Intergenerational Long Term Effects of Preschool - Estimates from a Structural Dynamic Programming Model *

James J. Heckman
Department of Economics
University of Chicago
Chicago, IL 60637
mailto:jheckman@uchicago.edu

Lakshmi K. Raut†
Social Security Administration
500 E Street, SW, 9^{th} Floor
Washington, DC 20254
mailto:Lakshmi.Raut@ssa.gov

Abstract

Using the NLSY79 (National Longitudinal Study of Youths, 1979) and the NLSY79 Children and Young Adults datasets, this paper formulates and then empirically estimates an altruistic model of parental preschool investment within a structural dynamic programming framework. The paper provides conditions for identification of the structural parameters of the dynamic programming model and carries out a Lucas-Critique free policy analysis using the estimated structural parameters. The paper empirically estimates the production processes for social, motivational and cognitive skills of a child and the role that the parental preschool investment plays in such production processes. It then examines the effect of a publicly provided preschool policy to disadvantaged children on their educational and labor market achievements, and also on the intergenerational long-term effect on social mobility, schooling mobility, and earnings inequality. The paper calculates the tax burden of such a social contract policy taking into account these intra- and inter-generational effects.

Keywords: Preschool Investment, Early Childhood Development, Intergenerational Social Mobility, College Mobility
JEL Classification Nos.: J24, J62, O15, I21

First Draft: December 2002
This Draft: November 2007

*An earlier draft under was presented at the Western Economic Association Meeting, 2007, University of Southern California, Indian Statistical Institute, The University of Nevada at Las Vegas, and California State University at Fullerton. Comments of the participants of these workshops, especially Juan Pantano's comments as a discussant are gratefully acknowledged

†Raut is an Economist at the Social Security Administration (SSA). This paper was prepared prior to his joining SSA, and the analysis and conclusions expressed are those of the authors and not necessarily those of SSA.

Intergenerational Long Term Effects of Preschool - Estimates from a Structural Dynamic Programming Model

1 Introduction

We formulate an altruistic model of parental preschool investment within a structural stochastic dynamic programming framework. The structural parameters of a structural dynamic programming model are not, in general, all statistically identified. If some parameters are not identified, the estimation of these parameters using the maximum likelihood procedure or any other optimizing procedures may cause serious computational problems. Moreover, the policy analysis based on unidentified parameter estimates is of very little content. In this paper we provide conditions for parametric and non-parametric identification of our structural dynamic programming model, and then use these estimated parameters to examine the effect of preschool on the production of cognitive and non-cognitive skills of children, their effects on school and labor market achievements, and the intergenerational long-term effects on social mobility, schooling mobility and earnings inequality.

In the past two decades in the US the income gap between the rich and the poor and the wage gap between the college educated and the non-college educated workers have been widening. The rate of returns from college graduation is substantially high. The children of poor socioeconomic status (SES) constitute a large proportion of the children who do not complete college. There are Federal education loan programs which provide subsidized loans to the children of poor SES to attend college. The interest rates of these loan programs are substantially lower than the rate of returns from college education. Yet there is not enough demand for these loans and a large percentage of the US children of poor SES do not graduate from college. The liquidity constraint is not a major reason why these kids do not attend college. For instance, Carneiro and Heckman [2002] show that only 4% of the US households are liquidity constrained in the provision of post secondary education. Equalizing education has remained a main policy in the US to reduce poverty and income disparities. Many are, however, highly skeptical about a positive answer to the basic question: ”Can we conquer poverty through school?”

There are many reasons for this skepticism. In the US, education up to high school level is virtually free. Yet many children of poor SES do not complete high school and many of them perform poorly in schools. This naturally beckons to the possibility that the poor quality of the public schools that the children of poor SES attend is the reason for such failings. Improving school quality will improve school performance of these children only marginally. Many empirical studies find that better school quality in terms of lower class size, higher public expenditures per pupil, improved curriculum, and higher desegregation have only marginal effects on school performance of the children of poor SES. See Hanushek [1986] for a survey of the studies along this line.

A growing consensus reached among educators, among media writers (see for instance, Taub [2002], among researchers in economics (see for instance, Heckman [1999] and Currie [2001]) and among researchers in sociology, psychology and education (see for in stance, Barnett [1995], Entwisle [1995], McCormick [1989], Schweinhart et al. [1993]) that children of poor SES are not prepared for college because they were not prepared for school to begin with. The most effective intervention for the children of poor SES should be directed at the preschool stage so that these children are prepared for school and college. The question is then does preschool has long-term positive effects on school performance and labor market success? This is the main issue we address in this paper.

There are two types of quantitative studies on this issue: One set of studies use data on high cost high quality pilot preschool programs such as the High Scope/Perry Preschool Program and the North Carolina Abecedarian Study. These studies find a substantial lasting effect of these programs on school performance and labor market outcomes. The participants in these programs are, however, a very small in number and are not representative of the US population.

The other set of studies use data on Head Start preschool program which is funded by the Federal government. It is available to the children whose parents earn incomes below poverty line. Not all eligible children are covered by the program, however. The quality of the program is very poor compared to the above mentioned pilot programs or most private preschool programs. Some studies find that the Head Start Preschool Program has no long-term effect on children's cognitive achievements and school performance especially for black children. Currie and Thomas [1995] carry out a careful econometric investigation and conclude that the benefits disappear for black children because most of the Head Start black children attend low quality public schools. But after controlling for the school quality, they find significant positive effects of Head Start Preschool Program. See Barnett [1995] for a survey of other studies on the long-term school effect of early childhood programs.

The above studies are not based on nationally representative samples of children, and most studies examine only the effect on school performance such as grade retention and high school and college graduation rates, and do not model parental choice of investing in their children's preschool. In this paper we formulate a model of parental investment in preschool that is guided by economic incentives. We show that preschool benefits children to acquire socialization and motivation skills, especially for the children of poor SES who live in poor HOME environments. We also show that the self-reported measures of motivation skills significantly improve school performance, and the socialization and motivation skills improve the life-time earnings of children, after controlling for their education level, innate ability, and family background. Heckman and Rubinstein [2001] use data on GED testing program in the US, after careful econometric analysis they show that noncognitive skills are important determinant of earnings and educational attainment. We formulate the parental preschool investment decision problem within an intergenerational altruistic framework. We use a mixed reduced form estimation method using the NLSY data and calibration method for some of the parameters, and then use these estimates to examine the long-term intergenerational economic effects of publicly providing preschool to children of poor SES.

The rest of the paper is organized as follows: Section 2 provides the basic decision making framework. Section 3 defines a few notations that are used oftenly in the paper. Section 4 develops our structural model and discusses its estimation strategy. Section 5 provides the identification conditions. Section 6 describes the precise estimation algorithm that we use. Section 7 provides the empirical specifications of the production processes of cognitive and non-cognitive skills and also reports the parameter estimates. Section 8 carries out policy analysis.

2 The Basic Framework

In this section we formulate an econometrically implementable model of preschool investment decision of an altruistic parent in a dynamic programming framework. The preschool investment decision of a parent depends on several other decisions at later stages of a child's life. While we describe each of these decision stages for a better understanding of our framework and for future work, in this paper, however, we restrict only to preschool investment decisions, taking all other decisions as exogenously given. We treat each parent-child pair independently within a family with more than one children. We assume parthenogenetic mode of biological reproduction in our model and with due respect to both genders, we address all individuals in male gender.

2.1 Individual Decision Problem

We assume that an individual's life comprises of several discrete periods during which important life-cycle events relevant to leaning and earning occur. While it may be more realistic to have finer divisions of these periods, for analytical tractability and given data limitations, we aggregate the whole life-cycle into four periods as follows: [0-4], [5-16], [17-25], [26--]. In each of these periods some educational and labor market decisions are made and outcomes are observed. During age [0-5], a parent invests in his child's preschool activities which develop the child's school readiness, and cognitive, social, and motivational skills. Let a denote the parental preschool investment decision. At the end of the preschool period, the child acquires a level of innate ability or cognitive skill τ, social skill σ motivational skill μ, self-esteem skill, η, and internal self-control skill ϕ. The levels of these skills that a child develop depend on various other childhood interventions, for instance, on the child-rearing practices at home, the nature of neighborhood in which the child grows up, and the level of schooling, cognitive, socialization and motivational skills of the parent. We do not, however, explicitly include these additional determinants of skill formation in this paper to keep computations manageable .

During ages [5-16], the child goes to school. The school performance at this stage depends on his level of τ, $\sigma$and μ that the child has acquired during the previous stage, on the quality of the school that he attends, and the type of neighborhood kids whom the child mingles with. It also depends on the parental home inputs such as how many hours the parent spend time with the child to do his homework, how many hours the child watches TV, and how stable and stimulating the relationships among the family members. Many of these are choice variables for the parent. We again do not include these factors to simplify our computations.

During ages [17-25] the child decides whether to complete college education or not, which depends on his parent's income, his learned and innate abilities. We take this decision exogenously given, and denote by a variable s. During ages [26-] he works, forms hid family with a child and decides how much to invest in his preschool.

We now formulate the optimal preschool investment choice problem of the parent. We assume that while an individual's permanent yearly earnings is an important determinant of the level of preschool investment, there are life cycle events other than ε_s that may influence this decision. We represent all these events by a random variable ε_p. Furthermore, individuals are assumed to differ in their taste which affect individual's choices. The taste variation is represented as a random variable ε_t. We bundle all these unobserved sources of heterogeneity among individuals into a vector ε. The state variables of our system are represented by the vector $z = (\tau, \sigma, \mu, \eta, \phi, s, \varepsilon)$. We denote the observable components of the state variable by $x = (\tau, \sigma, \mu, \eta, \phi, s)$ and use the notation $z = (x, \varepsilon)$ to represent the above information. For any variable w, we adopt the convention of using w if it refers to a parent and w' if it refers to his child.

