain the black-white wage gap. These papers were especially fashionable from the 1970s through the 1990s and come to a number of different conclusions. Card and Krueger (1992) find the black-white wage differential fell from about 40% in 1960 to 25% in 1980. They attribute one-fifth of this decrease to the change in the quality of education for black children over that time. Grodsky and Pager (2001) look at the occupational structure of black and white earnings and find that, within the private sector, racial wage differentials increase as workers move up the career ladder, even controlling for individual characteristics. They find that this is not the case in the public sector. Neal and Johnson (1996) rely on the Armed Forces Qualification Test, taken between the ages of 16-18, as a measure of skill and find that it can explain most of the black-white wage gap, although they leave room for some (small) level of labor market discrimination. They suggest that the wage gap failed to narrow after 1980 because of a significant lingering skills gap.

Urzua (2008) attempts to control for both observed and unobserved ability, allowing the schooling decision to be endogenous to the model. He finds that unobserved cognitive ability has a large impact on the realized level of education, but still leaves the majority of the wage gap unexplained.

Recent work on labor market discrimination against black workers suggests that it still exists and can take quite insidious forms. Charles and Guryan (2008) test the model presented by Becker (1957) that predicts that the wages of black workers will be lower in areas with more racial prejudice. They present robust findings that this is indeed the case with labor market discrimination especially affecting black workers in areas of the country in which the marginal white worker has more prejudicial views. They calculate that one quarter of the black-white wage gap is due to this form of discrimination. In an experiment to test the possible discrimination of employers, Bertrand and Mullainathan (2004) sent out identical resumes to job advertisements with either traditionally white or black sounding names. They found that the resumes with black-sounding names, such as Jamal Jones or Lakisha Washington, get 50% fewer callbacks than did the same resumes with white-sounding names, such as Emily Walsh or Greg Baker. Because this was a controlled randomized experiment, the authors conclude that racial discrimination is still prevalent in the U.S. labor market.

As Urzua (2008) says, his analysis is unable to explain the source of differences in the unobserved ability between races. He postulates that such differences could be due to "unmeasured racial differences in early family environment including prenatal family environment, unmeasured racial differences in early schooling environment, cultural differences between groups, biological/genetic differences between groups, or, most likely, a combination of all of these." As noted above, Card and Krueger (1992) believe that an increase in the relative quality of education helped reduce the black-white wage gap after the end of Jim Crow laws. Maxwell (1994) finds evidence that the quality of schooling is the main driver of differences in skills between black and white workers and claims that it can explain as much as two-thirds of the wage gap. Her explanation is that black students come out of compulsory schooling with much lower basic skills than do their white counterparts, leading to lower levels of post-secondary education and lower wages. In fact, some research finds that once family background and educational ability (such as test scores at the end of high school, for example) are controlled for, black students go on to get at least as much education as their white counterparts (e.g. Bennett and Lutz, 2009).

While the model in this paper allows for both discrimination and racial differences in ability, it also allows a role for behavioral differences in aspirations. There has been considerable debate, starting with Fordham and Ogbu (1986) about whether or not successful black students are ostracized by their peers for "acting white" and are therefore less likely to strive for academic success. Cook and Ludwig (1996), for example, find little support for the idea that black students suffer greater penalties for academic success than do white students. Kao and Tienda (1998), on the other hand, find that black and Hispanic students find it more difficult to maintain their high educational aspirations than do white students. They point to the lower socio-economic status of these students as a reason it may be more difficult to achieve their educational goals.

Both Austen-Smith and Fryer (2005) and Akerlof and Kranton (2000) have built theoretical models in which identity in a group can lead a person to choose what appear to be suboptimal investments in education in order to avoid social costs that outweigh the income benefits. The model presented in this paper does not directly punish members in the disadvantaged group for investing more in education, as in Austen-Smith and Fryer (2005), but proposes that members in different groups may measure "success" differently. The reference level of consumption, c_0, can be interpreted as when "failure" or a loss turns into "success" or a gain. The loss-averse agents in this model do all they can to avoid a loss. But it is quite possible that the point at which failure turns into success will differ across groups. While the advantaged group wants to do at least as well as their parents (so that $a^A \geq 0$), it is possible that the disadvantaged group, perhaps due to generations of discrimination or social mores that are the result of discrimination, measure success at a lower level. That is, they believe they are consuming at a gain even at a lower consumption level than their parents (so that $a^D < 0$). In this way, the model is similar in spirit to Akerlof and Kranton (2000) who allow behavior to differ based on how people identify with a certain group.

Is it reasonable for black children to expect to do worse than their white counterparts? Mazumder (2008) measures racial differences based on where children start in the income distribution. While most born at the very bottom of the income distribution will do better than their parents no matter their race (having nowhere to go but up), this becomes less likely for black children born higher in the income distribution. For black children born in the 20-40 percentile range, only about half will exceed their parents' position in the income distribution, compared to between 60-80% of white children. Less than 40% of black children born in percentiles 40-50 will exceed their parents' position while white children have a better than 50% chance of doing so. Mazumder also calculates the probability of exceeding one's parents' position in the income distribution by 20 or more percentage points. For white children born in the bottom quarter of the income distribution this probability is about 60%. For black children the comparable number is only 40%. Meanwhile, 46% of the white children born in the bottom *half* of the income distribution will exceed the income range of their parents while only 22% of black children will.

This section suggests that there are a number of likely causes for the differences in black and white labor market outcomes. While it appears that overt labor market discrimination has decreased substantially over the last fifty years, there exists the possibility that such discrimination still exists, at least to some extent. It also seems quite plausible that the continuing legacy of Jim Crow is not in overt wage discrimination but rather in lower quality primary and secondary schooling that is the result of continued de facto segregation. This leads to lower basic skills for black students which then results in less completed education and lower earnings. Finally, it is possible that the expectations for labor market success are lower for black workers than for their white counterparts, leading them to decrease their reference level of consumption and define down what constitutes success. In this model, a lower level of reference consumption will lead black students to get less education, leading to lower earnings. Each of these three explanations is explored in the next section, first alone, and then in combination with the others.

2.4 Calibration Results

2.4.1 Parameter Values and Calibration Targets

The calibration methodology is similar to Malloy (2010). In that paper, I calibrated the model using the parameters in the wage function to match the education distribution and median income levels by education to the population of white, full-time workers. In this section I use those calibrated values as the starting point and focus on the differences between the black and white workers, using the three parameters in this model that allow for group differences. That is, I start with the model calibrated for white workers and experiment with different values for the parameters δ, representing possible labor market discrimination, $\bar{\psi}^D$, representing possible lower earnings ability, and a^D, representing the idea that the disadvantaged group may have lower aspirations than the advantaged group. It is, of course, possible that the reason black workers have different labor market outcomes is that they face a different wage function than do white workers. However, in order to focus on the mechanisms

Table 2.1: Parameter Values

Parameter	Value	Description	Source/Explanation
λ	2.25	Loss aversion parameter	Tversky and Kahneman (1992)
μ	0.01	CARA parameter for gains	Babcock et al (1993)
ν	0.8	CARA parameter for losses	Strong loss aversion
$\beta = R = (1+r)$	1	Rate of time preference & interest rate	Simplicity
ϕ_1	0.0	Linear term in cost of education	Estimated from Dept. of Education,
ϕ_2	0.3	Quadratic term in cost of education	National Ctr. for Education Statistics
e, ψ	[0,1]	Range for education and earning ability	Normalization
$\bar{\psi}$	0.5	Average ability level	
σ_ψ	0.15	Standard deviation of earning ability, ψ	3 st. dev. above & below the mean
ρ	0.0	Heritability of earning ability	Simplicity

explicit in this model, I make the simplifying assumption that the underlying wage function is the same for both groups.

The main parameter values are presented in table 2.1. I use a CARA parameter of 0.01 for gains. I use Tversky and Kahneman's (1992) value for the loss aversion parameter, λ, of 2.25. Other estimates of this parameter in the literature generally find it to be between 2 and 3. Given that strong loss aversion requires that the CARA parameter for losses, ν, be less than $\ln(\lambda)$, I let ν be the maximum value of $\ln(\lambda) \approx 0.8$. Because there are only two periods, I allow both β and the gross interest rate R to equal one. This is done without loss of generality so long as we make the standard assumption that $\beta R = 1$. In order to estimate the direct education costs, I use the Department of Education's National Center for Education Statistics, then normalize against the median income of a high school graduate, assuming each period is approximately 20 years. I set the direct cost of the first two levels of education, equivalent to a high school dropout and high school graduate, at zero as most students in the United States have access to free education through the twelfth grade.

The target moments for white full-time workers are presented in table 2.2 and the calibrated wage parameters are presented in table 2.3. The left column of table 2.2 shows the fraction of white workers in each education category, calculated using

Table 2.2: Target and Calibrated Moments for White Population

Moment	Target*	Model
Education Distribution		
High School Dropouts	7.9%	8.8%
High School Grads	28.5%	23.8%
Some College/Assoc. Degree	27.8%	27.5%
Bachelor's Degree	23.1%	31.0%
Master's Degree	9.2%	7.2%
Prof. Degrees/PhD	3.5%	1.9%
Median Income by Education Level (HS grad = 1.0)		
High School Dropout Wage	0.71	0.75
High School Grad Wage	1.00	1.00
Some College/Assoc. Degree Wage	1.19	1.57
Bachelor's Degree Wage	1.60	1.65
Master's Degree Wage	1.80	2.10
Prof. Degrees/PhD Wage	2.71	4.04

* Target is for white, full-time workers.

Table 2.3: Calibrated Parameter Values

Parameter	Calibrated Value
α_1: linear term on education	0.285
α_2: quadratic term on education	0.190
ζ: interaction term (education*ability)	0.655
κ: linear term on ability	1.450
Minimum marginal return to education	3.1%
Average marginal return to education	8.4%
Maximum marginal return to education	13.9%
Max-Min Wage Ratio	13.2

CPS data from 2007, as well as the median income for each education level expressed relative to the median for a high school graduate. The right column of table 2.2 shows the comparable moment implied by the model. The model does a fairly good job matching for the middle of the education and income distribution and struggles to match the education distribution at the high end. Table 2.3 also reports the marginal return to an extra year of education as a check on the calibration. These numbers vary between 3.1% and 13.9%, with an average of 8.4%, which are reasonable values according to the literature of wages and education (e.g. Lemieux, 2006).

