The Pennsylvania State University

The Graduate School

College of the Liberal Arts

UNCOVERING THE CAUSAL PATHWAYS OF HEALTH AND EDUCATION CHOICES

A Dissertation in
Economics
by
Andrew B. Rice

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2010

The dissertation of Andrew B. Rice was reviewed and approved[1] by the following:

Edward J. Green
Professor of Economics and Mathematics
Dissertation Advisor, Chair of Committee

Edward Coulson
Professor of Economics

Joris Pinkse
Professor of Economics

Duane Alwin
McCourtney Professor of Sociology and Demography

Vijay Krishna
Distinguished Professor of Economics
Director of Graduate Studies in Economics

[1] Signatures are on file in the Graduate School.

Abstract

This dissertation concerns the relationship between education and health. While this is not a new topic, the focus of this research is to explicitly model the education and health decision process and estimate such a model. Compared to current literature which focuses on statistical modeling as opposed to economic modeling of the decision, this project allows for deeper intuitive and policy analysis.

Chapter one reviews the Economics literature on the relationship between health and education. It then discusses the current issues in the literature and proposes a solution to some of those problems. This chapter also provides a simple economic model of health and education choice from which a more complicated dynamic model is derived in chapter two.

Chapter two proposes a dynamic model of health and education choice allowing for unobserved heterogeneity. This model is then estimated using data from the Health and Retirement Study. Results from the estimation suggest that an exogenous increase in education only increases expectation of life by approximately one third of the amount typically reported in the current

literature.

Chapter three performs various robustness checks on the structural estimation described in chapter one. One test is to simulate data from the estimated model and perform statistical analysis analogous to what the current literature would perform on real data. This analysis shows that while the model predicts a far less mortality decline from a year of school, simulated data still generates the same results as real data when put to the standard analysis. The other major test performed in this chapter is to vary the amount of unobserved heterogeneity allowed in the model to see how sensitive it is to this number.

Chapter four performs various policy experiments using the model estimated in chapter one. It examines the cost and benefits of compulsory education policy, college cost policy, and smoking policies. While all of these policies have various effects, this chapter only looks at the health benefits of these levers and compares the various cost benefit ratios. Ultimately the conclusion is that even with the reduced benefit of education compared to the current literature, education investment as a health policy is most likely effective. However, policies that target health decisions, especially smoking are superior in terms of cost benefit ratios.

Contents

List of Figures

List of Tables

Acknowledgments

This project would not have gotten off the ground without the patience and inspiration of my advisor Edward Green. Ed Coulson and Joris Pinkse provided valuable advice on the direction in which to look to solve of many technical and data issues. I am grateful to Duane Alwin for opening up my awareness to other ideas presented outside the field of Economics. I am in great debt to Ken Wolpin for allowing me to attend his lectures about empirical modelling. Funding for this dissertation was supplied in part by the Human Capital Foundation as well as Bates White LLC. Any remaining errors are of course my own.

Dedication

For my wife who patiently put her life on hold while I toiled on this project.

Chapter 1

Introduction

This dissertation examines the well known, robust link between higher educational attainment and lower mortality rates. According to the National Center for Health Statistics, "The age-adjusted death rate for those with less than 12 years of education was 650.4 deaths per $100,000$ U.S. standard population, 36.2 percent higher than the rate of 477.6 deaths per $100,000$ U.S. standard population for those with 12 years of education and 3.2 times the rate for those with 13 years of education or more."(Kung et al., 2008) At every age and level of education, incremental education lowers mortality at every subsequent age. A rule of thumb is that this divergence of mortality rates translates into a raw statistical relationship that life expectancy rises by about a year and a half for every incremental year of education.

The correlation between longevity and educational attainment has been known and studied for decades. Kitagawa and Hauser (1973) are widely regarded as the originators of modern work on this topic, albeit with a focus on socioeconomic status. By using education as a main indicator of socioeconomic status the authors inadvertently showed the correlation. A host of other studies throughout the years have reinforced that this correlation is present and strong even controlling for many other factors(see Deaton and Paxson (2001) for a direct study or Grossman (2005) for a review of the literature).

The predominant theoretical framework of these studies has been an application of Becker (1962), in which a persons health status is represented as a stock of human capital. This application was suggested by Ben-Porath (1967) and was worked out in detail by Grossman (1972). Grossman considered the role of education in detail, modeling it as a complementary investment that increases the rate of return on other human capital investments specific to health.

Of course, the rule of thumb mentioned above is only a statistical correlation, which could reflect any combination of the following four mechanisms:

1. that education gives people lower mortality by making them better able to maintain their health than they would otherwise be;

2. that education gives people lower mortality by providing them with stronger incentives to maintain their health than they would otherwise have;

3. that, conversely, (an expectation of) lower mortality disposes people to acquire more education than they would otherwise acquire, or

4. that some third attribute leads people both to acquire high education and also to make other decisions that reduce their mortality.

Mechanisms 1 and 2 speak to forward causation: more education causes better health outcomes. Mechanism 3 relates to what some call reverse causation. This mechanism has not been widely considered in the literature, however the model presented in this paper will allow for it. Mechanism 4 being mostly operative would suggest that the association between education and health is purely due to self selection into better health and more education. The next section discusses how the literature has tried to examine these mechanisms.

1.1 Instrumental Variables and Natural Experiment Literature

For health economists it seems that for the most part the prior expectation is that the health education gradient is generated primarily through mechanisms 1 and 2. Grossman's seminal model of health choice specifically assumes that education is a productivity enhancing input into a health production function Grossman (1972). The theory that the association may come mostly from selection into better health and more education is widely attributed to Fuchs (1982). Specifically Fuchs postulates that time discounting could be a root cause for the association. The argument is that those individuals who discount the future heavily will neither invest in the long term payoff of extra schooling; nor will they invest in costly but beneficial health choices like not smoking or exercising since these carry little value for a myopic person.

If Fuchs' argument is correct, simply running a linear regression assuming education causes better health outcomes will lead to the over estimation of the effect of education on health. To address this issue there is a literature that uses instrumental variables to try and control for this selection. For instance, Sander (1995) uses family background and the region in which an individual lives at at age 16 as instruments for education in a regression on the probability of smoking cessation. This literature finds that education and healthy behavior is still quite correlated even after attempting to control for endogenaity. These type of studies unfortunately have the flaw that the instruments can be suspect. Ideally an instrument would only be correlated with education and not health, however given the strong association in the data this is unlikely to be true for any observable characteristic. For instance family background is likely related to education, but it is also likely related to health behaviors. This will again bias the effect of education on health in a standard regression.

The most recent literature takes a different approach. Since instruments that are correlated with education but not health outcomes are likely hard to find in the data, an interesting identification strategy is to find a policy that exogenously changes educational attainment (a so called quasi-natural experiment) and see if there is a health response to that policy. Lleras-Muney (2005) uses compulsory schooling laws in the United States as instruments for exogenous shifts in education in a two stage least squares approach to estimating the health returns to education. Lleras-Muney finds that even in this framework, the effect of health on education does not diminish, suggesting that mechanism 4 may not be important. Similar research strategies have been employed with mixed results. Arendt (2005) (Denmark), and Spasojevic (2003) (Sweden) find that education remains a strong determinant of mortality or health outcomes. Reinhold and Jürges (2009)(West Germany) Clark and Royer (2007) (UK), and Albouy and Lequien (2009) (France) come to the opposite conclusion.

Outside of the compulsory schooling papers, there have been a number of attempts at using other schooling policies. Kenkel et al. (2006) uses variation in GED policies by state as an instrument to see if there is a difference between the health behaviors of high school drop outs and GED recipients. de Walque (2007) uses Vietnam draft lottery numbers as an instrument for exogenous schooling shift and finds that extra schooling decreases smoking behavior.

All of the aforementioned literature is concerned with eliminating bias problems caused by the possibility of unobserved heterogeneity (mechanism 4). As is noted in Cutler and Lleras-Muney (2006), even if mechanism 4 has no effect on the correlation, little is known about the difference between the direct effects of education on health (mechanism 1), and the incentive effects (mechanism 2). One reason for this is that while quasi-natural experiments are well suited to examine if causation exists, they do not technically tell us much about pathways of causation. For that we need more detailed economic modeling; of which is presented in this paper.

It seems that policy makers have responded to literature's uncertainty about mechanisms by trying to cover all of their bases. The current proposal for the next decade's national health objectives (Healthy People 2020) reflect the power of this association by including several goals related to education. Of interest to this research are goals ECBP HP20201 and ECBP HP202011. ECBP HP20201 focuses on increasing general high school completion as a health policy. ECBP HP202011 focuses on a very narrow set of skills to promote in public schools. If schooling has a direct cause on health levels we would expect direct skill enhancement to be an appropriate and productive policy. If schooling mostly affects health through the other mechanisms, then general education would be the best policy. General education is however quite expensive and it would be helpful to know why general education enhances health levels in this case.

The research reported in this dissertation contributes to the literature in a several ways. It uses individual panel data from the Health and Retirement Study on health decisions instead of aggregate data as in Lleras-Muney (2005). This is the first study in which the decision process is explicitly modeled and can apportion weights to the different mechanisms that may generate the health and education gradient as seen in data. In this regard the paper draws on the models of Rust (1987) for structure and Arcidiacono and Jones (2003) to allow for permanent unobserved heterogeneity across a finite mixture of types.[1] The estimation procedure used in this dissertation is an extension of

[1] The general method for dealing with finite type mixtures in this setting was first put forth by Heckman and Singer (1984) and later used in dynamic discrete choice models starting with Keane and Wolpin (1997). Arcidiacono and Jones (2003) greatly reduces the computational burden of this model feature for the class of models descendant from Rust

Arcidiacono et al. (2007) which also uses HRS data to explain health decisions, however the current research focuses on the relationship between endogenous education choice and health behaviors, whereas the former research does not allow for education choice.

1.2 A Motivating Example

To understand how variations in time preference may be important in understanding how education affects aggregate mortality figures, we present a simple two period discrete choice model. This model retains several characteristics that will be present in a more rigorous empirical study: investment in education reduces early life utility (think of lost wages or dislike of studying) and increases end of life utility; investment in health reduces utility in early life and increases the probability of survival in later life; and traditional time preference modeling looks enters into a life-cycle model the same way death probability might.

There are two time periods $t = 1, 2$ and an agent is characterized by a *time preference parameter* $\beta \in [0, 1]$. In the first period an agent has to make two choices: an dichotomous education level, $\overline{e}$ or $\underline{e}$, and a dichotomous health investment level, $\overline{h}$ or $\underline{h}$. For notational clarity any underlined variable is the low value and any variable with an upper bar is a higher value. There are explicit costs to these decisions and for simplicity we assume that $\underline{e} = \underline{h} =$

(1987).

0 and that c_h and c_e are the respective marginal costs of a unit of health investment and a unit of education. Thus a choice of $\overline{e}$ has a cost of $c_e\overline{e}$ and a choice of $\overline{h}$ has a cost of $c_h\overline{h}$. Wages depend on age as well as education choice so that we can write wages as $w(e,t)$. These wages have the property that $w(\overline{e},1) < w(\underline{e},1) = w(\underline{e},2) < w(\overline{e},2)$. Finally, there is a survival probability $s(h)$ that determines the ex ante chance of mortality in the second period given a level of health investment. We assume that with high investment an individual lives to the second period for sure, $s(\overline{h}) = 1$, and that if a person does not invest in health related issues they face an uncertain future life span, i.e. $s(\underline{h}) < 1$

The agent makes all decisions in the first period at which point she realizes the explicit costs of her decisions and the wage in period one. At period two if the agent is alive she generates her second period utility, and if not generates a zero utility. This leads to an agent choosing between high and low education, and choosing between high and low health investment to maximize the following expression:

$$w(e,1) - c_h h - c_e e + \beta s(h) w(e,2)$$

Since there are only four choices we can list all of the payoffs in a small payoff matrix:

Due to the simplicity of the model, if we give some numeric values to the variables we can easily solve for the optimal choice given a particular $s(\underline{h})$ and β. The following parameter values were derived from the Panel Study of

Table 1.1: Payoffs for Descriptive Model

	$\underline{h}$	$\overline{h}$
$\underline{e}$	$w(\underline{e},1)+\beta s(\underline{h})w(\underline{e},2)$	$w(\underline{e},1)-c_h\overline{h}+\beta w(\underline{e},2)$
$\overline{e}$	$w(\overline{e},1)-c_e\overline{e}+\beta s(\underline{h})w(\overline{e},2)$	$w(\overline{e},1)-c_e\overline{e}-c_h\overline{h}+\beta w(\overline{e},2)$

Income Dynamics: The solution to this problem for all values of $s(\underline{h})$ and β is

Table 1.2: Parameter Values for Descriptive Model

$w(\underline{e},1)$	$w(\underline{e},2)$	$w(\overline{e},1)$	$w(\overline{e},2)$	c_h	c_e	$\underline{e}$	$\overline{e}$
700	700	583	1124	100	50	12	16

shown in figure 1.1. Note that in general as the marginal value of health investment goes up ($s(\underline{h})$ goes down) more and more β values emit a high health investment. This graph also highlights a fundamental identification issue: relatively high levels of β with low levels of marginal value of health investment emit the same decision as low levels of β with high marginal values of health investment. This suggests any empirical strategy may need information about either β or expected mortality risk in order to have an identified model.

For a thought experiment also consider a case where an agency has taken away the education choice and imposes a requirement of $\overline{e}$. This is a crude approximation of the compulsory schooling law changes that Lleras-Muney (2005) and others use. The decision is reduced to only a health investment choice, where an individual will choose $\overline{h}$ if and only if $w(\overline{e},1)-c_e\overline{e}+\beta s(\underline{h})w(\overline{e},2) < w(\overline{e},1)-c_e\overline{e}-c_h\overline{h}+\beta w(\overline{e},2)$.