We assume that given his parental preschool investment decision a, and a realization of his parent's state variables $z = (x, \varepsilon)$, the components of a child's state variable, $\tau', \sigma', \mu', \eta', \varphi'$ and ε' are generated stochastically by the following conditional probability density functions:

$$\begin{gathered} q_{\tau}\left(d\tau'|\tau, s, a\right) \\ q_{\sigma}\left(d\sigma'|\tau', \tau, \sigma, \mu, s, a\right) \\ q_{\mu}\left(d\mu'|\tau', \tau, \sigma, \mu, s, a\right) \\ q_{\eta}\left(d\eta'|\tau', \tau, \sigma, \mu, s, a\right) \\ q_{\varphi}\left(d\varphi'|\tau', \tau, \sigma, \mu, s, a\right) \\ q_{s}\left(ds'|\tau', \sigma', \mu', s, a\right) \\ g\left(d\varepsilon'|\tau', \sigma', \mu', s'\right) \end{gathered} \tag{1}$$

In the above specifications of the conditional probabilities, the conditioning variables conform to what we know about the production processes of these state variables. We will discuss details of each production process in section 7.2. Given the density functions in the system of Equations (1), the transition probability measure $p\left(dx', d\varepsilon'|x, \varepsilon, a\right)$ over the states of our system is determined.

Given a parent's $z = (x, \varepsilon)$, we assume that his life-time average annualized earnings is $w(x, \varepsilon)$. Let A be the set of all possible choices that any individual may make. We assume it to be an ordered set. Assume that the annualized average cost to parent for making a preschool investment choice a is $\theta(a)$, $a \in A$. The annualized consumption given a choice a is then $c(w, a) \equiv w - \theta(a)$ .The choices of a parent with observable characteristics x are restricted to the set $A(x, \varepsilon) \equiv \{a \in A | c(w(x, \varepsilon), a) > 0\}$. The choice a yields direct utility from life-time annualized consumption and indirect utility through its effect on child outcome and welfare, as represented in the following Bellamn equation corresponding to the parent's preschool investment decision problem

$$V(x, \varepsilon) = \max_{a \in A(x, \varepsilon)} u(x, \varepsilon, a) + \beta \int V\left(x', \varepsilon'\right) p\left(dx', d\varepsilon'|x, \varepsilon, a\right) \tag{2}$$

where $V(.)$ is the intergenerational welfare function known in the dynamic programming literature as the value function, $u(.)$ is the felicity index of yearly permanent consumption over the whole lifetime of the parent, and the parameter β measures the degree of parental altruism towards the child.

Under general regularity conditions on $u(.), p\left(dx', d\varepsilon' | x, \varepsilon, a\right)$, and β the value function $V\left(x, \varepsilon\right)$, and a measurable optimal decision rule $h^*\left(x, \varepsilon\right)$ exist (see, for instance, Bhattacharya and Majumdar(1989, Theorem 3.2).

Given $u(.), p\left(dx', d\varepsilon' | x, \varepsilon, a\right)$, and β satisfying the regularity conditions, we carry out a Lucas-Critique free policy evaluation by examining a policy's effect on the individual optimal decision, a, on the intergenerational welfare level V and also examine the intergenerational long-run aggregate effect on the economy by aggregating individual choices with respect to the long-run , also known as, invariant, population distribution of the equilibrium transition probability distribution $p\left(dx', d\varepsilon' | x, \varepsilon, a^*\left(x, \varepsilon\right)\right)$.

To be able to do this, we need to estimate the structural parameters. Our data consists of a sample of parent-child pairs with information on parent's observable state x, child's observable state x', parent's permanent income w, and parent's preschool investment decision a. Suppose a vector of parameters ξ_p specifies the probability distributions in Eq. (1), i.e., given ξ_p, the transition probability distribution $p\left(dx', d\varepsilon' | x, \varepsilon, a\right)$ is determined. Our problem is then to statistically estimate the structural parameters $\zeta = \left\{u(.), \xi_p, \beta\right\}$ given observable information on a random sample of parent-child pairs $y = \{(x_i, x_i'), a_i\}_{i=1}^{n}$ such that the predicted behaviors of the sample from the model are close to observed behavior. We denote the log-likelihood function of the sample by $\mathcal{L}_y(\zeta)$. Estimation of the model involves two steps: For a given ζ, calculate the probability distribution of the endogenous variables $a_i | x_i$ and $x_i' | x_i, a_i$ using the model to form the log-likelihood of the sample $\mathcal{L}_y(\zeta)$ and then use an appropriate estimation procedure to choose a ζ.

Two questions need be addressed to that end: First, is the computation of the likelihood $\mathcal{L}_y(\zeta)$, which involves solving the dynamic programming problem in Eq. (2) repeatedly for each (x, ε), feasible with the currently available computing technology, especially when ε is a continuous multivariate random variable? Second, are the structural parameters of the model identified (the precise definition of identification stated later)?

The answer to both questions is in general no. Following the literature, we make simplifying assumptions to transform the above structural dynamic programming problem into a random utility model of discrete choices. We will show that these assumptions greatly simplify the computation, and the identification conditions. Given those assumptions, we will see that (1) the structural parameters ξ_p determine the transition distribution $p(x'|x,a)$ of the observable state variables, which is the mixture distribution of the original transition distribution, more specifically $p(x'|x,a) = \int p(x', \varepsilon'|x, \varepsilon, a)\ d\varepsilon|x\ d\varepsilon'|x'$, and that (2) the optimal choice probabilities $P(a|x), a \in A(x), x \in X$ that is used to define the observed discrete choices depend on ξ_p through $p(x'|x,a)$. Given that optimal choice a is treated as an exogenous variable in the estimation of $p(x'|x,a)$, maximization of joint likelihood of two components is more efficient. To make estimation task computationally manageable, however, again following the trend in the literature, in place of ξ_p, we take an estimate of $p(x'|x,a)$ as a fixed parameter in the vector of parameters ζ, and in place of β, we calibrate β from other studies, and then form the likelihood of the sample of observed discrete choices $a_i|x_i$ for identification and estimation of the remaining parameters.

3 Notations

In the rest of the paper, our parameter vector is $\zeta = \{u(x,a), p(x'|x,a), \beta\}, a \in A(x), x \in X$ where $p(x'|x,a)$ and β are fixed. Denote by Ξ the set of all such parameter values. We denote by $\mathcal{L}_y(\zeta)$ the log-likelihood of the sample of observed choices $y = \{a_i|x_i, i = 1...n\}$. The log-likelihood function $\mathcal{L}_y(\zeta)$ is defined given a set of conditional choice probabilities $\{P(a|x), a \in A(x), x \in X\}$ which depend on ζ.

Let J_x denote the number of elements in the feasible choice set $A(x)$. Denote by $J = \sum_{x \in X} J_x$. Assume that X is a finite ordered set of M elements.

Denote by $F(a) = [f(x'|x,a)]_{x',x \in X}$ the $J_x \times J_{x'}$ conditional transition probability matrix given a choice $a \in A(x)$ where the element $f(x'|x,a)$ corresponding to the row x and the column x' is the probability of the child moving to state x' given that his parent is from the state x and he had made a choice $a \in A(x)$. We denote by $F(x,a)$ the row vector of $F(a)$ corresponding to the parent's state x.

The vector of conditional choice probabilities denoted by $\mathcal{P} = \{P(a|x), a \in A(x), x \in X\}$ is ordered by the primary index of ordering in X and the secondary index of the ordering in A. For each x, the component vector of conditional choice probabilities $\{P(a|x), a \in A(x)\}$ belongs to a $J_x - 1$ dimensional simplex. The vector $\mathcal{P}$ of all conditional choice probabilities is a member of $\Re^{MJ}_{++}$ but restricted to the interior of the M-fold cross product of the $J_x - 1$ dimensional simplices, which we denote by $\triangle$.

For any function $v(x,a)$, its vector representation is a $J \times 1$ vector v (i.e., with the same symbol v) in which function values $v(x,a)$'s are ordered in the same way as $\mathcal{P}$. For any scalar or a vector function $w(x)$, we denote by w (again using the same symbol w to denote it) the values of $w's$ stacked in rows in the same order as in the ordered set X.

For any random vector or a random variable $w(x,a)$, we denote its expectation with respect to a by $\bar{w}(x)$, i.e., $\bar{w}(x) \equiv \sum_{a\in A(x)} w(x,a)P(a|x)$, (with the convention that when w is a random vector, the product in this summation is element-by-element). Define the $M\times J$ matrix Π derived from a vector of conditional choice probabilities $\mathcal{P}$ by

$$\underset{M\times J}{\Pi} = \begin{pmatrix} P(a=1|x_1) & ... & P(a=J|x_1) & ... & 0 & ... & 0 \\ 0 & ... & 0 & ... & 0 & ... & 0 \\ 0 & ... & 0 & ... & P(a=1|x_M) & ... & P(a=J|x_M) \end{pmatrix}$$

and the transition matrices in matrix notation as a $J\times M$ matrix F as,

$$\underset{J\times M}{F} = \begin{pmatrix} f(x_1'|x_1,a=1) & ... & f(x_M'|x_1,a=1) \\ & ... & \\ f(x_1'|x_1,a=J_{x_1}) & ... & f(x_M'|x_1,a=J_{x_1}) \\ & ... & \\ f(x_1'|x_M,a=1) & ... & f(x_M'|x_M,a=1) \\ & ... & \\ f(x_1'|x_M,a=J_{x_M}) & ... & f(x_M'|x_M,a=J_{x_M}) \end{pmatrix}$$

4 Structural Estimation

Following Rust[1994] we make the following simplifying assumptions to transform the original model in Eq. (1) to a random utility model, and then explore various computational simplifications that recently appeared in the literature.

We assume that $w(x,\varepsilon)$ and hence $A(x,\varepsilon)$ does not depend on ε, i.e., $w()$ does not contain any unobservable idiosyncratic shocks, i.e., ε_p and ε_s are absent in our model. However, we assume that ε represents a taste shifter for individual preferences and constitutes our only source of unobserved heterogeneity, the specific nature of which is stated formally in the following assumption.

Assumption 1 *$u(x,\varepsilon,a) = u(x,a) + \varepsilon(a)$, and support of $\varepsilon(a)$ is the real line for all $a \in A(x)$.*

We also make the following additional assumptions.