Table 2.4 presents target moments for black full-time workers. Note that the

Table 2.4: Target and Calibrated Moments for Black Full-Time Workers

Moment	Target*
Education Distribution	
High School Dropouts	7.7%
High School Grads	33.2%
Some College/Assoc. Degree	32.7%
Bachelor's Degree	18.6%
Master's Degree	6.2%
Prof. Degrees/PhD	1.7%
Median Income by Education Level (White HS grad = 1.0)	
High School Dropout Wage	0.65
High School Grad Wage	0.83
Some College/Assoc. Degree Wage	1.02
Bachelor's Degree Wage	1.33
Master's Degree Wage	1.64
Prof. Degrees/PhD Wage	2.21

* Target is for black, full-time workers.

median income for each worker is expressed in terms of the median income for a white high school graduate. As noted above, black workers have a slightly lower fraction of high school dropouts than white workers, but a larger fraction with a a high school degree or some level of college short of a bachelor's degree. Comparing tables 2.4 and 2.2, we see that black workers earn less at every level of education than do white workers.

The first calibration exercise I undertake is attempting to match the target moments of the black population in table 2.4 by using only one of the three available parameters reflecting group differences. These results are presented in sections 4.3-4.5. In section 4.6, I see if combinations of two or more types of group differences can improve the model fit. But first I see what explanations such as labor market discrimination and borrowing constraints can do in a model without loss aversion.

2.4.2 Traditional Explanations in a Standard Model

In Malloy (2010) I find that a model with habit formation and preferences that are concave everywhere cannot generate a correlation of income across generations unless there are severe borrowing constraints. In the face of borrowing constraints, the model does generate some intergenerational persistence in income as the poorest cannot afford to go to school beyond a free high school education. In this section I investigate whether or not the standard model with wage discrimination and/or borrowing constraints can match the target moments of the black worker education distribution and median income levels and what these models imply for the intergenerational transition rates between income quintiles.

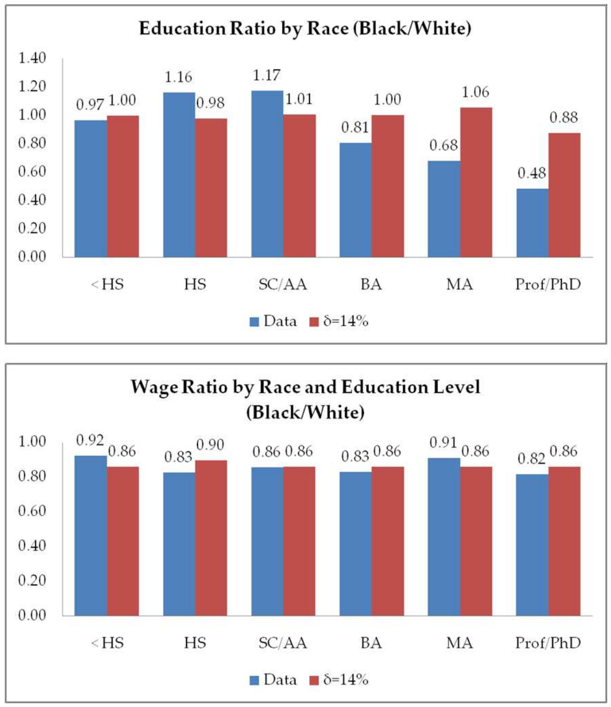
The only difference between the standard model with habit formation and the model presented above with loss aversion is that the utility function is now concave for all values of z_t:

$$U(z_t) = \frac{1 - e^{-\mu z_t}}{\mu}$$

As in equation (5) above, I allow the disadvantaged group to face possible labor market discrimination. In addition, the model will allow for the possibility that the bottom portion of the disadvantaged group does not have access to financial markets and therefore cannot borrow the cost of education. While there are of course no current legal restrictions that bar black students from borrowing (there are, in fact, laws prohibiting racial discrimination), there is some evidence that black students are less likely to avail themselves of the financial resources available to pay for college[2], in part because parents who have not gone to college have less information about student financial aid options. If this is the case, a borrowing constraint that limits the very poorest may be a reasonable approximation of reality.

[2] One such study was done by the United Negro College Fund's Patterson Institute commissioned by Sallie Mae. A summary of results can be found here: http://pattersonresearchinstitute.org/SallieMaeFindings.htm

Figure 2.6: Education and Median Income Ratios with Standard Utility, Habit Formation, and Wage Discrimination



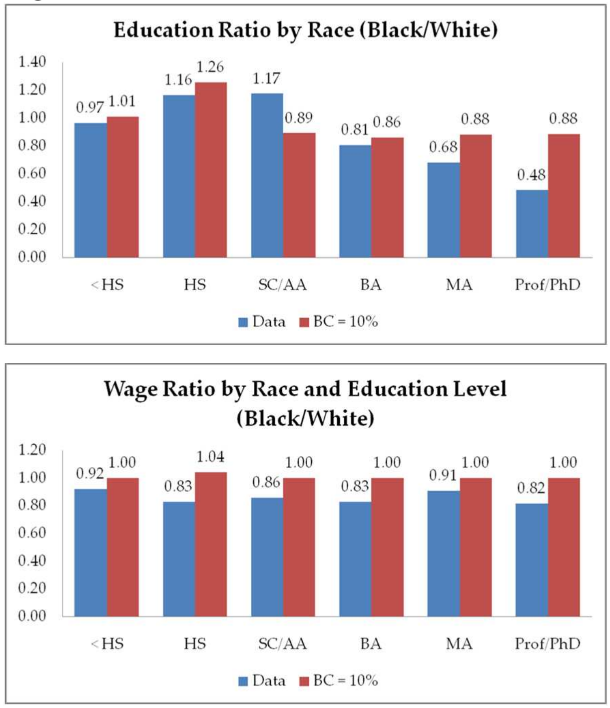
Using only the wage discrimination parameter, I find that a value of $\delta = 0.14$ does the best job of matching the target moments. While the model is able to match the black-white ratio of the median income levels by education, it is not able to match the black-white ratio of the share of workers at each education level. The moments observed in the data and generated by the model are given in figure 2.6. The share of the disadvantaged workers in each education category is only marginally different than for advantaged workers.

While Malloy (2010) finds that the standard model with habit formation and no wage discrimination is not able to generate intergenerational transition rates between income quintiles that match the data, the introduction of wage discrimination helps generate realistic transition rates. In fact, the standard model, with $\delta = 0.14$, is able to generate a poverty trap in the disadvantaged group for the bottom quintile of 36.4% and for the bottom two quintiles of 53.4%. The corresponding moments in the data are 43.7% and 67.6%, respectively. While this poverty trap is not quite large enough to match the data, it is larger for the bottom 20% than that implied by the model with loss aversion, as shown in the next section. The model also generates a smaller jump up rate for disadvantaged workers of 13.8% (compared to 4.1% in the data), but the fall down rate for this group is much too large at 35.9% (compared to 21.6% in the data). Similarly, the affluence net is too small at 13.4% for disadvantaged workers, compared to 21.3% in the data. Interestingly, while the correlation of intergenerational income for the advantaged group is basically zero, for the disadvantaged group this correlation is slightly negative at -0.013.

Borrowing constraints have the opposite effect of wage discrimination in the model. That is, the model with borrowing constraints is able to come close to matching the ratios of worker shares in each education category, but is unable to match the ratios of median income at each education level. This should not be surprising as workers in the model without discrimination are paid based on their productivity. The disadvantaged at each education level are equally productive as their advantaged counterparts so that the median income levels will only differ when there is a large difference in which types of worker receive each level of education. The observed and simulated moments when the bottom 10% of the disadvantaged group is unable to borrow to finance education are given in figure 2.7.

Given that the model with borrowing constraints is unable to match the median income by education data, it is not surprising that it does less well in matching the intergenerational transition rates between income quintiles as well. In this version of the model there is no poverty trap for those in the disadvantaged group born into the bottom 20% of the income distribution and only a modest one of 46.9% for those born into the bottom 40%. There is no significant difference in the affluence net or the rate at which children fall down from the top quintile to the bottom quintile between the two groups. The only transition rate this version of the model is able to come close to matching is the rate at which poor black children jump up from the bottom quintile

Figure 2.7: Education and Median Income Ratios with Standard Utility, Habit Formation, and Borrowing Constraints



to the top. In the data this rate is 4.1% while the model with borrowing constraints generates a value of 8.2%.

In this version of the model with standard concave utility and habit formation, a combination of wage discrimination and borrowing constraints does a fair job of matching the target moments. The main weakness is that it is unable to match the much lower share of black workers at the three highest levels of education. For example, in this version of the model, with wage discrimination of 14% and in which the bottom 5% of the disadvantaged group cannot borrow, the ratio of disadvantaged-advantaged workers at the top education level is 0.80, as opposed to 0.48 in the data.

The version of the model with both wage discrimination and a borrowing constraint does a fairly good job of matching the key intergenerational income transition rates. For example, it generates a poverty trap for the bottom 20% of the disadvantaged group of 39.8% (compared to 43.7% in the data) and 59.7% for the bottom 40% (compared to 67.6% in the data). It still generates too low of an affluence net (13.8%

vs. 21.3%) and the fall down rate is once again larger than in the data. The jump up rate is 9.1% (compared to 4.1% in the data). As we will see, compared to this model, the model with loss aversion generates too low of a poverty trap for the bottom 20%, but does a better job in matching the affluence net and jump up and fall down rates for black families. The overall correlation of intergenerational income in this version of the model is a modest 0.028 while for the disadvantaged group it is 0.072.

2.4.3 Wage Discrimination

In this section I explore the effect of labor market discrimination, as captured by the parameter δ in the wage function (5), on education and median income levels of the disadvantaged group. Wage discrimination decreases the marginal benefit of an extra year of education and reduces the numbers of the disadvantaged group that invest in the higher levels of schooling.

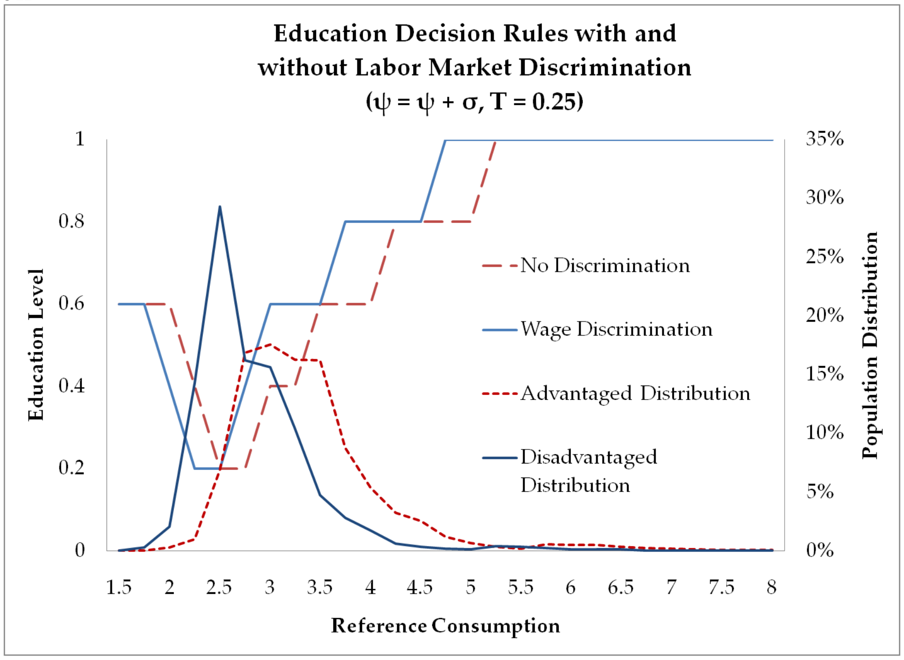
2.4.3.1 Decision Rule

One implication of the loss aversion model discussed in Malloy (2010) is the existence of a U-shaped education decision rule. Because agents have loss-averse preferences, reference consumption is important in determining how much education each person will invest in. Those in the middle of the reference consumption distribution may be able to avoid a loss in the first period by investing in less education and working more. Those with higher reference consumption will be more likely to face a loss in the first period no matter what level of education is chosen, and so will group the loss in the first period and invest in the level of education that will maximize lifetime resources. Those at the bottom of the income distribution will find it easier to avoid a loss in the first period (as their reference level of consumption is low), and so will invest in more education than those to their right in the reference consumption distribution.

