Estimating the Technology of Cognitive and Noncognitive Skill Formation*

Flavio Cunha
The University of Chicago

James Heckman
The University of Chicago,
American Bar Foundation
and University College Dublin

Susanne Schennach
The University of Chicago

First draft, March 2006
This draft, June 27, 2006

*This paper was generously supported by the National Science Foundation (SES-0241858, SES-0099195, SES-0452089); the National Institute of Child Health and Human Development (R01HD43411); and the Committee for Economic Development with a grant from the Pew Charitable Trusts. This research was also supported by the Children's Initiative, a project of the Pritzker Family Foundation at the Harris School of Public Policy Studies at the University of Chicago; and the American Bar Foundation. Rodrigo Pinto provided valuable research assistance. We have benefited from comments received from Gary Becker, Lars Hansen, Kevin Murphy, Petra Todd, and Ken Wolpin; as well as from participants at the Yale Labor/Macro Conference (May 2006) and University of Chicago Applications Workshop (June 2006). A website containing supplementary material is available at http://jenni.uchicago.edu/elast-sub.

Abstract

This paper formulates and estimates a model of the evolution of child cognitive and noncognitive skills as determined by parental investments at different stages of the life cycle. We estimate the elasticity of substitution between contemporaneous investments and inherited stocks of skills to assess the benefits of early investment in children compared to later investment. We develop a nonlinear factor analysis and use it to establish nonparametric identification of the technology of skill formation. A by-product of our approach is a framework for the evaluation of childhood interventions that avoids reliance on arbitrarily scaled test scores. Since any monotonic transformation of a test score is also a valid test score, "value added" analyses of test scores have no clear interpretation. We develop a general nonparametric solution to this problem by anchoring test scores and other outcome measures in adult outcomes with interpretable scales.

Keywords: cognitive skills; noncognitive skills; dynamic factor analysis; anchoring test scores; parental influence

JEL codes C31, J13

Flavio Cunha
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
flavio@uchicago.edu
phone: 773-256-6216

James J. Heckman
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
jjh@uchicago.edu
phone: 773-702-0634
fax: 773-702-8490

Susanne Schennach
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
smschenn@uchicago.edu
phone: 773-702-8199
fax: 773-702-8490

1 Introduction

This paper estimates the multistage technology governing the formation of cognitive and noncognitive skills in childhood through parental investments. The different stages of the technology correspond to different developmental phases of the life cycle of the child. We identify and estimate key substitution parameters that determine the importance of early parental investment for subsequent lifetime achievement, and the costliness of later remediation if early investment is not undertaken. Understanding the evolution of cognitive and noncognitive skills is important for understanding the sources of success in adult life.[1] A large body of research documents the importance of cognitive skills for social and economic success.[2] An emerging body of research shows the parallel importance of noncognitive skills.[3]

We develop a new nonlinear factor analysis and establish nonparametric identification of the technology of skill formation. We develop a method for measuring skill formation in childhood that avoids direct reliance on test scores as measures of final outcomes. Since any monotonic transformation of a test score is still a valid test score, widely used "value added" measures relating the effects of inputs on changes in test scores are intrinsically arbitrary. We anchor test scores by examining the effects of inputs on test scores as they affect cardinal measures of adult achievement such as the probability of graduating from high school.

This paper builds on the theoretical analysis of Cunha and Heckman (2003) and Cunha, Heckman, Lochner, and Masterov (2006). They extend the model of Becker and Tomes (1979, 1986) which assumes a one period childhood to a multiperiod model of childhood. Their model organizes a large body of empirical evidence demonstrating the importance of investment in early childhood and the costliness of later remediation if early investments are not made. It also organizes the evidence on the development of animals.[4]

Our empirical analysis builds on the previous research of Cunha and Heckman (2006). Those authors estimate a dynamic factor model that exploits cross equation restrictions (covariance restrictions in linear systems) to secure identification of a linear version of a multistage technology for child investment using a version of dynamic linear state space models.[5] The idea underlying their approach and this paper is that

[1] See Cunha, Heckman, Lochner, and Masterov (2006).

[2] See Herrnstein and Murray (1994), Murnane, Willett, and Levy (1995), and Cawley, Heckman, and Vytlacil (2001).

[3] See Heckman, Stixrud, and Urzua (2006) and the references they cite.

[4] See Knudsen, Heckman, Cameron, and Shonkoff (2006).

[5] See Shumway and Stoffer (1982) and Watson and Engle (1983).

cognitive and noncognitive skills and parental investments are low dimensional latent variables. Building on the analysis of Jöreskog and Goldberger (1975) and Carneiro, Hansen, and Heckman (2003), they use a variety of measurements related to skills and investments to proxy latent skills and investments. With enough measurements relative to the number of latent skills and investments, it is possible to identify the latent state space dynamics generating the evolution of skills using cross-equation restrictions.

To achieve linear estimating equations, Cunha and Heckman (2006) assume that the elasticity of substitution parameter between early and late investments is one. This paper identifies a more general nonlinear technology by extending linear state space and factor analysis to a nonlinear setting. This extension allows us to identify crucial elasticity of substitution parameters governing the trade-off between early and late investments.

Drawing on the analyses of Schennach (2004a) and Hu and Schennach (2006), we produce nonparametric identification of the technology of skill formation. We relax many of the strong independence assumptions for the error terms in the measurement equations that are maintained in Cunha and Heckman (2006) and Carneiro, Hansen, and Heckman (2003). We relax the linearity of the technology assumption used by Cunha and Heckman (2006) and Todd and Wolpin (2003, 2005) so that we allow inputs to interact in producing output. We generalize the factor analytic index function models used by Carneiro, Hansen, and Heckman (2003) to allow for more general functional forms for measurement equations. We solve the problem of defining a scale for the output of childhood investments from multiple sources by anchoring test scores using the adult outcomes of the child, which have a well defined cardinal scale. We normalize the latent variables that generate test scores by determining how the latent variables predict adult outcomes.[6] This approach sets the scale of the test scores and latent variables in an interpretable metric. Using this metric, analysts can meaningfully interpret changes in output and conduct interpretable value added analyses. Cunha and Heckman (2006) develop a class of anchoring functions invariant to affine transformations. In this paper we consider a more general class of monotonic transformations and present a new analysis of identification of the anchoring equations and the technology of skill formation.

The plan of this paper is as follows. Section 2 summarizes and extends the theoretical analysis of Cunha and Heckman (2003), Cunha, Heckman, Lochner, and

[6] Cawley, Heckman, and Vytlacil (1999) present an analysis that anchors test scores in earnings outcomes.

Masterov (2006), and Cunha and Heckman (2006) in order to motivate our empirical analysis and interpret the evidence. Section 3 presents our identification analysis. Section 4 discusses our particle filtering estimation strategy. Section 5 discusses our data and estimates. Section 6 concludes.

2 A Model of Cognitive and Noncognitive Skill Formation

We analyze a model with multiple stages of childhood, $t = 1, 2, \ldots, T$, $T \geq 2$, followed by adult working life, $t = T+1, T+2, \ldots, T+S$. Adult outcomes are produced by cognitive skills, θ^C, and noncognitive skills, θ^N. The influential analysis of Becker and Tomes (1979, 1986) assumes only one period of childhood ($T = 1$) and considers one output associated with "human capital" that can be interpreted as a composite of cognitive and noncognitive skills. We denote by I_t^k the parental investments in child skill k at stage t, $k = C, N$ and $t = 1, 2, \ldots, T$. Investments can include the parental choice of schools and schooling quality.

Skills evolve in the following way. Each agent is born with initial conditions $\theta_0 = \left(\theta_0^C, \theta_0^N\right)$. Family environmental and genetic factors may influence these initial conditions (see Olds, 2002, and Levitt, 2003). At each stage t, $\theta_t = \left(\theta_t^C, \theta_t^N\right)$ denotes the vector of skill or ability stocks. The technology of production of skill k in period t depends on the stock of skills at date t and investment at t, I_t^k:

$$\theta_{t+1}^k = f_t^k\left(\theta_t, I_t^k\right), \tag{2.1}$$

for $k = C, N$ and $t = 1, 2, \ldots, T$. We assume that f_t^k is twice continuously differentiable, increasing and concave in I_t^k. In this model, stocks of skills help to produce next period skills and affect the productivity of investments. Stocks of cognitive skills can promote the formation of noncognitive skills and *vice versa* because θ_t is an argument of (2.1).

Direct complementarity between the stock of skill l and the productivity of investment I_t^k in producing skill k in period $t+1$ is defined as:

$$\frac{\partial^2 \theta_{t+1}^k}{\partial I_t^k \partial \theta_t^l} > 0, \quad t = 1, \ldots, T, \quad l, k = C, N.$$

Period t stocks of abilities and skills promote acquisition of skills in $t+1$ by making investment more productive. Students with greater early cognitive and noncognitive abilities are more efficient in later learning of both cognitive and noncognitive skills. The evidence from the early intervention literature suggests that the enriched early environments of the Abecedarian, Perry and CPC programs promote greater efficiency in learning in high schools and reduce problem behaviors.[7]

Technology (2.1) is sufficiently rich to capture the evidence on learning in rodents and macaques documented by Meaney (2001) and Cameron (2004), respectively.[8] Emotionally nurturing early environments producing θ^N create preconditions for later cognitive learning. Emotionally secure young animals actively explore their environments and learn more quickly than less secure animals. This is an example of complementarity.