Figure 1.2 shows the optimal decisions in this case. Note that for all points,

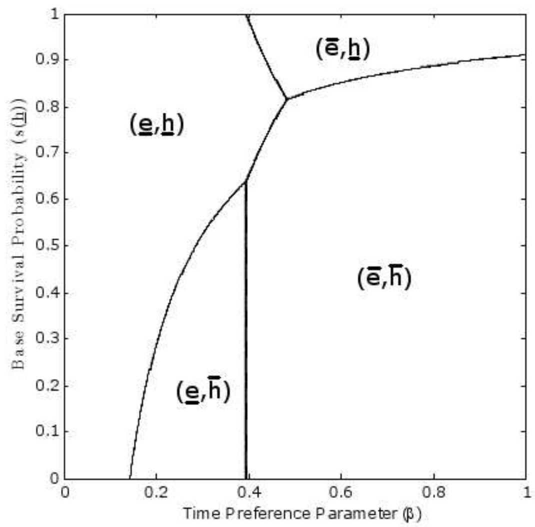


Figure 1.1: Cutoff regions (unregulated)

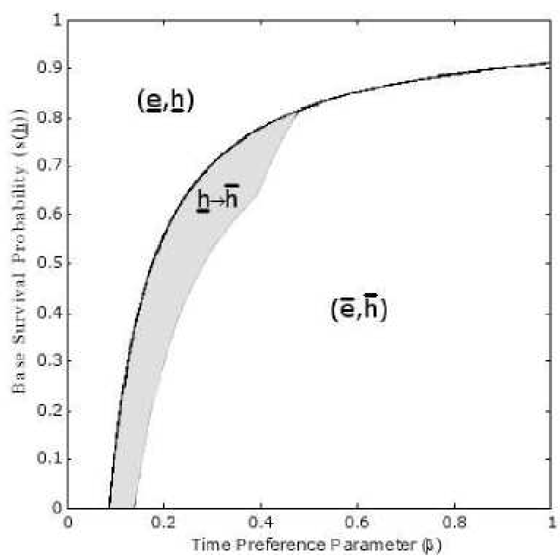


Figure 1.2: Cutoff regions (compulsory education)

health investment is at least as much as in the no policy case, and at some points what was once a $\underline{h}$ decision is now a $\overline{h}$ decision. In this example, compulsory schooling causes a reduction in mortality without education having any direct ability to do so. Instead education changes the incentives of certain types of people such that they choose to invest more in health and thus live longer.

1.3 Agenda of Dissertation

The dissertation will formalize and estimate an individual decision model in chapter 2, evaluate the robustness of the estimation in chapter 3, and examine the policy implications of these exercise in chapter 4. The goal of the project is to gain insight into the mechanics of how education and health are related and to then evaluate several policies that could be implemented depending on a policy maker's goal.

Chapter 2

A Structural Estimation of Health and Education Choice

As discussed in chapter 1, the goal of this dissertation is to explore the mortality education relationship through the lens of an economic model. In order for this exercise to have any meaning, the decision model must satisfy three features: it should be able to allow for several pathways of the relationship, it should not deviate from traditional decision models, and it must be simple enough to solve and estimate. The reason for the first is so that the data does not "force" the model into one particular pathway of a relationship because the true pathway does not exist. The second is an attempt to stay as agnostic as possible so that the model does not favor one pathway over another. The last feature is just a matter of feasibility.

This chapter sets forth a decision model that satisfies these features and estimates it using a large national set of data from a panel of retirees in the United States.

2.1 Decision Model

Consider a world in which an individual optimizes over his lifespan using the following structure common to the dynamic discrete choice estimation literature.[1] An individual begins his decision life at time $t = 0$ and lives until at most T discrete periods. To pin this to the data assume that $t = 0$ relates to age 14, each period is two years and $T = 40$ so that the model covers from ages 14 to 96. Each individual is characterized by a pure time preference parameter $\beta \in [0, 1]$.[2] In each period there are J_t number of mutually exclusive options available. In order to reduce the choice space, I assume that at time 0 each agent makes an education choice $e \in \{8, \ldots, 17\}$ and then follows through with that decision. For all $t \geq 1$ there are only health choices about smoking and drinking so that $J_t = 4$.[3]. Let d_{jt} take the value of 1 if the j^{th} option is taken at time t and 0 otherwise and $d_t = [d_{0t}, \ldots, d_{J_t t}]$. Since these options

[1]This setup extends the model in Arcidiacono et al. (2007) to allow for endogenous education choice.

[2]The distinction between pure time preference and discount rate in this model related to an individual's expectation of mortality. The intention of β is to capture a preference for time rather than combine that with a subjective risk of reaching the next period.

[3]There are four possible joint values of the two dichotomous variables for smoking and heavy drinking

are defined as mutually exclusive:

$$\forall\ t,\ \sum_{j=1}^{J_t} d_{jt} = 1$$

All of the information needed to make a decision at time t is contained in a parameter vector θ and a state X_t which evolves over time according to a Markov transition function $F(X_{t+1}|X_t, d_t, \theta)$. I assume that X_t is composed of both an observable component x_t and an unobservable component ϵ_t so that $X_t = [x_t, \epsilon_t]$. An agent knows the full state X_t at time t, however any outside observer cannot see ϵ_t which will therefore not be in a data set while x_t may be. Each agent is rational, such that ex ante beliefs about future states are consistent with the transition function F. For tractability I assume the following about the structure of the transition function:

$$F(X_{t+1}|X_t, d_t, \theta) = G(\epsilon_{t+1}|\epsilon_t, d_t)F_x(x_{t+1}|x_t, d_t, \theta) \tag{CI}$$

which simply states that the unobserved state variables evolve independently from the observed state variables conditional on decisions.[4] Specification of the transitions is explained further in section 2.1.1

I assume that preferences can be characterized by a time separable utility function with payoff $U_t(X_t, d_t)$ at time t. Furthermore, denoting the time

[4]Note that this does not necessarily imply that x_t and ϵ_t are independent: since both are known to an agent at the time of decision making, d_t may (and in this setup will) be a function of both x_t and ϵ_t.

invariant payoff function for option j as $u_j(x_t, \theta)$; U_t takes the following form:

$$U_t(X_t, d_t, \theta) = \sum_{j=0}^{J_t} d_{jt}\left[u_j(x_t, \theta) + \epsilon_{jt}\right] \tag{AS}$$

2.1.1 Function Specifications

For clarity, partition the parameter vector as $\theta = \{\theta^d, \theta^e, \theta^y, \theta^s\}$, where θ^d is the portion of the parameter vector used in the per period utility functions, θ^e affects education decisions, θ^y will be used in wage transition functions, and θ^s is used in the survival probabilities.

Each option in the health choice has its own parameter values, and the intercept is interpretable as the difference in utility level between option j and a baseline option (in this case the no smoking and no heavy drinking option is the baseline). For all $t > 0$, u_j takes the following specification:

$$u_j(x_t, \theta) = \theta^d_{j0} + \theta^d_{j1} ln(y_t) + \theta^d_{j2}\mathbb{1}(qsmoke) + \theta^d_{j2}\mathbb{1}(qdrink)$$

For $t = 0$ (the education choice period) u_j is specified as:

$$u_j(x_0, \theta) = \theta^e_0\mathbb{1}(edu_j \leq 12)edu_j + \theta^e_1\mathbb{1}(12 < edu_j \leq 16)edu_j + \theta^e_2\mathbb{1}(edu_j \geq 17)edu_j$$

where $\mathbb{1}(x)$ is an indicator function returning 1 if the statement x is true. I also specify the form of $F_x(x_{t+1}|x_t, d_t, \theta)$, specifically the transitions of unknown future states such as next period income and next period mortality state. I have assumed that there are no unobservable state variables in these transitions, so I simply induce transition probabilities from the data using the following regression forms.

Survival probabilities take a simple logit form:

$$\Pr[survive_{t+1}|X_t, d_t, \theta] = S(X_t, d_t, \theta) = \frac{1}{1 + \exp(-x_{st}\theta^s)}$$

Where:

$$x_{st}\theta^s = \theta_0^s + \theta_1^s age_t + \theta_2^s edu + \theta_3^s \mathbb{1}(drink_t) + \theta_4^s \mathbb{1}(ex - smoker) + \theta_5^s \mathbb{1}(smoke_t)$$

Income transition probabilities are backed out from a log-normal regression function of the form:

$$\begin{aligned} y_t = {} & \theta_0^y + \theta_1^y edu + \theta_2^y age + \theta_3^y age^2 + \theta_4^y y_{t-1} \\ & + \theta_5^y \mathbb{1}(12 < edu \leq 16) + \theta_6^y age \mathbb{1}(edu \geq 17) + \nu \end{aligned}$$

2.1.2 Solution

Given the prior environment, a rational agent simply solves the following discounted lifetime problem:

$$\max_{(\rho_0,\ldots,\rho_T)} E_\rho \left[\sum_{t=0}^{T} \beta^t U_t \left(X_t, d_t, \theta \right) | X_0 \right]$$

Where ρ_t is a policy rule that transforms a state to a decision, i.e. $d_t = \rho_t(X_t)$. The rationality assumption implies that the beliefs used in the expectation operator E_ρ are consistent with the transition probabilities for uncertain states according to the Markov function F. It also implies that agents make decisions considering the ex ante expected payoff from education and health investments (as opposed to their ex post realizations).

2.2 Estimation

According to Rust (1987), given assumptions (AS) and (CI) above and assuming that ϵ_t comes from the type I extreme value distribution, this model can be formulated as a dynamic multinomial logit and estimated by maximum likelihood. Specifically, with the distributional assumption on ϵ_t, the value functions for each option can be written as follows:

$$V_{jT}(x_T;\theta) = u_{jT}\left(x_T, \theta^d\right) \tag{2.1}$$

$$\begin{aligned} V_{jt}(x_t;\theta) =& u_{jt}\left(x_t, \theta^d\right) + \beta S(x_t, d_j, \theta^s) \\ & \times \int \left(\ln \left[\sum_{k=1}^{J_{t+1}} \exp\left\{ V_{k(t+1)}\left(x_{t+1};\theta\right)\right\}\right]\right) F_x\left(dx_{t+1}|x_t, d_j, \theta\right) \end{aligned} \tag{2.2}$$

The dynamic multinomial logit form then implies choice probabilities of the form:

$$\Pr(d_{jt} = 1|x_t) = \frac{\exp\left\{V_{jt}\left(x_t;\theta\right)\right\}}{\sum_{k=0}^{J_t-1} \exp\left\{V_{kt}\left(x_t;\theta\right)\right\}} \tag{2.3}$$

Given θ and the specifications listed above, the likelihood can be calculated for each state and choice for each observation. If each individual is indexed by i in the data where $i = 1, \ldots, N$ and is observed from model period $\underline{t}_i$ to $\overline{t}_i$ then the contribution to the log likelihood from observation i is:

$$\begin{aligned} L_i(\theta) = \ln\left(\Pr\left(d_{ij0}|x_0, \theta\right)\right) &+ \sum_{t=\underline{t}_i+1}^{\overline{t}_i} \ln\left(\Pr\left(d_{ijt}|x_t, \theta\right)\right) \\ &+ \sum_{t=\underline{t}_i+1}^{\overline{t}_i} \ln\left(\Pr\left(x_t|x_{t-1}, d_{t-1}, \theta^y, \theta^s\right)\right) \end{aligned} \tag{2.4}$$

In principle the parameters of this objective function could be jointly estimated by full information maximum likelihood (FIML). The number of iterations needed to optimize a log-likelihood function increases rapidly in the number of parameters needed to be estimated. To estimate this model by FIML would require many iterations, and each one would have to solve the dynamic choice problem which is computationally expensive. For tractability Rust (1987) suggests maximizing the log likelihood $\sum_{i=0}^{N} l_i(\theta)$ via a two step process. Since the log likelihood is additively separable, the state transition log likelihood is maximized first by choosing θ^y and θ^s. Then taking θ^y and θ^s as given θ^e and θ^d are then chosen to maximize the decision variable log likelihood.

2.2.1 Unobserved Heterogeneity

In order to study the effects of heterogeneity on health and education choices, this model must be extended to allow for heterogeneous agents. As noted in

Arcidiacono and Jones (2003), unobserved heterogeneity spoils the separability that allows for two step procedure described above. Instead one must maximize the likelihood function by optimizing over the full parameter space, which given the scope of this problem is computationally infeasible. Arcidiacono and Jones (2003) specify, and Arcidiacono et al. (2007) implement an algorithm that reintroduces the ability to estimate the parameter space in steps. A brief explanation of how that is achieved in this model is presented below. For more details see the prior mentioned papers.

Following Heckman and Singer (1984) and Keane and Wolpin (1997) assume that there a finite number of discrete types of agents $m = 1, \ldots, M$.[5] These types may affect transition probabilities as well as preferences. If type were observed equation (2.4) becomes:

$$L_i(\theta, m) = \ln\left(\Pr\left(d_{ij0}|x_0, \theta, m\right)\right) + \sum_{t=\underline{t}_i+1}^{\overline{t}_i} \ln\left(\Pr\left(d_{ijt}|x_t, \theta, m\right)\right) + \sum_{t=\underline{t}_i+1}^{\overline{t}_i} \ln\left(\Pr\left(x_t|x_{t-1}, d_{t-1}, \theta^y, \theta^s, m\right)\right) \qquad (2.5)$$

Unfortunately the type of an individual is unobserved so more structure is needed. Define π_i^m as the probability that individual i is of type m. The

[5] In theory there can be as many types as individuals in the sample, in practice this is typically set at some low number. In this case I use 2 types

contribution to the log likelihood from an observation then becomes:

$$L_i(\theta) = \ln\left(\sum_{m=1}^{M} \pi_i^m \left(l_{eim}(\theta) \prod_{t=\underline{t}_i+1}^{\bar{t}_i} l_{dimt}(\theta) l_{ysimt}(\theta^y, \theta^s)\right)\right) \tag{2.6}$$

This form is computationally problematic since it destroys the additive separability of the log likelihood. To maximize (2.6) would require an optimization routine to cover the full parameter space. Instead Arcidiacono and Jones (2003) suggests the following:

1. Allow time invariant individual data z_i to affect π_i^m using the following specification
$$\pi_i^m(\gamma) = \frac{\exp(\gamma_m z_i)}{1 + \exp(\gamma_m z_i)} \tag{2.7}$$

2. Calculate the conditional probability of individual i being type m as:
$$P_i^m = \frac{\pi_i^m(\gamma) l_{eim}(\theta) \prod_{t=\underline{t}_i+1}^{\bar{t}_i} l_{dimt}(\theta) l_{ysimt}(\theta^y, \theta^s)}{\sum_{j=1}^{M} \pi_i^j(\gamma) l_{eij}(\theta) \prod_{t=\underline{t}_i+1}^{\bar{t}_i} l_{dijt}(\theta) l_{ysijt}(\theta^y, \theta^s)} \tag{2.8}$$

3. Calculate the type-qqqqweighted log likelihood of the sample
$$\sum_{i=1}^{N} \sum_{m=1}^{M} P_i^m L_i(\theta, m) \tag{2.9}$$

Equation (2.9) reintroduces separability and thus θ can be optimized in two steps as before. The algorithm to estimate the full model is as follows: pick a starting value for θ^d and θ^e and choose θ^y and θ^s to maximize the state transition log likelihood. Given this value for θ choose γ to maximize equation (2.6). Calculate π_i^m and then P_i^m and then maximize equation(2.9) in two steps. Repeat this process until the log likelihood converges.