Figure 2.8 graphs the education decision rule for an agent with earning ability one standard deviation above the mean and a bequest of $T = 0.25$ facing no wage discrimination, and that for an identical agent who does face labor market discrimination (so that $\delta > 0$). Both groups exhibit the U-shaped decision rule. However, the decision rule for agents facing discrimination is shifted to the left. This implies that those facing wage discrimination will in some cases choose a higher (or equal) level of education than those who do not face such discrimination, all else equal. While this may seem counterintuitive, it does make sense within the model. Groups facing discrimination earn lower wages at every level of education. Those with higher wages may be able to avoid a loss in the first period by investing in a lower level of education, while those facing lower wages are more likely to be facing a loss in any event, and thus optimally invest in more education so as to avoid (or minimize) a loss in the second period. As can be seen in the graph, there are also areas (at the bottom of the reference consumption distribution) in which the disadvantaged group chooses a lower level of education. In this case, because of the lower wages, it does not make sense to invest in more education, either because it would lead to a loss or because it would not reduce the loss due to the lower marginal benefit of education on wages.

These results, perhaps surprisingly, agree with the results from the literature (described in section 3), which show that when controlling for educational ability and family background, black students are actually likely to choose more education than their white counterparts. The story suggested by this model is that disadvantaged agents are more likely to face unavoidable losses in period one due to wage discrimination, and therefore optimally choose to invest in more schooling to avoid losses in period two. Figure 2.8 also graphs the resulting population distribution by reference consumption for both groups. Because of the lower wages due to discrimination, the distribution of the disadvantaged group by reference consumption is shifted to the left compared to the advantaged population. This shift is why the disadvantaged group

Figure 2.8: Education Decision Rule and Population Distribution with Wage Discrimination

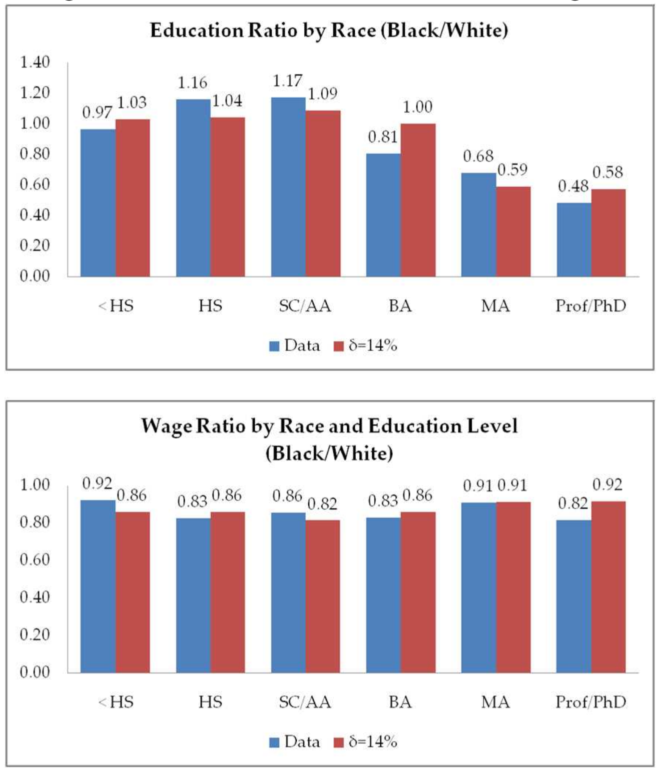


tends to get less education despite the fact that a disadvantaged student will actually invest in more education than an advantaged student of equal ability and bequest at higher levels of reference consumption.

2.4.3.2 Matching Education and Wage Ratios

The goal of the calibration in this section is to try to match the black-white ratios of education levels and median income by education, using only the labor market discrimination parameter δ. As it happens, a value of $\delta = 0.14$ does the best in matching the two distributions separately, using a simple Euclidean norm of the differences. Both the data (from the CPS) and the model simulation are graphed in figure 2.9. In the education distribution (the top of figure 2.9), the model correctly predicts that few in the disadvantaged group will invest in the two highest levels of education. Because the marginal benefit obtaining more education is lower when δ is high, only those at the very top of the earning ability distribution will find higher

Figure 2.9: Target Moments and Calibration for Wage Discrimination



education worthwhile. The model is less successful in replicating the significantly lower fraction of black workers who have finished a four-year college degree. While the model predicts more black workers in the high school graduate and some college/Associate's degree categories, about the same fraction in both groups earn a bachelor's degree. The disadvantaged group does have more high school dropouts than in the data, although as discussed above, the data for full-time noninstitutionalized workers does not capture much of what is happening at the bottom of the black income/education distribution.

The model does a fairly good job of matching the median income levels by education group. Figure 2.9 shows that in the middle four education levels, the model is within four percentage points of the data. The wages for disadvantaged workers in the model are too low for the high school dropouts and too high for those with a PhD or professional degree. Note that the equilibrium median wage in the model is not simply 14% lower for the disadvantaged group at all education levels. The two highest education levels, for example, have equilibrium wages nine percent and eight percent lower than the advantaged group. This is because only disadvantaged workers with very high earning ability will find it worthwhile to invest in these education levels, so that the median earning ability at these education levels will be higher in the disadvantaged group than for the advantaged group.

2.4.3.3 Intergenerational Income Transition Rates

In computing the model's implications for intergenerational mobility, I must make an assumption on the share of disadvantaged agents in the labor force, since this share will affect the cutoffs for income quintiles. Black people make up approximately 12% of the total U.S. population. However, the non-Hispanic white population is only about 68%. I run two simulations, one with a disadvantaged population of about 12% and one with a share of 25%.

The first thing to note from the simulations is that, as in Malloy (2010), the model generates a positive correlation of (log) wages between parents and children despite the fact that earning ability is assumed to be independent across generations. For the advantaged group the correlation of income is 0.146 and for the disadvantaged group 0.138. The correlation of income pooling across groups is 0.164 in the simulation assuming that the disadvantaged group is 12% of the total and 0.173 in the simulation assuming a 25% share of disadvantaged workers. These numbers are well below the estimated intergenerational correlation of income of around 0.6 found by Mazumder (2005) and Mulligan (1997).

Table 2.5 shows the model's intergenerational transition rates between income quintiles, along with their empirical counterparts, found in Mazumder (2008). I present two measures of the poverty trap and affluence net. The first, Poverty Trap 20 and Affluence Net 20, measures the likelihood of staying in the bottom (top) 20

Table 2.5: Simulated Intergenerational Transition Rates with Wage Discrimination Only ($\delta = 0.14$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	25.5%	18.9%	17.6%	24.9%	19.6%
Poverty Trap 40	44.6%	67.6%	52.8%	42.9%	73.3%	49.0%	38.7%	69.8%	48.2%
Affluence Net 20	38.9%	21.3%	37.8%	45.5%	21.7%	44.5%	46.8%	30.2%	45.0%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	42.7%	50.8%	59.2%	41.9%	57.1%
Jump Up	10.6%	4.1%	7.4%	11.5%	5.3%	10.5%	13.8%	6.7%	11.8%
Fall Down	10.4%	21.6%	10.9%	17.0%	23.8%	17.3%	17.0%	22.3%	17.6%

* Mazumder (2008)

percent of the income distribution if you are born in the bottom (top) 20 percent. The second measure, Poverty Trap 40 and Affluence Net 40, give the same likelihoods if born in the bottom (top) 40% of the income distribution. I also present the Jump Up rate, defined as the fraction going from the bottom 20% to the top 20%, and the Fall Down rate, the fraction going from the top 20% to the bottom 20%.

As can be seen from the table, the simulation does fairly well in matching the transition rates. In particular, it replicates the fact that black children have a much higher fall down rate than jump up rate, as well as the fact that they face a larger poverty trap and smaller affluence net than white children. There are two major discrepancies between the model and the data. First, the model does a poor job of matching the fall down rate for the white population. In the simulation the fall down rate for the advantaged group is 17.0% as opposed to only 10.4% in the data. Second, the model does a much better job of matching the 40% measure of the poverty trap and affluence net than it does of matching the more narrow 20% measures. For example, the poverty trap for the disadvantaged group is only around 25% in the simulation when measured with just the bottom quintile, compared to 43.7% for black families in the data. However, the Poverty Trap 40 measure is 67.6% in the data for black children, while the model generates measures of 73.3% and 69.8% in the two simulations. Similarly for the advantaged group, the wider measure of the affluence net does a better job of matching the data than does the more narrow measure.