Adult human capital h is a combination of different stage $T+1$ skills:

$$h = g\left(\theta^C_{T+1}, \theta^N_{T+1}\right), \tag{2.2}$$

where g is increasing in $\left(\theta^C_{T+1}, \theta^N_{T+1}\right)$.[9] This specification assumes that there is no comparative advantage in the labor market or in other areas of social performance.[10]

To fix ideas and provide a framework that motivates our econometric analysis, consider a *CES* version of the technology where we assume that $\theta^C_t, \theta^N_t, I^C_t$, and I^N_t, are scalars:

$$\theta^C_{t+1} = B_C\left\{\psi^C_C\left(\theta^C_t\right)^\alpha + \psi^C_N\left(\theta^N_t\right)^\alpha + \left(1-\psi^C_C-\psi^C_N\right)\left(I^C_t\right)^\alpha\right\}^{\frac{1}{\alpha}}, \quad t = 1,\ldots,T, \tag{2.3}$$

$$\theta^N_{t+1} = B_N\left\{\psi^N_C\left(\theta^C_t\right)^\sigma + \psi^N_N\left(\theta^N_t\right)^\sigma + \left(1-\psi^N_C-\psi^N_N\right)\left(I^N_t\right)^\sigma\right\}^{\frac{1}{\sigma}}, \quad t = 1,\ldots,T, \tag{2.4}$$

where $B_k > 0$, $1 \geq \psi^k_C \geq 0$, $1 \geq \psi^k_N \geq 0$, and $1 \geq 1-\psi^k_C-\psi^k_N \geq 0$, $k = C, N$. $\frac{1}{1-\alpha}$ is the elasticity of substitution in the inputs producing θ^C_{t+1} and $\frac{1}{1-\sigma}$ is the elasticity of substitution of inputs in producing θ^N_{t+1} where $\alpha \in (-\infty, 1]$ and $\sigma \in (-\infty, 1]$.

I^N_t and I^C_t are direct complements with (θ^C_t, θ^N_t) irrespective of the substitution parameters α and σ, except in limiting cases. Focusing on the technology for producing θ^C_{t+1}, when $\alpha = 1$, the inputs are perfect substitutes (the elasticity of substitution is infinite). In this case, the inputs θ^C_t, θ^N_t and I^C_t can be ordered by their relative

[7]See Currie and Blau (2006), and the survey by Cunha, Heckman, Lochner, and Masterov (2006).

[8]See also the work of Greenough, Black, and Wallace (1987).

[9]We do not analyze post-childhood investment in this paper. In principle our technology can account for this.

[10]We thus rule out one potentially important avenue of compensation that agents can specialize in tasks that do not require the skills in which they are deficient. Cunha, Heckman, Lochner, and Masterov (2006) present a more general task function that captures the notion that different tasks require different combinations of skills and abilities.

productivity in producing θ_{t+1}^C. The higher ψ_C^C and ψ_N^C, the higher the productivity of θ_t^C and θ_t^N respectively.

When $\alpha = -\infty$, the elasticity of substitution is zero. All inputs are required in the same proportions to produce a given level of output so there are no possibilities for technical substitution, and $\theta_{t+1}^C = \min\left\{\theta_t^C, \theta_t^N, I_t^C\right\}$. In this version of the technology, investments in periods prior to t as embodied in θ_t^C, θ_t^N are a *bottleneck* for investment in $t+1$. Full compensation for adverse environments before period t through investment in t is impossible.

To complete the *CES* example, assume that adult human capital in period $T+1$, denoted h, is a *CES* function of the two skills accumulated at stage $T+1$,

$$h = \left\{\psi^h \left(\theta_{T+1}^C\right)^{\upsilon} + \left(1-\psi^h\right)\left(\theta_{T+1}^N\right)^{\upsilon}\right\}^{\frac{1}{\upsilon}}, \tag{2.5}$$

where $\psi^h \in [0,1]$, and $\upsilon \in (-\infty, 1]$. In this parameterization, $\frac{1}{1-\upsilon}$ is the elasticity of substitution across different skills in the production of adult human capital. Equation (2.5) reminds us that the market, or life in general, requires the use of multiple skills. Heckman, Stixrud, and Urzua (2006) show the importance of both cognitive and noncognitive skills in explaining a variety of market and nonmarket outcomes.

For the special case where $\alpha = \sigma = \upsilon$ and childhood lasts two periods ($T = 2$), and assuming no period "0" investments, we obtain adult human capital in terms of investments and initial endowments as

$$h = \left\{\begin{array}{c} \tau_{2C}\left(I_2^C\right)^{\alpha} + \tau_{2N}\left(I_2^N\right)^{\alpha} + \tau_{1C}\left(I_1^C\right)^{\alpha} \\ +\tau_{1N}\left(I_1^N\right)^{\alpha} + \tau_{0C}\left(\theta_0^C\right)^{\alpha} + \tau_{0N}\left(\theta_0^N\right)^{\alpha} \end{array}\right\}^{\frac{1}{\alpha}}, \tag{2.6}$$

where

$$\begin{aligned}
\tau_{2C} &= \psi^h B_C^{\alpha}\left(1-\psi_C^C-\psi_N^C\right), \\
\tau_{2N} &= \left(1-\psi^h\right) B_N^{\alpha}\left(1-\psi_C^N-\psi_N^N\right), \\
\tau_{1C} &= \left[\psi^h B_C^{2\alpha}\psi_C^C + \left(1-\psi^h\right)\left(B_N B_C\right)^{\alpha}\psi_C^N\right]\left(1-\psi_C^C-\psi_N^C\right), \\
\tau_{1N} &= \left[\psi^h \left(B_C B_N\right)^{\alpha}\psi_N^C + \left(1-\psi^h\right) B_N^{2\alpha}\psi_N^N\right]\left(1-\psi_C^N-\psi_N^N\right), \\
\tau_{0C} &= \psi^h B_C^{\alpha}\left[\left(\psi_C^C\right)^2 B_C^{\alpha} + \psi_N^C\psi_C^N B_N^{\alpha}\right] \\
&\quad + \left(1-\psi^h\right) B_N^{\alpha}\left(\psi_C^N\psi_C^C B_C^{\alpha} + \psi_N^N\psi_C^N B_N^{\alpha}\right), \text{ and} \\
\tau_{0N} &= \psi^h B_C^{\alpha}\left(\psi_C^C\psi_N^C B_C^{\alpha} + \psi_N^C\psi_N^N B_N^{\alpha}\right) \\
&\quad + \left(1-\psi^h\right) B_N^{\alpha}\left[\psi_C^N\psi_N^N B_C^{\alpha} + \left(\psi_N^N\right)^2 B_N^{\alpha}\right].
\end{aligned}$$

The τ parameters capture the contribution of inputs to final output. Consider τ_{1C} and fix ψ^h, the share of final output due to cognitive skills. The greater the productivity in producing output (B_N and B_C), and the higher the productivity of I^C in producing cognitive skills (the higher $1 - \psi_C^C - \psi_N^C$), the higher τ_{1C} which is the productivity of I_1^C on the stock of skills in period 3. Early investments raise the productivity of later investments by augmenting the stock of skills and facilitating the productivity of later investments. Initial conditions operate in a fashion similar to that of early investments.

For the case of a single skill ($\theta_{Ct} = \theta_{Nt}$ for all t) and a scalar investment, technology (2.6) simplifies to

$$h = \left\{\tau_2 (I_2)^\alpha + \tau_1 (I_1)^\alpha + \tau (\theta_0)^\alpha\right\}^{\frac{1}{\alpha}}, \tag{2.7}$$

where τ_2, τ_1 and τ_0 are defined appropriately.[11] When α is small, low levels of early investments I_1, or low levels of early endowments, are not easily remediated by late investments I_2 in producing adult human capital. The other face of *CES* complementarity is that when α is small, high early investments should be followed with high late investments in order for the early investments to be effective. In the extreme case when $\alpha \to -\infty$, (2.7) converges to $h = (\min\{I_1, I_2, \theta_0\})$. The Leontief case contrasts with the case of perfect *CES* substitutes, which arises when $\alpha = 1$: $h = \tau_2 I_2 + \tau_1 I_1 + \tau_0 \theta_0$. When we impose the further restriction that $\tau_2 = \tau_1$, we generate the model that is implicitly assumed in the literature on human capital investments that collapses childhood into a single period. If $\tau_2 = \tau_1 = \tau_0$, we obtain full symmetry of endowments and inputs. In the special case $\varepsilon_2 = \varepsilon_1$, the output of adult human capital is determined by the total amount of human capital investment, regardless of how the investment is distributed across periods of childhood. In the case of perfect substitutes, it is possible, in a physical productivity sense, to compensate for early investment deficits by later investments, although it may not be economically efficient to do so.

Cunha and Heckman (2006) analyze the optimal timing of investment using a special version of the technology embodied in (2.7). Adapting their analysis to our problem, the ratio of early to late investments varies as a function of α, τ_2, and τ_1. For a model where parents maximize the net present value of child wealth, parents decide how much to invest in period "1," I_1, how much to invest in period "2," I_2, and how much to transfer in risk-free assets, given total parental resources. Parents are assumed to be unable to extract resources from their children.