Unobserved heterogeneity is supposed to influence agents decisions, so type must enter into the specification of the technology and utility functions. For this project heterogeneity enters in several ways. The wage transition regression and survival regression include type dummies. The intercept of the utility function for each option varies by type. In some specifications the discount factor β will be allowed to vary by type.

The time invariant data used to separate out types and help control for unobserved differences in initial conditions include: wave 1 smoking and drinking choice, subjective probabilities of surviving to age 75 divided by calculated life table probabilities (as a measure of relative optimism about health), whether or not the individual's mother lived past 70, and mother's education.

2.3 Data

The primary source of data for the estimation is the RAND version (I) of the Health and Retirement Study (HRS) sponsored by the National Institute of Aging. The HRS is a large longitudinal study of individuals 50 years of age and older along with their spouses or partners. The RAND data have been cleaned and documented to clear up some inconsistencies in the raw HRS data and add to usability. The survey over samples African-Americans, Hispanics, and residents of Florida. The first wave of the HRS was conducted in 1992 and has been followed up on every two years, with the latest published wave for the HRS being the 2006 wave. The sample is divided into several cohorts for which interviewing started during different waves. This paper focuses on the so called HRS cohort since it has data for 8 waves (16 years of the life cycle). The HRS cohort contains 13434 individuals: a sample of individuals born between 1931 and 1941 and their spouses. The HRS cohort was chosen over the AHEAD cohort (a more mature sub-sample in the HRS data) because the AHEAD cohort is likely to have had significant mortality even before the survey and as such is probably a highly selected sample of healthier people.

So as to not confuse age, period, and cohort effects; we assume that all individuals in the HRS cohort born between 1931 and 1941 have the same cohort experience (i.e. the structural estimates should be the same) Future research may be able to use more of the data, however for this project we want to use as large a life span as possible.

The HRS has a full range of demographic, financial, health, family, and work related variables. Mortality rates in various sub-populations are estimable because there has been significant mortality in the sample since the first survey wave. By wave 8 of the HRS 24.7% of the HRS cohort has died. Since a key component of any dynamic model is future *expectations* rather than actual outcomes, I use the fact that the HRS contains a set of questions asking people what they think is their probability of surviving until ages 75 and 85 to help separately identify mortality expectations from other future discounting mechanisms. These data, while relatively noisy, have been shown to vary appropriately with known risk factors, and respond reasonably to new information (Hurd and McGarry, 2002).

Currently this study only examines the life experience of males. For HRS cohort it is quite clear that the education and health choice incentives differ for males and females especially in the wage dimension. Dealing with males simplifies the model by not having to worry as much about labor force participation and non pecuniary benefits from schooling. Also the males and females in this particular sample are (at least in part) married to each other. Since the choices of spouses are likely to be correlated in some unknown way, modeling these decisions would be prohibitively costly.

There are 4594 males in the HRS cohort born between 1931 and 1941.[6] Of these 1402 are lost to attrition not caused by a mortality event. It is assumed that education decision making starts at age 14, so any observation with an education below 8th grade is dropped. Dropping some missing values takes the sample to 2293 males covering 14402 person-waves.

Summary statistics for the data used is in tables 2.1 and 2.2. Income household income which includes all income from all sources including a spouses income. Age is in years at time of interview. Note that the average age is falling in the sample. This is due to a mortality effect that is biased towards later life. Heavy drinking is defined as drinking more than two drinks per day, and

[6]There are more males in the cohort born outside of this range due to marriage and data collection error. These observations are dropped since it is unlikely that they are making decisions in the same structural environment as the cohort this study is focused on.

Table 2.1: Summary Statistics: Time Variant Data

Variable	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
	Wave 2		Wave 3		Wave 4		Wave 5	
Log income	10.715	0.939	10.796	0.897	10.771	0.943	10.784	0.917
Age	57.416	3.174	59.404	3.181	61.26	3.171	63.159	3.167
Heavy Drinker	0.072	0.258	0.072	0.259	0.063	0.243	0.058	0.234
Ex-smoker	0.489	0.5	0.5	0.5	0.527	0.499	0.545	0.498
Smoker	0.244	0.429	0.229	0.42	0.196	0.397	0.173	0.379
	Wave 6		Wave 7		Wave 8			
Log income	10.756	0.914	10.804	0.874	10.811	0.820		
Age	65.283	3.162	67.187	3.18	69.135	3.141		
Heavy Drinker	0.065	0.246	0.066	0.249	0.07	0.255		
Ex-smoker	0.551	0.498	0.572	0.495	0.584	0.493		
Smoker	0.162	0.368	0.135	0.342	0.121	0.326		

seems to have little trend. Smoking has a tremendous drop over the course of the sample starting at over 24% of respondents smoking down to just 12%.

Table 2.2: Summary Statistics: Time Invariant Data

Variable	Mean	Std. Dev
Mother lived past 70	0.772	0.419
Wave 1 Smoking	0.227	0.419
Wave 1 Heavy Drinker	0.079	0.270
E(Prob Survive 75)/life table	1.057	0.453
Mothers Education	9.934	3.254
Wave 1 Ex Smoker	0.477	0.500
Years of Education	13.20	2.543

For information about family health that may be useful for health expectations the HRS has data on how old the individual's mother was when she died (or age if alive), as well as a subjective probability of surviving. Mothers age is turned into a dummy variable which takes the value of 1 if an individual's mother lived past 70. The subjective probability of survival in the HRS has been shown to correspond fairly well with actual mortality.[7] This study uses subjective probability of living to 75 divided by the United States life table probability. One could consider this variable a relative optimism variable. If it is over 1, an individual is relatively optimistic about their survival chances given their age compared to the rest of their peers.

Table 2.3 is where the correlation between education and mortality shows up strongest. Period by period mortality is small, however after 14 years it is

[7] Hurd and McGarry (2002) present evidence that this variable responds well to health shocks and mortality risks.

Table 2.3: Percentage of Population Remaining

	Wave						
Education	2	3	4	5	6	7	8
8	0.96	0.92	0.88	0.80	0.72	0.67	0.65
9	0.96	0.89	0.86	0.82	0.78	0.73	0.70
10	0.96	0.91	0.89	0.85	0.81	0.76	0.72
11	0.96	0.90	0.86	0.80	0.76	0.71	0.69
12	0.97	0.94	0.91	0.88	0.84	0.81	0.77
13	0.97	0.95	0.93	0.90	0.85	0.83	0.79
14	0.97	0.92	0.88	0.84	0.79	0.77	0.73
15	0.99	0.95	0.93	0.90	0.87	0.84	0.79
16	0.98	0.94	0.93	0.91	0.88	0.85	0.81
17	0.99	0.96	0.94	0.93	0.90	0.87	0.85

clear that those who have more education are surviving at a much higher rate than those with less. The next section will set out a formal model that will try and describe these data.

2.4 Results

The model is estimated under two separate specifications. Both specifications allow for unobserved heterogeneity as described above. They are different in the discount factor that is used in each case. The fist model assumes that both types have a one year rate of $\beta = 0.95$, and the second assumes that type 1 has a one year discount rate of $\beta = .75$ and type 2 has a one year $\beta = 0.95$.[8] There are four sets of parameters to estimate: survival function parameters

[8]The values were chosen based on a grid search for the best fit of the data.

(table 2.4), wage transition parameters (table 2.5), utility parameters (table 2.6), and type probability parameters (table 2.7).

The survival function and wage transition function makeup the beliefs that an agent has in the model about the value of the future. These of course can only be estimated from observed data, and if an agent has different subjective beliefs than what the data suggest, the solution to the dynamic programming problem will be incorrect. Therefore, I assume that the observed mortality transitions and wage transitions correspond with the subjective beliefs of the agents in the model.

Table 2.4: Survival Function Parameters

β	0.95		Heterogeneous	
	Coeff.	Std. Err.	Coeff.	Std. Err.
constant	7.7037	0.7806	7.7472	0.7816
education	0.0536	0.0196	0.0541	0.0197
age	-0.0724	0.0108	-0.0726	0.0108
heavy drinking	0.1158	0.2018	0.1031	0.2007
ex-smoker	-0.5993	0.1397	-0.6025	0.1395
smoker	-1.1153	0.2204	-1.1472	0.2205
type 1	0.0436	0.1800	0.0072	0.1811

The survival transition estimates seem to have appropriate magnitudes as well as signs with the exception of heavy drinking. Of particular interest to this paper is the coefficient of education on survivability. It seems as if considering unobserved heterogeneity in the survival function has no effect or may actually increase the effect of education on two period survival probabilities. If education had no direct effect on mortality this coefficient should be zero. Note that because of the non linearity of the survival function, as age increases (and therefore drags down the survival probability) the marginal effect of survival from smoking increases, thus increasing the cost of smoking as one ages.

Unobserved type is not significant in the survival function suggesting differences in survival rates come through choice rather than chance. If there were a difference between types in the survival function, this would suggest that different types have different baseline survival probabilities which would affect the choice decision. A significant value of this parameter then could be evidence for reverse causation (those that have higher subjective survival probabilities may invest more or less in various human capital; or mechanism 3). As such this model does not lend evidence to the theory that reverse causation is operative.

Table 2.5: Wage Transition Parameters

	0.95		Heterogeneous	
	Coeff.	Std. Err.	Coeff.	Std. Err.
constant	6.1012	1.1266	6.1009	1.1267
education	0.0437	0.0073	0.0436	0.0073
age	-0.0683	0.0356	-0.0683	0.0356
age^2	0.0864	0.0200	0.0873	0.0201
lagged income	0.5898	0.0101	0.5898	0.0101
college	0.0280	0.0311	0.0279	0.0311
postgrad	0.0897	0.0480	0.0900	0.0480
type 1	0.0005	0.0003	0.0005	0.0003

The wage parameters suggest a concave wage-age profile with relatively high persistence as is typically seen in wage data.[9] Of note is that education changes the growth rate of wages so that higher educated people have steeper wage paths. Again the heterogeneity parameter for this transition is insignificant.

The utility parameters in table 2.6 seem to be what one would expect given the model. Compared to non-smoking and not heavy drinking, smoking and not heavy drinking has a large utility benefit. Just heavy drinking is utility enhancing but not as much as just smoking. Choosing to drink and smoke has a utility benefit; just not as much as smoking alone. There is a fairly large penalty for quitting smoking attesting to the large persistence in smoking in the data. Without this parameter there is not enough stickiness to the smoking choice. The education parameters can be interpreted as cost of foregone wages on top of psychic cost of an extra year of education. It seems reasonable to expect that the cost of college is higher than the cost of high school on a per year basis.

The type parameters in table 2.7 seem to suggest that only smoking in wave 1 has an appreciable effect on type probabilities. The type estimation is semi-parametric in the sense that it puts more weight to type probabilities that maximize the log likelihood and then adjusts the type parameters to match these weights. The high coefficient on the constant implies that the nonparametric type selection dominates the parametric part of the algorithm. More research is necessary here to determine which variables would best affect

[9]See Heckman et al. (2008) for more information on this phenomenon.

Table 2.6: Utility Parameters

		0.95		Heterogeneous	
		Coeff.	Std. Err.	Coeff.	Std. Err.
smoking=1, heavy driking=0	constant	6.2928	0.0195	5.6504	0.0642
	income	0.8440	0.0000	0.8709	0.0001
	type 1	1.7667	0.0022	1.5187	0.0015
smoking=0, heavy driking=1	constant	1.5459	0.1805	0.5977	0.1799
	income	1.0059	0.0004	1.0458	0.0004
	quit smoke	-4.2414	0.0296	-3.8925	0.0291
	type 1	-5.1518	0.0061	-4.9940	0.0049
smoking=1, heavy driking=1	constant	3.4997	0.1758	2.0986	0.1874
	income	0.8928	0.0004	0.9507	0.0004
	type 1	0.0805	0.0197	1.1472	0.0087
education	high school/yr	-0.4827	0.0005	-0.4955	0.0005
	college/yr	-0.9863	0.0002	-1.0045	0.0002
	post-graduate/yr	-1.0773	0.0001	-1.0916	0.0002
log likelihood		-29407		-29012	

Table 2.7: Type 2 Parameters

	0.95	Heterogeneous
constant	8.2010	8.2099
oldmom	0.0117	0.0120
wave 1 smoking	0.1141	0.1251
wave 1 drinking	0.0686	0.0627
wave 1 outlook (75)	-0.0030	-0.0091
mothers education	-0.0003	-0.0006
wave 1 ex-smoker	0.0094	0.0109
avg prob type 1	0.1695	0.1828
avg prob type 2	0.8305	0.8172

type probabilities instead of just letting the choice probability estimation shift the type weights.

2.5 Model Fit

One comparison between the model and the data that is striking is the proportion of the sample that remain alive by education after all of the waves of the HRS.

Table 2.8 compares the actual survival proportions in the data to the predicted survival proportions from the heterogeneous model broken down by education. Note the strong association between higher education and survival proportion in the data. The model however does not estimate a very high direct marginal health return to education. Since the data and the model predictions match fairly well, the model must predict less healthy behavior from

Table 2.8: Proportion Alive After 8 Waves

	% Remaining	
Education	Data	Predicted
8	0.65	0.66
9	0.70	0.63
10	0.72	0.67
11	0.69	0.73
12	0.77	0.75
13	0.79	0.78
14	0.73	0.73
15	0.79	0.77
16	0.81	0.79
17	0.85	0.83

less educated individuals.

Table 2.9 shows aggregate smoking and drinking choice across types. In all versions of the model Type 1 is less likely to smoke and drink as well as have higher education. Since improved education and improved health choices are correlated across types it is clear that unobserved heterogeneity plays some role in the overall correlation.

To examine how well this model fits the data table; 2.10 compares predicted choice probabilities to the choice proportions in the data. Generally speaking the heterogeneous discount specification fits better, although both over-predict smoking in the early waves which drops off too quickly. In the data drinking drops and then recovers, the heterogeneous model also fits this pattern better than the single discount rate model.