2.4.4 Lower Aspirations

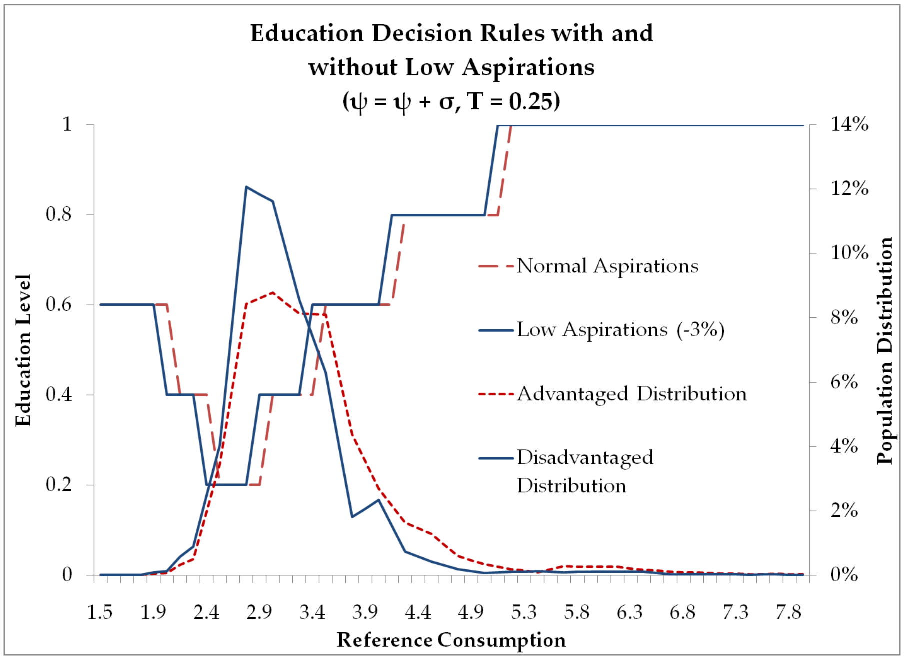
If people in the disadvantaged group have lower aspirations, represented by $a^D < 0$, then they will measure utility against a lower reference point than their (otherwise identical) peers in the group with standard aspirations ($a^A = 0$). The empirical rationale for including this possibility in the model was discussed in section 2.3 which indicated that black children born in the bottom half of the income distribution were less likely than their white counterparts to do better than their parents. If this expectation has been internalized within the culture, then it may be perfectly understandable that black students reduce their expectations of labor market success. This, in turn, will affect the equilibrium level of education they choose and will affect how likely they are to escape the poverty trap or be caught by the affluence net.

The model with $a^D < 0$ can capture the possibility that lower education levels among the black population are due to the effect of a different identity among black students and different aspirations because of this identity. In this version of the model, there is nothing intrinsically different in the disadvantaged population, such as lower earning ability or lower wages due to labor market discrimination, to make them choose less education. But because they self-identify with the group, they have lower expectations of success, reducing the equilibrium level of education.

2.4.4.1 Decision Rule

Figure 2.10 plots the education decision rule for agents with ability one standard deviation above the mean and a bequest of $T = 0.25$, both with normal aspirations and with low aspirations ($a^D = -0.03$). Lower aspirations have the same qualitative

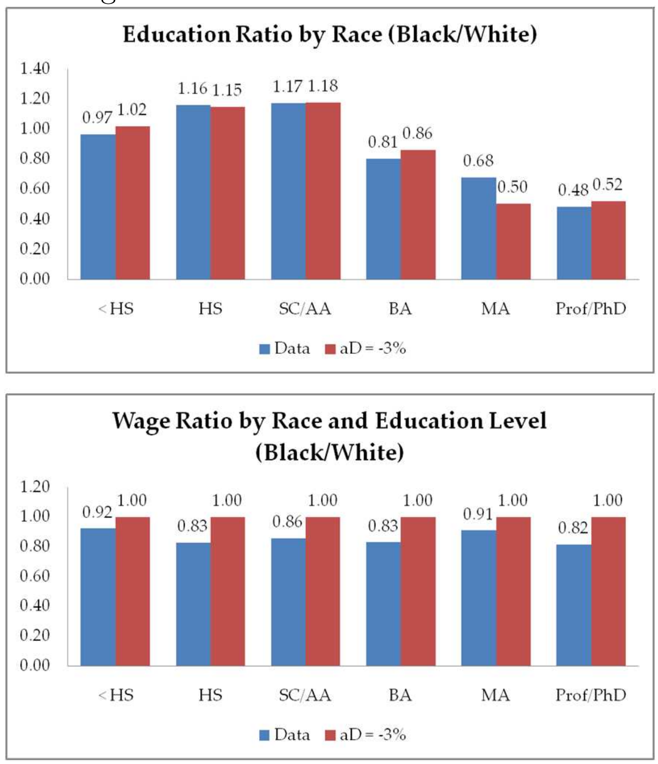
Figure 2.10: Education Decision Rule and Population Distribution with Low Aspirations



effect as wage discrimination, shifting the curve representing the utility-maximizing level of education to the left. At higher levels of reference consumption, those with lower aspirations will choose the same or higher levels of education. This is because it is easier for these people to avoid a first period loss than for those with the same earnings ability and bequest but with a higher level of reference consumption. Disadvantaged workers at the bottom of the consumption distribution, meanwhile, will choose the same or lower levels of education. The lower reference level of consumption means that they may be able avoid a loss in the first period with a lower level of education.

As with wage discrimination, the model's result that some students with lower aspirations are likely to get more education could help explain why some empirical studies find that, controlling for education ability and family background, black students choose more education than their white counterparts. But as we will see in the next section, the equilibrium result of low aspirations is that the disadvantaged group will wind up unconditionally with less education. Figure 2.10 shows this by

Figure 2.11: Target Moments and Calibration for Low Aspirations



plotting the distribution of each group by reference consumption. The distribution of the disadvantaged group is shifted to the left, although less so than with wage discrimination of 14% in figure 2.8.

2.4.4.2 Matching Education and Wage Ratios

This section searches for a value of a^D to match the distribution of education and median income for black workers. I find that $a^D = -0.03$ provides the best fit. Figure 2.11 presents the results for this case.

The calibration does fairly well at matching education levels with the exception of implying too few Master's degrees in the disadvantaged group. With lower aspirations, agents are able to consume at their reference level of consumption with lower levels of education. Thus, as the level of aspirations falls (or as a^D becomes more negative), fewer and fewer workers will invest in the highest levels of education. Table 6 shows

Table 2.6: Education Distribution for Aspiration Levels

Education Level	Data*	$a^D = 0\%$	$= -1\%$	$= -2\%$	$= -3\%$	$= -4\%$	$= -5\%$
High school dropout	7.7%	8.8%	9.0%	8.9%	9.0%	9.3%	9.1%
High School Diploma	33.2%	23.8%	25.3%	25.6%	27.3%	28.9%	45.3%
Some College/Associate's	32.7%	27.5%	29.5%	31.0%	32.4%	37.5%	31.1%
Bachelor's Degree	18.6%	31.0%	29.0%	28.3%	26.7%	21.9%	12.5%
Master's Degree	6.2%	7.2%	5.7%	5.0%	3.6%	2.0%	1.7%
PhD/Professional Degree	1.7%	1.7%	1.6%	1.3%	1/0%	0.4%	0.4%

* Black, full-time workers, 25-64 years old, CPS

the distribution of education for $a^D = -1\%, -2\%, ..., -5\%$. The share of high school dropouts stays stays roughly constant as a^D becomes more negative, but the level of Bachelor's degrees, Master's degrees, and PhD/Professional degrees fall steadily. As a result, the shares of workers with only a high school degree or some level of college short of a Bachelor's degree increase. One possibility, not currently explored in this paper, is that the degree of lower aspirations, or the value of a^D, varies by reference consumption level. It could be that those in the disadvantaged group at the top of the consumption distribution have only slightly lower levels of aspirations than the advantaged group while those at the bottom of the distribution have much lower levels of aspirations. Allowing for heterogeneity in aspiration levels may enable the model to do a better job in explaining the education distribution of black full-time workers.

While the model with lower aspirations can match racial differences in education, it completely fails in matching the median income by education level. Because there is no wage discrimination in this simulation, the only way to change the median income level is to significantly change the ability of those who are investing in each education level. That is, unless the median earning ability of the median worker at each level of education changes, there will be no change in the median income. When aspirations are lower by 3%, as in figure 2.11, we can see that wages for the disadvantaged group are the same as those in the advantaged group. While this is not true for all levels

Table 2.7: Simulated Intergenerational Transition Rates with Low Aspirations Only ($a^D = -0.03$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	18.0%	17.6%	17.6%	18.5%	17.8%
Poverty Trap 40	44.6%	67.6%	52.8%	46.7%	48.8%	47.0%	46.7%	50.6%	47.6%
Affluence Net 20	38.9%	21.3%	37.8%	46.1%	39.0%	45.5%	45.5%	39.6%	44.5%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	49.6%	51.2%	51.4%	49.2%	51.0%
Jump Up	10.6%	4.1%	7.4%	10.4%	10.1%	10.3%	11.5%	10.0%	11.2%
Fall Down	10.4%	21.6%	10.9%	17.2%	18.5%	17.3%	17.0%	16.9%	17.0%

* Mazumder (2008)

of a^D, the median wage is just as likely to be (slightly) higher as it is to be (slightly) lower. This makes complete sense given that the underlying distribution of earnings ability is the same in each group. In order to explain the consistently lower wages for black workers at every level of education, we will have to look elsewhere.

2.4.4.3 Intergenerational Income Transition Rates

Despite the fact that the model with only lower aspirations is unable to match the median income by education level, it still is able to generate slight differences in intergenerational transition rates between groups that are in the right direction. This is because the lower aspirations cause the education distribution for the disadvantaged group to shift to the left due to the leftward shift in reference consumption plotted in figure 2.10. This shift makes it more likely that some disadvantaged agents born in the bottom 20% or 40% of the income distribution will stay poor as agents with low reference consumption will invest in less education than they would with standard aspirations as shown in figure 2.10. Table 2.7 gives the intergenerational transition rates when aspirations are lower for the disadvantaged group by 3%. The disadvantaged group has a (slightly) higher poverty trap and affluence net than the advantaged group in the simulations, but not by enough to match the data.

Table 2.8: Simulated Intergenerational Transition Rates with Low Aspirations Only ($a^D = -0.05$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	16.5%	17.4%	17.6%	17.0%	17.4%
Poverty Trap 40	44.6%	67.6%	52.8%	42.9%	64.8%	46.9%	40.6%	65.7%	48.4%
Affluence Net 20	38.9%	21.3%	37.8%	45.5%	26.7%	44.1%	45.5%	31.8%	43.7%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	43.4%	50.9%	59.2%	51.0%	58.1%
Jump Up	10.6%	4.1%	7.4%	11.5%	10.5%	11.4%	11.5%	10.2%	11.2%
Fall Down	10.4%	21.6%	10.9%	17.0%	17.1%	17.2%	17.0%	17.7%	17.1%

* Mazumder (2008)

Table 2.8 provides the transition rates when aspirations for the disadvantaged group are lower by 5%. From table 2.6, this level of a^D generates too large a shift in the education distribution, especially at the higher levels of education. From table 2.8, however, this level of lower aspirations does a much better job of matching the poverty trap for black children born in the lower 40% of the income distribution. It also reduces the affluence net for the disadvantaged group born in the top 20% of the income distribution, although not far enough to match the data. Overall, this table does lend support to the idea that lower aspirations may be one of the factors driving the poverty trap among black families.