[11] Set $\psi^h = 1$ or 0 to obtain the coefficients, where, e.g., for $\psi^h = 1$: $\tau_{0C} = \tau_0$, $\tau_{1C} = \tau_1$, and $\tau_{2C} = \tau_2$.

For an interior solution, assuming that the prices of investments are the same in both periods, they obtain

$$\log\left(\frac{I_1}{I_2}\right) = \left(\frac{1}{1-\alpha}\right)\left[\log\left(\frac{\tau_1}{\tau_2}\right) - \log(1+r)\right]. \tag{2.8}$$

Figure 1 plots the ratio of early to late investment as a function of τ_1/τ_2 for different values of α. *Ceteris paribus*, if $\frac{\tau_1}{\tau_2(1+r)} > 1$, the greater the *CES* complementarity, (i.e., the lower α), the lower the growth of investments over time. In the limit, if investments complement each other strongly, optimality implies that they should be equal in both periods. *Ceteris paribus*, the higher τ_1 relative to τ_2, the higher first period investments should be relative to second period investments. Intuitively, if early investments have a substantial impact in determining future stocks of human capital, optimality implies that early investments should also be high relative to later investments. The higher the interest rate, the lower is $\frac{I_1}{I_2}$. This reflects the opportunity costs of investing today relative to investing tomorrow.

An augmented version of (2.8) remains valid as an interior solution to a dynamic general equilibrium overlapping generations model with two periods of childhood and two periods of adulthood, where parents cannot leave debts to children and parents themselves are credit constrained.[12] The effects of early constraints on later investment decisions will depend on the *CES*-complementarity or substitutability of investment across ages. When investments are very substitutable, families will tend to respond to early constraints by re-allocating investments to later periods. In this case, investments during constrained periods should decline, while investments at later ages should increase to partially offset any reductions in human capital. On the other hand, when investments are highly complementary over time, any reduction in early investments makes later investments less productive. If investments are strongly complementary, investment may decline at all stages in response to constraints that only bind for a few stages.

[12] Following Cunha, Heckman, Lochner, and Masterov (2006), let u denote the per period parental preferences for consumption, and c_1 and c_2 denote consumption of the parents when the adults parents are old and young, respectively, we obtain an additional term in (2.8):

$$\log\left(\frac{I_1}{I_2}\right) = \left(\frac{1}{1-\alpha}\right)\left[\log\left(\frac{\tau_1}{\tau_2}\right) - \log(1+r) - \log\frac{u'(c_1)}{u'(c_2)}\right].$$

In our overlapping generations model, period 1 for the child is the first period of adulthood for the parents. Period 2 for the child is the period of old age for the parents.The additional term is the growth rate of the marginal utility of consumption, which reflects the severity of early credit constraints. Investment in children will increase at a faster rate (or, decrease at a slower rate) over the life cycle among constrained families than among unconstrained families.

While these examples are useful for building intuition about the importance of the elasticity of substitution for determining the optimal timing of lifecycle investments, they oversimplify the analysis of skill formation. It is implausible that the elasticity of substitution for skills in final output is the same as the elasticity of substitution for investments in skills, and that a common elasticity of substitution governs the inputs into cognitive and noncognitive skills.

In our econometric analysis, we allow different elasticities of substitution to govern the technologies generating cognitive and noncognitive skills, and for both to be different from the elasticity of substitution for cognitive and noncognitive skills in final output. Thus, we do not impose the assumption that $\alpha = \sigma$, nor do we require that either α or σ equals υ. The goal of this paper is to estimate the elasticity parameters governing the trade-offs among investments and endowments in producing adult human capital, and to explore their implications.

3 Identifying the Technology using Dynamic Factor Models

Identifying and estimating technology (2.1) is a challenging task. Both the inputs and outputs can only be proxied, producing the problem of measurement error. General nonlinear specifications of technology (2.1) raise additional problems regarding measurement error in latent variables for nonlinear systems. In this paper, we build on the analysis of Cunha and Heckman (2006) to estimate the full technology presented in Section 2.

In the notation of Section 2, the goal of this paper is to identify the evolution of both cognitive and noncognitive outcomes, as defined by a version of (2.1) augmented to include unobserved (by the econometrician) productive inputs η_t^C and η_t^N and the stock of the mother's cognitive and noncognitive skills (θ_M^C and θ_M^N, respectively):

$$\begin{aligned} \theta_{t+1}^C &= f_t^C\left(\theta_t^C, \theta_t^N, I_t^C, \theta_M^C, \theta_M^N, \eta_t^C\right) \\ \theta_{t+1}^N &= f_t^N\left(\theta_t^C, \theta_t^N, I_t^N, \theta_M^C, \theta_M^N, \eta_t^N\right), \ t = 1, \ldots, T, \text{ given } \theta_0 = \left(\theta_0^C, \theta_C^N\right). \end{aligned} \tag{3.1}$$

We assume that the technologies are monotone increasing in their arguments and twice continuously differentiable. Specifically, we seek to accomplish the following tasks. (1) To determine how stocks of cognitive and noncognitive skills at date t affect the stocks of skills at date $t+1$, examining both self productivity (the effects of θ_t^N on θ_{t+1}^N, and θ_t^C on θ_{t+1}^C) and cross productivity (the effects of θ_t^C on θ_{t+1}^N and the effects of θ_t^N on θ_{t+1}^C) at each stage of the life cycle. (2) To develop a non-linear dynamic factor model where we proxy $\theta_t = \left(\theta_t^N, \theta_t^C\right)$ by vectors of measurements on skills which can include test scores as well as outcome measures. In our analysis, test scores and parental inputs are indicators of latent skills and latent investments. We account for measurement errors in output and input measures. (3) To estimate the elasticities of substitution for the technologies governing the production of cognitive and noncognitive skills. (4) To anchor the scale of test scores using adult outcome measures instead of relying on test scores as a measure of output. This enables us to compare the impact of diverse inputs on a common measure of output.

Our identification analysis proceeds in the following way. We start with a model where the measurements are linear and separable in the latent variables, as in Cunha and Heckman (2006). We establish identification of the joint distribution of the latent variables without imposing their strong independence assumptions about the measurement errors. With the joint distribution of latent variables in hand, we nonparametrically identify technology (3.1) given alternative assumptions about η_t^k. We then extend the analysis to identify a model with general anchors for the latent variables in adult outcomes to make their scale interpretable, and to compare the impacts of different inputs on different technologies based on a common metric. Finally, we relax the assumption of separability in the latent variables in the measurement equations.

3.1 Identifying the Distribution of the Latent Variables

We assume that θ_t^C and θ_t^N is not directly observed. Instead, we observe a vector of measurements, such as test scores and behavioral measures, $Y_{j,t}^C, Y_{j,t}^N$ for $j = 1, 2, \ldots, m_t^k$, $k = C, N$, respectively. We assume that measurements are separable additive functions of the latent inputs θ_t^k,

$$Y_{j,t}^k = \mu_{j,t}^k + \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k \text{ for } j = 1, \ldots, m_t^k, \_t = 1, \ldots, T, \text{ and } k = C, N, \tag{3.2}$$

and set $\alpha_{1,t}^k = 1$ and $E\left(\theta_t^k\right) = 0$ for $k = C, N$, $t = 1, \ldots, T$. Such normalizations are required to set the scale and the location of the factors. We relax the separability assumption below.

We also assume that parental investments I_t^k are not directly observed. We observe a vector of measurements $X_{j,t}^k$. We represent these by

$$X_{j,t}^k = \mu_{j,t}^{X^k} + \beta_{j,t}^k I_t^k + \varepsilon_{j,t}^{I^k} \text{ for } j = 1, \ldots, m_t^{I^k}, t = 1, \ldots, T, \text{ and } k = C, N \tag{3.3}$$

and we set $\beta_{1,t}^k = 1$ and $E\left(I_t^k\right) = 0$ for $k = C, N$, $t = 1, \ldots, T$. These normalizations set the scale and the location of I_t. Let θ_M^C and θ_M^N denote mother's latent cognitive and noncognitive skills, respectively, which are also assumed to be unobserved by the econometrician. We have a vector of measurements on mother's cognitive and noncognitive abilities, M_j^k for $j = 1, \ldots, m_{M^k}$,

$$M_j^k = \mu_j^{M^k} + \delta_j^k \theta_M^k + \varepsilon_j^{M^k}, \quad k = C, N. \tag{3.4}$$

We normalize $\delta_1^k = 1$ and $E\left(\theta_M^k\right) = 0$ for $k = C, N$. We seek to recover the joint distribution of the latent variables, as well as the associated factor loadings in the measurement equations. We use this joint distribution as an input for identifying the general stage-specific technology functions f_t^k for $t = 1, \ldots, T$, $k = C, N$. We also secure nonparametric identification of the distributions of η_t^k, for $t = 1, \ldots, T$ and $k = C, N$, under the assumptions given below. For simplicity we do not discuss identification of the means of the measurements because under the assumptions presented in the next section, this is straightforward. We can include regressors in the model through the $\mu_{j,t}^k$, $\alpha_{j,t}^k$, $\mu_{i,j}^{X^k}$, $\beta_{j,t}^k$, $\mu_j^{M^k}$, δ_j^k. If the measurement equations are linear and separable, we only need to assume that, conditional on the regressors for a given k and t, the $\varepsilon_{j,t}^k$ are uncorrelated over $j = 1, \ldots, m_t^k$ and with θ_t^k. We make parallel assumptions for $\varepsilon_{j,t}^{I^k}$ and $\varepsilon_j^{M^k}$. They are uncorrelated with θ_m^k, I_t^k, and θ_t^k and with each other. The uncorrelatedness assumptions are much weaker than the independence assumptions invoked in Cunha and Heckman (2006). We now establish identifiability of the factor loadings and the joint distribution of the latent variables.