Table 2.11 describes the model's fit to the education distribution of the

Table 2.9: Choices by Type

	Mean Edu.		Wave 2	Wave 3	Wave 4	Wave 5
		$\beta = 0.95$				
Type 1	9.38	Smoking	0.59	0.54	0.50	0.45
		Drinking	0.42	0.46	0.48	0.52
Type 2	12.95	Smoking	0.28	0.20	0.14	0.09
		Drinking	0.04	0.04	0.04	0.04
		Heterogeneous β				
Type 1	9.37	Smoking	0.59	0.54	0.50	0.45
		Drinking	0.41	0.44	0.47	0.51
Type 2	13.09	Smoking	0.27	0.19	0.13	0.08
		Drinking	0.07	0.07	0.06	0.06
				Wave 6	Wave 7	Wave 8
		$\beta = 0.95$				
Type 1	-	Smoking		0.40	0.35	0.30
		Drinking		0.56	0.59	0.63
Type 2	-	Smoking		0.05	0.03	0.01
		Drinking		0.04	0.05	0.05
		Heterogeneous β				
Type 1	-	Smoking		0.40	0.35	0.30
		Drinking		0.54	0.58	0.62
Type 2	-	Smoking		0.05	0.03	0.01
		Drinking		0.05	0.05	0.06

data. The model has a hard time fitting the data without heterogeneous discounting. Chapter 3 discusses how different formulations of the model perform.

Table 2.10: Choice Probabilities

		Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Data	smoke=0, drink=0	0.715	0.726	0.767	0.791	0.794	0.820	0.828
	smoke=1, drink=0	0.213	0.202	0.171	0.151	0.141	0.114	0.102
	smoke=0, drink=1	0.041	0.046	0.038	0.036	0.044	0.045	0.051
	smoke=1, drink=1	0.031	0.026	0.025	0.022	0.021	0.021	0.019
	smoking %	0.244	0.229	0.196	0.173	0.162	0.135	0.121
	heavy drinking %	0.072	0.072	0.063	0.058	0.065	0.066	0.070
0.95	smoke=0, drink=0	0.584	0.646	0.698	0.740	0.768	0.794	0.808
	smoke=1, drink=0	0.339	0.274	0.222	0.176	0.143	0.114	0.089
	smoke=0, drink=1	0.039	0.046	0.050	0.057	0.064	0.071	0.084
	smoke=1, drink=1	0.037	0.034	0.030	0.027	0.024	0.021	0.018
	smoking %	0.377	0.308	0.252	0.203	0.167	0.135	0.108
	heavy drinking %	0.077	0.079	0.080	0.083	0.089	0.092	0.103
Heterogeneous	smoke=0, drink=0	0.613	0.668	0.714	0.750	0.774	0.796	0.807
	smoke=1, drink=0	0.288	0.236	0.195	0.159	0.133	0.108	0.086
	smoke=0, drink=1	0.042	0.048	0.052	0.059	0.067	0.073	0.088
	smoke=1, drink=1	0.056	0.047	0.039	0.032	0.027	0.022	0.018
	smoking %	0.344	0.284	0.234	0.191	0.159	0.130	0.105
	heavy drinking %	0.098	0.095	0.091	0.091	0.093	0.096	0.106

Table 2.11: Education Distribution

scooling	data	0.95	Heterogeneous
8	0.046	0.076	0.079
9	0.039	0.074	0.074
10	0.048	0.093	0.089
11	0.040	0.142	0.135
12	0.346	0.242	0.232
13	0.063	0.065	0.064
14	0.107	0.068	0.068
15	0.037	0.069	0.071
16	0.123	0.069	0.071
17	0.152	0.103	0.119
High School	0.519	0.627	0.607
College	0.329	0.271	0.274
Post-grad	0.152	0.103	0.119

2.5.1 Life Expectancy From Exogenous Education

The since the model predicts mortality relatively well one would expect that the model generates a large correlation between life expectancy and education. Specifically define life expectancy at age 14 as:

$$e_{14} = \sum_{t=14}^{47} 2 \cdot E[S(X_t, d_t, \theta)]$$

Where the expectation is taken with respect to optimal choice probabilities and the state evolution. Using the heterogeneous model the value for the sample is 71.12. This value is a bit high most likely because the data does not contain the very high mortality ages that push down life expectancy. It is easy in this

Table 2.12: Marginal Life Expectancy Gains From an Extra Year of Education

edu	Marg. LE
8	1.42
9	1.07
10	0.71
11	0.42
12	0.35
13	0.14
14	0.16
15	0.14
16	0.15
17+	0.13

model to vary education exogenously. For every observation just add one more year of education without changing the parameters of the model and rerun the life expectancy calculation. The counterfactual value is 71.44. Since the model has nothing to say about adding education to postgraduates, the 0.31 average years of life added does not include an effect from the postgraduates in the sample.

The current literature suggests that this value should be close to 1.5 years for an exogenous shift in education. One theory is that the health returns to education are concave in education. Prior research has only focused on exogenous shifts in education on very low levels of education. If the returns are negligible for high levels of education and strong for low levels; that would account for the disparity. Table 2.12 shows the predicted life expectancy gain at age 25 for each level of education. It shows strong decreasing marginal returns to education, and is consistent with the current literature. If it is true that the current literature is only measuring exogenous shifts in education at low levels of education as described in chapter 1 then the reported result that the marginal increase in life for one year of education being 1.5 years is close to the 1.42 of grade 8 in this model. This result is explored more carefully in chapter 3.

2.6 Conclusion

This chapter has set forth a standard life cycle model adapted to fit the health and education lifetime decision. Results suggest that this model can fit various features of the data fairly well. They also call into question the common result that education causes one and a half years of life gain; as this model suggests that an exogenous shift in education will result in an average of only one third of a year of life. Further chapters will explore this result more and examine some policy implications.

Chapter 3

Estimation Robustness

Structural estimation has two major advantages: it allows for more advanced policy and counterfactual analysis, and it allows economic theory to have a say in estimation of data. The down side is that it is computationally intensive and can feel like a "black box" from which results are pulled. The goal of this chapter is to examine the estimation and results described in chapter 2 with the intent of validating them in order to perform policy analysis with the model in chapter 4.

3.1 Simulations

The solution method to the model described in chapter 2 results in a series of decision rules that map an optimal choice from the information that the model has at any given state. These rules, while for the most part intuitive, are hard to describe in any systematic way. This is the reason why these models can feel like an opaque system. Indeed, counter intuitive results are difficult to convincingly explain since one cannot just follow the model through and see what is happening. To alleviate this problem sometimes simulations will be used in this chapter. The technical details of how a simulation is performed in this model is can be found in appendix A.1.

Intuitively a simulation of the model simply creates multiple copies of the sample in the data at some time period and then draws random noise according to the model specification. Using this noise it determines what every simulated individual would choose based on the estimated model and information at the base period from the data and records their choice. Based on this choice, the simulation calculates the mortality probability of each individual and determines a random mortality event based on this probability. If the simulated individual survives, the simulation determines their state of being in the next period and then repeats the choice and mortality simulation. This continues until the simulated has completed its full life cycle.

3.2 Validation Against Existing Results

Since this project was started in response to a growing literature using quasi-natural experiments stemming from Lleras-Muney (2005) one way to validate the model would be to simulate data from the model described in chapter 2 using the parameter estimates and see if similar results can be be generated. If it can then it could be argued that the model does well in describing the rules that individuals use when making choices related to education and health.

3.2.1 Technical Review of Current Methods

The standard model currently used to test the causal effect of education on health is summarized in this section. Typically the data must contain individuals that have been forced to obtain more schooling or identifiable groups of individuals of whom some subset have been forced to obtain more schooling than they otherwise would have. The most common type of data are those data that include individuals that live in a region that has had a change in minimum compulsory schooling laws. In this case, observations are grouped by state of birth, and a subset of the individuals in each group have been forced to obtain more schooling (it is unknowable which observations were forced to more schooling without asking each individual directly).

These data must also contain mortality information in at least two time periods and personal characteristics of observations. If this type of data is available a two stage least squares estimation procedure can be performed using the following model:

$$d_{itr} = b_d + e_i\pi_d + X_{it}\beta_d + W_r\delta_d + \alpha_{dr} + \epsilon_{ditr}$$

$$e_{ir} = b_e + c_r\pi_e + X_i\beta_e + W_r\delta_e + \alpha_{er} + \epsilon_{eitr}$$

Where d_{it} is whether or not individual i from region r is alive at time t, b is a constant, e is an education level, X is a set of individual characteristics, W is a set of region characteristics, α is a set of region dummies, and c is a set of dummies that describe the compulsory schooling laws that individual i was subject to.

The fist stage is to estimate the second equation and then use the predicted results as the values for in the second stage. In the second stage the first equation is estimated and then standard errors are corrected. Note that identification of this model requires exclusion restrictions: in this case meaning the first stage include some variables that are not included in the second stage. Only c is excluded from the second stage. Furthermore if c were correlated with an unobservable that is also correlated with d, this exercise will result in biased results. For instance states may have a general sentiment of better health and more education that is not captured in W.

Upon estimating a model similar to the one above Lleras-Muney (2005) concludes by reporting that at the average, a one year increase in education can account for a decrease in mortality rate of 1.3 over a 10 year span (or $\pi_d = -0.013$).

3.2.2 Simulation of the Model

To compare the estimated model to Lleras-Muney's results, it must be simulated. This is done according to the procedure outlined in appendix A.1 starting from $t = 0$ (age 14). The only exception to this procedure is that if a simulant dies before their respective individual would have entered the sample, that simulant is kept until it dies again after the entry point of the individual. Each observation is simulated 50 times for a total number of $114,650$ simulations.

Lleras-Muney (2005) uses cross sections ten years apart. To perform a similar analysis I will take cross sections from the simulated data at 1992 and 2002. Unfortunately I do not have access to state of birth in the HRS. I do however have information on census division of birth. As is further examined in section 4.2 this is not the best grouping since census divisions have similar compulsory schooling laws in the 1950's. The flexibility of the structural model helps here. I randomly choose several regions and increase its minimum education to grade 11. That information is used as the exclusion restriction in the first stage thus simulating an exogenous law change.

This method ensures that the policy instruments are strictly exogenous in the simulation. One must keep in mind however that in the real data the state policies are likely endogenous insofar as laws are made with the wishes of the people in mind. If state policy is endogenous in the real data it would likely bias the results upward and as such I would expect the simulated data to produce two stage least squares estimates that are lower than the real data.

3.2.3 2SLS Estimation

The model is simulated such that all simulants that are born in census division 3, 4, 5, and 6 have had an exogenous policy increasing the minimum grade required to grade 11. First the simulated data is estimated using a simple linear probability model specified as follows:

$$d_i = b + e_i\pi_d + X_i\beta_d + W_{ir}\delta + \alpha_{ir} + \epsilon_i$$

X includes age, mother's education, and if the mother lived past age 75. W includes information about the census division as obtained from the 1960 United States Census. These variables are the percentage individuals in an living in an urban area, percentage of workers employed in manufacturing, and percentage of foreign born individuals in the division. These variables were chosen to match as closely as possible to the estimations in Lleras-Muney (2005). The actual values of these variables are listed in appendix A.2.

This particular regression does not attempt to control for the endogeneity problem of education in this regression. The standard result reported by Lleras-Muney is a value between -0.011 and -0.017 on the education coefficient in this type of model.[1] Table 3.1 lists the results of this estimation on data simulated from the structural model after estimation. The results from the linear probability model run on simulated data show an effect of education on mortality comparable to the current literature.

The next step is to use the information on simulated policy to run a two stage least squares estimation using the policy as an instrument as described in the previous section. In the prevailing literature the value of the coefficient on education generally increases in magnitude to between -0.017 and -0.060

[1] The variation comes from performing the estimation on different samples.

Table 3.1: Linear Probability Regression of Simulation

Variable	Coeff.	Std. Err.	pval
Education	-0.0142	0.0005	0.0000
Age	0.0096	0.0004	0.0000
% Urban	0.0009	0.0003	0.0010
% Manufacturing	-0.0010	0.0002	0.0000
% Foreign	-0.0006	0.0007	0.4440
Mother's Edu	0.0001	0.0004	0.8680
Mother Lived > 75	0.0062	0.0030	0.0390
Constant	-0.3273	0.0297	0.0000

the standard errors become larger.[2]

Table 3.2: First Stage Estimates (2SLS)

Variable	Coeff.	Std. Err.	pval
Age	-0.0147	0.0021	0.0000
% Urban	0.0252	0.0021	0.0000
% Manufacturing	-0.0295	0.0025	0.0000
% Foreign	0.0233	0.0045	0.0000
Mother's Edu	0.2557	0.0022	0.0000
Mother Lived > 75	0.1152	0.0176	0.0000
Comp School	0.2369	0.0324	0.0000
Constant	10.0590	0.1913	0.0000

Table 3.2 shows the first stage of this procedure. The key item to note is that the policy variable is positive and significant, suggesting that that is creating an exogenous increase in the average of education. This is expected

[2] Lleras-Muney (2005) uses both aggregate census data and samples from individual panel data sets. The lower bounds on the point estimates generally come from an estimation using individual data as opposed census data which resembles this exercise most closely.

since the model is forcing a subset of the simulants to have more education than they would have had they been born in a different census division.

Table 3.3: Second Stage Estimates (2SLS)

Variable	Coeff.	Std. Err.	pval
Education	-0.0281	0.0236	0.2330
Age	0.0094	0.0005	0.0000
% Urban	0.0011	0.0004	0.0140
% Manufacturing	-0.0012	0.0004	0.0060
% Foreign	-0.0004	0.0008	0.6490
Mother's Edu	0.0036	0.0061	0.5500
Mother Lived > 75	0.0078	0.0040	0.0530
Constant	-0.1789	0.2538	0.4810

Table 3.3 displays the estimation results from the second stage of the 2SLS estimation. The coefficient on education increases and has larger standard errors just as in real data. Indeed this mirrors almost exactly the type of results the prevailing literature reports, even thought the structural model estimates a far weaker effect of education on health.

3.3 Marginal Effect of Each Year

In chapter 2 it was shown that the predicted life expectancy gain from one extra year of education decreased as the level of education obtained went up. This is a standard decreasing marginal benefits argument. It is not immediately clear how the model can generate such an outcome. The answer becomes apparent when looking at the difference in life expectancy for each type.

During estimation the model apportions a probability that an individual is of a certain type. It turns out in the model it is more likely that an individual is type 2 the lower the education level. If it were the case that type 2 individuals have a high benefit from an exogenous shift in education and type 1 individuals have a low benefit from extra education regardless of level a downward sloping marginal benefit curve can be constructed.

For example compare grade eight to grade sixteen. The model will apportion almost full weight to type 2 if the observation attended only grade eight. Likewise it will apportion almost all weight to type 1 if an observation attended grade 17. So an experiment that shifts grade eight observations by one year has all the characteristics of increasing only type 2 observations by one year. The grade sixteen experiment will be pushing almost only type 1 individuals up one year. To see if this is the case the type distribution by education, and the life benefit by type need to be examined.