2.4.5 Lower Earning Ability

Some studies find that black students do not have the same skills as their white counterparts. For example, Neal and Johnson (1996) find that the distribution of scores for black students in the Armed Forces Qualification Test (AFQT) is shifted down compared to their white counterparts. One plausible explanation is that black students get lower quality early education than their white counterparts. As discussed in section 2.3, Card and Krueger (1992) find that the reduction in the black-white wage gap from 1960 to 1980 can partially be explained by an increase in the quality of education for black students. Maxwell (1994) finds that controlling for the quality of education can explain much of the black-white wage gap.

In this section, I examine a version of the model that allows for a different distribution of earning ability, ψ, between the two groups. I allow for three possible differences in the distribution: a mean shift; a shift in the maximum level; and a shift in the minimum level. A shift in the mean would correspond to a ubiquitous difference in the quality of education between black and white students. A decrease in only the minimum level would represent the idea that the main cause of black-white differences is the very bottom of the earning ability distribution. Regardless of the type of shift, the assumption of independence of earning ability across generations ($\rho = 0$) is maintained in these experiments.

2.4.5.1 Decision Rule

Perhaps the most important thing to note when the distribution of earning ability is different between the two groups is that such differences will have no effect whatsoever on the education decision rule. That is, for a given earning ability, bequest, and reference level of consumption, the chosen level of education will be exactly the same whether a person is in the advantaged group or the disadvantaged group. The major change, discussed in the next section, will be the equilibrium distribution of education. If there are fewer people at the high end of earnings ability, there will be fewer people who choose to invest in professional degrees or PhDs. If there are more people at the lower end of the earning ability distribution, there will be more high school dropouts. But no matter what the education distribution, the wages of each worker will still depend only on education and earning ability, not on the group to which they belong.

This result of the model should be entirely obvious once stated. In this model there is no reason for someone from the disadvantaged group (with the same reference consumption and bequest) to get more or less education as compared to someone from the advantaged group with the same earning ability. So long as there is no wage discrimination, wages of each group at each education level will be the same, and the only thing that will change is the distribution of education levels within each population.

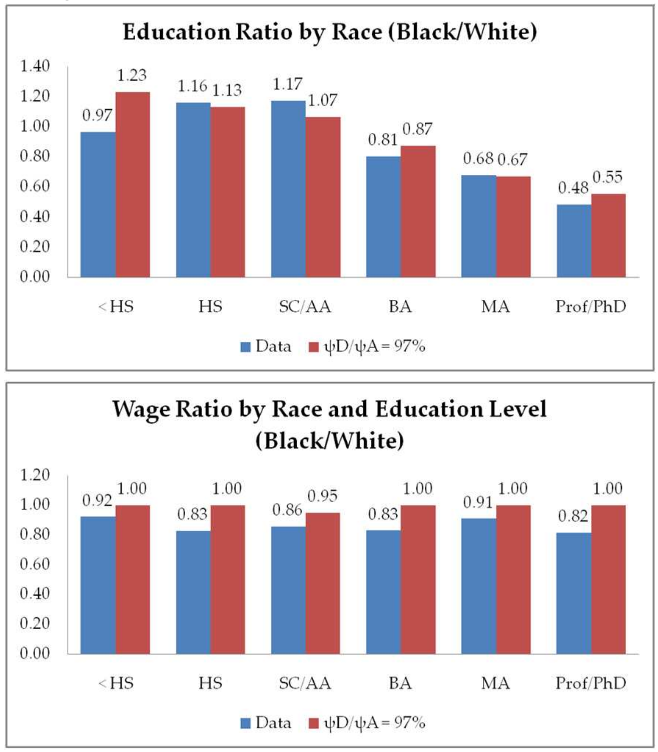
2.4.5.2 Matching Education and Wage Ratios

The distribution of ψ^D that does the best job of matching the target moments for the black full-time working population is one that drops the maximum value earnings ability by approximately 6%, keeps the minimum value the same, and therefore reduces the mean ability by approximately 3%. The calibration results are graphed in figure 2.12. The major discrepancy compared to the observed education distribution is that there are more high school dropouts. As explained above, this may not truly be a weakness given the number of black people who are not in the labor force and do not have a high school degree. Although the distribution of ψ for disadvantaged workers has the same minimum as for advantaged workers, there is still a fatter left hand tail of the distribution for disadvantaged workers as the standard deviation stays the same for the two groups so that more disadvantaged agents in the model will hit up against the minimum[3]. Anybody to the left of a certain value of ψ will find it optimal to drop out of high school and start work as soon as possible.

This simulation, similar to the case with lower aspirations only, does a poor job of matching the median income by education level. As show in figure 2.12, the median income levels are almost exactly the same in both the advantaged and disadvantaged groups. The only difference is a slight decrease at the some college/Associate's degree level because of a slightly larger population in that group that is shifted slightly to the left in terms of earning ability due to the different distribution. Simulations with other leftward shifts in the distribution of earning ability for the disadvantaged group

[3]This is without loss of generality as the cutoff for those who find it optimal to not invest in any additional education is above the minimum value of ability.

Figure 2.12: Target Moments and Calibration for Lower Earning Ability



showed similar results on relative wages, sometimes with an education group having slightly higher median income, sometimes slightly lower, but usually the same.

These simulation results challenge the findings discussed above that attribute much of the black-white wage gap to lower ability in black workers, whether due to unobserved cognitive ability, as in Urzua (2008), or lower educational quality, as in Maxwell (1994). The only way that lower earning ability could reduce the median income at a given education level in this model would be for those with lower earning ability in the disadvantaged group to get *more* education than their counterparts in the advantaged group. In this case, the wage would be lower, controlling for education, for the disadvantaged group because they would have lower inherent earnings ability for a given level of education. But this can only be true if the education decision rule differs by group, perhaps because of different aspirations.

Table 2.9: Simulated Intergenerational Transition Rates with Lower Earning Ability Only ($\frac{\bar{\psi}^D}{\bar{\psi}^A} = 0.97$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	19.9%	17.9%	17.6%	19.6%	18.1%
Poverty Trap 40	44.6%	67.6%	52.8%	46.7%	53.4%	47.6%	46.7%	53.6%	48.4%
Affluence Net 20	38.9%	21.3%	37.8%	46.1%	39.8%	45.5%	45.5%	41.5%	44.8%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	46.3%	50.9%	51.4%	46.2%	50.4%
Jump Up	10.6%	4.1%	7.4%	10.4%	8.4%	10.1%	11.5%	9.1%	10.9%
Fall Down	10.4%	21.6%	10.9%	17.2%	20.4%	17.5%	17.0%	21.2%	17.8%

* Mazumder (2008)

2.4.5.3 Intergenerational Income Transition Rates

Shifting the distribution of the earning ability of the disadvantaged group to the left implies a larger population that will have low earnings, especially at the bottom two levels of education. This shift will also reduce the number at the top who will find it optimal to invest in the highest levels of education. Table 2.9 shows the intergenerational transition rates when disadvantaged workers have a mean earning ability three percent lower than advantaged workers. As expected, the poverty trap is higher for the disadvantaged group, although it is not as high as in the data or as in the simulation with only wage discrimination (table 2.6). In addition, both measures of the affluence net are lower for the disadvantaged group, as is the jump up rate. However, none of the measures are as close to the data as in the simulation with only wage discrimination and the model is unable to match the median income levels by education. These results, along with the results for median income by education level, would seem to question whether or not lower quality primary and secondary education (or anything else that affects earning ability during childhood) is really the driving force behind the black-white wage gap.

2.4.6 Allowing More than One Difference Between Groups

The goal of this section is to see if the model can more accurately match the target moments using either two or all three of the group differences in combination. Given the data discussed in section 2.3 there is reason to think that the black-white wage gap may have more than one cause. It's possible, for example, that wage discrimination could lead to lower aspirations as members of the disadvantaged group see that they are not treated fairly in the labor market. This would not be completely irrational as membership in the disadvantaged group is passed down from parent to child so that both face the same level of labor market discrimination. However, if shared experience of discrimination leads to a group identity, as in Akerlof and Kranton (2000), then it may also lead to lower aspirations and a defining down of success.

It is also possible that members of the disadvantaged group receive lower quality education and are, on average, less productive than member of the advantaged group. In this case, it is possible that employers use a rule of thumb when setting wages that leads to labor market discrimination. In this case, the discrimination would not be rational. As shown in the previous section, when investment in education is endogenous and there is no labor market discrimination, members of the disadvantaged group who invest in a particular level of education will be virtually identical in terms of ability to those in the advantaged group who choose the same level of education. But if employers do not understand this and there is a pervasive belief that the disadvantaged group is less productive because they receive lower quality education, this belief could be enough to sustain some level of wage discrimination.

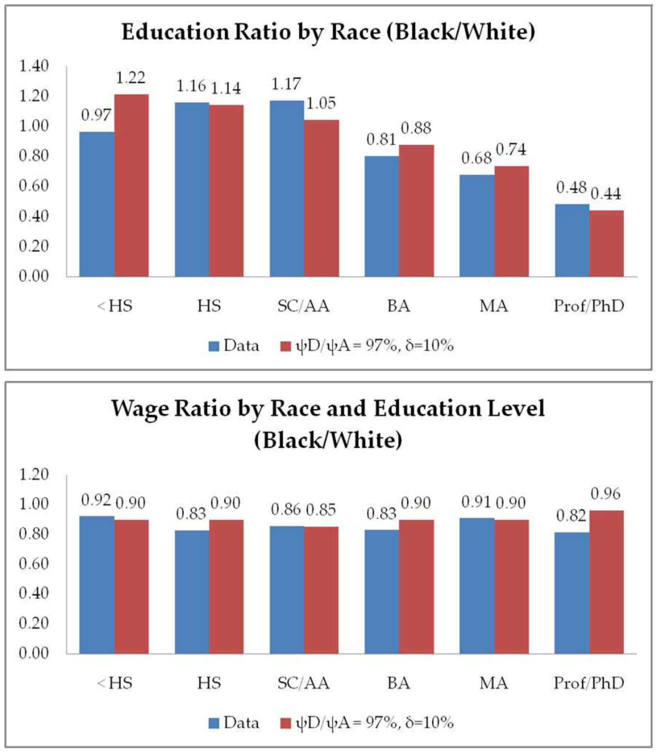
2.4.6.1 Calibration

A model without wage discrimination, so that $\delta = 0$, but with lower aspirations and a lower average earning ability for the disadvantaged group does no better in matching the target moments than a model with only one or the other. This should not be surprising as both of these versions of the model fail to match the lower median income levels for all education groups for black workers that we see in the data. Both of these mechanisms push the disadvantaged group to invest less in education so that the model can do a fair job in matching the education distribution of black workers, but they cannot account for the lower income levels conditioned on education. In order for the model to match relative incomes, we need a version with wage discrimination.