Assumption 2 *The transition probability $p(x',\varepsilon'|x,\varepsilon,a) = g(\varepsilon'|x')f(x'|x,a)$, for some density function g with finite first moment, and twice continuously differentiable.*

Assumption 3 *The set of observable individual characteristics* $X = \{x^1, ..x^M\}$ *is a finite ordered set.*

Under assumptions 1 - 3 , we have

$$V(x,\varepsilon) = \max_{a\in A(x)} u(x,a) + \varepsilon(a) + \beta \sum_{x'\in X} \int V(x',\varepsilon') g(d\varepsilon'|x') f(x'|x,a) \tag{3}$$

Denote the value function after integrating out the unobservable component of the state variable by $v(x) \equiv \int V(x,\varepsilon) g(d\varepsilon|x)$. Integrating both sides of Equation (3) with respect to the conditional density $g(d\varepsilon|x)$, and utilizing this notation for $v(x)$, we have

$$v(x) = \int \max_{a\in A} [\tilde{v}(x,a) + \varepsilon(a)] g(d\varepsilon|x) \tag{4}$$

where

$$\begin{aligned} \tilde{v}(x,a) &\equiv u(x,a) + \beta \sum_{x'\in X} v(x') f(x'|x,a) \\ &= u(x,a) + \beta F(x,a).v \end{aligned} \tag{5}$$

Eq. (4) above is a random utility model in which the function $\tilde{v}(x,a)$ measures the common utility that an individual of observable characteristics x derive from a choice $a \in A(x)$. Denote by

$$\Omega(x,a) = \{\varepsilon | \tilde{v}(x,a) + \varepsilon(a) \geq \tilde{v}(x,a') + \varepsilon(a'), \text{for all } a' \in A(x)\} \tag{6}$$

the set of individuals with observed characteristics x who made a as their optimal choice. The conditional choice probabilities are then given by

$$P(a|x) = \int_{\Omega(x,a)} g(d\varepsilon|x). \tag{7}$$

By partitioning the domain of integral (4) into disjoint regions, $\Omega(x,a), a \in A(x), x \in X$, and then integrating we have the following,

$$\begin{aligned} v(x) &= \sum_{a\in A(x)} P(a|x) \left[u(x,a) + \frac{\int_{\Omega(x,a)} \varepsilon(a) g(d\varepsilon|x)}{P(a|x)} + \beta \sum_{x'\in X} v(x') f(x'|x,a) \right] \\ &= \sum_{a\in A(x)} P(a|x) [u(x,a) + e(x,a) + \beta F(x,a)\cdot v] \text{ ...(*)} \\ &= \bar{u}(x) + \bar{e}(x) + \beta \bar{F}(x) \cdot v \end{aligned} \tag{8}$$

where

$$e(x,a) \equiv \int_{\Omega(x,a)} \varepsilon(a)\, g(d\varepsilon|x) \,/\, P(a|x) \tag{9}$$

in line (*) is the conditional expectation of the component $\varepsilon(a)$ of the random vector ε given x and a. Writing the above in matrix notation, we have

$$v = \bar{u} + \bar{e} + \beta \bar{F} \cdot v \equiv \Phi(v,\zeta) \tag{10}$$

Let $v(\zeta)$ be a fixed point $v = \Phi(v,\zeta)$ of the map $\Phi(v,\zeta)$ for given $\zeta \in \Xi$, and denote by $\mathcal{P}(v)$ the conditional choice probabilities in Eq. (7) for given a value function v. Then the computation of the likelihood of the sample is simplified to the computation of the fixed point of the above map $\Phi(v,\zeta)$. Following the line of argument in Rust [1994] it can be shown that for each ζ, there exists a unique fixed point $v(\zeta)$ and it is the limit of the following iterative process:

For a given $\zeta \in \Xi$,

Step 1: start with an initial $v_0 \in \Re^M$.

Step 2: Compute $\mathcal{P} \in \triangle$ using Eq. (7), and then compute $v_1 = \Phi(v_0,\zeta)$ from Eq. (10). If $v_1 = v_0$ stop, denote this common value as $v(\zeta)$. Retain $\mathcal{P}(v(\zeta))$ to calculate the log-likelihood of the sample $\mathcal{L}_y(\mathcal{P}(v(\zeta)))$. If $v_1 \neq v_0$ go to Step 1 setting the initial $v_0 \equiv v_1$.

We denote the above iterative process symbolically as $\underset{v=\Phi(v,\zeta)}{\text{Solve}} \mathcal{L}_y(\mathcal{P}(v(\zeta)))$. Given the solution v and the corresponding vector of conditional choice probabilities $\mathcal{P}(v(\zeta)$ of this iteration process, the log-likelihood is of the sample is now defined and the maximum likelihood estimation procedure could be carried using the log-likelihood function. For convenience of exposition, we represent the above two stages of finding a maximum likelihood estimate of the structural parameters symbolically as

$$\underset{\zeta\in\Xi}{\text{Max}} \circ \underset{v=\Phi(v,\zeta)}{\text{Solve}} \mathcal{L}_y(\mathcal{P}(v(\zeta))) \tag{11}$$

In the above algorithm, the computation of $P(a|x)$, and $e(x,a)$ involve multi-dimensional numerical integration, which may make computations extremely slow. Both computational tasks are, however, substantially simplified under the following assumption:

Assumption 4 *The components of ε are independently and identically distributed as extreme value distribution with location parameter 0 and scale parameter 1.*

McFadden (1981) has shown that under Assumption 4., $e(x,a) = (\lambda - \ln P(a|x))$,where λ is the Euler-Mascheroni constant, with a numerical value of $\lambda = 0.57721566$, and the conditional choice probability $P(a|x)$ has the following Logit representation,

$$P(a|x) = \frac{e^{\tilde{v}(x,a)}}{\sum_{a \in D} e^{\tilde{v}(x,a')}} \tag{12}$$

The above strategy of computational simplification was pioneered by Rust [1987].

The computational burdens could be, however, further simplified as follows: From Eq. (10) it follows that $v = \left[I_M - \beta \bar{F}\right]^{-1} [\bar{u} + \bar{e}]$. Substituting this in Eq. (5), we have

$$\tilde{v}(x,a) = u(x,a) + \beta F(x,a) \left[I_M - \beta \bar{F}\right]^{-1} [\bar{u} + \bar{e}] \tag{13}$$

It is easy to see that given $\mathcal{P}_0 \in \triangle$, the right hand side of the above and hence a new vector of conditional choice probabilities say $\mathcal{P}_1 \in \triangle$ in Eq. (12) can now easily be computed. We represent this relationship for each structural parameter $\zeta \in \Xi$ by $\mathcal{P}_1 = \Psi(\mathcal{P}_0, \zeta)$. Following the line of argument in Aguirregabiria and Mira (2002) it is easy to show that for each $\zeta \in \Xi$, there exists a unique fixed point $\mathcal{P}(\zeta)$ to the mapping $\Psi(\mathcal{P}_0, \zeta)$, and starting from any initial $\mathcal{P}_0 \in \triangle$ the iterative process $\mathcal{P}_{n+1} = \Psi(\mathcal{P}_n, \zeta)$, $n \geq 0$ converges to the fixed point $\mathcal{P}(\zeta) \in \triangle$. We symbolically denote this iteration solution process as $\text{solve}_{\mathcal{P}=\Psi(\mathcal{P},\zeta)}$. Then the maximum likelihood estimation procedure can be again symbolically represented as

$$\underset{\zeta \in \Xi}{\text{Max}} \circ \underset{\mathcal{P}=\Psi(\mathcal{P},\zeta)}{\text{Solve}} \mathcal{L}_y(\mathcal{P}(\zeta)) \tag{14}$$

A form of the the above procedure is used by Hotz and Miller (1993). Notice that the iteration step in Eq. (14) is in the conditional choice probability space which is of dimension $J - M$. This iteration process combines computations of P and v of the previous iteration step into one step and may not, however, lead to any reduction in the number of numeric operations and thus may not be any more efficient than the previous procedure.

Aguirregabiria and Mira (2002), however, introduced a further computational simplification by interchanging the order of those two computations in Eq. (14) as follows:

$$\underset{\mathcal{P}=\Psi(\mathcal{P},\zeta^*(\mathcal{P}))}{\text{Solve}} \circ \underset{\zeta \in \Xi}{\text{Max}} \mathcal{L}_y(\mathcal{P}(\zeta)) \tag{15}$$

where $\zeta^*(\mathcal{P})$ is the argmax of the inner maximization problem for a given $\mathcal{P} \in \triangle$. Aguirregabiria and Mira provide evidence on computational efficiency of their maximum likelihood estimation procedure and also study the asymptotic distribution theory of their maximum likelihood estimator. We follow this estimation procedure in section 6 after addressing the identification issues.

5 Identification of Structural Parameters

In the previous section we saw that given $\zeta \in \Xi$, there exists a unique likelihood function $\mathcal{L}_y(P(\zeta))$. To be able to estimate $\zeta \in \Xi$, the model should be identified in the sense that

$$\mathcal{L}_y(\zeta) = \mathcal{L}_y(\zeta') \text{ } a.e. \text{ if and only if } \zeta = \zeta', \tag{16}$$

the a.e. is with respect the dominant probability measure defining the likelihood of the sample. Following Prakasa Rao (1992), we say that our model is *globally identified* if the relationship in Eq. (16) holds for any two $\zeta, \zeta' \in \Xi$, and is *locally identified* around a particular parameter $\zeta \in \Xi$, if the relationship in Eq. (16) holds for all $\zeta' \in \Xi$ in a neighborhood of ζ.