Table 3.4 details the average type probabilities assigned to each education group. Notice how as education grows it is more and more likely that an individual is assigned to type 2. So any composite average will start trending quickly towards the type 2 benefit as education increases.

Table 3.5 shows the stark difference in life expectancy from each type at each level of education. Type 2 individuals are generally healthier and thus have less to gain from any education benefit. There is still some decreasing benefit from education by type, but this combined with the type distribution by education shifting towards type 2 exaggerates the aggregated curve.

Table 3.4: Type Distribution by Education

edu	Type 1	Type 2
8	0.35	0.65
9	0.36	0.64
10	0.37	0.63
11	0.25	0.75
12	0.20	0.80
13	0.23	0.77
14	0.17	0.83
15	0.11	0.89
16	0.07	0.93
17	0.05	0.95

Table 3.5: Education Benefit by Type

edu	Type 1	Type 2
8	3.15	0.55
9	2.45	0.28
10	1.71	0.22
11	1.21	0.21
12	1.14	0.20
13	0.47	0.11
14	0.39	0.13
15	0.40	0.11
16	0.68	0.12
17	0.46	0.11

3.4 Unobserved Heterogeneity

Adding unobserved heterogeneity into the model introduces considerable complexity in both the computation and intuitive interpretation of results. In chapter 2 the model and estimation presented used a form of heterogeneity in which there are only a small number of types. If we imagine that every individual has unobserved characteristics that may matter, we would ideally like a model that has as many types as individuals. Unfortunately in a dynamic model this puts insurmountable stress on the computation of parameters.

Since in these type of models not accounting for unobserved heterogeneity in the data can seriously bias the results; a compromise is necessary. Typically the number of unobserved types ranges from two to four with the hope that that can capture most of the relevant variation in unobserved type. The results reported in section 2.4 were obtained by setting the number of types to 2. In this section I report the results and some analysis about the results when there is no unobserved heterogeneity, and when there are four types. These are compared to the two type results.

3.4.1 Two Types

For convenience the two type parameter values are listed here to serve as a baseline for comparison to the other model specifications described below. Since the model fits better with heterogeneous discount rates, the baseline will be the two type model with a two year discount rate of $\beta^2 = .9$ for type 1 and $\beta^2 = .56$ for type 2. Table 3.6 repeats the parameter values from the survival function of the two type model. The important parameters in this estimation are the education and type 1 parameters. Education is small but significant and the type parameter is insignificant.

Table 3.6: Survival Function Parameters: Heterogeneous 2 Type Model

	Coeff.	Std. Err.
constant	7.7472	0.7816
education	0.0541	0.0197
age	-0.0726	0.0108
heavy drinking	0.1031	0.2007
ex-smoker	-0.6025	0.1395
smoker	-1.1472	0.2205
type 1	0.0072	0.1811

Table 3.7: Wage Transition Parameters: Heterogeneous 2 Type Model

	Coeff.	Std. Err.
constant	6.1009	1.1267
education	0.0436	0.0073
age	-0.0683	0.0356
age^2	0.0873	0.0201
lagged income	0.5898	0.0101
college	0.0279	0.0311
postgrad	0.0900	0.0480
type 1	0.0005	0.0003

Table 3.7 lists the results for the wage regression for the two type model. Keep in mind that this is an AR(1) model in log wages. Each value is interpreted as the effect on the difference between two wage periods. The results of this regression are consistent with the common parabolic wage path in the labor literature.

Table 3.8 simply lists the utility parameters estimated from the two type

Table 3.8: Utility Parameters Heterogeneous 2 Type Model

		Coeff.	Std. Err.
smoking=1, heavy driking=0	constant	5.6504	0.0642
	income	0.8709	0.0001
	type 1	1.5187	0.0015
smoking=0, heavy driking=1	constant	0.5977	0.1799
	income	1.0458	0.0004
	quit smoke	-3.8925	0.0291
	type 1	-4.9940	0.0049
smoking=1, heavy driking=1	constant	2.0986	0.1874
	income	0.9507	0.0004
	type 1	1.1472	0.0087
education	high school/yr	-0.4955	0.0005
	college/yr	-1.0045	0.0002
	post-graduate/yr	-1.0916	0.0002
log likelihood		-29407	

model. A full discussion of these results is presented in section 2.4.

3.4.2 One Type

The two type model uses two year discount rates of 0.9 and 0.56. To get a reference point and to see why heterogeneity must be accounted for, a one type model has been estimated twice: once with a discount rate of .9 and once with a discount rate of 0.56.

Table 3.9: Utility Parameters, 1 Type, $\beta^2 = 0.9$

		Coeff.	Std. Err.
	cons	31.4198	0.0198
smoking=1, heavy driking=0	income	-0.2325	0.0000
	cons	-12.3373	0.0329
smoking=0, heavy driking=1	income	1.4867	0.0001
	quit smoke	-2.5587	0.0289
	con	2.9286	0.1263
smoking=1, heavy driking=1	income	0.9391	0.0003
	high school/yr	0.3283	0.0005
education	college/yr	-0.1568	0.0003
	post-graduate/yr	-0.2403	0.0003
log likelihood		-29407	

Table 3.9 shows the utility parameters when $\beta^2 = .9$ and table 3.10 shows the parameters when $\beta^2 = .56$. In both cases the model has a hard time fitting the data since it is hard for a rational model to reconcile strong investment in the future (education) and high rates of detrimental behavior. This shows up

Table 3.10: Utility Parameters, 1 Type, $\beta^2 = 0.56$

		Coeff.	Std. Err.
smoking=1, heavy driking=0	cons	36.617	0.010
	income	-0.639	0.000
smoking=0, heavy driking=1	cons	-8.361	0.079
	income	1.268	0.000
	quit smoke	-2.928	0.029
smoking=1, heavy driking=1	con	19.827	0.133
	income	0.047	0.000
education	high school/yr	0.551	0.000
	college/yr	0.163	0.000
	post-graduate/yr	0.163	0.000
log likelihood		-29407	

in the positive utility from a year of education and the very high (in relation to the two type model) parameter on the utility values for smoking choices. In order to come close to generating the education distribution in the data the model has to set education as a positive utility choice. This clearly goes against the idea that education generally causes dis-utility at the time of education and then generates a future payoff.

Since there are no heterogeneous types in these models, both the survival transition functions and the wage transition models will be the same for each specification. These would be the same no matter what specification is used as long as there is only one type (an artifact of the particular estimation method used).

Table 3.11: Survival Parameters, 1 Type

	Coeff.	Std. Err.
constant	7.760	1.1260
education	.0542	0.0072
age	-.0727	0.0356
heavy drinking	0.101	0.0101
ex-smoker	-0.603	0.0311
smoker	-1.153	0.0480

Table 3.11 lists the survival parameters of both models. Note that the value of education in the survival function is similar to the two type model. This is some evidence that there is little heterogeneity in the pure effect of education on health. Indeed the only real effect of heterogeneity here is to increase the health effects of smoking. This is due to the fact that some of the effect of smoking is being attributed to being an unhealthy type as opposed to just smoking.

Table 3.12: Wage Parameters, 1 Type

	Coeff.	Std. Err.
constant	6.16388	1.12656
education	0.04659	0.00724
age	-0.07114	0.03557
age^2	0.00052	0.00028
lagged income	0.59420	0.01006
college	0.02121	0.03109
postgrad	0.08473	0.04802

Finally table 3.12 lists the wage parameters of the one type model under both specifications. Again this is not too different from the two type model.

3.4.3 Four Types

In the previous section it was shown that one type models seem to capture raw wage and survival transitions that the two type model captures. The difference came in the utility parameters. In this section the four type specification is examined.

The four type model is an attempt to see if more can be gained from adding types without running into computational difficulties. The two type model is good at showing the difference between a particular low investment and high investment set of types. There may be differences even within these groups. For the best comparison between two and four type models I set type 1 and type 3 to have $\beta^2 = .9$; and type 2 and type 4 to have $\beta^2 = .56$. This creates a model where type 1 and 3 relate to the first type of the two type model and type 2 and 4 relate to type 2 through the discount rate.

Tables 3.13 and 3.14 show the survival and wage transition parameters respectively. Somewhat surprisingly there is still little difference between the two type model. Almost all of the type parameters are insignificant suggesting that all of the heterogeneity enters through the utility parameters.

Table 3.15 describes the utility parameters of the four type model. The only qualitatively different result is that the smoking and heavy drinking parameter dropped significantly (thus reducing its probability of being chosen). The general patterns are similar: smoking is unconditionally preferred; quit-

Table 3.13: Survival Parameters, 4 Types

	Coeff.	Std. Err.
constant	7.849	0.789
education	0.053	0.020
age	-0.074	0.011
heavy drinking	0.351	0.231
ex-smoker	-0.589	0.139
smoker	-1.109	0.195
type 2	0.000	0.184
type 3	-0.643	0.649
type 4	-0.378	0.262

Table 3.14: Wage Parameters, 4 Types

	Coeff.	Std. Err.
constant	6.171	1.127
education	0.044	0.007
age	-0.069	0.036
age^2	0.000	0.000
lagged income	0.591	0.010
college	0.026	0.031
postgrad	0.089	0.048
type 2	-0.067	0.030
type 3	-0.058	0.066
type 4	-0.079	0.027

ting cigarettes is costly; heavy drinking is not as favorable as smoking (even though its health effects seem quite a bit less). With one caveat the types are interesting, in relation to type 1; type 3 (which has the same discount rate) likes smoking less and drinking more; type 2 (which has a lower discount rate)

Table 3.15: Utility Parameters, 4 Types

		Coeff.	Std. Err.
smoking=1, heavy driking=0	cons	8.726	0.017
	income	0.802	0.000
	type 2	1.121	0.025
	type 3	-9.564	0.941
	type 4	-1.499	0.010
smoking=0, heavy driking=1	cons	-7.152	0.416
	income	1.006	0.001
	quit smoke	-3.884	0.041
	type 2	-2.423	1.001
	type 3	12.881	0.132
	type 4	8.482	0.015
smoking=1, heavy driking=1	con	4.068	0.152
	income	0.908	0.000
	type 2	0.011	0.038
	type 3	2.783	2.064
	type 4	-0.954	0.014
education	high school/yr	-0.373	0.000
	college/yr	-0.908	0.000
	post-graduate/yr	-1.000	0.000
log likelihood			

is more likely to smoke and less likely to drink; and type 4 is much more likely to drink.

The one caveat here is that the standard errors on some parameters are expanding in some cases greatly. This is the result of a practical limit on the number of types that can be incorporated into the model. In general it seems that adding another type to the same discount rate simply allows more flexibility in the model to separate drinkers from smokers. The expense is that the estimates start becoming unstable as more types are added. It seems like four types is the limit here. As we will see in the next section this has little impact on the overall choice probabilities and as such the two type model will be used in the policy chapter.

3.4.4 Heterogeneity Discussion

While the parameter values make sense for the four type model, the choice probabilities just do not do as well as the two type model in fitting the data. Table 3.16 shows the choice probabilities for the four type model and compares it to the data and the two type model. Clearly the four type model has trouble fitting the choice data. The question then is: what gives the four type model a likelihood advantage? It comes from fitting the model slightly better on the education choice. Table 3.17 shows the distribution of education for the four type model. It seems to do slightly better predicting the education distribution but not enough to justify this as a superior model.

While it is clear that using only one type to try and fit the data results in unrealistic parameter values, the difference between two types and 4 types is not so clear. Computationally two types is vastly preferred as the time needed to compute estimations and simulations increases rapidly in the number of types.[3] The two specifications do not offer much difference in performance (in

[3]For instance the estimation of a two type model once specified as desired may take

Table 3.16: Choice Probabilities: 4 Type

		Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Data	smoke=0, drink=0	0.715	0.726	0.767	0.791	0.794	0.820	0.828
	smoke=1, drink=0	0.213	0.202	0.171	0.151	0.141	0.114	0.102
	smoke=0, drink=1	0.041	0.046	0.038	0.036	0.044	0.045	0.051
	smoke=1, drink=1	0.031	0.026	0.025	0.022	0.021	0.021	0.019
	smoking %	0.244	0.229	0.196	0.173	0.162	0.135	0.121
	heavy drinking %	0.072	0.072	0.063	0.058	0.065	0.066	0.070
Two Type	smoke=0, drink=0	0.613	0.668	0.714	0.750	0.774	0.796	0.807
	smoke=1, drink=0	0.288	0.236	0.195	0.159	0.133	0.108	0.086
	smoke=0, drink=1	0.042	0.048	0.052	0.059	0.067	0.073	0.088
	smoke=1, drink=1	0.056	0.047	0.039	0.032	0.027	0.022	0.018
	smoking %	0.344	0.284	0.234	0.191	0.159	0.130	0.105
	heavy drinking %	0.098	0.095	0.091	0.091	0.093	0.096	0.106
Four Type	smoke=0, drink=0	0.499	0.571	0.633	0.686	0.728	0.765	0.789
	smoke=1, drink=0	0.318	0.262	0.218	0.179	0.150	0.123	0.102
	smoke=0, drink=1	0.032	0.034	0.036	0.040	0.044	0.048	0.056
	smoke=1, drink=1	0.152	0.133	0.113	0.095	0.078	0.064	0.053
	smoking %	0.469	0.396	0.331	0.274	0.228	0.186	0.154
	heavy drinking %	0.184	0.167	0.149	0.135	0.122	0.112	0.109

Table 3.17: Education Distribution: 4 Type

scooling	Data	2 Type	4 Type
8	0.046	0.079	0.072
9	0.039	0.074	0.071
10	0.048	0.089	0.088
11	0.040	0.135	0.137
12	0.346	0.232	0.246
13	0.063	0.064	0.059
14	0.107	0.068	0.066
15	0.037	0.071	0.071
16	0.123	0.071	0.075
17	0.152	0.119	0.114
High School	0.519	0.627	0.607
College	0.329	0.271	0.274
Post-grad	0.152	0.103	0.119

some cases the four type is worse). Which gives confidence in using the two type model for policy analysis (since the model must be solved and simulated a number of times the two type model is preferred).

3.5 Conclusion

A dynamic structural estimation of health and education can generate the same type of data patterns that generate results in the current literature of

48 hours on eight computer cores running at 2Ghz. The four type model in theory would take at least 4 days simply from doubling the amount of computations needed to solve the model. On top of that there are more parameters to estimate so the minimization algorithm requires more model solutions to converge. In this case the 4 type model once specified as desired takes around 10 days to compute in practice.