The calibrated model that allows for a shift in the earning ability of the disadvantaged group and wage discrimination does a better job of matching the target moments than the version with only the shift in earning ability. These results are presented in figure 2.13. This figure shows results for a model assuming a drop in the minimum value of ψ^D of 6%, so that the average is reduced by approximately 3%, and a value of $\delta = 10\%$. There are two things of interest to note here. First, this version does no better than the experiment that relied solely on wage discrimination. In fact, its fit (measured by the norm of the differences between the data and the simulation) is marginally worse. Second, the level of wage discrimination that provides the best fir is actually less than in the calibration with only wage discrimination ($\delta = 10\%$ as opposed to $\delta = 14\%$). The lower earning ability in some sense reinforces the wage discrimination so that we need lower levels of overt wage discrimination to match the target moments from the data.

A model with wage discrimination and lower aspirations (but no shift in the distribution of earning ability) does a better job of matching the target moments than any of the other versions that rely on one or two difference mechanisms. The target moments with values of $\delta = 11\%$ and $a^D = -2\%$ are presented in figure 2.14. The main weakness in this experiment is that it implies relatively too many people in the disadvantaged group investing in a Bachelor's degree and too few in some level of college less than a BA. But once again the calibration shows that less overt

Figure 2.13: Target Moments and Calibration for Lower Earning Ability and Wage Discrimination



discrimination is needed to fit the data when one of the other mechanisms in the model works along with discrimination.

Finally, I use all three parameters available in the model to see if I can improve on the results using just wage discrimination and aspirations. The parameter values that do the best job of matching the target moments are a wage discrimination value of 12%, lower aspirations of 1%, and lower earning ability of 1.6%. Interestingly, this calibration is actually slightly worse than the version with only wage discrimination and low aspirations. It does a slightly better job of matching the median income levels, but not as good a job of matching the education distribution. Results are presented in figure 2.15.

Figure 2.14: Target Moments and Calibration for Lower Aspirations and Wage Discrimination

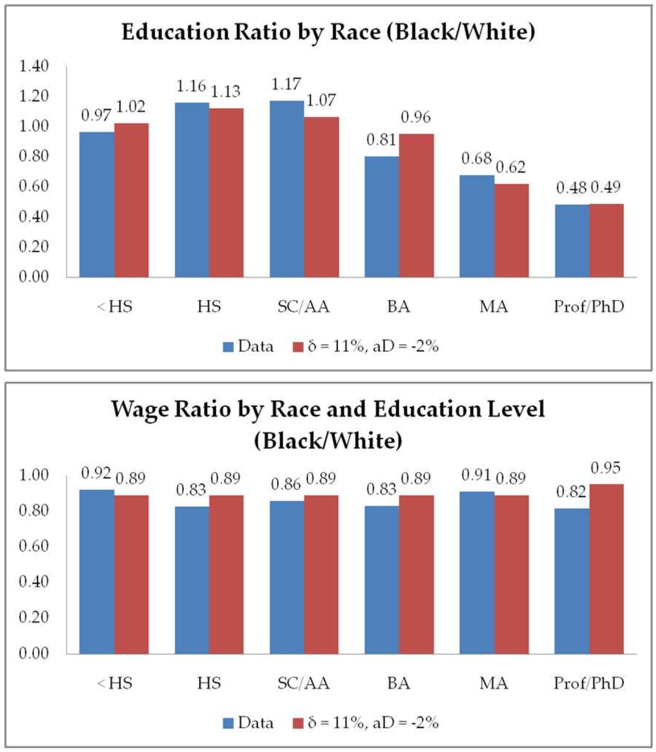


Figure 2.15: Target Moments and Calibration for Lower Aspirations, Lower Ability, and Wage Discrimination

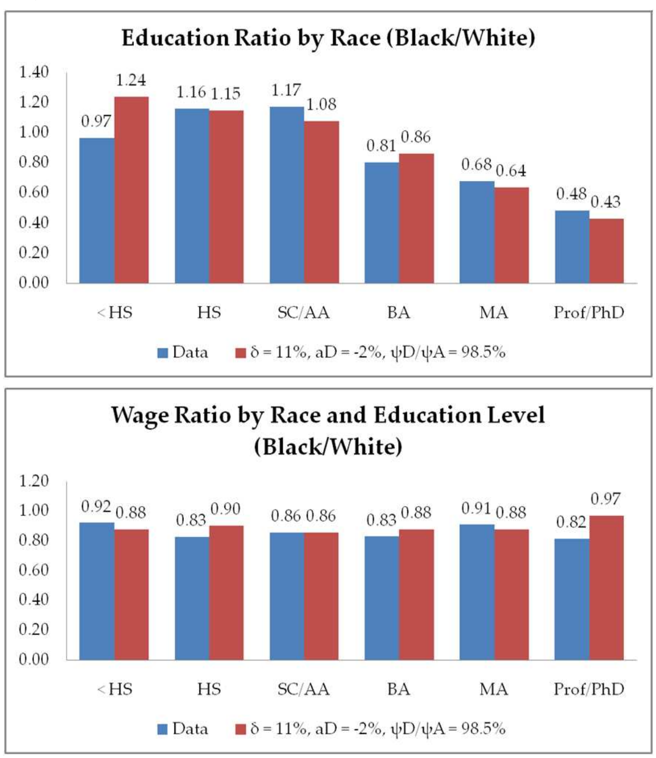


Table 2.10: Simulated Intergenerational Transition Rates with Wage Discrimination ($\delta = 0.10$) and Lower Earning Ability ($\frac{\bar{\psi}^D}{\bar{\psi}^A} = 0.97$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	25.8%	19.0%	17.6%	26.5%	20.3%
Poverty Trap 40	44.6%	67.6%	52.8%	42.9%	74.4%	49.1%	40.6%	68.1%	49.3%
Affluence Net 20	38.9%	21.3%	37.8%	45.4%	23.0%	43.3%	46.8%	32.5%	45.2%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	46.7%	51.1%	59.2%	43.8%	57.1%
Jump Up	10.6%	4.1%	7.4%	11.5%	5.6%	10.4%	13.8%	6.6%	11.6%
Fall Down	10.4%	21.6%	10.9%	17.0%	25.3%	17.5%	17.0%	26.9%	18.1%

* Mazumder (2008)

2.4.6.2 Intergenerational Income Transition Rates

The simulations with more than one factor do a fairly good job of matching the intergenerational transition rates between income quintiles. Specifically, they increase the poverty trap of the bottom quintile for the disadvantaged group (although none of the models are able to generate a high enough poverty trap for the bottom 20% to match the data). The multi-factor models are also able to reduce the affluence net, reduce the jump up rate, and increase the fall down rate for the disadvantaged group. However, these results are only marginally better (in terms of fit measured by the norm of the differences) than the simulation with only wage discrimination. Table 2.10 gives the transition rates for the model with wage discrimination and a shift in the earning ability of the disadvantaged group. The fall down rate is actually a bit high, although that seems to be a general weakness of the model as it currently stands. The model is still not able to generate a small enough affluence net for those in the disadvantaged group born in the top 40% of the income distribution. Such a low affluence net for black families seems to be one of the more curious, and important, aspects of the data[4].

The transition rates for the simulation using both wage discrimination and lower

[4] While the affluence nets for black families in Mazumder (2008) using the NLSY seem very small, Hertz (2005) found even smaller affluence nets using the PSID.

Table 2.11: Simulated Intergenerational Transition Rates with Wage Discrimination ($\delta = 0.11$) and Low Aspirations ($a^D = -0.02$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	21.5%	18.2%	17.6%	26.9%	20.1%
Poverty Trap 40	44.6%	67.6%	52.8%	42.9%	72.1%	48.8%	40.6%	63.5%	47.6%
Affluence Net 20	38.9%	21.3%	37.8%	45.5%	31.2%	44.8%	46.8%	33.4%	45.5%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	41.3%	50.8%	59.2%	43.0%	57.3%
Jump Up	10.6%	4.1%	7.4%	11.5%	5.0%	10.5%	13.8%	6.6%	11.8%
Fall Down	10.4%	21.6%	10.9%	17.0%	23.9%	17.4%	17.0%	21.4%	17.4%

* Mazumder (2008)

Table 2.12: Simulated Intergenerational Transition Rates with Wage Discrimination ($\delta = 0.12$), Low Aspirations ($a^D = -0.02$), and Lower Earning Ability Only ($\frac{\bar{\psi}^D}{\bar{\psi}^A} = 0.984$)

	Data*			Simulation (D = 12%)			Simulation (D = 25%)		
Transition Rate	White	Black	Total	Adv.	Disadv.	Total	Adv.	Disadv.	Total
Poverty Trap 20	24.9%	43.7%	33.5%	17.6%	26.5%	19.2%	17.6%	26.6%	20.3%
Poverty Trap 40	44.6%	67.6%	52.8%	42.9%	71.8%	48.6%	40.6%	64.6%	48.2%
Affluence Net 20	38.9%	21.3%	37.8%	45.5%	21.3%	44.5%	46.8%	27.2%	45.0%
Affluence Net 40	54.3%	31.3%	52.8%	51.4%	43.4%	50.9%	59.2%	43.7%	57.4%
Jump Up	10.6%	4.1%	7.4%	11.5%	4.7%	10.3%	13.8%	6.0%	11.4%
Fall Down	10.4%	21.6%	10.9%	17.0%	29.8%	17.6%	17.0%	27.2%	17.9%

* Mazumder (2008)

aspirations are presented in table 2.11. There is, in fact, very little difference between these rates and those in table 2.10. Wage discrimination appears to be the main driving force behind intergenerational mobility in this model. The transition rates with all three factors are given in table 2.12.