It is well-known that in general the structural parameters are not identified in structural dynamic programming problems (see Rust [1994]). To find reasonable conditions for identification, note from Eq. (6) that the optimal choices are invariant if we add a location m and divide both sides by a scale factor $\sigma > 0$, for each $x \in X$. Thus it follows that we can recover the utility function only up to a scale which and location. Given this fact, we restrict the one period utility function $\{u(x,a), a \in A(x)\}$ to lie in a $J_x - 1$ dimensional open submanifold of $\Re^{J_x}$for each $x \in X$. We take each possible utility vector $\{u(x,a), a \in A(x), x \in X\}$ to lie in the cross product (or equivalently in the direct sum, if we view $\Re^{J_x}$ to be embedded in $\Re^J$) of these $J_x - 1$ dimensional submanifolds over all $x \in X$. There are many such manifolds, and up to diffeomorphisms they are all equivalent. We define one such manifold $\mathcal{U}$ using a map $\varphi : \triangle \ni \mathcal{P} \mapsto u \in \Re^J$ (which reads as, φ takes a member $\mathcal{P}$ in $\triangle$ to a member u in $\Re^J$) defined by

$$u = \left[I_J + \beta F \left(I_M - \beta \bar{F} \right)^{-1} \Pi \right]^{-1} [\tilde{v} - \tilde{e}] \equiv \varphi(\mathcal{P}) \tag{17}$$

where $\tilde{v}(x,a) = \ln P(a|x)$ and $\tilde{e} = \beta F \left(I_M - \beta \bar{F} \right)^{-1} \Pi e$. Take $\mathcal{U} = \varphi^{-1}(\triangle)$. It can be shown that the set $\mathcal{U}$ is an $J - M$ dimensional smooth manifold. Given parameters β, and F fixed, we restrict our parameter space Ξ to be such that the u-component of a parameter vector $\zeta \in \Xi$ is restricted to lie in $\mathcal{U}$. The most general non-parametric family that we can restrict our parameters u to lie in is $\mathcal{U}$. Our nonparametric identification issue boils down to the question, under what conditions can we identify our structural model in this non-parametric family of $\mathcal{U}$? Theorem 1 addresses this, using the following assumption

Assumption 5 , *Given transition probabilities* F, *the degree of altruism parameter* β *is such that (1)* $0 \leq \beta < 1$ *and (2)* $I_J + \beta F \left(I_M - \beta \bar{F} \right)^{-1} \Pi$ *is of full rank.*

Note that there always exist such $\beta's$ at least near $\beta = 0$ and that $\beta = 1$ will violate condition (2) since in that case $I_M - \beta\bar{F}$ is not invertible as each row will add-up to zero.

Theorem 1 (Non-parametric Identification) *Suppose the components β, and F of the parameter vector are fixed. Let $\mathcal{P} \in \triangle$ be a vector of conditional choice probabilities that satisfy Assumption 5. Then there exists a unique utility function $\{u(x,a), a \in A(x), x \in X\} \in \mathcal{U}$ that generates $\mathcal{P}$ as the optimal solution to the choice problem in Eq. (2). Furthermore, the model in Eq. (2) is globally or locally non-parametrically identified depending on whether Assumption 5 holds globally or locally.*

Proof. Let $\mathcal{P} \in \triangle$ be a vector of conditional choice probabilities that satisfy Assumption 5. Note that writing Eq. (13) in matrix notation, we have $\tilde{v} = \left[I_J + \beta F \left(I_M - \beta\bar{F}\right)^{-1} \Pi\right] u + \beta F \left(I_M - \beta\bar{F}\right)^{-1} \Pi e$, where $\bar{F}$ is the expectation of $F(a)$ with respect to $\mathcal{P}$. Taking $\tilde{v}(x,a) \equiv \ln P(a|x)$, and denoting by $\tilde{e} = \beta F \left(I_M - \beta\bar{F}\right)^{-1} \Pi e$, we have

$$u = \left[I_J + \beta F \left(I_M - \beta\bar{F}\right)^{-1} \Pi\right]^{-1} [\tilde{v} - \tilde{e}] \tag{18}$$

Thus by Assumption 5, for each $\mathcal{P}$ there exists a unique $u \in \mathcal{U}$.

We now prove the second part regarding the nonparametric identification. Note that the data on distribution of choices given a fixed number of individuals $n(x)$ (a positive integer) for each observed value of individual characteristics $x \in X$ can be summarized as an ordered vector y defined similar to $\mathcal{P}$ by $y = (n(a|x), a \in A(x), x \in X)$ where $n(a|x)$ is the number of individuals who chose $a \in A(x)$ given their characteristics $x \in X$. The likelihood of the sample can be written as follows

$$\begin{aligned}
L_y(\mathcal{P}) &= \prod_{x\in X} \frac{n(x)!}{\prod_{a\in A(x)} n_a(x)!} \exp\left(\sum_{x\in X} n(x) \ln\left(1 - \sum_{a=1}^{J_x-1} P(a|x)\right)\right) \times \\
&\quad \exp\left(\sum_{x\in X}\sum_{a=1}^{J_x-1} n(a|x) \ln\left(\frac{P(a|x)}{1 - \sum_{a=1}^{J_x-1} P(a|x)}\right)\right) \\
&= h(y)\, g(\eta) \exp(y'\eta), \text{ where } \eta = (\eta(a|x), a \in A(x), x \in X), \text{with} \\
\eta(a|x) &= \ln\left(\frac{P(a|x)}{1 - \sum_{a=1}^{J_x-1} P(a|x)}\right), \text{and } g(\eta) = -\sum n(x) \ln\left(1 + \sum_{a=1}^{J_x-1} \exp \eta(a|x)\right),
\end{aligned}$$

and $h(y)$ is the multiplicative component in the first expression. It follows from the above that $L_y(\mathcal{P})$ is an exponential distribution. The determinant $\det(\mathfrak{I}(\mathcal{P}))$ of the Fisher information matrix $\mathfrak{I}(\mathcal{P})$ of $L_y(\mathcal{P})$ at any parameter vector $\mathcal{P} \in \triangle$ can be shown to be $\det(\mathfrak{I}(\mathcal{P})) = \left[\prod_{x\in X}\prod_{a=1}^{J_x-1} P(a|x)\right]^{-1}$, which is always > 0 since each $P(a|x) > 0$.

Since $\det(\mathfrak{I}(\mathcal{P}))$ is a continuous function of $\mathcal{P}$, there exists a neighborhood of $\mathcal{P}$ in $\triangle$ such that the Fisher information matrix is of full rank for all $\mathcal{P}$ in that neighborhood. Moreover, note that the function $g(\eta)$ is continuously differentiable in η. Hence by Prakash Rao[1992, Theorem 6.3.2], we have that for any $\mathcal{P}'$ in a neighborhood of $\mathcal{P}$, we have $L_y(\mathcal{P}) = L_y(\mathcal{P}')$ $a.e. \Leftrightarrow \mathcal{P} = \mathcal{P}'$. But $u = \varphi(\mathcal{P})$ in Eq. (17) is a 1-1 function from $\triangle$ to $\mathcal{U}$ around $\mathcal{P} \in \triangle$ that satisfies Assumption 5. Hence for any $\zeta \in \Xi$ such that the corresponding $\mathcal{P}(\zeta)$ satisfies Assumption 5, there exists a neighborhood of ζ in Ξ such that for any ζ' in that neighborhood, $L_y(\mathcal{P}(\zeta)) = L_y(\mathcal{P}(\zeta'))$ $a.e. \Leftrightarrow \zeta = \zeta'$. Hence the model in Eq. (2) is locally nonparametrically identified around a ζ whose associated $\mathcal{P}(\zeta)$ satisfies Assumption 5. It is also clear that if Assumption 5 is true for all $\mathcal{P} \in \triangle$, the model in Eq. (2) is also globally identified. ■

The conditional choice probabilities $\mathcal{P} = \{P(a|x), a \in A, x \in X\}$ are nothing but the aggregate demand functions of discrete choices $a \in A$ as a function of individual characteristics $x \in X$., The characteristics $x \in X$ is acting like a price of the Marshallian demand function. Nonparametric identification problem in our set-up can be viewed as the well-known aggregation problem of the consumer theory: Given a system of demand functions $\mathcal{P} \in \triangle$, when does there exists a utility function $u(x, a)$ that generates $\mathcal{P}$ as the optimal solution of problem in Eq. (2). The above theorem provides conditions for an analogous aggregation problem in the present context of structural dynamic programming problem.

Suppose instead of most general non-parametric utility specifications for the parameter vector ζ, we parametrize u (and possibly also β,but F is fixed) to have a parametric form $\zeta : \Theta \to \Xi$, where $\Theta \subset \Re^k$, $k < J - M + 1$ is an open set. When can we identify such parametric models? To state our sufficient condition for this, we recall a definition from the Differential Geometry. A map $f : \Theta \to \triangle$ is an *immersion* at $\theta \in \Theta$, an open subset of $\Re^k$, if the differential map $df_\theta : \Re^k \to T_{f(\theta)}(\triangle)$ is injective, i.e., one-to-one, where $T_{f(\theta)}(\triangle)$ is the tangent space of the manifold $\triangle$ at $f(\theta)$.

Theorem 2 (Parametric Identification) *Let $\Theta \subset \Re^k$ be an open set. Let $\zeta : \Theta \to \Xi$ denotes a family of parametric models. A parametric model is locally identified at $\theta \in \Theta$ if and only if the map $\mathcal{P}(\zeta(\theta)) : \Theta \to \triangle$ is an immersion at θ. The parametric model is globally identified if and only if the map $\mathcal{P}(\zeta(\theta))$ is an injective map.*

Proof. Since$\mathcal{P}(\zeta(\theta))$ is an immersion at θ, there exists a neighborhood around θ in Θ such that $\mathcal{P}(\zeta(\theta))$ is one-one in this neighborhood. For any θ' in this neighborhood of θ, $\mathcal{L}_y(\mathcal{P}(\zeta(\theta))) = \mathcal{L}_y(\mathcal{P}(\zeta(\theta')))$ a.e. implies $\mathcal{P}(\zeta(\theta)) = \mathcal{P}(\zeta(\theta'))$ since $\mathcal{L}_y(\mathcal{P})$ is globally identified in the parameter space $\triangle$ by theorem 1). Hence $\theta = \theta'$ since $\mathcal{P}(\zeta(\theta))$ is 1-1 in this neighborhood. The second part follows immediately. ■

6 Econometric implementation

In this paper we follow the structural estimation procedure of Aguirregabiria, V. and P. Mira (2002) by parameterizing $u_\theta(x,a) = \theta_0 w(x) - \theta_1 a$, where θ_0 is the marginal utility of annualized lifetime earnings, and θ_1/θ_0 is the preschool investment cost in the unit of earnings w. Note that while $u_\theta(x,a)$ is not identified, because for each $x \in X$, the ordered vector $\{u_\theta(x,a), a \in A(x)\}$ should belong to an one dimensional subspace of $\Re^2$, in this specification u lies in a two-dimensional manifold instead. It is interesting to note, however, that the preschool investment cost θ_1/θ_0 is identified.

Our estimation procedure is as follows: First compute F the transition probability matrix from data of the type (x_i, x_i') of the observable states for individuals. Assume a parametric form of the utility function $u_\theta(x,a)$, where $\theta \in \Re^k$.