1.5 years of life expectancy per year of education. The same model however only suggests an average increase from a true exogenous increase in education to be approximately 0.33 years of life per year of education. The discrepancy comes from both marginal effects being different than average effects, and this author also questions the assumption that compulsory schooling laws are exogenous to the model in the current literature.

In general it seems that adding types to the structural model quickly puts strains on the practical estimation of this model. The minimum number of types necessary to get a stable estimation should be used in order to best fit the data. In this case jumping from two to four types has little practical gain but large practical costs. For that reason the rest of the dissertation uses only the two type model.

Chapter 4

Is Education a Health Policy Lever?

On top of all of the other benefits education (see Behrman and Stacey (1997) for a review of the social benefits of education) education seems to have a qualitatively large effect on health. Given this fact and the prior literature that seems to support the hypothesis that education causes health improvements, it is natural to think about using national education policy as a lever to improve public health.

Lleras-Muney (2005) and the literature that follows attempts to isolate the effect of an *exogenous* increase in education in her sample. That is to say we are interested in the effect of life expectancy for different years of education *only* from education. To see why this is difficult to isolate, consider two different situations: the first is the level of education one would choose with no outside intervention, and the second is the level of education one higher than the no intervention choice. The level of education chosen by any individual is clearly driven by social status and possibly a whole host of other effects so variation in educational attainment and variation in health behaviors and outcomes may be jointly determined by other factors. The strategy used in the past is to find a way to look at a forced change in schooling and assume that this year of school has nothing to do with any other factors. This has potential problems since a policy may have different impacts at different levels of schooling which could be varying with other factors that affect health.

As an example, imagine a policy that increases the minimum level of education required by law from grade seven to grade eight. First we have to control for whatever decision process created the original decision of leaving school at grade seven, and then we have to see what the effect of that additional year of school would have for *that particular type of person*. If we cannot generalize this effect to other types of people then all we have identified is a "local average treatment effect" (Imbens and Angrist, 1994). In this case we have only shown the effect of a policy of forcing seventh graders to take an additional year of schooling.

To avoid these problems I will lean on an economic model that has been estimated to fit a large national data set. Keane (2009) argues that this approach to answering policy questions does not have the interpretive drawbacks that the prior literature has, at the expense of some additional assumptions and complicated computational modeling. The details of this estimation can be found in chapter 2.

As noted in chapter 1 one goal in the national health policy platform is to increase educational attainment. This chapter will address four policies that are related to the education and health decisions that generate the strong correlation between both outcomes. These four policies are: changing compulsory schooling laws, subsidizing college, late life lump sum payments, and simply addressing health behavior directly.

4.1 Education Cost and Value of a Statistical Life

In order to do cost benefit analysis we must know how much a policy would cost should it be implemented. Some relevant variables for the policies that will be presented are the cost of public education, the cost of college education, and the statistical monetary value of a year of life.

The cost of education that is used will be $6,692 for secondary education in 1992 dollars (since all of the policies will have to be dealing with schooling in the 50's). Higher education cost will be values at $14,652. Details of these values are in appendix A.3.

Value of life is a hotly debated topic with values ranging from several hundred thousand dollars to tens of millions of dollars. Clearly placing a dollar amount on life is subject to various interpretational assumptions which this dissertation will remain silent on. For policy purposes it helps to have a number for costs and benefits to be compared.

Viscusi (2008) discusses in depth various different values taken by regulatory agencies (between $1 and $8 million) and strongly argues for a value of $7 million based on his view of the literature and prior work. Ashenfelter (2006) performs an estimation using data on speed limit changes in the united states to come up with an estimate with a lower bound of $1.6 million and an upper bound of $6 million. In light of the consistent orders of magnitude of modern estimates and levels at which current policy makers choose as the value of a statistical life I will use a value of $7 million.[1] I must note however that small variations in these values can potentially significantly change the cost benefit analysis of any policy. As long as these values are consistent across all of the simulations it will give a basis for comparison. The actual costs and benefits have no real meaning on a simulated cohort.

4.2 Life Gains From Compulsory Schooling

In order to align with the current literature, I will first examine the effect of additional compulsory schooling requirements. The the effect of a compulsory schooling law will be maximized when it affects the most people. In chapter 2 I examined a case in which everyone is required to obtain one more year of education than they would have absent any law. This is of course an irrelevant policy since it is impossible to enforce such a law. For the most part states

[1] I will also use $100,000 in 2008 dollars as the value of a statistical life as in Viscusi (2008) when necessary.

have instituted complicated laws mandating either a certain number of years of education or a minimum age at which schooling is no longer required.[2]

The sample of people used to estimate the model in chapter 2 was born between 1931 and 1941 meaning that the relevant compulsory schooling laws would be the ones in effect from 1945 to 1960.

Tables 4.1 and 4.2 list the various minimum ages at which schooling is compulsory by state and year (Edwards, 1978). Since the HRS does not have data on state of birth but does have data on census region of birth, table 4.3 groups these policies into averages by census region. Note that the mode policy during this time frame is 16 years and there is not much variation. This suggests that even given data on state of birth, a difference in differences estimator as used by Lleras-Muney (2005) may not have enough state variation in this period for inference. It does however suggest a baseline minimum schooling level of tenth grade in most states.

One could imagine that a national minimum education level was passed such that everyone is required by law to obtain at least eleven years of education. For the sake of simplicity assume that these laws are perfectly enforceable. If an observation has over the minimum education level no action needs to be taken; they would not have been affected by such a law. If an observation has below the minimum level of education in reality I assume that observation would be affected by the law and be forced to attend more schooling (as if

[2]Some states have or have had combined rules such as must go to school until sixteen years of age or have had passed at least eighth grade. See Lleras-Muney (2002) for details on state laws from 1915 to 1939, Angrist and Krueger (1991) for laws from 1960 to 1980, and Edwards (1978) for state laws from 1939 to 1960 (the relevant period of the HRS sample).

Table 4.1: Minimum Age At Which Schooling is Mandatory: 1939-1960

Year	'39 – '40	'44 – '45	'54 – '55	'59 – '60
Alabama	16	16	16	16
Arizona	16	16	16	16
Arkansas	16	16	16	16
California	16	16	16	16
Colorado	16	16	16	16
Connecticut	16	16	16	16
Delaware	16	17	16	16
Florida	16	16	16	16
Georgia	14	14	16	16
Idaho	18	18	16	16
Illinois	16	16	16	16
Indiana	16	16	16	16
Iowa	16	16	16	16
Kansas	16	16	16	16
Kentucky	16	16	16	16
Louisiana	14	15	16	16
Maine	16	16	16	15
Maryland	16	16	16	16
Massachusetts	16	16	16	16
Michigan	16	16	16	16
Minnesota	16	16	16	16
Mississippi	16	16	16	
Missouri	14	14	16	16
Montana	16	16	16	16
Nebraska	16	16	16	16
Nevada	18	18	18	17
New Hampshire	16	16	16	16
New Jersey	16	16	16	16
New Mexico	16	16	17	17
New York	16	16	16	16
North Carolina	14	14	16	16
North Dakota	17	17	17	16

Table 4.2: Minimum Age At Which Schooling is Mandatory: 1939-1960 (continued)

Year	'39 – '40	'44 – '45	'54 – '55	'59 – '60
Ohio	18	18	18	18
Oklahoma	18	18	18	18
Oregon	16	16	18	18
Pennsylvania	18	18	17	17
Rhode Island	16	16	16	16
South Carolina	16	16	16	
South Dakota	17	16	16	16
Tennessee	16	16	17	17
Texas	16	16	16	16
Utah	18	18	18	18
Vermont	16	16	16	16
Virginia	15	16	16	
Washington	16	16	16	16
West Virginia	16	16	16	16
Wisconsin	16	16	16	16
Wyoming	17	16	16	17

Table 4.3: Years of Compulsory Schooling by Census Region

Year	'39 – '40	'44 – '45	'54 – '55	'59 – '60
New England	16	16	16	15.83
Mid Atlantic	16.67	16.67	16.33	16.33
EN Central	16.4	16.4	16.4	16.4
WN Central	16	15.86	16.14	16
S Atlantic	15.38	15.63	16	16
ES Central	16	16	16.25	16.33
WS Central	16	16.25	16.5	16.5
Mountain	16.88	16.75	16.63	16.63
Pacific	16	16	16.67	16.67

they had complied with the law).

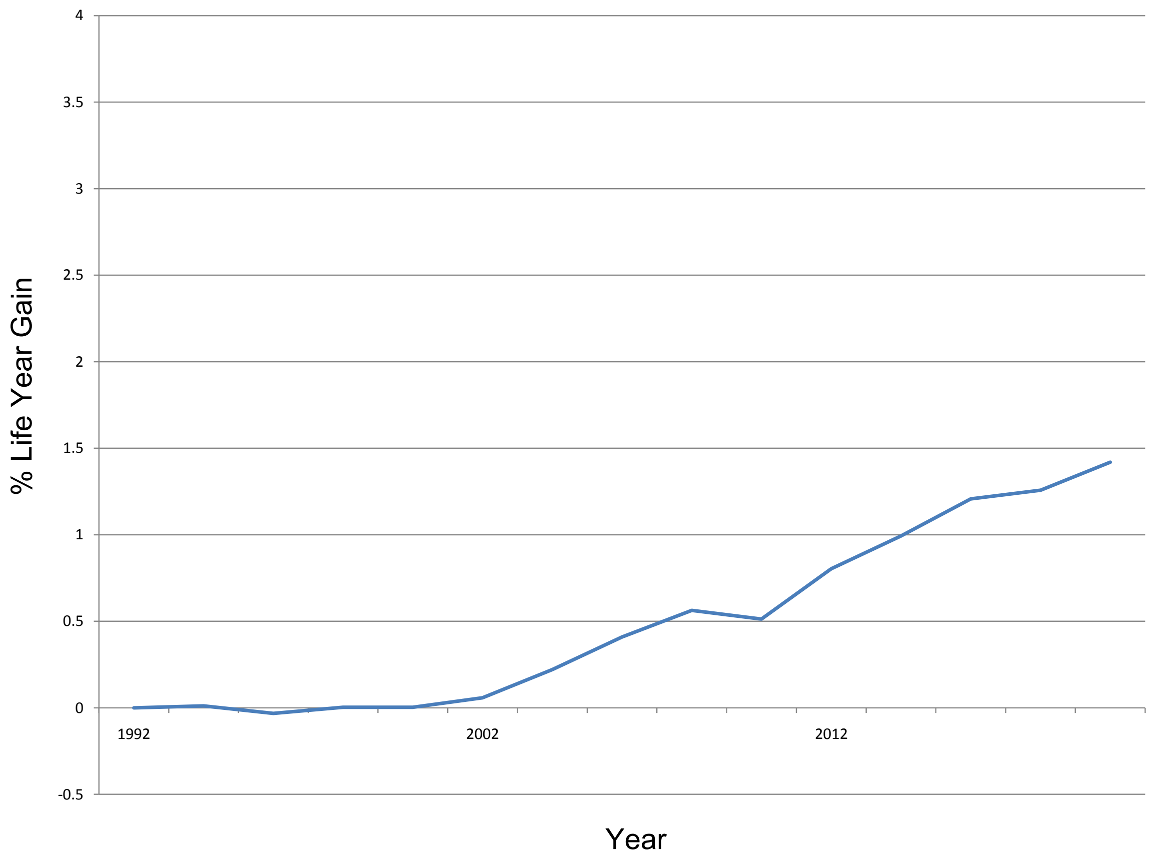
This policy can be simulated and compared with a simulation of the model with no policy modifications.[3] Specifically it can compare the simulated number of life years a base HRS cohort will live to a simulated number of life years an HRS cohort would have had if it were subjected to the education policy.

Figure 4.1 shows the mortality results of this exercise. The data has 2,293 individuals simulated 50 times each for a simulated cohort size of 114,650. The education policy seems to have increases survival rates, however to compare to the rest of this chapter it will be necessary to compute cost benefit ratios for all for the policies. In this case the policy increases the number of years of education obtained by the simulated cohort by 30,276. At a value of $6,692 per year of education this would imply a cost of $202,606,992. The benefit is 10,814 life years over the life span of the simulated cohort for a total undiscounted value of $707,505,950 (1992 dollars) using the statistical value of a life year as described in the prior section.

Since the benefit of such a program happens decades after implementing the program I discount the benefit value. Using a discount rate of 0.95 for policy purposes the discounted total benefit of the policy is $106,048,247 with a benefit to cost ratio of 0.52. For this policy to have a benefit to cost ratio of 1 (break even) the discount rate would have to be 0.967.

[3]The details of these simulations are in appendix A.1.

Figure 4.1: Life Year Gain With a National Compulsory Schooling Law



4.3 Subsidizing College

Another way to interpret the Healthy People 2010 goal of increased education is to try and increase college level attainment. Typically policy levers at the college level come from decreasing various attainment costs. These type of policies have different effects than education mandates. There is a direct and an indirect effect of changing college costs in a dynamic model. The direct effect is simply the change in incentives from increasing the value of attaining a certain year of schooling. The indirect comes from changing the option values of each level of schooling prior to the level targeted by a policy.

For example: imagine if the first year of college became completely free. The direct effect would be that those students who were on the margin between choosing 12 and 13 years of school would choose 13 and the total number of students acquiring at least 13 years of education would go up. The indirect effect can be seen in grade 12 attainment: since the value of the 13^{th} year went up and grade 12 is a pre-requisite for grade 13, the incentives for completing grade 12 also go up to keep the option of attending grade 13 open.

Since prior chapters have made a case that college levels of education should have little effect on life expectancy one might think that a policy targeting college years may not be effective at increasing health outcomes. Indeed subsidizing college levels of attainment may have only a small direct effect on mortality, however the indirect effect of increasing the incentives for lower years of education should enhance the health effects of the policy. The following counterfactual experiment examines two cases: the first is the example above where the first year of college is free, and the second is just halving the education cost for college.