3.2 The Identification of the Factor Loadings and of the Joint Distributions of the Latent Variables

We first establish identification of the factor loadings under the assumption that $m_t^k \geq 3$, $k = C, N, I^C, I^N, M^C, M^N$. Without loss of generality, we focus on $\alpha_{j,t}^C$ and note that similar expressions can be derived for the loadings for the remaining latent factors.

Since $\left\{Y_{j,t}^C\right\}_{j=1}^{m_t^C}$ is observed, we can compute $Cov\left(Y_{i,t}^C, Y_{j,t}^C\right)$ from the data for all i, j pairs. Let $\sigma_{\theta_t^C}^2$ denote the variance of θ_t^C, $t = 1, 2, \ldots, T$, and note that it may change over time. Since we have normalized $\alpha_{1,t}^C = 1$, we obtain:

$$Cov\left(Y_{1,t}^C, Y_{2,t}^C\right) = \alpha_{2,t}^C \sigma_{\theta_t}^2. \tag{3.5}$$

In addition,

$$Cov\left(Y_{1,t}^{C}, Y_{3,t}^{C}\right) = \alpha_{3,t}^{C}\sigma_{\theta_t}^{2} \quad \text{and} \quad Cov\left(Y_{2,t}^{C}, Y_{3,t}^{C}\right) = \alpha_{2,t}^{C}\alpha_{3,t}^{C}\sigma_{\theta_t}^{2}. \tag{3.6}$$

Taking ratios,

$$\frac{Cov\left(Y_{2,t}^{C}, Y_{3,t}^{C}\right)}{Cov\left(Y_{1,t}^{C}, Y_{3,t}^{C}\right)} = \alpha_{2,t}^{C}, \quad \alpha_{3,t}^{C} \neq 0, \sigma_{\theta_t}^{2} \neq 0, \quad \text{and} \quad \frac{Cov\left(Y_{2,t}^{C}, Y_{3,t}^{C}\right)}{Cov\left(Y_{1,t}^{C}, Y_{2,t}^{C}\right)} = \alpha_{3,t}^{C}, \quad \alpha_{2,t}^{C} \neq 0, \sigma_{\theta_t}^{2} \neq 0.$$

We can identify $\alpha_{2,t}^{C}$ and $\alpha_{3,t}^{C}$ from the ratios of covariances. Proceeding in the same fashion, we can identify $\alpha_{j,t}^{C}$ for $j = 2, 3, \ldots, m_t^C$, $t = 1, \ldots, T$ up to the normalization $\alpha_{1,t}^{C} = 1$. Then, using for example (3.5), we can identify $\sigma_{\theta_t^C}^{2}$ for all $t = 1, 2, \ldots, T$.

Once the parameters $\alpha_{1,t}^{C}, \alpha_{2,t}^{C}, \ldots, \alpha_{m_t^C,t}^{C}$ are identified (up to the normalization $\alpha_1^C = 1$), we can rewrite (3.2), assuming $\alpha_{j,t}^{C} \neq 0$, as:

$$\frac{Y_{j,t}^{C}}{\alpha_{j,t}^{C}} = \frac{\mu_{j,t}^{C}}{\alpha_{j,t}^{C}} + \theta_t^C + \frac{\varepsilon_{j,t}^{C}}{\alpha_{j,t}^{C}}, \; j = 1, 2, \ldots, m_t^C. \tag{3.7}$$

In this form, it is clear that the known quantities $Y_{j,t}^{C}/\alpha_{j,t}^{C}$ play the role of repeated error-contaminated measurements of θ_t^C. Hence, we can employ the analysis of Schennach (2004a,b) to find a closed form expression for the joint distribution of the latent factors, collectively denoted by

$$\theta = \left(\left\{\theta_t^C\right\}_{t=1}^{T}, \left\{\theta_t^N\right\}_{t=1}^{T}, \left\{I_t^C\right\}_{t=1}^{T}, \left\{I_t^N\right\}_{t=1}^{T}, \theta_M^C, \theta_M^N\right).$$

Although the availability of numerous indicators for each latent factor is helpful in improving the efficiency of the estimation procedure, the identification of the model can be secured (after the factor loadings are known) if only two measurements of each latent factor are available. Since we have at least two different measurements for each latent factor, we can define, without loss of generality, the following two vectors

$$\begin{aligned} W_1 &= \left(\left\{Y_{1,t}^C/\alpha_{1,t}^C\right\}_{t=1}^{T}, \left\{Y_{1,t}^N/\alpha_{1,t}^N\right\}_{t=1}^{T}, \left\{X_{1,t}^{I^C}/\beta_{1,t}^C\right\}_{t=1}^{T}, \left\{X_{1,t}^{I^N}/\beta_{1,t}^N\right\}_{t=1}^{T}, M_1^C/\delta_1^C, M_1^N/\delta_1^N\right)' \\ W_2 &= \left(\left\{Y_{2,t}^C/\alpha_{2,t}^C\right\}_{t=1}^{T}, \left\{Y_{2,t}^N/\alpha_{2,t}^N\right\}_{t=1}^{T}, \left\{X_{2,t}^{I^C}/\beta_{2,t}^C\right\}_{t=1}^{T}, \left\{X_{2,t}^{I^N}/\beta_{2,t}^N\right\}_{t=1}^{T}, M_2^C/\delta_2^C, M_2^N/\delta_2^N\right)'. \end{aligned}$$

These vectors consist of the first and the second measurements for each factor, respectively. The corresponding measurement errors are

$$\omega_1 = \left(\left\{ \varepsilon_{1,t}^{C} \right\}_{t=1}^{T}, \left\{ \varepsilon_{1,t}^{N} \right\}_{t=1}^{T}, \left\{ \varepsilon_{1,t}^{I^C} \right\}_{t=1}^{T}, \left\{ \varepsilon_{1,t}^{I^N} \right\}_{t=1}^{T}, \varepsilon_1^{M^C}, \varepsilon_1^{M^N} \right),$$

$$\omega_2 = \left(\left\{ \varepsilon_{2,t}^{C}/\alpha_{2,t}^{C} \right\}_{t=1}^{T}, \left\{ \varepsilon_{2,t}^{N}/\alpha_{2,t}^{N} \right\}_{t=1}^{T}, \left\{ \varepsilon_{2,t}^{I^C}/\beta_{2,t}^{C} \right\}_{t=1}^{T}, \left\{ \varepsilon_{2,t}^{I^N}/\beta_{2,t}^{N} \right\}_{t=1}^{T}, \varepsilon_2^{M^C}/\delta_2^{C}, \varepsilon_2^{M^N}/\delta_2^{N} \right).$$

Identification of the distribution of θ is obtained from the following theorem. Let L denote the total number of latent factors (here $4T+2$). We let "$\perp$" denote orthogonality. We keep conditioning on the X implicit.

Theorem 1 *Let W_1, W_2, θ, ω_1, ω_2 be random variables taking values in $\mathbb{R}^L$ and related through*

$$W_1 = \theta + \omega_1,$$
$$W_2 = \theta + \omega_2.$$

If, for $k, k' = 1, \ldots, L$,

$$E\left[\omega_{1,k} | \theta_k, \omega_{2,k}\right] = 0 \tag{3.8}$$

$$\omega_{2,k} \perp \left(\theta, \omega_{2,k'}\right) \text{ for } k \neq k', \tag{3.9}$$

then the density of θ is given by:

$$p_\theta(\theta) = (2\pi)^{-L} \int e^{-i\chi\cdot\theta} \frac{E\left[e^{i\chi\cdot W_2}\right]}{\prod_{k=1}^{L} E\left[e^{i\chi_k W_{2,k}}\right]} \prod_{k=1}^{L} \exp\left(\int_0^{\chi_k} \frac{E\left[iW_{1,k} e^{i\zeta_k W_{2,k}}\right]}{E\left[e^{i\zeta_k W_{2,k}}\right]} d\zeta_k \right) d\chi, \tag{3.10}$$

provided all the above expectations exist.

Proof. *Combine the univariate nonparametric identification analysis of Theorem 1 from Schennach (2004b) with the multivariate moment identification result in Schennach (2004a).* ■

Note that θ can include elements that are perfectly measured. In that case, the corresponding elements of W_1 and W_2 are the same. An alternative set of assumptions producing identification is as follows. These assumptions allow for correlation over time in the measurement errors in a given type of measure of skills.[13]

[13] These conditions are informally sketched in Schennach (2004a, footnote 11).