1. Start with an initial $J \times 1$ vector of probabilities $\mathcal{P}_0 \in \triangle$.
2. Maximize the likelihood $\mathcal{L}(\theta; \mathcal{P}_0) = \prod_{i=1}^{n} P_0(a_i|x_i,\theta)$, where

$$\begin{aligned} P(a_i|x_i,\theta) &= \frac{e^{\tilde{v}(x,a;\theta)}}{\sum_{a\in D} e^{\tilde{v}(x,a';\theta)}} \\ \tilde{v}(x,a) &= u_\theta(x,a) + \beta F(x,a)\left[I_M - \beta\bar{F}\right]^{-1}\left[\bar{u}_\theta + \bar{e}\right] \end{aligned}$$

3. Given θ^* in step 2, compute $\mathcal{P}_1 = \{P(a|x,\theta^*), x \in X, a \in A\} \in \triangle$ from the above formula.
4. If $||\mathcal{P}_1 - \mathcal{P}_0|| < \varepsilon$ stop, else set$\mathcal{P}_0 = \mathcal{P}_1$ go to step 2.

We have used the public domain Sun Java programming language to implement the above estimation procedure and for all other computational tasks.

7 Empirical Findings

7.1 The Dataset and Variables

For our analysis we use the NLSY79 dataset and the NLSY79 Children and Young Adults. The NLSY79 dataset contains a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979, i.e., these sampled individuals represent a population born in the 1950s and 1960s, and living in the United States in 1979. These individuals are interviewed annually. The dataset has records of school and labor market experiences of these individuals and also the information on their cognitive and non-cognitive traits. We, however, also need information on most of these variables for the parents of the respondents. This dataset does not have much information on respondents' parents. So we link this dataset with the NLSY79 Children and Young Adults dataset. The child survey dataset includes longitudinal assessments of each child's cognitive, attitudinal and social, motivational, academic and labor market experiences.

Two other important datasets in this area of research are based on pilot preschool programs are High/Scope Perry Preschool Study and the Carolina Abecedarian Project. These are small scale pilot programs with small number of participants. Data from these programs contain school performance information but the labor market outcome data is weak. While these datasets are good for studying the effect of preschool program on school performance and labor market success, these datasets are not appropriate to estimate parents' preschool investment decision since the participants were selectively chosen. For details on the High/Scope Perry Preschool Study see Schweinhart et al. [1993] and on the Abecedarian Project, see Campbell et al. [1998].

More recently PSID Child Supplement began to collect data on a nationally representative sample of children. This dataset will enable one to link a child's school success and the labor market outcomes to a child's preschool experiences regarding the child rearing methods, home environment, teaching methods followed in schools. While the dataset contains the school performance of these children but the sampled cohort will have data on labor market outcomes only many years later in the future.

7.2 Production of social and motivational skills

We show in the next two subsections that motivation and socialization skills are important determinants of earning and learning. In this section we consider the production process of these two skills.

The literature in sociology, psychology, early childhood development and physiology suggest that early childhood investment is the most crucial input for development of cognitive, social and motivational skills. The studies in these literatures link school success to home environment, child rearing practices, neighborhood type in which the kid is raised. For instance, the Coleman report [1966] and subsequent studies find that family capital, which captures family tradition and values towards economic success and education, and social capital, which captures the benefits of social bonds, social norms, social networks, the social bonds between adults and children and among children in a neighborhood are of immense value during a child's growing up. These factors affect parental choices of preschool investment and child rearing methods which in turn determine a child's cognitive abilities and social abilities such as motivation and sociability that affect their learning and earning. Physiology literature produces ample evidence that the human brain develops extremely rapidly during age [2-4], and the type of stimulations regarding health and learning that the child experience during this period is a critical determinant of a child's cognitive, social and motor developments. Child psychology literature also points out that a structured preschool stimulation also boosts a child's self-confidence, school preparedness, parents' and teachers' assessment of the child's ability. These in turn create a conducive learning environment for the child over many more years of schooling beginning with the elementary school. See Entwisle [1995], and Barnett [1995] for more on these issues.

We construct the variables of our study as follows:

Early childhood inputs and home environment: We take father's and mother's education levels to measure family background. The NLSY dataset has poor measures of respondent's early childhood inputs. It has only a binary variable containing information on whether the respondent had preschool (does not include Head Start) experience or not. We treated individuals with Head Start experience as no preschool. Notice that this will lead to underestimation of the effect of preschool investment. We use the revised AFQT score to measure innate ability.

Socialization skill (σ): Each respondent were asked how social towards others he/she felt at age 6, expressed in the scale of 1 to 4, the highest number represents most social. We create a binary sociability variable by assigning the value 1 if a respondent reported a value of 3 or 4 and assigning 0 otherwise.

Motivation skill (μ): The educational goal (μ) is the grade that the respondent in 1979 expected to achieve.

Rosenberg measure of Self-esteem skill (η): It measures the positiveness with which individuals regard themselves, i.e., a positive sense of self. Six questions were taken from the classic Rosenberg (1965) scale in the NLSY surveys. There is, however, no well accepted definition of adequate self-esteem. Based on the distribution, we divided the 25-point scale by treating a score of 20 or greater indicated a high self-esteem and assign a value 1 to η and a value 0 to η otherwise.

Perlin measure of internal self-control (ϕ): This measures to what extent individuals believe that their life chances are under their control (Perlin et al. 1981). This is similar to Rotter scale of self-control. The respondents were asked seven questions yielding scores ranging from 0 to 28. We assign a value 1 representing a high sense of self-control to respondents with a score between 23 and 28 inclusive, otherwise we assign a a value 0.

We estimated Logit models for the cognitive and non-cognitive skills for the child sample. These parameter estimates are then used to fix an estimate of the transition probability $p(x'|x,a)$. We report table 1 the parameter estimates for specifications in which only the significant regressors (x and a). In our structural maximum likelihood estimations, however, we have reported sensitivity of parameter estimates for this specification and speciations in which we have used both significant and insignificant parameter estimates for $p(x'|x,a)$.

Table 1: Logit model of cognitive and non-cognitive skills.

	Talent τ	Socialization σ	Motivation (Education goal) μ	Internal Self-Control(Perlin): ϕ	Self-Esteem (Rosenberg):η	College* s
Intercept	-3.587 (3.46)	1.164 (11.60)	0.428 (0.92)	-1.106 (13.56)	0.672 (8.96)	-4.694 (18.22)
Own Talent		0.835 (5.55)	1.036 (7.60)	0.402 (4.62)	0.596 (5.28)	1.877 (12.08)
Parent's Talent	1.707 (15.91)					
σ :Socialisation				0.243 (3.09)		0.477 (3.28)
μ : Motivation (Education Goal)						2.726 (17.28)
ϕ : Internal Self-Control (Perlin)			0.503 (2.78)			0.443 (2.26)
η : Self-Esteem (Rosenberg)		0.372 (3.45)	0.325 (3.35)	0.380 (4.26)	0.551 (6.08)	1.245 (4.94)
Parents' Grade	1.814 (1.75)					1.339 (5.17)
Preschool	0.424 (4.47)	0.310 (3.03)	0.190 (2.09)			0.668 (3.72)

Notes: * Attributes in the first column are those of the individuals, and estimated using the 1979 youth sample.

From table 1 it is clear that after controlling for parents' grade, preschool experience has significantly positive effect on socialization and all measures of motivation skills except the Rotter's scale of self control. The estimates in the table also show that innate ability has strong positive effect on all measures of motivation skills but has no significant effect on socialization skills. Socialization skills are created in the family, preschool and neighborhood inputs.

It will be interesting to see if preschool has stronger positive effect on socialization and motivation skills of children of poorer SES. If so, then the preschool could be a used to compensate for the better HOME environment that the well to do counterpart of these children have, and through preschool we can achieve a higher equality of opportunities by equalizing the differences in starting social, motivation, cognitive and motor skills of the children.

7.3 An Augmented Earnings Function - Role of socialization and motivation skills

In this section we examine the effect of social and motivation skills together with the effect of innate ability and grades on earnings. The previous studies included only innate ability, schooling level and school quality as the main determinants of earnings. While preschool investment is an important determinant of these skills, we also included preschool binary variable as one of the regressors in the earnings function to see if it has an independent effect. In our specification, we included two dummy variables, High School (taking value 1 if a respondent had the high school degree) and College (if a respondent graduated from college). These dummy variables together with grade variable are to capture the earnings premiums for graduating from high school and college. Since we included AFQT score which is a reasonably good measure of one's innate ability, we do not have the ability biases in our estimates. We use the yearly earnings data to estimate the model.

Table 1 shows the parameter estimates of this augmented earnings function. The first column is for all three races together and the next three columns give the estimates for the Hispanics, Blacks and the Whites ethnic groups separately. It is clear from the estimates that after controlling for innate ability, family background and the schooling level, the measures of socialization and motivation skills have significant positive effect on earnings for all ethnic groups. Preschool has independent positive effect only for blacks. It is also interesting to note that a college graduate earns 8.35% higher returns in the overall population, and for Blacks and Hispanics this premium is even higher, slightly above 10%. The sociability skills are significant only for White but not for Black and Hispanic workers.

Table 2: Determinants of earnings -- role of cognitive and non-cognitive skills (from the parent sample)

Variables	All Races	Hispanic	Black	White
Intercept	2.369 (31.28)	2.355 (13.64)	0.813 (4.44)	2.613 (27.27)
Own Talent	0.005 (32.28)	0.004 (8.79)	0.006 (12.07)	0.003 (15.67)
Grade	0.054 (22.82)	0.037 (7.83)	0.088 (14.08)	0.057 (18.43)
Dummy for High School	0.065 (8.22)	0.048 (2.82)	0.028 (1.52)	0.095 (9.07)
Dummy for College	0.088 (7.61)	0.097 (2.83)	0.109 (3.59)	0.084 (6.30)
Age	0.319 (70.49)	0.306 (29.04)	0.354 (32.92)	0.314 (56.03)
Square of Age	-0.004 (51.59)	-0.004 (20.84)	-0.004 (24.53)	-0.004 (40.96)
Mother's grade	0.000 (0.30)	0.011 (4.29)	0.016 (4.35)	-0.003 (1.12)
Father's grade	0.007 (5.74)	0.004 (1.77)	-0.006 (2.29)	0.012 (6.93)
Dummy for preschool	0.001 (0.15)	-0.048 (2.32)	0.060 (3.10)	0.007 (0.63)
Socialization	0.013 (1.90)	-0.026 (1.58)	0.025 (1.46)	0.014 (1.65)
Motivation (education goal)	0.002 (1.16)	0.016 (3.52)	0.007 (1.37)	0.007 (2.72)
Self-esteem(Rosenberg)	0.018 (16.32)	0.026 (9.51)	0.018 (6.49)	0.018 (13.50)
Internal self-control(Perlin)	0.024 (21.07)	0.032 (11.49)	0.026 (9.36)	0.019 (13.46)
Gender	-0.512 (74.98)	-0.491 (30.43)	-0.365 (21.98)	-0.578 (68.77)
R^2	0.381	0.396	0.375	0.383
n	81,005	13,769	15,972	51,264

Notes: Absolute t-values are in parentheses.