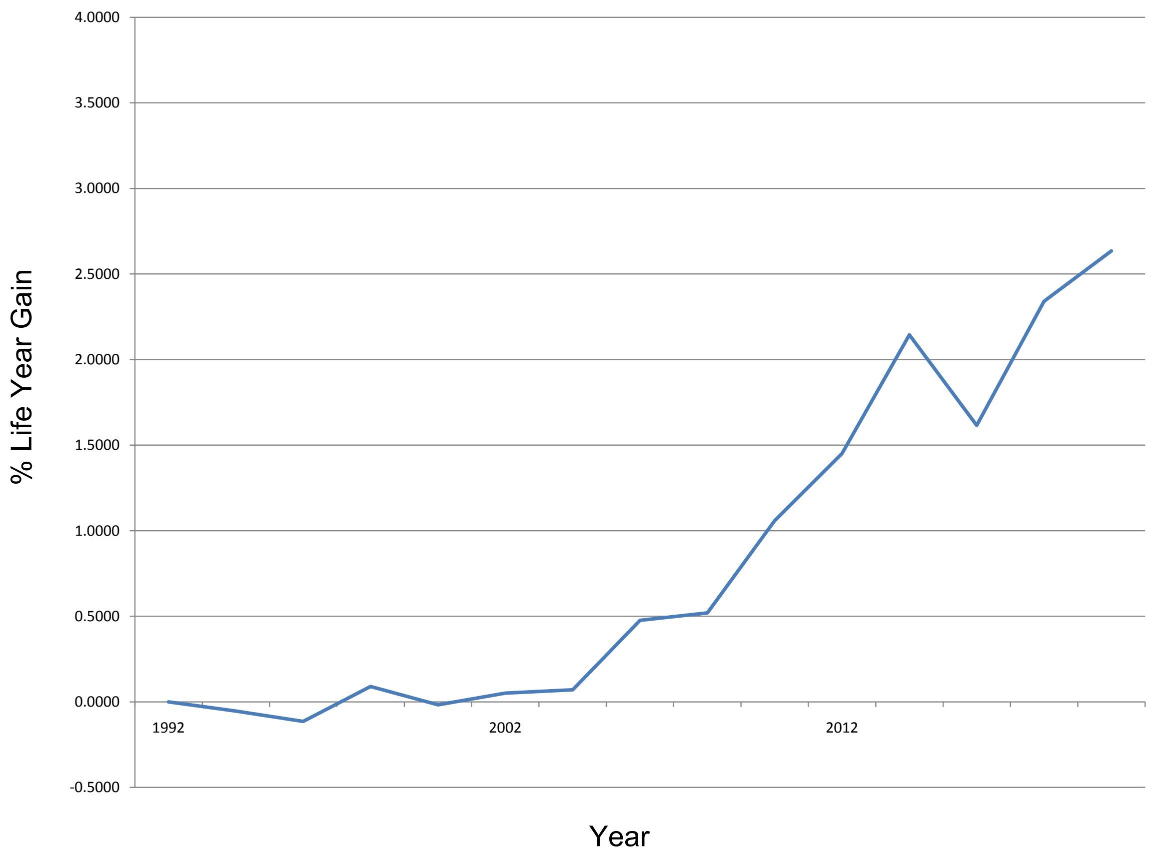
Table 4.4: Education Shift: Free 13^{th} Year

Year	Baseline	Policy
8	0.08	0.07
9	0.07	0.06
10	0.08	0.06
11	0.13	0.08
12	0.23	0.13
13	0.07	0.10
14	0.07	0.11
15	0.07	0.11
16	0.08	0.11
17+	0.12	0.18

Table 4.4 highlights the education effect of just changing the incentives for one particular year of college. Across the board it reduces the amount of people terminating before college and even increases the attendance percentage for years after the thirteenth. The former effect is described above, the latter effect is simply because making the thirteenth year free decreases the total lifetime cost of education for everyone that would have obtained more than thirteen years.

Figure 4.2 shows the survival differential between the policy and the baseline. This policy generates substantial life gain, but it also generates substantial increases in education across the board and this us quite costly. The discounted total increase in life years is valued at $201,172,930 but the cost of

Figure 4.2: Life Year Gain With a Fully Subsidized Thirteenth Year of Schooling



this level of education shift is valued at $1,306,730,254 in 1992 dollars. The benefit to cost ratio in this simulation is 0.154. Break even would require a discount rate of 0.9999.

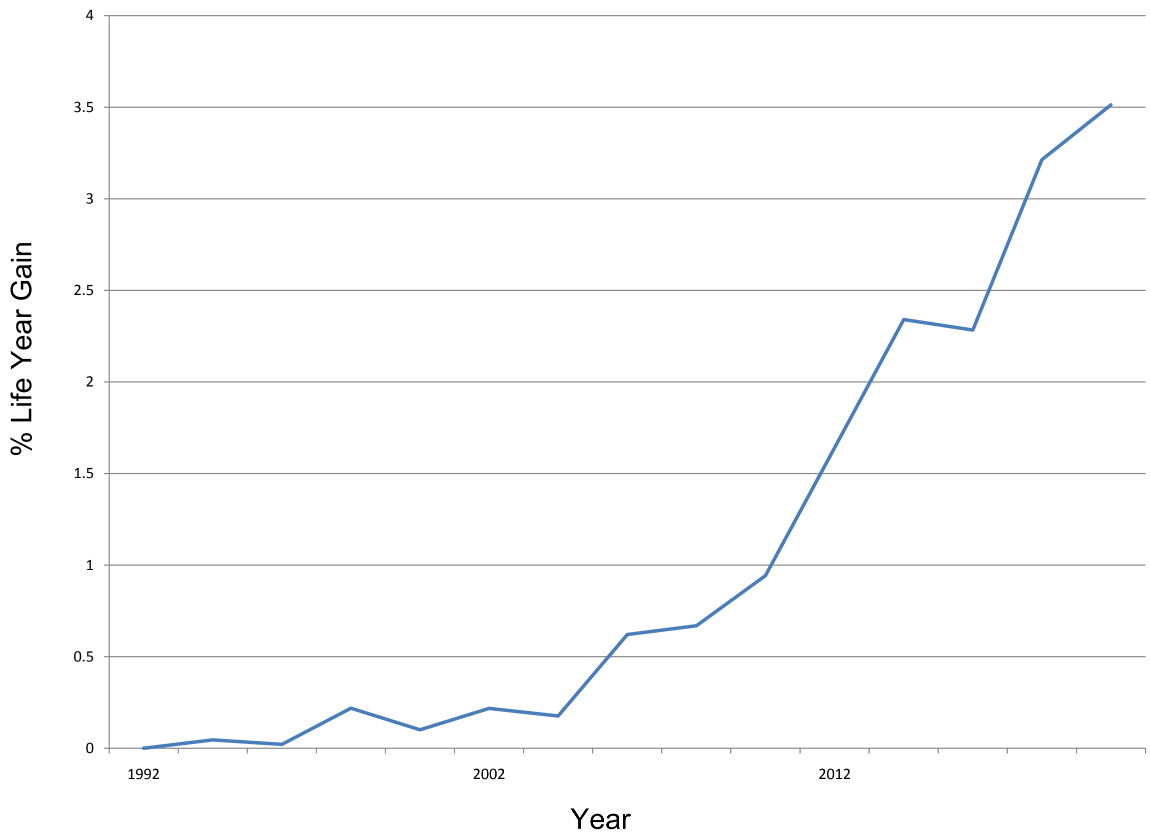
Halving the cost of an undergraduate degree expectedly has an even more pronounced effect. The education shift is shown in table 4.5.

Table 4.5: Education Shift: Half College Cost

Year	Baseline	Policy
8	0.08	0.07
9	0.07	0.05
10	0.08	0.05
11	0.13	0.06
12	0.23	0.09
13	0.07	0.04
14	0.07	0.06
15	0.07	0.11
16	0.08	0.18
17+	0.12	0.30

Figure 4.3 shows the life gain in this sample to be quite large. However this particular policy increases education so much that the benefit to cost ratio is about the same as the first college policy. In this policy the value of the life years gained is $470,508,707 but the 1955 cost of the additional education is $2,569,930,196 for a benefit to cost ratio of 0.183. A discount rate of 0.995 would make this a break even policy. This is a similar result to Cunha and Heckman (2007) who notes that while subsidizing college can have quite large effects, in the end there are usually other policies that are more cost effective.

Figure 4.3: Life Year Gain When College is at Half Estimated Utility Cost



4.4 Attacking Smoking Directly

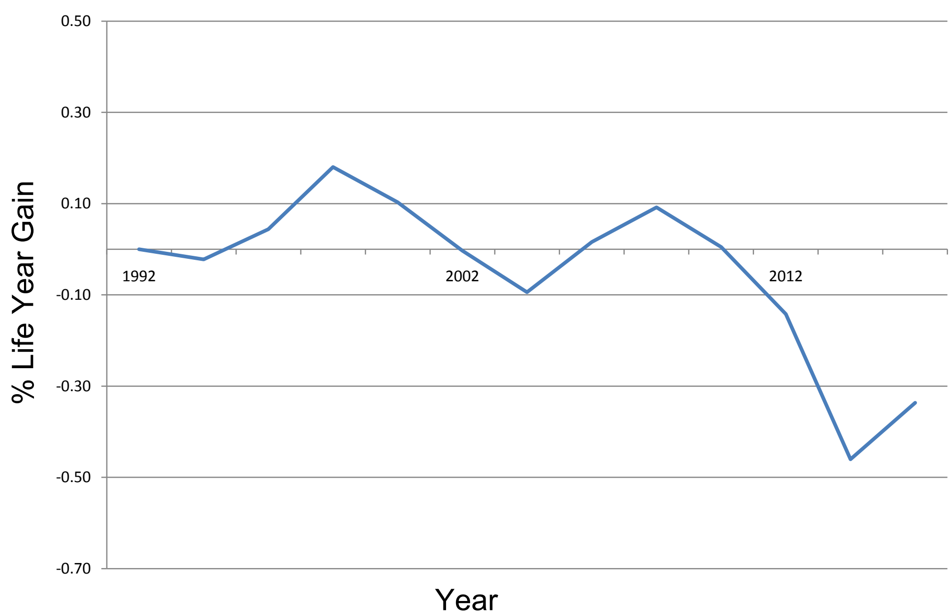
It is well known that smoking is related with strong decreases in both expected years of life and quality adjusted life years. Various studies of the personal cost to smoking (ignoring externalizes) that include life loss show a staggering cost per cigarette.[4] Since this is the case it could be the case that a strong anti smoking program or tax on consumption has a higher health benefit than investing in education from the top. This comparison is not exactly fair (smoking is so detrimental), but it is done as an exercise for comparison against the other policies cost effectiveness.

Figure 4.4 shows the life year gain from a $1000 tax every period a person smokes after 1992. This is relatively ineffective on these people since if they are smoking at this age they really enjoy it. This is the classic case of an inelastic good.

The next experiments are either a cessation program, or general anti smoking programs. I set the effectiveness at 10%. In the simulation these experiments simply take a random draw and if the 10% region is selected and the simulant originally would have smoked; that simulant decides not to in that period. For cessation programs the draw is only taken if the simulant smoked last period and is about to smoke this period. General anti smoking policies do not discriminate. The difference is that cessation programs can be targeted at those who are smoking, and general anti-smoking campaigns might be treating individuals that would never smoke. Keep in mind that the cost values,

[4]Some estimates have been as high as $200 per pack of cigarettes Viscusi (2008)

Figure 4.4: Life Year Gain With a $1000 Tax



while in the right ballpark are just rough estimates for comparisons purpose. A complete analysis of the costs and benefits of various smoking programs is outside of the scope of this project. All smoking cost and benefit numbers have been adjusted to 1992 dollars.

I use a value of \$3.50 per capita per year as the cost for a general anti smoking program that is 10% effective. Inflating this across all years of the simulation generates a cost of \$21,897,356. The life year gain equates to a value of \$177,020,095 for a benefit to cost ratio of 8.08.

I use a value of \$275 per smoker per year as the cost for a general smoking cessation program that is 10% effective. Inflating this across all years of the simulation generates a cost of \$55,299,081. The life year gain equates to a value of \$311,999,394 for a benefit to cost ratio of 5.64.

4.5 Comparison of Policies and Conclusion

Table 4.6: Comparison of Benefit to Cost Ratios ($\beta = 0.95$

Policy	B/C
Comp School	0.52
Free Freshman	0.15
Half College	0.18
Smoking Program	9.50
Smoking Cessation	3.49

Table 4.6 compares the different programs at the base line $\beta = 0.95$ level. While the anti-smoking programs are rough estimates of value, the big take-