Theorem 2 *Let W_1, W_2, θ, ω_1, ω_2 be random variables taking values in $\mathbb{R}^L$ and related through*

$$\begin{aligned} W_1 &= \theta + \omega_1, \\ W_2 &= \theta + \omega_2. \end{aligned}$$

If

$$E\left[\omega_1 | \theta, \omega_2\right] = 0, \tag{3.11}$$

$$\omega_2 \perp \theta, \tag{3.12}$$

then the density of θ is given by:

$$p_\theta\left(\theta\right) = \left(2\pi\right)^{-L} \int e^{-i\chi\cdot\theta} \exp\left(\int_0^\chi \frac{E\left[iW_1 e^{i\zeta\cdot W_2}\right]}{E\left[e^{i\zeta\cdot W_2}\right]} \cdot d\zeta\right) d\chi, \tag{3.13}$$

provided all the above expectations exist. Note that the innermost integral is the integral of a vector-valued field along a continuous path joining the origin and the point $\chi \in \mathbb{R}^L$ and that the outermost integral is over the whole $\mathbb{R}^L$ space.

Proof. *See Appendix A.*

Armed with the density of θ, we turn to the identification of the technology of skill formation.

3.3 Identification of the Technology Function

If the density of θ is known, we can identify a general nonseparable technology function,

$$\theta_{t+1}^k = f_t^k\left(\theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N, \eta_t^k\right), \; t = 1, \ldots, T, \tag{3.14}$$

for $k = C, N$, if, for example, we assume that η_t^k is independent over time and independent of all the other inputs including $\eta_t^{k'}$, $k \neq k'$. Note that even if θ were perfectly observed, we could not separately identify the distribution of η_t^k and the function f_t^k because, without further normalization, a change in the density of η_t^k can be undone by a change in the function f_t^k.

One solution to this problem is to assume that (3.14) is additively separable in η_t^k. Another way to avoid this ambiguity is to normalize η_t^k to have a uniform density on $[0, 1]$. Any of the normalizations suggested by Matzkin (2003, 2007) could be used. Assuming η_t^k is uniform $[0, 1]$, we show that f_t^k for $k = C$ or N is nonparametrically identified, by noting that, from the knowledge of p_θ, we can calculate, for any $\bar{\theta} \in \mathbb{R}$,

$$\Pr\left[\theta_{t+1}^k \leq \bar{\theta} | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right] \equiv G\left(\bar{\theta} | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right).$$

We identify the technology of skill formation by setting

$$f_t^k\left(\theta_t^C, \theta_t^N, I_t^k, \eta_t^k, \theta_M^C, \theta_M^N\right) = G^{-1}\left(\eta_t^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right)$$

where $G^{-1}\left(\eta_t^k | \theta_t^N, \theta_t^C, I_t^k, \theta_M^C, \theta_M^N\right)$ denotes the inverse of $G\left(\bar{\theta} | \theta_t^N, \theta_t^C, I_t^k, \theta_M^C, \theta_M^N\right)$ with respect to its first argument, i.e. the value $\bar{\theta}$ such that $\eta_t^k = G\left(\bar{\theta} | \theta_t^N, \theta_t^C, I_t^k, \theta_M^C, \theta_M^N\right)$. By construction, this operation produces a function f_t^k that generates outcomes θ_{t+1}^k with the appropriate distribution, because a random variable is mapped into a uniformly distributed variable under the mapping defined by its own cdf.

The more traditional separable technology function with zero mean disturbance, $\theta_{t+1}^k = f_t^k\left(\theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right) + \eta_t^k$, is covered by our analysis by defining

$$f_t^k\left(\theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right) \equiv E\left[\theta_{t+1}^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right],$$

where the expectation is taken under the density $p_{\theta_{t+1}^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N}$, which can be calculated from p_θ. The density of η_t^k conditional on all variables is identified from

$$\begin{aligned} & p_{\eta_t^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N}\left(\eta_t^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right) \\ & \quad = p_{\theta_{t+1}^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N}\left(\eta_t^k + E\left[\theta_{t+1}^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right] | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N\right), \end{aligned}$$

since $p_{\theta_{t+1}^k | \theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N}$ is known from p_θ. Our analysis of the identification of the production function with missing inputs is more general than that of Olley and Pakes (1996), who also consider use of proxies to measure unobserved inputs. They assume access to perfect proxies to measure unobserved inputs, whereas we allow for imperfectly measured proxies, i.e. measurement error.

We now show how to anchor θ_{T+1}^C and θ_{T+1}^N in adult outcomes.

3.4 Anchoring Skills in an Interpretable Metric

It is common in the empirical literature on child investment to measure outcomes in units of test scores. However, test scores are arbitrarily scaled. To gain a better understanding of the relative importance of cognitive and noncognitive skills and their interactions, and of investments at different stages of the life cycle, it is desirable to anchor these skills in a common scale.

To do so we model the effect of period $T+1$ cognitive and noncognitive skills on high school graduation rates:

$$\Pr\left[D=1|\theta^C_{T+1},\theta^N_{T+1}\right]=g\left(\theta^C_{T+1},\theta^N_{T+1}\right),$$

where $D=1$ indicates graduation from high school ($D=0$ otherwise). Any other function of period $T+1$ latent skills can be used that satisfies our conditions (e.g. earnings or the probability of occupational choice). This function corresponds to the "h" in Section 2. We keep the conditioning on the regressors implicit. To establish identification of $g\left(\theta^C_{T+1},\theta^N_{T+1}\right)$, we include the dummy D in the vector θ. Since we assume that the dummy D is measured without error, the corresponding element of the two repeated measurement vectors W_1 and W_2 are identical and equal to D. Theorem 1 implies that the joint density of D, θ^C_t and θ^N_t is identified, thus making it possible to identify $\Pr\left[D=1|\theta^C_{T+1},\theta^N_{T+1}\right]$.

We can now extract two separate "anchors" $g^C\left(\theta^C_{T+1}\right)$ and $g^N\left(\theta^N_{T+1}\right)$ from the function $g\left(\theta^C_{T+1},\theta^N_{T+1}\right)$, by integrating over one of the variables, e.g.

$$\begin{aligned} g^C\left(\theta^C_{T+1}\right) &\equiv \int g\left(\theta^C_{T+1},\theta^N_{T+1}\right)p_{\theta^N_{T+1}}\left(\theta^N_{T+1}\right)d\theta^N_{T+1}, \\ g^N\left(\theta^N_{T+1}\right) &\equiv \int g\left(\theta^C_{T+1},\theta^N_{T+1}\right)p_{\theta^C_{T+1}}\left(\theta^C_{T+1}\right)d\theta^C_{T+1}, \end{aligned}$$

where the marginal densities, such as $p_{\theta^N_T}\left(\theta^N_{T+1}\right)$, are identified from our preceding analysis. We condition on the regressors. The "anchored" skills, denoted by $\tilde{\theta}^k_t$, are defined as

$$\tilde{\theta}^k_t=g^k\left(\theta^k_t\right),\ k=C,N,\quad t=1,\ldots,T.$$

We assume that conditional on the regressors, the $g^k\left(\theta^k_t\right)$ are invertible. We combine the identification of the anchoring functions with the identification of the technology function $f^k_t\left(\theta^C_t,\theta^N_t,I^k_t,\eta^k_t,\theta^M_C,\theta^M_N\right)$ established in the previous section to prove that the technology function expressed in terms of the anchored skills — denoted by $\tilde{f}^k_t\left(\tilde{\theta}^C_t,\tilde{\theta}^N_t,I^k_t,\eta^k_t,\theta^M_C,\theta^M_N\right)$ — is also identified. To do so, we redefine the technology function to be, for $k=C,N$,

$$\tilde{f}^k_t\left(\tilde{\theta}^C_t,\tilde{\theta}^N_t,I^k_t,\theta^C_M,\theta^N_M,\eta^k_t\right)\equiv g^k\left(f^k_t\left(\left[g^C\right]^{-1}\left(\tilde{\theta}^C_t\right),\left[g^N\right]^{-1}\left(\tilde{\theta}^N_t\right),I^k_t,\theta^C_M,\theta^N_M,\eta^k_t\right)\right),$$

where $\left[g^k\right]^{-1}(\cdot)$ denotes the inverse of the function $g^k(\cdot)$. It is easily established that

$$\begin{aligned}
\tilde{f}_t^k\left(\tilde{\theta}_t^C, \tilde{\theta}_t^N, I_t^k, \theta_M^C, \theta_M^N, \eta_t^k\right) &= \tilde{f}_t^k\left(g^C\left(\theta^C\right), g^N\left(\theta_t^N\right), I_t^k, \theta_M^C, \theta_M^N, \eta_t^k\right) \\
&= g^k\left(f_t^k\left(\left[g^C\right]^{-1}\left(g^C\left(\theta_t^C\right)\right), \left[g^N\right]^{-1}\left(g^N\left(\theta_t^N\right)\right), I_t^k, \theta_M^C, \theta_M^N, \eta_t^k\right)\right) \\
&= g^k\left(f_t^k\left(\theta_t^C, \theta_t^N, I_t^k, \theta_M^C, \theta_M^N, \eta_t^k\right)\right) \\
&= g^k\left(\theta_{t+1}^k\right) = \tilde{\theta}_{t+1}^k,
\end{aligned}$$

as desired. Hence, $\tilde{f}_t^k$ is the equation of motion for the anchored skills $\tilde{\theta}_t^k$ that is consistent with the equation of motion f_t^k for the original skills θ_t^k.