7.4 Estimation of Schooling Level

We consider two specifications of the schooling policy $s^*\left(\tau', \sigma', \mu', \varepsilon_s^{'}, h\right)$ in this paper. In the first specification, we assume that the schooling level is a continuous variable. We specify the optimal reaction function $s^*\left(\tau', \sigma', \mu', \varepsilon_s^{'}, h\right)$ as a linear function. We assume that the random variable $\varepsilon_s^{'}$ constitutes the error term and satisfies all the assumptions of the OLS model[1] The parameter estimates from this model are shown in table 3 for all ethnic groups together, and also separately for the Hispanic, Black and White populations. We included the socialization and motivation skills together with innate ability, family background as measured by parent's education level.

It is clear from the estimates that the main determinant of grade is the innate ability measured by AFQT score. After controlling for family background, we also find that motivation measures have significant positive effect on schooling level. Out of the three measures of motivation, the measure μ_2 based on the expected the grades that the respondent desired to attain while very young turns out to be the most important one. The sociability skill has, however, no effect on the schooling level.

In our second specification we consider only two levels of schooling: college and more ($s = 1$), and no college ($s = 0$). Again we assume that $s^*\left(\tau', \sigma', \mu', \varepsilon_s^{'}, h\right)$ is linear, and that $\varepsilon_s^{'}$ constitutes the error term and is normally distributed. This gives us a Probit model of college enrollment. We use a subset of the above regressors in this specification and use these estimates to calibrate our basic model in equation (2). The parameter estimates are reported in table 4. Here again the innate ability, motivation, preschool and the college status of parents (which takes 1 if at least one parents had some college, and 0 otherwise) turn out to have significant positive effects on the probability of college enrollment.

7.5 Optimal Parental Preschool Investment Decision

We assume that the state variables s, τ, σ, μ are binary, the random variable ε is continuous which is observed by the decision maker but not by the econometrician, and the preschool investment decision a is a binary variable, taking value 1 when parents decide to invest in preschool and 0 otherwise. For most children, we have two parents but in our model we have assumed one parent. We could take mother as the parent. We have instead used both parent's information as follows: We construct parent's binary schooling variable s by assigning s = 1 if the average grades of two parents is more than 12, otherwise s = 0. We assume that τ is

[1] More generally we could assume that $E\left(\varepsilon_s^{'} | \tau', \sigma', \mu', h\right) = 0$ and use GLS method to correct for heteroskedasticity.

Table 3: Determinants of grade -- role of cognitive and non-cognitive skills (from the parent sample)

Variables	All Races	Hispanic	Black	White
Intercept	2.753 (13.24)	3.324 (5.95)	3.245 (7.66)	2.335 (8.53)
Own Talent	0.028 (29.55)	0.035 (11.38)	0.030 (12.85)	0.028 (22.19)
Mother's grade	0.059 (6.48)	0.027 (1.41)	0.113 (6.03)	0.076 (5.43)
Father's grade	0.028 (3.77)	-0.002 (0.12)	0.004 (0.28)	0.067 (6.44)
Dummy for Preschool	0.266 (4.85)	-0.086 (0.57)	0.227 (2.25)	0.301 (4.25)
σ : Socialization	0.037 (0.83)	-0.128 (1.04)	0.119 (1.34)	0.051 (0.94)
μ : Motivation (Education Goal)	0.458 (40.75)	0.425 (13.93)	0.348 (14.98)	0.468 (31.43)
η : Self-Esteem (Rosenberg)	0.035 (5.10)	0.042 (2.10)	0.069 (4.83)	0.020 (2.32)
η : Internal self-control(Perlin)	0.034 (4.78)	0.055 (2.75)	0.010 (0.73)	0.034 (3.87)
Gender	0.182 (4.27)	0.126 (1.06)	0.464 (5.47)	0.117 (2.24)
R^2	0.560	0.488	0.541	0.585
n	5,782	1,012	1,218	3,552

biologically inherited and it is not influenced by preschool investment. We create the binary variable τ assigning value 1, i.e., an individual is highly talented if the AFQT score of the individual is 70 or higher, and assigning value 0 otherwise.

The estimate of the preschool investment cost depends on the calibrated value of the altruism parameter β as can be seen from table 5. Schweinhart et al. took average yearly preschool cost to be \$6178 per year. Consistent with their study, we take calibrate the altruism parameter to $\beta = 0.35$ for our analysis to be consistent with their cost estimate. The optimal preschool investment decision and the value function are shown respectively in column 3 and 4 of table 6.

We consider a public policy of providing preschool to children of poor socio economic status (SES) in all periods. We define a parent to fall in the poor SES if his earnings is less than 70 percent of the average earnings in economy. This will incur a per capita cost, but such policy may also improve social mobility, earnings inequality and to a higher level of per capita long-run earnings. We examine if the gain from per capita earnings can outpace the cost of providing such a social insurance program. We also look at its within generation effect on earnings, and on intergenerational social and college mobility.

Table 4: Maximum likelihood parameter estimates given two different estimates of $p\left(x'|x,a\right)$ and altruism parameter $\beta = 0.35$.

	Parameter Estimates given $p\left(x'|x,a\right)$ using	
	only significant $x's$	all $x's$
Marginal utility $(\hat{\theta}_0)$ from average earnings	6.729 (5.136)	8.452 (5.646)
Utility cost $(\hat{\theta}_1)$ of Preschool investment	4.636 (9.391)	4.761 (10.077)
Annualized cost in dollars $1000 \times (\hat{\theta}_1/\hat{\theta}_0)$	689.032	563.229
Percent of poor SES population	28.39	36.30
Per capita cost to society ('000 dollars)	0.196	0.204
Per capita change in earnings ('000 dollars)	0.310	0.432
Log-likelihood	-1039.626	-1037.236

Note: Absolute value of t-statistics are in parentheses.

Table 5: Sensitivity of maximum likelihood estimates with variations of the altruistic parameter β.

β	$\hat{\theta}_0$	t-stat of $\hat{\theta}_0$	$\hat{\theta}_1$	t-stat of $\hat{\theta}_1$	annualized per capita costs and benefits in $ '000: costs: parent	costs: tax payer	benefits: $\triangle\bar{w}$
0.030	82.313	4.484	4.430	8.572	0.054	0.015	0.313
0.070	35.151	4.566	4.458	8.673	0.127	0.036	0.313
0.110	22.274	4.647	4.485	8.775	0.201	0.057	0.313
0.150	16.255	4.729	4.512	8.877	0.278	0.079	0.312
0.190	12.761	4.810	4.538	8.979	0.356	0.101	0.312
0.230	10.476	4.892	4.563	9.082	0.436	0.124	0.311
0.270	8.863	4.973	4.588	9.185	0.518	0.147	0.311
0.310	7.661	5.055	4.613	9.288	0.602	0.171	0.310
0.350	6.729	5.136	4.637	9.391	0.689	0.196	0.310
0.390	5.985	5.218	4.660	9.495	0.779	0.221	0.309
0.430	5.376	5.299	4.682	9.599	0.871	0.247	0.308
0.470	4.867	5.380	4.704	9.704	0.967	0.274	0.308
0.510	4.435	5.462	4.726	9.809	1.065	0.302	0.307
0.550	4.064	5.543	4.746	9.914	1.168	0.332	0.306
0.590	3.741	5.624	4.766	10.020	1.274	0.362	0.306
0.630	3.456	5.705	4.786	10.126	1.385	0.393	0.305
0.670	3.204	5.787	4.804	10.233	1.500	0.426	0.304
0.710	2.978	5.868	4.822	10.340	1.619	0.460	0.303
0.750	2.774	5.949	4.839	10.448	1.744	0.495	0.302
0.790	2.590	6.030	4.856	10.556	1.875	0.532	0.301
0.830	2.421	6.111	4.872	10.664	2.012	0.571	0.300
0.870	2.267	6.192	4.887	10.774	2.156	0.612	0.299
0.910	2.125	6.273	4.901	10.883	2.306	0.655	0.298
0.950	1.994	6.354	4.915	10.994	2.465	0.700	0.297
0.990	1.872	6.435	4.928	11.105	2.633	0.747	0.296