Figure 4.5: Life Year Gain With a 10% Effective Smoking Program

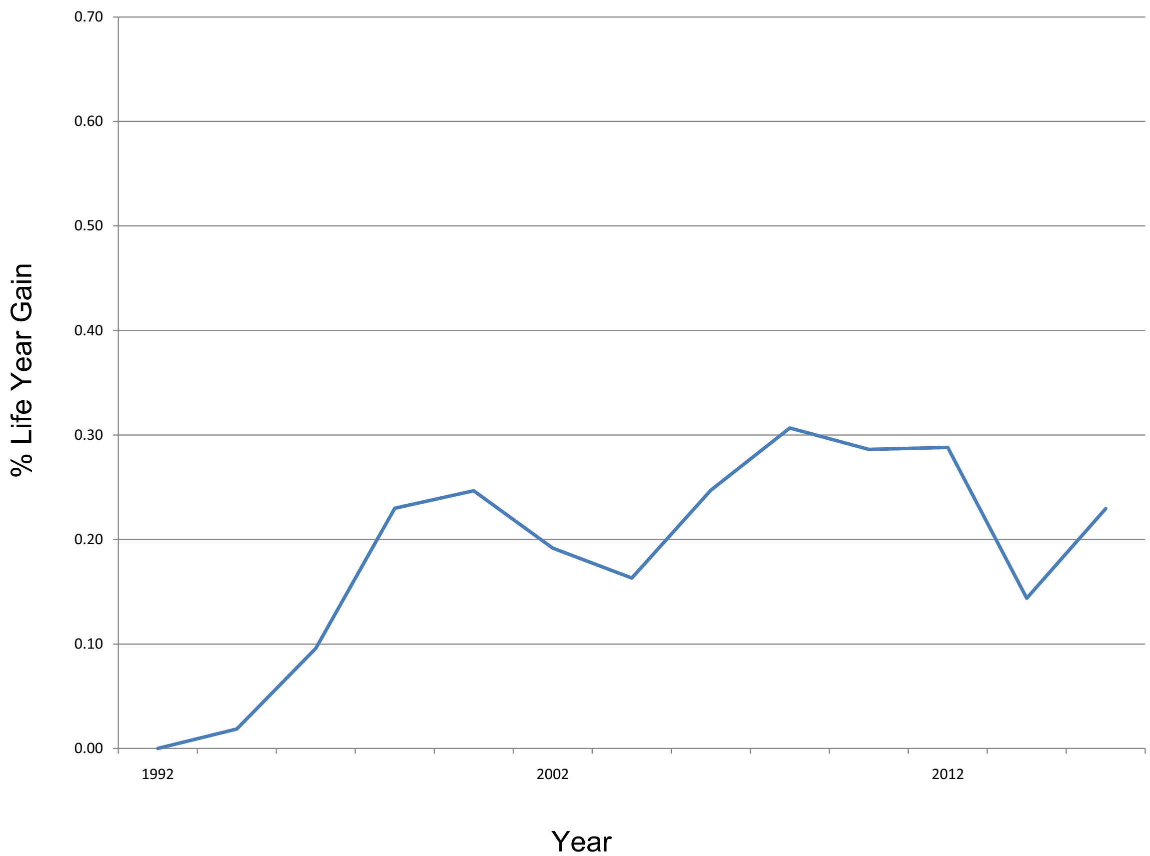


Figure 4.6: Life Year Gain With a 10% Effective Smoking Cessation Program

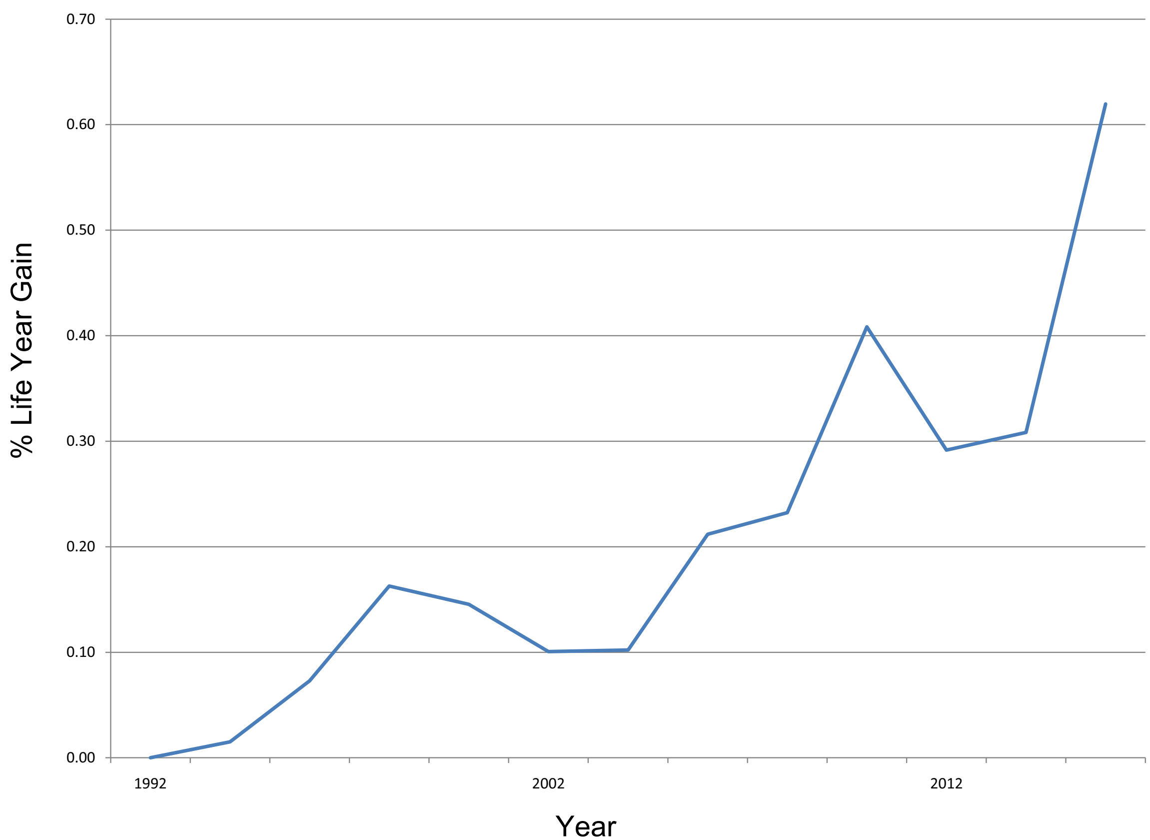


Table 4.7: B/C Ratios Under Different β

β	CS	Free 13	$\frac{1}{2}$ College	Gen Smoke	Smoke Quit
0.9999	3.46	0.99	1.20	8.96	6.25
0.98	1.64	0.47	0.57	8.60	6.00
0.95	0.52	0.15	0.18	8.08	5.64
0.90	0.07	0.02	0.02	7.26	5.06
0.75	0.00	0.00	0.00	5.04	3.52

away here is the magnitude differences between the smoking programs and the education programs.

If the value of β is a concern table 4.7 shows the various benefit to cost ratios at various different values of β. Of note is that the schooling policies are quite sensitive to discount rates. This is because the life benefit is discounted for many years to compare to the education cost in these policies. The smoking policies are not as sensitive, but even for the most forgiving discount rates are significantly higher than education policies. With a discount rate of 0.9 the smoking policies have benefit to cost ratios that are orders of magnitude larger than the education policies.

If indeed the goal is to improve public health, this research argues that attacking the problems directly is bound to be more cost effective. This of course ignores any external benefits from education (it also ignores the reduction of negative externalizes from smoking). It is not to say that both policies should not be explored. Indeed if patience levels are high in the population, many education policies have attractive health benefits. This research only claims that the highest value policies seem to target health behaviors directly.

Appendix

A.1 Simulation

A typical simulation from period t is performed using the following algorithm:

1. **Draw random shocks:** There are several places in which random variables need to be drawn: a draw from a uniform over $(0,1)$ to determine type, 9 type 1 extreme value draws are taken for education shocks, four type 1 extreme value draws are taken for each period from t to T, a draw from a normal distribution with mean 0 and standard deviation depending on the estimation value is drawn for each time period for wage shocks, and a draw from a uniform $(0,1)$ is made for each time period to determine mortality in that period. This is done for each observation s times, where s is the number of simulations run.

2. **Simulate the type:** Using the time invariant data for observation i pull the conditional type distribution from the model. Use the random

shock to determine what type simulant is is.

3. **Simulate the education choice:** Conditional on type and invariant data calculate the contemporaneous utility from each education choice, add in the shock for each choice, and add in the future value of each choice. Pick the highest value of these choices and set that education for observation is.

4. **Simulate wage path:** Conditional on type and education choice for is use estimated wage equation and a starting wage to simulate the next wage. Using the next wage and another shock determine the second wage. Continue until time T and repeat for every simulant.

5. **Simulate health decision:** Starting at time t using the education choice, wage draw, and type, and four utility shocks for simulant is choose the optimal health choice. Set this as the last decision made and calculate this again for time $t + 1$. Repeat until time T.

6. **Simulate mortality:** Given the simulated type, health decision, education decision, and wage path, calculate the survival probability in period t for simulant is. Use the uniform random draw drawn for that simulant at that time to determine if the simulant dies at that period.

These steps generate multiple full sets of data for every observation. The default number of simulations in this dissertation is 50.[1] Since there are 2293

[1] It is noted in the section if a different number is used

observations in the data, this generates a data set with 114,650 simulated individuals.

A.2 2SLS Simulated Estimation Details

Table A.1 lists the statistics used for each census division in section 3.2.3. Manufacturing was constructed by determining the percentage of workers in the 1960 public use micro sample were employed in a 300 level industry (manufacturing). The other two variables are calculated directly from the public use micro sample. All values are in percentages.

Table A.1: 1960 Census Division Statistics

Division	% Urban	% Manufacturing	% Foreign
New England	76	20	10
Mid Atlantic	81	16	11
EN Central	73	23	05
WN Central	58	08	02
S Atlantic	57	08	02
ES Central	48	09	00
WS Central	67	06	02
Mountain	67	06	03
Pacific	81	14	08

Table A.2 lists the summary statistics for the individual characteristics used in section 3.2.3.

Table A.2: Simulant Individual Summary Statistics

Variable	Mean	Std. Dev.
Mother Lived 75	.818	.386
Mother's Education	9.93	3.11

A.3 Policy Costs and Benefit Valuations

The general smoking program costs are derived from the California smoking policies put in place in the early 1990's as described in Pierce et al. (1998). The program reduced smoking rates by approximately 10% at a value of 3.35 per capita per year in 1992.

Elixhauser (1990) reviews several studies on smoking cessation programs. The various programs have wide ranging success rates from 5%-30% and per patient costs between $22 and $400 (1984 dollars). The costs and effectiveness of these type of programs seems to be subject to large uncertainty. So as not to give an unfair advantage to this type of policy I choose conservative estimates of $275 and a 10% success rate.

Viscusi (2008) uses a value of a statistical life year of $100,000 in relation to a $7 million value of a statistical life in 2008. Deflated to 1992 values, I use $65,425 per life year.

Schultz (1960) presents a detailed look at expenditures on education in the United States in 1956. This research includes loss of income from school attendance in the cost of schooling. For secondary schooling he calculates a value of $1421 per year per student in costs in 1956 (adjusted to 1960 dollars).

For college spending he calculates \$3111 per student per year in 1956 (also in 1960 dollars). Inflated to 1992 dollars these values increase to \$6,692 and \$14,652 respectively.

Bibliography

Albouy, V. and L. Lequien, "Does compulsory education lower mortality?," *Journal of Health Economics*, 2009, *28* (1), 155–168.

Angrist, J.D. and A.B. Krueger, "Does compulsory school attendance affect schooling and earnings?," *The Quarterly Journal of Economics*, 1991, *106* (4), 979–1014.

Arcidiacono, P. and J.B. Jones, "Finite Mixture Distributions, Sequential Likelihood and the Em Algorithm," *Econometrica*, 2003, *71* (3), 933–946.

– , H. Sieg, and F. Sloan, "Living Rationally Under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking," *International Economic Review*, 2007, *48* (1), 37–65.

Arendt, J.N., "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification," *Economics of Education Review*, 2005, *24* (2), 149–160.

Ashenfelter, O., "Measuring the Value of a Statistical Life: Problems and Prospects*," *The Economic Journal*, 2006, *116* (510), C10–C23.

Becker, G.S., "Investment in Human Capital: A Theoretical Analysis," *The Journal of Political Economy*, 1962, *70* (S5), 9.

Behrman, J.R. and N. Stacey, *The social benefits of education*, Univ of Michigan Pr, 1997.

Ben-Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," *The Journal of Political Economy*, 1967, *75* (4), 352.

Clark, D. and H. Royer, "The effect of education on adult mortality and health: Evidence from the United Kingdom," *University of Florida (mimeo)*, 2007.

Cunha, F. and J. Heckman, "The technology of skill formation," *The American Economic Review*, 2007, pp. 31–47.

Cutler, D.M. and A. Lleras-Muney, *Education and Health: Evaluating Theories and Evidence*, National Bureau of Economic Research Cambridge, Mass., USA, 2006.

de Walque, D., "Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education," *Journal of Health Economics*, 2007, *26* (5), 877–895.

Deaton, A. and C. Paxson, "Mortality, education, income, and inequality among American cohorts," *Themes in the Economics of Aging*, 2001, pp. 129–65.

Edwards, L.N., "An empirical analysis of compulsory schooling legislation, 1940-1960," *Journal of Law and Economics*, 1978, *21* (1), 203–222.

Elixhauser, A., "The costs of smoking and the cost effectiveness of smoking-cessation programs," *Journal of public health policy*, 1990, *11* (2), 218–237.

Fuchs, V.R., "Time Preference and Health: An Exploratory Study," *Economic Aspects of Health*, 1982, p. 93.

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *The Journal of Political Economy*, 1972, *80* (2), 223–255.

– , "Education and Nonmarket Outcomes," Working Paper 11582, National Bureau of Economic Research August 2005.

Heckman, J. and B. Singer, "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica: Journal of the Econometric Society*, 1984, pp. 271–320.

Heckman, J.J., L.J. Lochner, and P.E. Todd, "Earnings Functions and Rates of Return," *Journal of Human Capital*, 2008, *2* (1), 1–31.

Hurd, M.D. and K. McGarry, "The Predictive Validity of Subjective Probabilities of Survival," *The Economic Journal*, 2002, *112* (482), 966–985.

Imbens, G.W. and J.D. Angrist, "Identification and estimation of local average treatment effects," *Econometrica: Journal of the Econometric Society*, 1994, *62* (2), 467–475.

Keane, M.P., "Structural vs. atheoretic approaches to econometrics," *Journal of Econometrics*, 2009.

– **and K.I. Wolpin**, "The Career Decisions of Young Men," *Journal of Political Economy*, 1997, *105* (3), 473–522.

Kenkel, D., D. Lillard, and A. Mathios, "The roles of high school completion and GED receipt in smoking and obesity," *Journal of Labor Economics*, 2006, *24* (3), 635–660.

Kitagawa, E.M. and P.M. Hauser, *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*, Harvard University Press, 1973.

Kung, Hsiang-Ching, Donna L. Hoyert, Jiaquan Xu, and Sherry L. Murphy, *Deaths: Final Data for 2005*, Vol. 56 of *National Vital Statistics Report* April 2008.

Lleras-Muney, A., "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939*," *The Journal of Law and Economics*, 2002, *45.*

– , "The relationship between education and adult mortality in the United States," *Review of Economic Studies*, 2005, pp. 189–221.

Pierce, J.P., E.A. Gilpin, S.L. Emery, M.M. White, B. Rosbrook, and C.C. Berry, "Has the California tobacco control program reduced smoking?," *Jama*, 1998, *280* (10), 893.

Reinhold, S. and H. Jürges, "Secondary School Fees and the Causal Effect of Schooling on Health Behavior," *Health Economics*, 2009.

Rust, J., "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, 1987, *55* (5), 999–1033.

Sander, W., "Schooling and quitting smoking," *The Review of Economics and Statistics*, 1995, *77* (1), 191–199.

Schultz, T.W., "Capital formation by education," *The journal of political economy*, 1960, *68* (6), 571–583.

Spasojevic, J., "Effects of Education on Adult Health in Sweden: Results from a Natural Experiment," *PhD diss., Graduate Center, City University of New York.*, 2003.

Viscusi, W.K., "How to value a life," *Journal of Economics and Finance*, 2008, *32* (4), 311–323.

Vita

Andrew B. Rice

Andrew Rice was born on November 6, 1980 in Fort Worth, Texas. He received a Bachelor of Science from the Pennsylvania State University in 2003. The next year he enrolled in the Economics department at the same university studying the health and education decisions of individuals. During his studies he was also an instructor of Economics for which he received the Harold F. Martin Graduate Assistant Outstanding Teaching Award.