Our reliance on measurement equations that are separable in the latent variables appears to confer a certain arbitrariness on the analysis. Carneiro, Hansen, and Heckman (2003) present an analysis for nonseparable measurement equations based on a separable latent index structure, but invoke strong independence and "identification-at-infinity" assumptions. We develop a more general approach for identifying the distribution of θ from general nonseparable measurement equations.

3.5 The identification of a general model for measurements

In this section, we consider a factor model of the general form

$$Z_j = a_j\left(\theta, \varepsilon_j\right) \text{ for } j = 1, \ldots, m, \tag{3.15}$$

where $m \geq 3$ and where the indicator Z_j is observed while the latent factor θ and the disturbance ε_j are not. We keep conditioning on the covariates implicit. The variables Z_j, θ, and ε_j are vectors of the same dimension. In our application, the observed vector of indicators and corresponding disturbances would be

$$\begin{aligned}
Z_j &= \left(\left\{Y_{j,t}^C\right\}_{t=1}^T, \left\{Y_{j,t}^N\right\}_{t=1}^T, \left\{X_{j,t}^{I^C}\right\}_{t=1}^T, \left\{X_{j,t}^{I^N}\right\}_{t=1}^T, M_j^C, M_j^N\right) \\
\varepsilon_j &= \left(\left\{\varepsilon_{j,t}^C\right\}_{t=1}^T, \left\{\varepsilon_{j,t}^N\right\}_{t=1}^T, \left\{\varepsilon_{j,t}^{I^C}\right\}_{t=1}^T, \left\{\varepsilon_{j,t}^{I^N}\right\}_{t=1}^T, \varepsilon_j^{M^C}, \varepsilon_j^{M^N}\right)
\end{aligned}$$

while the vector of unobserved latent factors is

$$\theta = \left(\left\{\theta_t^C\right\}_{t=1}^T, \left\{\theta_t^N\right\}_{t=1}^T, \left\{I_t^C\right\}_{t=1}^T, \left\{I_t^N\right\}_{t=1}^T, \theta_M^C, \theta_M^N\right).$$

The functions $a_j(\cdot,\cdot)$ for $j = 1, \ldots, m$ in equation (3.15) are unknown. It is necessary to normalize one of them (without loss of generality, $a_1(\cdot,\cdot)$) in some way to achieve identification, as established in the following theorem.

Theorem 3 *The distribution of θ in Equation (3.15) is identified under the following conditions:*

1. $p_{Y_2|Y_1,Y_3,\theta}(Y_2|Y_1,Y_3,\theta) = p_{Y_2|\theta}(Y_2|\theta)$ *and* $p_{Y_1|Y_3,\theta}(Y_1|Y_3,\theta) = p_{Y_1|\theta}(Y_1|\theta)$.

2. $p_{Z_1|Z_3}(Z_1|Z_3)$ *and* $p_{Z_1|\theta}(Z_1|\theta)$ *each form a complete family of distributions (indexed by* Z_3 *and* θ*, respectively).*

3. *The density* $p_{Z_2|\theta}(Z_2|\theta)$ *is bounded uniformly in* Z_2 *and* θ*.*

4. *Whenever* $\theta \neq \tilde{\theta}$, $p_{Z_2|\theta}(Z_2|\theta)$ *and* $p_{Z_2|\theta}\left(Z_2|\tilde{\theta}\right)$ *differ at least over a set of positive probability.*

5. *There exists a known functional* M*, mapping a density to a vector, that has the property that* $M\left[p_{Z_1|\theta}(\cdot|\theta)\right] = \theta$.

Proof. *See Appendix A.*

The first set of conditional independence assumptions is implied by the conditional independence assumptions traditionally made in the standard linear factor model. A vector of correctly measured variables C can trivially be handled by including C in the list of conditioning variables in all densities invoked in the assumptions. Theorem 3 will then imply that $p_{\theta|C}(\theta|C)$ is identified. Since $p_C(C)$ is identified it follows that $p_{\theta,C}(\theta,C) = p_{\theta|C}(\theta|C)p_C(C)$ is also identified.

Versions of Assumption 2 appear in the nonparametric instrumental variable literature (e.g. Newey and Powell (2003), Darolles, Florens, and Renault (2002)). Intuitively, the requirement that $p_{Z_1|Z_3}(Z_1|Z_3)$ form a complete family demands that the density of Z_1 vary sufficiently as Z_3 varies (and similarly for $p_{Z_1|\theta}(Z_1|\theta)$). More formally, it must be possible to express any density as a (perhaps uncountably infinite) linear combination of $p_{Z_1|Z_3}(Z_1|Z_3)$ for different values of Z_3.

Assumption 3 is a standard regularity condition frequently invoked to justify maximum likelihood estimators. Assumption 4 is automatically satisfied, for instance, if θ is univariate and $a_2(\theta,\varepsilon_2)$ is strictly increasing in θ. However it holds much more generally, since $a_2(\theta,\varepsilon_2)$ is nonseparable, the distribution of Z_2 conditional on θ can change with θ, thus enabling Assumption 4 to be satisfied even if $a_2(\theta,\varepsilon_2)$ is not strictly increasing in θ.

The last assumption specifies how the observed Z_1 is used as a reference to "anchor" the scale of the unobserved θ. The most common choice of the functional M

would be the mean, the mode, the median, or any other well-defined measure of location (thus allowing for nonclassical measurement error). One way to satisfy this assumption is to normalize $a_1(\theta, \varepsilon_1)$ to be equal to $\theta + \varepsilon_1$, where ε_1 has zero mean. Many other nonseparable functions can also satisfy this assumption. With the distribution of $p_\theta(\theta)$ in hand, we can identify the technology using the analysis of section 3.3.

Note that Theorem 3 *does not* state that the distributions of the errors ε_j or that the functions $a_j(\cdot, \cdot)$ are fully identified. In fact, it is always possible to alter the distribution of ε_j and the dependence of the function $a_j(\cdot, \cdot)$ on its second argument in ways that cancel each other out, as has often been noted in the nonseparable literature (e.g. Matzkin, 2003, 2007). However, this ambiguity does not prevent the identification of the distribution of θ.

We now present our method for implementing the identification analysis and the model to produce the empirical estimates reported in this paper.

4 Estimation

Although we have established nonparametric identification of the technology of skill formation, we use parametric maximum likelihood to estimate the model and we do not estimate under the most general possible conditions given our parametric assumptions. Our approach is motivated by two observations. First, a fully nonparametric approach is too data hungry to apply to samples of the size we have at our disposal, because the convergence rates of nonparametric versions of our estimators are slow. Second, solving a high-dimensional dynamic factor model is a computationally demanding task that can only be made manageable by invoking parametric restrictions. Our approach offers the advantage of providing a simple way of obtaining standard errors. For simplicity in the empirical work reported in this paper, we will further limit our analysis to the separable factor model for measurements. However, as the preceding discussion illustrates, this is not a fundamental restriction of our approach.

We now develop our likelihood function. As before, $p(\theta)$ denotes the density of θ. Although we do not directly observe θ, we observe the measurements of it, Z. Let $z_{t,j,h}^k$ denote the measurement j associated with the factor θ_t^k for person h in period t. Sample size is H. Let $\varepsilon_{t,j,h}^k$ denote the measurement error associated with the measurement $z_{t,j,h}^k$. Let $p_{\varepsilon_{j,t}^k}$ denote the density function of $\varepsilon_{t,j,h}^k$. We can write the likelihood in terms of ingredients we can measure or identify as:

$$p(Z) = \prod_{h=1}^{H} \int \cdots \int p(\theta) \prod_{t=1}^{T} \prod_{k \epsilon \{C,N,I,M^C,M^N\}} \prod_{j=1}^{m_t^k} p_{\varepsilon_{j,t,h}^k} \left(z_{j,t,h}^k - \mu_{j,t,h}^k - \alpha_{j,t,h}^k \theta_t^k \right) d\theta_t^k. \tag{4.1}$$

This is maximized subject to a parametric technology constraint and the normalizations on the measurements discussed in section 3.1. We assume that the measurement error ε_{jt}^k is classical, so that $p_{\varepsilon_{jt}^k|\theta_t^k}\left(\varepsilon_{jt}^k|\theta_t^k\right) = p_{\varepsilon_{jt}^k}\left(\varepsilon_{jt}^k\right)$. This considerably reduces the number of terms needed to form the likelihood.[14]

In principle, one can estimate the parameters in $\left\{\mu_{j,t}^k, \alpha_{j,t}^k\right\}$, the parameters of the technology, and the $p(\theta)$ by maximizing (4.1) directly. In order to do that, one can approximate $p(Z)$ by computing the integrals numerically in a deterministic fashion. However, if the number of integrals is very large, a serious problem arises. The number of points required to evaluate the integrals is very large. For example, if we have three latent variables and four time periods, so that $T = 4$, then $\dim(\theta) = 12$ and we have to compute an integral of dimension twelve to obtain the function $p(Z)$. This implies computing about seventeen million points for each individual h if we pick four evaluation points for each integral we have to approximate numerically. The rate of convergence of the numerical approximation decreases with $\dim(\theta)$. Therefore, in order to obtain good approximations of $p(Z)$ even in the case with three factors and four time periods, we would need more than only 4 points of evaluation for each integral.

We avoid this problem by relying on simulation methods, which offer the well-known advantage of enabling the approximation of these integrals with an accuracy that only depends on the number of simulation draws, regardless of the actual dimension of the underlying variables.