Table 6: Equilibrium Solution

$[\tau, \sigma, \mu, \eta, \phi, s]$	p_0	Earnings	$P_b(a=1\|x)$	$P_a(a=1\|x)$	$optV_b$	$optV_a$	p_b^*	p_a^*
[0, 0, 0, 0, 0, 0]	0.1797	4.1520	0.0735	1.0000	60.1545	62.5914	0.0088	0.0075
[0, 1, 0, 0, 0, 0]	0.0013	4.3993	0.0740	1.0000	62.1775	64.6075	0.0351	0.0334
[0, 0, 1, 0, 0, 0]	0.0029	4.7769	0.0735	1.0000	64.3592	66.7961	0.0155	0.0136
[0, 1, 1, 0, 0, 0]	0.0000	5.0241	0.0740	1.0000	66.3822	68.8122	0.0571	0.0544
[0, 0, 0, 0, 1, 0]	0.0029	6.2648	0.0843	1.0000	76.4401	78.9237	0.0226	0.0195
[0, 1, 0, 0, 1, 0]	0.0000	6.5120	0.0848	1.0000	78.5247	80.9992	0.0920	0.0866
[1, 0, 0, 0, 0, 0]	0.0665	6.7053	0.1312	1.0000	79.0355	82.0031	0.0001	0.0001
[0, 0, 0, 1, 0, 0]	0.3029	6.8013	0.0822	1.0000	78.6578	81.1644	0.0043	0.0038
[0, 0, 1, 0, 1, 0]	0.0000	6.8896	0.0843	1.0000	80.6448	83.1284	0.0297	0.0246
[1, 1, 0, 0, 0, 0]	0.0006	6.9525	0.1320	1.0000	81.0655	84.0271	0.0009	0.0010
[0, 1, 0, 1, 0, 0]	0.0060	7.0485	0.0827	1.0000	80.6867	83.1864	0.0178	0.0170
[0, 1, 1, 0, 1, 0]	0.0000	7.1369	0.0848	0.0812	82.7293	83.0784	0.1014	0.0897
[1, 0, 1, 0, 0, 0]	0.0019	7.3301	0.1312	0.1235	83.2402	83.6627	0.0002	0.0002
[0, 0, 1, 1, 0, 0]	0.0342	7.4261	0.0822	0.0778	82.8625	83.2858	0.0071	0.0062
[1, 1, 1, 0, 0, 0]	0.0000	7.5774	0.1320	0.1246	85.2701	85.6776	0.0015	0.0015
[0, 1, 1, 1, 0, 0]	0.0013	7.6734	0.0827	0.0785	84.8914	85.2989	0.0261	0.0244
[1, 0, 0, 0, 1, 0]	0.0010	8.8180	0.1416	0.1348	95.3049	95.6369	0.0004	0.0004
[0, 0, 0, 1, 1, 0]	0.0215	8.9140	0.0910	0.0871	94.9085	95.2359	0.0115	0.0100
[1, 1, 0, 0, 1, 0]	0.0000	9.0653	0.1420	0.1357	97.3925	97.7091	0.0030	0.0033
[0, 1, 0, 1, 1, 0]	0.0003	9.1613	0.0913	0.0877	96.9992	97.3112	0.0470	0.0441
[1, 0, 0, 1, 0, 0]	0.1705	9.3545	0.1409	0.1330	97.5396	97.9225	0.0001	0.0001
[1, 0, 1, 0, 1, 0]	0.0010	9.4429	0.1416	0.1348	99.5096	99.8416	0.0004	0.0004
[0, 0, 0, 0, 0, 1]	0.0092	9.4477	0.1346	0.1259	100.6712	101.0250	0.0002	0.0002
[0, 0, 1, 1, 1, 0]	0.0048	9.5389	0.0910	0.0871	99.1132	99.4406	0.0128	0.0105
[1, 1, 0, 1, 0, 0]	0.0063	9.6018	0.1416	0.1340	99.5746	99.9434	0.0006	0.0007
[1, 1, 1, 0, 1, 0]	0.0000	9.6901	0.1420	0.1357	101.5972	101.9138	0.0027	0.0028
[0, 1, 0, 0, 0, 1]	0.0000	9.6949	0.1344	0.1260	102.7032	103.0454	0.0011	0.0013
[0, 1, 1, 1, 1, 0]	0.0000	9.7862	0.0913	0.0877	101.2039	101.5158	0.0428	0.0373
[1, 0, 1, 1, 0, 0]	0.0285	9.9794	0.1409	0.1330	101.7442	102.1272	0.0001	0.0001
[0, 0, 1, 0, 0, 1]	0.0006	10.0725	0.1346	0.1259	104.8759	105.2297	0.0042	0.0047
[1, 1, 1, 1, 0, 0]	0.0006	10.2266	0.1416	0.1340	103.7792	104.1481	0.0008	0.0009
[0, 1, 1, 0, 0, 1]	0.0000	10.3198	0.1344	0.1260	106.9078	107.2501	0.0250	0.0293
[1, 0, 0, 1, 1, 0]	0.0146	11.4673	0.1474	0.1409	113.7605	114.0576	0.0002	0.0003
[0, 0, 0, 0, 1, 1]	0.0013	11.5604	0.1193	0.1131	117.1245	117.4035	0.0015	0.0016
[1, 1, 0, 1, 1, 0]	0.0000	11.7145	0.1476	0.1415	115.8532	116.1366	0.0020	0.0021
[0, 1, 0, 0, 1, 1]	0.0000	11.8077	0.1185	0.1127	119.2084	119.4756	0.0101	0.0116
[1, 0, 0, 0, 0, 1]	0.0006	12.0009	0.1672	0.1580	122.9650	123.2108	0.0000	0.0000
[1, 0, 1, 1, 1, 0]	0.0054	12.0921	0.1474	0.1409	117.9652	118.2623	0.0002	0.0002
[0, 0, 0, 1, 0, 1]	0.0409	12.0969	0.1260	0.1183	119.4114	119.7248	0.0001	0.0001

Table 7: Continuation of Table 6.

$[\tau, \sigma, \mu, \eta, \phi, s]$	p_0	Earnings	$P_b(a=1\|x)$	$P_a(a=1\|x)$	$optV_b$	$optV_a$	p_b^*	p_a^*
[0, 0, 1, 0, 1, 1]	0.0000	12.1853	0.1193	0.1131	121.3292	121.6082	0.0257	0.0272
[1, 1, 0, 0, 0, 1]	0.0000	12.2482	0.1666	0.1578	125.0072	125.2455	0.0003	0.0004
[1, 1, 1, 1, 1, 0]	0.0000	12.3394	0.1476	0.1415	120.0579	120.3413	0.0015	0.0015
[0, 1, 0, 1, 0, 1]	0.0010	12.3442	0.1254	0.1181	121.4449	121.7480	0.0009	0.0010
[0, 1, 1, 0, 1, 1]	0.0000	12.4325	0.1185	0.1127	123.4131	123.6803	0.1394	0.1546
[1, 0, 1, 0, 0, 1]	0.0000	12.6258	0.1672	0.1580	127.1697	127.4155	0.0006	0.0007
[0, 0, 1, 1, 0, 1]	0.0076	12.7218	0.1260	0.1183	123.6161	123.9295	0.0030	0.0033
[1, 1, 1, 0, 0, 1]	0.0000	12.8730	0.1666	0.1578	129.2119	129.4501	0.0068	0.0077
[0, 1, 1, 1, 0, 1]	0.0006	12.9690	0.1254	0.1181	125.6496	125.9527	0.0175	0.0200
[1, 0, 0, 0, 1, 1]	0.0006	14.1137	0.1384	0.1324	139.1534	139.3505	0.0003	0.0003
[0, 0, 0, 1, 1, 1]	0.0257	14.2097	0.1081	0.1030	135.7709	136.0161	0.0012	0.0013
[1, 1, 0, 0, 1, 1]	0.0000	14.3609	0.1373	0.1316	141.2364	141.4260	0.0035	0.0039
[0, 1, 0, 1, 1, 1]	0.0010	14.4569	0.1071	0.1024	137.8560	138.0911	0.0080	0.0091
[1, 0, 0, 1, 0, 1]	0.0165	14.6502	0.1465	0.1391	141.5275	141.7475	0.0000	0.0000
[1, 0, 1, 0, 1, 1]	0.0003	14.7385	0.1384	0.1324	143.3581	143.5551	0.0039	0.0043
[0, 0, 1, 1, 1, 1]	0.0032	14.8345	0.1081	0.1030	139.9756	140.2208	0.0169	0.0176
[1, 1, 0, 1, 0, 1]	0.0013	14.8974	0.1458	0.1387	143.5690	143.7823	0.0003	0.0004
[1, 1, 1, 0, 1, 1]	0.0000	14.9858	0.1373	0.1316	145.4411	145.6307	0.0426	0.0476
[0, 1, 1, 1, 1, 1]	0.0006	15.0818	0.1071	0.1024	142.0606	142.2957	0.0898	0.0972
[1, 0, 1, 1, 0, 1]	0.0029	15.2750	0.1465	0.1391	145.7322	145.9522	0.0005	0.0006
[1, 1, 1, 1, 0, 1]	0.0003	15.5223	0.1458	0.1387	147.7737	147.9870	0.0059	0.0067
[1, 0, 0, 1, 1, 1]	0.0247	16.7629	0.1203	0.1157	157.6138	157.7899	0.0003	0.0003
[1, 1, 0, 1, 1, 1]	0.0013	17.0102	0.1191	0.1148	159.6961	159.8658	0.0034	0.0039
[1, 0, 1, 1, 1, 1]	0.0054	17.3878	0.1203	0.1157	161.8185	161.9945	0.0034	0.0037
[1, 1, 1, 1, 1, 1]	0.0000	17.6350	0.1191	0.1148	163.9008	164.0705	0.0369	0.0410

8 Economic Benefits from Public Provision of Preschool

We have shown that investment in preschool enhances certain skills that are important for learning and earning. We have also seen that the parents of poor SES do not invest in their children's preschool. If preschool is publicly provided for the children of poor SES, it will have many economic benefits: It will increase social mobility, it will reduce income inequality, it will improve college enrollment rate, it will improve the community or criminal behavior, and it will also bring higher tax revenues because more workers will be earning high wages. It is important to note that the magnitude of the effect of publicly provided preschool will depend on if the social protection will be available to all future generations or it is just a one time policy.

While looking at the magnitude of the estimated economic benefits below, it is important to keep in mind that the effects that we report are underestimated for many reasons: First, we have treated the Head Start children same as children without preschool. Second, the preschool programs that the respondents went into were the ones that existed during the sixties. The quality of preschool programs ever since has improved significantly and thus the effects of current preschool programs will be much higher than the estimates that we have.

Note that since ε does not affect earnings, the optimal a depends only on the observable component x of a parent's state variable, i.e. optimal preschool plan is $a(x)$. In the absence of the social contract suppose the parents follow the optimum preschool investment plans $a(x)$ as shown in table 6. The invariant distribution of the corresponding transition matrix $\{p(x')|x, a(x), x \in X\}$ is shown in table 6 under the heading $P_b(a = 1|x)$. The interpretation of this invariant distribution is as follows: If $P_b(a = 1|x)$ is the distribution of population over the observable states of generation t, and the parents of generation t follow the optimal preschool investment plan $a(x)$, the distribution of population of the next generation will also be $P_b(a = 1|x)$.