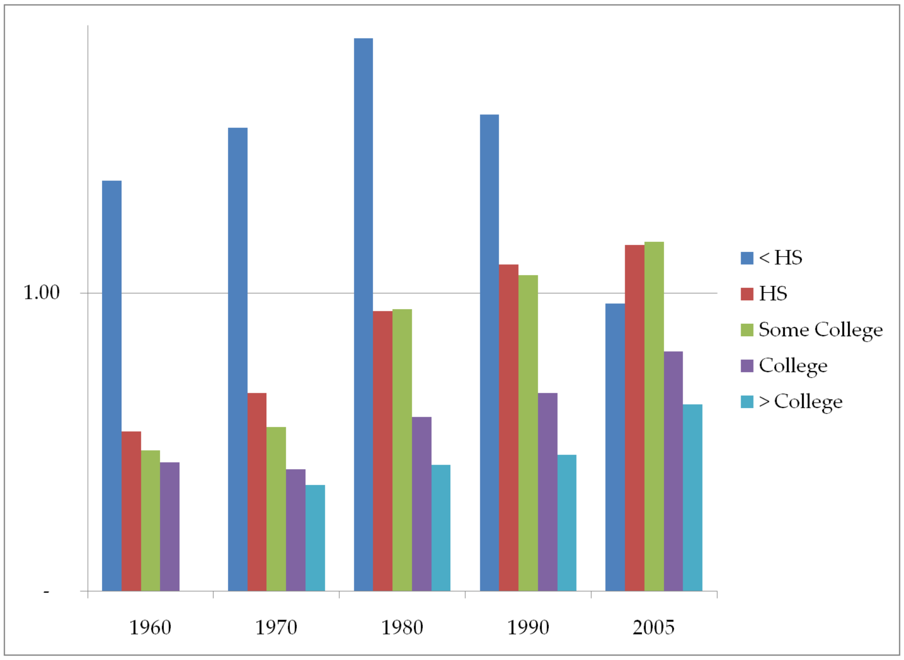
2.5 Generational Transitions

Between the implementation of the Jim Crow system of segregation at the end of the nineteenth century and the beginning of the Civil Rights movements, many black workers faced both severe labor market discrimination and a lower quality education system than their white counterparts. As noted above, Card and Krueger (1992) found that the black-white wage gap was about 40% in 1960 in the midst of the Civil

Rights movement. This had fallen to 25% by 1980 and is currently about 14% today (controlling only for the level of education), according to the CPS. Education rates for black and white workers were quite different in 1960 than they are today, both in absolute terms and in relative terms. The black-white education gap according to the U.S. Census from 1960-2005 is graphed in figure 2.16. One of the reasons for the relative difference was no doubt discrimination. Black students in many areas of the country simply did not have easy access to post-secondary education. Another reason, documented in Card and Krueger (1992), was the fact that the quality of education, especially in the segregated South, was worse for black students than white students. Finally, given that the wage gap was so high, there was less reason for black students to continue to invest in education as the marginal benefit of doing so was lower. One peculiarity of the data is that the ratio of black high school dropouts to white high school dropouts actually increased after the Civil Rights legislation of the 1960s, which made it illegal to discriminate in the labor market based on race. While the drop out rate for workers from both groups fell (from 56.8% to 17.4% for white workers and from 78.3% to 32.3%) from 1960 to 1980, it fell much faster for white workers, so that the ratio rose from 1.38 in 1960 to 1.86 in 1980.

If lower aspirations do indeed play a role in the level of education investment, then even in the rosiest post-Civil Rights scenario the education distribution for black and white workers would not equalize in the first generation. One reasonable question is how many generations it would take for black workers to invest in the same levels of education as white workers if labor market discrimination went from 40% to 0% after the civil rights legislation in the 1960s. It's difficult to accurately estimate the value of labor market discrimination that would match the education distribution in the 1960s as the underlying wage function, and especially the return to education, has changed so substantially. But we can ask, if the civil rights legislation was completely successful in wiping out labor market discrimination and reduced overt discrimination

Figure 2.16: Black-White Education Ratio, 1960-2005 (1.00 = Equality)

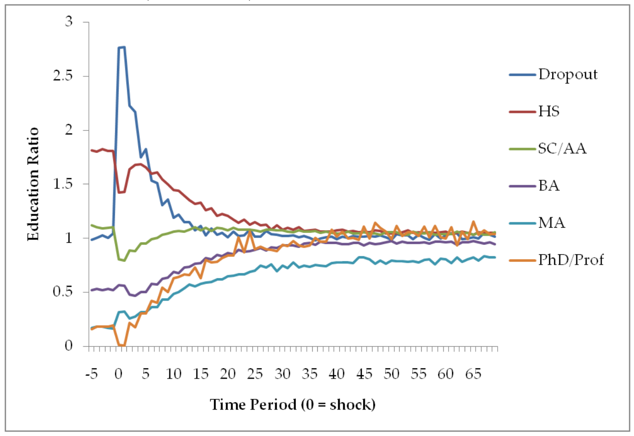


Source: Current Population Survey, U.S. Census

to zero, how long would it take for the education distribution of black workers to equalize to that of white workers?

I assume that the economy starts in a steady state with wage discrimination of 40% ($\delta = 0.4$ in the model) and then experiences an immediate and permanent decline in δ to zero. At first, I assume that aspirations are constant across the two groups so that $a^A = a^D = 0$. With loss aversion and reference consumption that looks to the previous generation, the elimination of discrimination has some unexpected consequences. Because the marginal benefit of each level of education has increased so dramatically, many agents in the disadvantaged group can now avoid a loss by choosing a lower amount of education. The relative number of dropouts, compared to the advantaged group, spikes in the first generation after the wage discrimination is eliminated and then begins slowly to fall to that of the advantaged group. There is a corresponding fall and then increase in the higher levels of education. This matches what we see in figure 2.16, but the process takes quite a long time in the model.

Figure 2.17: Disadvantaged-Advantaged Education Ratios Before and After Elimination of Wage Discrimination ($\delta = 0.4$)

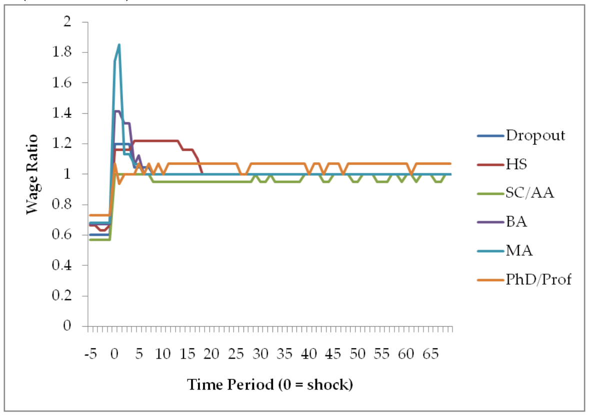


It takes 16 generations for the level of Bachelor's degrees and doctoral degrees in the disadvantaged population to equal that of the advantaged group. The transition process is pictured in figure 2.17 which graphs the disadvantaged-advantaged ratio of the percentage of each group at each education level.

The other question in the transition is what happens to median wages at each education level. Interestingly, in the model, the median wage of the disadvantaged group becomes immediately higher than that of the advantaged group at most education levels. This is because of the change in the ability ranges of the disadvantaged group that invest in each level of education. For example, 24% of the population decides to invest in the lowest level of education immediately following the shock. In this population are workers who have much higher inherent earning ability than the 8-9% of the advantaged population that invests in this level of education. It takes six generations for the median income levels of each group to be within 5% of each other, but this is because the median wages of the disadvantaged group need to come *down*, rather than up. This process is pictured in figure 2.18.

The above simulations assume that aspirations are the same for both the advantaged and disadvantaged group. The reference consumption level is inherited from the previous generation, so reference consumption is lower initially for the disadvantaged

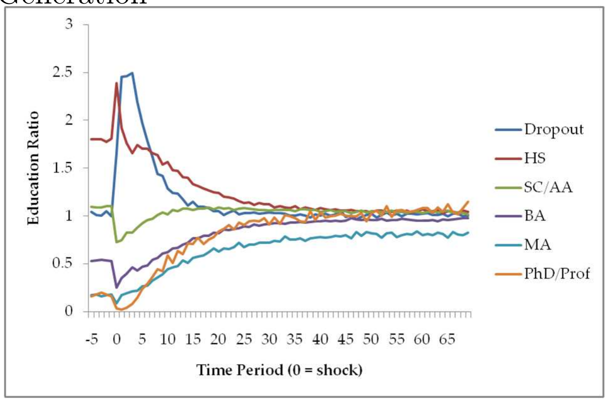
Figure 2.18: Disadvantaged-Advantaged Wage Ratio Before and After Elimination of Wage Discrimination ($\delta = 0.4$)



workers due to the legacy of lower wages, but there are no permanent differences in aspirations. If lower aspirations become entrenched in the disadvantaged group, then the results in section 2.4 imply that the education distribution will never equalize between the groups. When aspirations are 3% lower for the disadvantaged group, it takes 43 generations for the disadvantaged group to reach 80% of the share of Bachelor's degrees in the advantaged group, and the share of PhD/professional degrees for disadvantaged workers never reaches 80% of the share of advantaged workers. On the other hand, there is little change in how median wage levels equalize. Within six generations, the median wage levels of the disadvantaged group come down to within 5% of the advantaged group.

Of course, it is unlikely that wage discrimination would disappear overnight. In addition, some of the black-white wage gap in 1960 was no doubt due to the lower quality of education for black students, as claimed by Card and Krueger (1992). Both the quality of education and cultural norms that lead to wage discrimination take time to change. One useful exercise would be to see how long it would take to reach equality of median income by education levels and the distribution of education, assuming a given half-life of discrimination. For instance, if the level of wage discrimination is 40% in generation 0, 20% in generation 1, and 10% in generation 2, then the half-life

Figure 2.19: Disadvantaged-Advantaged Education Ratio with Half-Life of Wage Discrimination of One Generation



of discrimination is one generation. The level of discrimination decreases by 50% every generation. If the half life is more than one generation it will take longer to reach equality whereas if it is less we will reach equality more quickly. When Card and Krueger were writing in the early 1990s, there appeared to be some level of optimism as the black-white wage gap had fallen fairly substantially in the twenty years from 1960 to 1980, partly due to increases in the quality of education. Now there seems to be less cause for optimism. For example, there is very little difference in the current black-white wage gaps for 25-34 year olds and for 35-44 year olds. And both of those wage gaps are only slightly lower than that of 45-54 year olds.

I simulate the model assuming a level of wage discrimination starting at 40% which decreases by half over the next four generations so that it is 20% in generation 0, 10% in generation 1, 5% in generation 2, and 0% in generation 3. The education convergence rate, pictured in figure 2.19, is somewhat slower than in the model in which wage discrimination disappears immediately. It takes 19 generations for the Bachelor's degree rate of the disadvantaged group to come within 80% of the advantaged group and 20 generations for the doctoral degree rate to do the same. Interestingly, after 70 generations, there is still a fairly significant difference in the education distribution, as the disadvantaged group has 15% more doctoral degrees and 18% fewer Master's degrees.

Wages, perhaps surprisingly, behave in much the same way as in the model in which discrimination was immediately eliminated. The only significant difference is that because wage discrimination is cut in half in each generation, the education distribution is different so that there is a spike first in high school graduates and college graduates, pushing up the median wage in these two groups more quickly. In period zero, when overt wage discrimination is still 20%, only the groups of high school dropouts and those with some college have median earnings less than 97% of the advantaged group. In period 2 when the discrimination is 5%, the median earnings of the disadvantaged group are higher at each education level than that of the advantaged group due to the shift in who is investing in each level of education.

Cutting wage discrimination in half over each generation may strike some as overly optimistic (and perhaps others as overly pessimistic). I next simulate the model with wage discrimination decreasing by 10% each generation for four generations. The dynamics are fairly close to the half-life model. The only difference is that the slower pace delays the spike in high school dropouts for an extra generation. This leads to lower initial levels of high school dropouts in the first two generations, but higher levels after that. After about ten generations, the simulations are virtually identical. Because labor market discrimination is decreasing more slowly, median wage levels take longer to equalize. This means lower initial median earnings for the disadvantaged group, especially at the lower levels of education. It takes another generation for these groups to pass the median earnings of the advantaged group due to the shift in the population that is investing in each level of education.