4.1 Nonlinear Filtering

The vector Z is a collection of all measurements, for all factors, for all periods. Let z_t denote the coordinates of Z that are specific to period t:

$$z_t = \left(\left\{ Y_{j,t}^C / \alpha_{j,t}^C \right\}_{j=1}^{m_t^C}, \left\{ Y_{j,t}^N / \alpha_{j,t}^N \right\}_{j=1}^{m_t^N}, \left\{ X_{j,t}^{I^C} / \beta_{j,t}^C \right\}_{j=1}^{m_t^{I^C}}, \left\{ X_{j,t}^{I^N} / \beta_{j,t}^N \right\}_{j=1}^{m_t^{I^N}} \right)'.$$[15]

[14] We can allow for serial correlation in the measurement errors, but at greater computational cost.
[15] Without loss of generality, we can assume that the z_t have been demeaned.

Define ε_t similarly. Let θ_t denote the coordinates of θ that are specific to period t: $\theta_t = \left(\theta_t^C, \theta_t^N, I_t^C, I_t^N\right)$. For each period t, we have two vector equations:

$$z_t = \theta_t + \varepsilon_t, \tag{4.2}$$

$$\theta_{t+1} = f_t\left(\theta_t, \theta_M^C, \theta_M^N, \eta_t\right). \tag{4.3}$$

We can express $p(Z)$ as

$$p(Z) = p(z_1) \prod_{t=2}^{T} p\left(z_t | z^{t-1}\right), \text{ where } z^{t-1} = (z_1, \dots, z_{t-1}).$$

Because of the nonlinearity of our model we cannot use Kalman filtering. We use particle filtering methods to obtain $p\left(z_t | z^{t-1}\right)$ for $t = 2, \dots, T$ (see Hammersley and Morton, 1954; Doucet, de Freitas, and Gordon, 2001). Note that:

$$\begin{aligned} p\left(z_t | z^{t-1}\right) &= \int p\left(z_t, \theta_t | z^{t-1}\right) d\theta_t = \int p\left(z_t | \theta_t, z^{t-1}\right) p\left(\theta_t | z^{t-1}\right) d\theta_t \\ &= \int p(z_t | \theta_t)\, p\left(\theta_t | z^{t-1}\right) d\theta_t. \end{aligned}$$

Thus,

$$p(z_1, \dots, z_T) = \prod_{t=1}^{T} \int p(z_t | \theta_t)\, p\left(\theta_t | z^{t-1}\right) d\theta_t. \tag{4.4}$$

From our assumption about measurement errors, we know that $p(z_t | \theta_t) = p(\varepsilon_t)$. The problem is to construct $p\left(\theta_t | z^{t-1}\right)$. Nonlinear filters are algorithms that, given $p\left(\theta_t | z^{t-1}\right)$, allow one to compute $p\left(\theta_{t+1} | z^t\right)$. Similar to the Kalman filter, nonlinear filtering breaks this task into two steps: update and prediction. The update step produces $p\left(\theta_t | z^t\right)$ given $p\left(\theta_t | z^{t-1}\right)$. To perform this update step apply Bayes' rule:

$$p\left(\theta_t | z^t\right) = p\left(\theta_t | z_t, z^{t-1}\right) = \frac{p\left(\theta_t, z_t | z^{t-1}\right)}{p\left(z_t | z^{t-1}\right)} = \frac{p\left(z_t | \theta_t, z^{t-1}\right) p\left(\theta_t | z^{t-1}\right)}{p\left(z_t | z^{t-1}\right)} = \frac{p(z_t | \theta_t)\, p\left(\theta_t | z^{t-1}\right)}{p\left(z_t | z^{t-1}\right)}, \tag{4.5}$$

where the normalizing constant is $p\left(z_t | z^{t-1}\right) = \int p(z_t | \theta_t)\, p\left(\theta_t | z^{t-1}\right) d\theta_t$, which depends on $p(z_t | \theta_t) = p(\varepsilon_t)$ as defined by the measurement equation.

The prediction step generates $p\left(\theta_{t+1} | z^t\right)$ given $p\left(\theta_t | z^t\right)$, using f_t to obtain the prediction density of θ_t using the Chapman-Kolmogorov equation:

$$p\left(\theta_{t+1} | z^t\right) = \int p(\theta_{t+1} | \theta_t)\, p\left(\theta_t | z^t\right) d\theta_t. \tag{4.6}$$

By combining update and prediction steps, one can calculate $p(\theta_{t+1}|\, z^t)$ given $p(\theta_t|\, z^{t-1})$ and we can write the likelihood (4.1) recursively as in (4.4). Further details on our implementation of particle filtering are presented in Appendix B.

5 Estimating the Technology of Skill Formation

We estimate the technology on a sample of 1053 white males from the Children of the NLSY/79 (CNLSY/79) sample. Starting in 1986, the children of the NLSY/1979 female respondents have been assessed every two years. The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-competence of the children as well as their home environment. Data were collected via direct assessment and maternal report during home visits at every biannual wave. Appendix C discusses the measurements we use to proxy investment and output.[16] Appendix Table C-1 presents summary statistics of our data. While we have rich data on home inputs the information on schooling inputs is not so rich. Consistent with results reported in Todd and Wolpin (2005), we find that the poorly measured schooling inputs are estimated to have only weak and statistically insignificant effects on outputs. We do not use schooling inputs in our reported specifications.

Our dynamic factor models allow us to exploit the wealth of measures on investment and outcomes available in these data. The dynamic factor models solve several problems. First, there are many proxies for parental investments in children's cognitive and noncognitive development. Applying the dynamic factor model, we let the data tell us the best combination of family input measures to use in predicting the levels and growth in test scores. Measured inputs that are not very informative on family investment decisions will have estimated factor loadings that are close to zero.

Second, the models have the additional advantage that they help us solve the problem of missing data which is pervasive in longitudinal data. Assuming that the data are missing at random, we integrate out the missing items from the sample likelihood. We now present and discuss our empirical results using the CNLSY data.

5.1 Empirical Results

We use a separable measurement system joined with a *CES* technology applied to the proxy data on investment and latent skills defined in Appendix C and at our website to estimate a time-invariant version of the technology of skill formation. The

[16] Additional information can be found on the website http://jenni.uchicago.edu/elast-sub.

normalizations on the measurements we use are presented in our online tables. The specific form of the *CES* technology we use is

$$\theta_{t+1}^{k+1} = B_k \left[\psi_N^k \left(\theta_t^N\right)^{\varphi_k} + \psi_C^k \left(\theta_t^C\right)^{\varphi_k} + \psi_I^k \left(I_t\right)^{\varphi_k} + \psi_{M_C}^k \left(\theta_M^C\right)^{\varphi_k} + \psi_{M_N}^k \left(\theta_M^N\right)^{\varphi_k}\right]^{1/\varphi_k} \eta_t^k, \tag{5.1}$$

where $\sum_{s \in [N,C,I,M_C,M_N]} \psi_s^k = 1$. The latent factors are required to be nonnegative to define the technology. The measurement equations are defined in terms of logs of the components of θ_t: $Z_t = R_t\beta_t + \alpha_t \ln \theta_t + \varepsilon_t$, where R_t are regressors (age of the child in each period and a constant). To form the estimates, we use logs of (5.1), $k = C, N$. We assume that the innovations and the initial conditions are normally distributed: $\ln \eta_t^k \sim N\left(0, \sigma_{\eta_k}^2\right)$ and $\ln \theta_0^k \sim N\left(0, \sigma_{\theta_0^k}^2\right)$, $k = C, N$. The η_t^ℓ are independent over t and are independent of $\eta_t^{\ell'}$ for $\ell' \neq \ell$. The $\left(\theta_0^C, \theta_0^N\right)$ are freely correlated. We normalize the means of all factors to be zero.

We assume that $\varepsilon_t \sim N(0, \sum_t)$, where $\sum_t$ is a diagonal matrix. We impose the condition that ε_t is independent from $\varepsilon_{t'}$ for $t \neq t'$. In our empirical analysis we have two regressors for every measurement equation: a constant and the age of the child at assessment.

We report both anchored results and unanchored results, using the nonlinear version of anchoring described in detail in Appendix D. The anchored results allow us to compare the productivity of investments and stocks of different skills at different stages of the life cycle on the anchored outcome. In this paper, we anchor on high school graduation rates. We first report results in the scale of standardized test scores. We discuss estimates in the scale of the probability of graduating from high school below. We normalize the scale of the investment factor I_t by using "trips to the theater".[17]

Table 1 shows the estimated parameter values and their standard errors in the unanchored system. From this table, we see that: (1) both cognitive and noncognitive skills show strong persistence over time; (2) noncognitive skills affect the accumulation of the next period's cognitive skills and cognitive skills affect the accumulation of the next period's noncognitive skills; (3) the estimated parental investment factor affects noncognitive skills slightly more strongly than cognitive skills, although the differences are not statistically significant; (4) the mother's ability affects both the child's cognitive and noncognitive ability; (5) the mother's noncognitive skills also affect test outcomes.[18]

---

[17] In a sensitivity analysis for a linear version of our model, Cunha and Heckman (2006) show the insensitivity of the estimates of the technology to alternative normalizations of the inputs.