8.1 Social Mobility

Given any transition matrix $p(x'|x, a(x))$ over the observable states, there exists a number of mobility measures in the literature. Sommers and Conlisk [1979] argued that out of the existing measures, $1 - \lambda_{\max}$ is the most appropriate measure of social mobility, where $\lambda_{\max}$ is the second highest positive eigenvalue of the transition matrix (the highest positive eigenvalue of a transition matrix is always 1). We use this measure of social mobility to examine how the introduction of the social contract would improve social mobility. Our estimate of the measure of social mobility before the social contract is 0.568759 and after introduction of social program it improves to 0.598074. The estimate of 0.568759for the measure is very close to the estimates found in other studies of social mobility in the US.

8.2 College Mobility

Denote by $Q^s = [q_{ij}], i, j = 1, 2$, the intergenerational college mobility matrix in which state 1 represents no college and state 2 represents college and higher. The element q_{ij} represents the probability that a child of a parent of college education status i will move to the college education status j. We report below the estimated college mobility matrices, the corresponding invariant distributions, and the estimates of the mobility measure before and after the introduction of social contract. These estimates indicate that the introduction of the social contract will increase college enrollment from a 32.90 percent to a 37.21 percent, i.e. a 4.31 percent increase for a child of non-college parent. And the percentage of college enrolled population will increase in the long-run from the rate of 48.17 percent without social contract to a higher rate of 51.18 percent with the social contract. That is, there will be about a 3.01% increase in college enrollments in the long-run.

College mobility statistics before introduction of social contract:

$$Q_b^s = \begin{bmatrix} 0.6710 & 0.3290 \\ 0.3541 & 0.6459 \end{bmatrix}, p_b^s = \begin{bmatrix} 0.518327 & 0.481673 \end{bmatrix}, 1 - \lambda_{\max,b}^s = 0.683070$$

College mobility statistics after introduction of social contract:

$$Q_a^s = \begin{bmatrix} 0.6279 & 0.3721 \\ 0.3550 & 0.6450 \end{bmatrix}, p_a^s = \begin{bmatrix} 0.488177 & 0.511823 \end{bmatrix}, 1 - \lambda_{\max,a}^s = 0.727102$$

8.3 Income Inequality

Preschool experience will increase the income of the children of poor SES and thus it will reduce the income gap between the rich and the poor. Using Gini-coefficient to measure the income inequality, we would expect that over time the income inequality will improve. In the long-run, the income distribution that one observes is the invariant distribution. Thus we compute the Gini-coefficient of income inequality for the invariant income distribution before introduction of the social contract and compare it with the Gini-coefficient for the invariant income distribution after introduction of the social contract. The estimated Gini-coefficients are respectively 0.2133 without the social contract, and 0.2087 with the social contract. The estimated Gini-coefficient of earnings 0.2133 turns out to be very close to the estimates found in other studies on US. We note that the social contract of publicly providing preschool to children of poor SES leads to a significant reduction in the inequality of long-term earnings.

8.4 Tax Burden of the Social Contract

Suppose the government provides preschool to the children of poor SES perpetually. We know that the size of the population of poor SES will become smaller over time. Thus the resource needs of the program will become smaller, and the tax revenues will become higher over time. We can look at the stream of these costs and benefits to the society and then compute the average per period costs and benefits to calculate the tax-burdens of the social contract. Applying the Ergodic theorem, however, this boils down to computing the costs and benefits of the invariant distribution that will result after the introduction of the social contract.

There are approximately 28.39 percent of the population will fall in the poor socio economic status using our definition. Thus the per capita cost to the economy in the long-run of the social contract is \$195.638 but gain in the per capita income due to the introduction of social contract is \$309.60, so there is a net gain to the economy. This net gain is based on a reasonable value of the altruistic parameter β. The simulation results in our sensitivity analysis shows that, the lower is the value of the altruism parameter β, the higher is the gain from the introduction of the social contract. The economic reason for this is quite obvious. When parents have lower altruism towards children, they will invest less on their children's preschool since such investment decreases their own felicity index and increase welfare of the children which got a lower weight when β has lower value. This estimate of net gain is based on calibrating the value of β to the cost data of a high cost program as noted earlier whose benefits are supposed to be higher than our estimated benefits. Thus this gain is an underestimate of the actual net benefit. Furthermore, our benefit calculation does not take into account other public savings such as savings from welfare assistance and savings to the criminal justice system and potential victims of crimes. If we incorporate these, the returns will be much higher. Using data from the High/Scope Perry Preschool Program Schweinhart et al. estimated a total benefit of \$7.16 from all these sources for each dollar spent on the preschool program.

References

Aguirregabiria, V. and P. Mira (2002), "Swapping the Nested fixed Point Algorithm: A Class of Estimators for Discrete Markov Decisions Models," *Econometrica*, vol.70(4):1519-1543.

Aughinbaugh, Alison [2001], “Does Head Start Yield Long-Term Benefits?”, *Journal of Human Resources*, Vol.36(4):641-665.

Barnett, W. Steven (1995), “ Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes”, *The Future of Children: Long-Term Outcomes of Early Childhood Programs*, Vol. 5(3):25-50.

Bhattacharya, R. and M. Majumdar [1989], "Controlled Semi-Markov Model -The Discounted Case," *Journal of Statistical Planning and Inference*, vol.21:365-381.

Campbell, F. A., R. Helms, J.J. Sparling and C. T. Ramey (1998) “Early Childhood Progrmas and Success in School: The Abecedarian Study” in (eds) W. Steven Barnett and Sarane Spence Boocock, “*Early Care and Education for Children in Poverty: Promises, Programs, and Long-Term Results*”, State University of New York Press, New York.

Carneiro, Pedro and James J. Heckman [2002], “The Evidence on Credit Constraints in Post-Secondary Schooling”, NBER WP#9055.

Coleman James et al. [1966], “*Equality of Educational Opportunity*,” Washington DC, U.S. GPO.

Consortium of Longitudinal Studies (1983) “As the Twig is Bent: Lasting Effects of Preschool Programs” Hillside, NJ: Erlbaum.

Cunha, Flavio, James J. Heckman, Lancel Lochner and Dimitry V. Masterov (2006), "Interpreting the evidence of life cycle skill formaiton," Chapter 12, in (eds) Eric A. Hanushek and Finis Wetch, "Handbook of the Economics of Education," Volume 1, Elsevier B.V.

Currie, Janet(2001) 'Early Childhood Education Programs', *Journal of Economic Perspectives,* vol.15(2):213-238.

Currie, Janet and Duncan Thomas(2000) 'School Quality and Longer Term Effect of Head Start', *Journal of Human Resources*, Vol. 35(4):755-74.

Currie, Janet and Duncan Thomas(1995) 'Does Head Start Make a Difference?', *American Economic Review*, Vol. 85, No. 3. (Jun., 1995), pp. 341-364.

Datcher-Loury, Linda(1989) 'Family Background and school Achievement among low income blacks', *Journal of Human Resources*, vol.24(3):528-544.

Entwisle, Dorris (1995)”The Role of Schools in Sustaining Early Childhood Program Benefits”, *The Future of Children: Long-Term Outcomes of Early Childhood Programs*, , Vol. 5(3).

Hanushek, Eric A. (1986), “The Economics of Schooling: Production and Efficiency in Public Schools”, *Journal of Economic Literature,* Vol. 24, No. 3. (Sep., 1986), pp. 1141-1177.

Heckman, James (2000) 'Policies to Foster Human Capital', *Research in Economics*, 54(1):3-56.

Heckman, James and Yona Rubinstein (2001), “The Importance of Noncognitive Skills: Lessons from the GED Testing Program,” *American Economic Review*, 91(2):145-149.

Heckman, James J., L. J. Lochner and C. Taber (1999), " General-Equilibrium Cost-Benefit Analysis of Education and Tax Policies," Chapter 14 in (eds) Gustav Ranis and Lakshmi K. Raut, " Trade, Growth and Development: Essays in Honor of Professor T.N. Srinivasan," New York: Elsevier Science B.V.

Hotz, J. and R. A. Miller (1993), "conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, Vol.60:497-529.

Judd, K. L. (1999), “The Parametric Path Method for solving perfect foresight models” (mimeo) Working Paper, Stanford University.

Loury, Glenn (1981) "Is Equal Opportunity Enough?" *American Economic Review*, Vol. 71, No. 2 (Proceedings).

McCormick, Kathleen [1989], “*An Equal Chance: Educating At-Risk Children to Succeed*,” Alexandria, Virginia: National School Board Associations.

McFadden, Daniel (1981), "Econometric Models of Probabilistic Choice," in (eds), C. Manski and D. McFaden "*Structural Analysis of Discrete Data with Econometric Applications*," Cambridge, MA: MIT Press.

Nishimura, Kazuo and Lakshmi Raut (2007), "School Choice and the Intergenerational Poverty Trap," *Review of Development Economics,* 11(2) , 412-420,2007.

Perlin et al. (1981)

Praskash Rao, B. L. S. (1992), "Identifiability in Stochastic Models: Characterization of Probability Distributions," New York: Academic Press, Inc.

Raut, Lakshmi (1995), “'Signaling Equilibrium, Intergenerational Social Mobility and Long Run Growth”' Draft presented at the Seventh World Congress Meeting of the Econometric Society, Tokyo, 1995 http://papers.ssrn.com/abstract=852924.

Rosenberg, Morris. (1965), "*Society and the Adolescent Self- Image. Princeton*," New Jersey: Princeton University Press.

Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica,*Vol. 55(5),999-1033.

Rust, John [1994], “Structural Estimation of Markov Decision Processes”, in (eds) R.F. Engle and D.L. MacFadden, “*Handbook of Econometrics, Volume IV*,” Elsevier Science B.V.

Schweinhart, Lawrence J., Helen V. Barnes, and David P. Weikart (1993) “*Significant Benefits: The High/Scope Perry Preschool Study Through Age 27*”, Ypsilanti, Michigan: The High/Scope Press.

Sommers, Paul. M. and John Conlisk [1979], “Eigenvalue Immobility Measures for Markov Chains”, *Journal of Mathematical Sociology*, vol.6:253-276.

Stokey N. and R. Lucas [1989], "*Recursive Methods in Economic Dynamics*", Harvard University Press, Cambridge, Massachusetts

Traub, James(2000) 'What No School Can Do', *The New York Times Magazine*, January 16, 2000, section 6, pp.52-57.