2.6 Conclusion

The model presented here does a good job of matching the education distribution of black workers and a fair job matching the intergenerational transition rates between income quintiles. The model struggles in simulating the very large poverty trap experienced by black families in the lowest quintile of the income distribution. One possibility suggested by the simulations is that those born into the bottom quintile have much lower aspirations than those born above them in the income distribution. This would suggest that the existence of a large poverty trap is a self-fulfilling cycle, needing more creative policies to help eliminate the cycle of poverty. If low aspirations are truly to blame, increasing the aspirations among students at all levels and making higher education more accessible would help reduce the poverty trap.

All three mechanisms in the model, wage discrimination, low aspirations, and lower earning ability, are able to shift the education distribution and do a fair job of matching the black-white education gap that we see in the data. However, only wage discrimination can explain the persistent black-white wage gap that we see at every education level in the data. Because education is treated as an endogenous variable so that each agent chooses the level of education to invest in to maximize utility, we do not see persistently lower wages controlling for education level when the model has only low aspirations and/or lower earning ability.

It could be argued that the black-white education gap and black-white earnings gap are simply the result of transitioning from a system of overt wage discrimination to one of equality. While simulations of this transition in the model lend some credence to this possibility for the education distribution, the results are much less plausible for median wage levels. For example, the model may help explain why we see a spike in the black-white ratio of high school dropouts in the generation after the adoption of Civil Rights legislation. However, in this model, once discrimination is eliminated in the labor market, median wage levels for each group either equalize, or if there is a large shift in the education distribution of the disadvantaged group due to the role of reference consumption, invert, so that the median income of the disadvantaged group is actually higher. These transitions towards equality in the model are so slow that they suggest the model may not be capturing how reference levels are truly formed. The generation of black workers after the end of Jim Crow would not have made such large gains as they did if they had been satisfied with the place their parents had occupied in the income distribution.

Before I would feel comfortable concluding that the black-white wage gap is due solely to wage discrimination, there are at least two mechanisms outside of this model that deserve further exploration. The first, popular in the labor market literature explaining sex discrimination, is the possibility that prejudice or preference leads black workers into different fields of study than their white counterparts. If these different fields have significant pay differences then it may not be overt labor market discrimination that is leading to the black white-wage gap. This doesn't mean that prejudice or discrimination may not be the root cause, but it may mean the policy solution is much more complicated.

The second possibility is that black students do in fact receive a lower quality education, but that this does not generate a shift in earning ability as measured by the parameter ψ in the model, but rather changes the return to higher education, as captured by the parameters α_1, α_2, and ζ in the wage function. It is possible, in this case, that we could see some level of a black-white wage gap when there is a shift in the underlying distribution of earning ability. I leave this possibility for future work.

Bibliography

[1] Akerlof, George A. and Rachel E. Kranton. "Economics and Identity." *The Quarterly Journal of Economics*, vol. 115, no. 3 (2000): 715-753.

[2] Austen-Smith, David and Roland G. Fryer, Jr. "An Economic Analysis of "Acting White."" *The Quarterly Journal of Economics*, vol. 120, no. 2 (2005): 551-583.

[3] Babcock, Bruce A., E. Kwan Choi, and Eli Feinerman. "Risk and Probability Premiums for CARA Utility Functions." *Journal of Agricultural and Resource Economics*, vol. 18, no. 1 (1993): 17-24

[4] Becker, Gary, S. and Nigel Tomes. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*, vol.4, no. 3 (1986): S1-S39.

[5] Bennett, Pamela R. and Amy Lutz. "How African American Is the Net Black Advantage? Differences in College Attendance Among Immigrant Blacks, Native Blacks, and Whites." *Sociology of Education,* vol. 82, no. 1 (2009): 70-100.

[6] Bowles, Samuel, Herbert Gintis, and Melissa Osborne Groves, *eds.* Unequal Chances: Family Background and Economic Success. New Jersey: Princeton University Press (2005).

[7] Bowman, David, Deborah Minehart, and Matthew Rabin. "Loss aversion in a consumption-savings model." *Journal of Economic Behavior & Organization*, vol. 38 (1999): 155-178.

[8] Brown, Ryan et al. "Family and Community Influences on Educational Outcomes Among Appalachian Youth." *Journal of Community Psychology*, vol. 37, no.7 (2009): 795-808.

[9] Card, David, and Alan B. Krueger. "School Quality and Black-White Relative Earnings: A Direct Assessment." *The Quarterly Journal of Economics*, vol. 107, no.1 (1992): 151-200.

[10] Connor, Helen. "Deciding For or Against Participation in Higher Education: the Views of Young People from Lower Social Class Backgrounds." *Higher Education Quarterly*, vol. 55, no. 2, (2001): 204-224.

[11] Cook, Philip J. and Jens Ludwig. "Weighing the "Burden of 'Acting White'": Are There Race Differences in Attitudes toward Education?" *Journal of Policy Analysis and Management*, vol. 16, no. 2 (1997): 256-278.

[12] De la Croix, David. "Growth dynamics and education spending: The role of inherited tastes and abilities." *European Economic Review,* vol. 45, (2001): 1415-1438.

[13] Dearden, Lorraine, et al. "Education Subsidies and School Drop-out Rates." *The Institute for Fiscal Studies,* WP05/11 (2007).

[14] Fordham, Signithia and John U. Ogbu. "Black students' school success: Coping with the "burden of 'acting white.""" *The Urban Review*, vol. 18, no. 3 (1986): 176-206.

[15] Genicot, Garance and Debraj Ray. "Aspirations, Inequality, Investment and Mobility." *Working Paper,* (2009).

[16] Grodsky, Eric and Devah Pager. "The Structure of Disadvantage: Individual and Occupational Determinants of the Black-White Wage Gap." *American Sociological Review,* vo1. 66, no. 4 (2001): 542-567.

[17] Hahn, Ryan D. and Derek Price. "Promise Lost: College-Qualified Students Who Don't Enroll in College." *A Report by Institute for Higher Education Policy,* November (2008).

[18] Harlow, Caroline W. "Education and Correctional Populations." *Bureau of Justice Statistics Special Report*, NCJ 195670, January 2003.

[19] Harmon, Colm and Ian Walker. "The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales." *Economica*, vol. 67 (2000): 19-35.

[20] Hertz, Tom. "Rags, Riches, and Race: The Intergenerational Economic Mobility of Black and White Families in the Unites States." in Unequal Chances: Family Background and Economic Success, Bowles, Samuel, Herbert Gintis, and Melissa Osborne Groves, *eds.* New Jersey: Princeton University Press (2005).

[21] Kahneman, Daniel and Amos Tversky. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, vol. 47 (1979): 263-291.

[22] Kobberling, Veronica and Peter P. Wakker. "An index of loss aversion." *Journal of Economic Theory*, vol. 122 (2005): 119-131.

[23] Lemieux, Thomas. "Post-secondary Education and Increasing Wage Inequality.", *The American Economic Review*, vol.96, no. 2 (2006): 1-23.

[24] Kao, Grace and Marta Tienda. “Educational Aspirations of Minority Youth.” *American Journal of Education*, vol. 106, no. 3 (1998): 349-384.

[25] Malloy, Liam C. “Loss Aversion, Education and the Distribution of Income.” *Working Paper*, (2010).

[26] Malloy, Liam C. “Low Aspirations, Loss Aversion, and Persistent Group Outcomes.” *Working Paper*, (2010).

[27] Maxwell, Nan L. “The Effect on Black-White Wage Differences of Differences in the Quantity and Quality of Education.” *Industrial and Labor Relations Review*, vol. 47, no. 2 (1994): 249-264.

[28] Mazumder, Bhashkar. “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data.” *The Review of Economics and Statistics*, vol. 87, no. 2 (2005): 235–255.

[29] Mazumder, Bhashkar. “Upward Intergenerational Economic Mobility in the United States.” Economic Mobility Project, Pew Charitable Trusts, 2008.

[30] McGrath, Daniel J. et al. “Breaking New Ground: Diverse Routes to College in Rural America.” *Rural Sociology,* vol. 66, no. 2 (2001): 244–267.

[31] Mulligan, Casey. Parental Priorities and Economic Inequality. Chicago: University of Chicago Press (1997).

[32] U.S. Department of Education, National Center for Education Statistics, 2006-07 and 2007-08 Integrated Postsecondary Education Data System (IPEDS), (2008).

[33] Neilson, William S. “Comparative Risk Sensitivity with Reference-Dependent Preferences.” *The Journal of Risk and Uncertainty*, vol. 24, no. 2 (2002): 131–142.

[34] Piketty, Thomas and Emmanuel Saez. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics*, vol. 118, no. 1 (2003): 1-39. Updated online through 2007 at http://elsa.berkeley.edu/~saez/.

[35] Rizzo, John A. and Richard J. Zeckhauser. “Reference Income, Loss Aversion, and Physician Behavior.” *The Review of Economics and Statistics*, vol. 85, no. 4 (2003): 909-922.

[36] Roderick, Melissa et al. “From High School to the Future: Potholes on the Road to College.” *Consortium on Chicago School Research at the University of Chicago*, March 2008.

[37] Shea, John. “Does parents’ money matter?” *Journal of Public Economics*, vol. 77, no. 2 (2000): 155-184.

[38] Silberberg, Alan et al. “On Loss Aversion in Capuchin Monkeys.” *Journal of the Experimental Analysis of Behavior*, vol. 89, no. 2 (2008): 145-155.

[39] Tovar, Patricia. "The Effects of Loss Aversion on Trade Policy: Theory and Evidence." *Journal of International Economics*, vol. 78, no. 1 (2009): 154-167.

[40] Tversky, Amos and Daniel Kahneman. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, vol. 5 (1992): 297-323.

[41] Urzua, Sergio. "Racial Labor Market Gaps: The Role of Abilities and Schooling Choices." *The Journal of Human Resources*, vol. 43, no. 4 (2008): 919-971.

[42] Winter, Christopher. "Accounting for the Changing Role of Family Income in Determining College Entry." *Institute for Empirical Research in Economics University of Zurich*, Working Paper No. 402 (2009).

[43] Waldkirch, Andreas, Serena Ng, and Donald Cox. "Intergenerational Linkages in Consumption Behavior." *The Journal of Human Resources*, vol. 39, no. 2 (2004): 355-381.

[44] Yogo, Motohiro. "Asset Prices Under Habit Formation and Reference-Dependent Preferences." *Journal of Business & Economic Statistics*, vol. 26, no. 2 (2008): 131-143.