[18] These results differ from those reported by Cunha and Heckman (2006), who impose $\sigma = \alpha = 1$

The elasticities of substitution between investments and stocks of skills are both below 1, with noncognitive investments technologically more substitutable across periods than cognitive investments. This finding is consistent with the evidence on plasticity of noncognitive skills and the lesser plasticity of cognitive skills discussed in Cunha, Heckman, Lochner, and Masterov (2006).

The dynamic factors are estimated to be statistically dependent. Table 2 shows the evolution of the correlation patterns across the dynamic factors. The correlation between cognitive and noncognitive skills grows over the life cycle. The correlation is 0.18 as early as ages 6 and 7, and it grows to around 0.27 at ages 12 and 13. There is growing contemporaneous correlation between noncognitive skill and the home investment. The correlation starts off at 0.26 at ages 6 and 7 and grows to 0.37 by ages 12 and 13. The same pattern is not true for the correlation between cognitive skills and home investments. That correlation starts off near zero at ages 6–7 and grows to 0.21 by ages 12–13. The correlation between the mother's cognitive and noncognitive skills is positive. They are also positively correlated with the lifecycle skills of the child and with home investments, although the pattern does not have an age trend.

5.1.1 Anchoring our Estimates on the Probability of High School Graduation

To circumvent the problem that test score units are intrinsically arbitrary, we anchor outcomes in terms of their effect on high school graduation. The CNLSY data do not provide meaningful earnings histories. Table 3 reports anchored estimates in the probit of high school graduation, as described in Appendix D. Compared to the unanchored case, anchoring increases the estimated elasticity of substitution for both estimated skills, especially for noncognitive skills. Both estimates are still below 1 ($\alpha \cong -0.25$, $\sigma \cong -0.12$). When we use the linear probability model, the estimated elasticities of substitution are slightly larger ($\alpha \cong 0.9$, $\sigma \cong 1.03$).[19] The qualitative conclusions of Table 1 survive. In the anchored case, we can meaningfully compare the effects of parental investments on childhood outcomes. It is interesting to see that parental investments have similar impacts on cognitive and noncognitive skills once

in estimating the model. We find that $\varphi^C = \alpha \cong -0.25$ and $\varphi^N = \sigma \cong -0.12$. They find no role for the mother's ability on child noncognitive skill, whereas we find a strong role. We find that cognitive skills affect the accumulation of next period noncognitive skills. They find stronger correlation patterns among latent skills than we report. Allowing for general forms of substitution affects the estimates.

[19] See our website http://jenni.uchicago.edu/elast-sub/.

we anchor on the probability of graduating from high school.

5.1.2 Estimating the Components of the Home Investment Dynamic Factor

A by-product of our analysis is that we can estimate an implicit "home score" to proxy parental investments. Given the multiplicity of parental inputs in the CNLSY, it is common for empirical analysts to simply add up the components and use the resulting average as a measure of home input. The conventional approach equally weights the ingredients of the home score shown in the leftmost column of Table 4. (These are called "ad hoc" weights in the Table.) From the factor loadings of our separable measurement system, we can estimate the weights on the components over the stages of the child's life cycle. The exact construction of the weights is given at the base of Table 4. Our endogenously determined home score weights the ingredients differently at different stages of the child's life cycle.

Two features of the estimates reported in Table 4 stand out. (1) The weights change with age. "Books in the home" and "child has special lessons" are prominent examples of measures with weights that change with age. (2) The estimated measurement error ("share of residual variance due to uniqueness") is substantial. The substantial measurement error in the components accounts for why *OLS* estimates of a linear technology are downward biased.[20]

5.2 Simulating the Model

To understand the implications of our model, it is helpful to simulate it. We consider two types of simulations. First, we consider the evolution of cognitive and noncognitive skills over the four age ranges used to estimate the model (6–7, 8–9, 10–11, and 12–13). Second, we consider the importance of initial conditions and subsequent parental investments in the school-going years on high school dropout decisions.

Figures 2a and 2b plot the densities of estimated cognitive (2a) and noncognitive (2b) skills for our sample using the estimates for the anchored model. Later cognitive skills stochastically dominate early cognitive skills. The growth in noncognitive skills is much more dramatic. Recalling that our initial condition refers to the preschool (before age 6) years, this evidence is consistent with the evidence reported in Cunha, Heckman, Lochner, and Masterov (2006) that cognitive skills are fairly well established by age 8, but that noncognitive skills are more malleable at later ages.

[20] See Cunha and Heckman (2006).

Table 5 uses estimates of the anchored technology reported in Table 3 to consider the relative importance of initial conditions (endowments at the age of school entry) and parental investments during the schooling years on the probability of high school graduation. We examine a disadvantaged population drawn from the bottom 10th percentile of the population distribution of initial endowments, parental investments, and maternal cognitive and noncognitive skills. We consider the effect of movements in endowments and investments from the bottom 10^{th} percentile to the top 10^{th} percentile of the population distribution. Such movements correspond to enriched family interventions of the sort conferred by the Perry Preschool Program.[21] We predict only a 29.1% high school graduation rate for this group. If we raise preschool endowments or initial conditions to the top 10th percentile, but keep parental investments and maternal attributes fixed at the bottom 10th percentile of the population distribution, we boost high school graduation rates by more than 24 percentage points. Starting at low initial conditions and raising parental investment in children during the ages 12–13 to the top 10th percentile, keeping all other investments and endowments at a low level, increases high school graduation by 13 percentage points. If investment at ages 10–11 and 12–13 is increased to the top 10% level, holding initial conditions and maternal endowments at the bottom 10% level, the same effect is produced on high school graduation as results from the increase in the endowment at school entry. Remediating low initial conditions by high levels of investment during later ages (high I_1, I_2, I_3) produces a graduation rate of 62.7%. Remediating low initial conditions and following this by high parental investments during the school age years raises the high school graduation rate to 85.5%.

These calculations demonstrate the importance of the early years in producing high school graduation and in making later investments productive. They only illustrate technological possibilities. Since we have not estimated preference parameters or ascertained the importance of credit constraints, we cannot conduct a full fledged welfare analysis.

6 Conclusion

This paper formulates and estimates a multistage model of the evolution of child cognitive ad noncognitive skills as determined by parental investments at different stages of the life cycle of children. We estimate the elasticity of substitution between contemporaneous investment and stocks of skills inherited from previous periods to

[21] See Cunha, Heckman, Lochner, and Masterov (2006).

determine the substitutability between early and late investments. We also determine the quantitative importance of early endowments and later investments in determining high school graduation. We account for the proxy nature of the measures of parental inputs and of outputs. We develop a new nonlinear factor analysis and use it to establish nonparametric identification of the skill technology. A by-product of our approach is a framework for the evaluation of childhood interventions that avoids reliance on arbitrarily scaled test scores. Since any monotonic transformation of a test score is also a valid test score, "value added" analyses of test scores have no clear interpretation. We develop a nonparametric approach to this problem by anchoring test scores in adult outcomes with interpretable scales.

Using measures of parental investment and child outcomes from the Children of the National Longitudinal Survey, we estimate key substitution parameters governing the substitutability between early and late investments in cognitive and noncognitive skills. We find greater malleability and substitutability for noncognitive skills than for cognitive skills, consistent with evidence reported in Cunha, Heckman, Lochner, and Masterov (2006). Early endowments determined in the pre-school years play a much more important role in determining cognitive ability than in determining noncognitive ability. We also demonstrate important differences in estimates that arise from the choice of alternative anchors.

Our empirical work imposes restrictions not required to establish identification of the model. A promising line of future work in which we are currently engaged is to estimate more general models. Of special interest is the estimation of technologies that vary by the stage of the child's life cycle. This would allow for critical and sensitive periods for parental investment more general than what is allowed by the technology estimated in this paper.[22] A second empirical extension is to relax the assumption of serially independent measurement error. Our identification analysis allows for serial dependence in measurement errors.

[22] See Knudsen, Heckman, Cameron, and Shonkoff (2006) for a discussion of critical and sensitive periods and the evidence from human and animal species. See also Cunha, Heckman, Lochner, and Masterov (2006).

A Proofs of Theorems

A.1 Proof of Theorem 2

Under the assumptions that $E[\omega_1|\theta,\omega_2]=0$ and that $\omega_2 \perp \theta$ we obtain

$$\begin{aligned}
\frac{E\left[iW_1 e^{i\zeta\cdot W_2}\right]}{E\left[e^{i\zeta\cdot W_2}\right]} &= \frac{E\left[i\left(\theta+\omega_1\right)e^{i\zeta\cdot W_2}\right]}{E\left[e^{i\zeta\cdot W_2}\right]} = \frac{E\left[i\left(\theta+E\left[\omega_1|\theta,\omega_2\right]\right)e^{i\zeta\cdot W_2}\right]}{E\left[e^{i\zeta\cdot W_2}\right]} \\
&= \frac{E\left[i\theta e^{i\zeta\cdot W_2}\right]}{E\left[e^{i\zeta\cdot W_2}\right]} = \frac{E\left[i\theta e^{i\zeta\cdot(\theta+\omega_2)}\right]}{E\left[e^{i\zeta\cdot(\theta+\omega_2)}\right]} = \frac{E\left[i\theta e^{i\zeta\cdot\theta}\right]E\left[e^{i\zeta\cdot\omega_2}\right]}{E\left[e^{i\zeta\cdot\theta}\right]E\left[e^{i\zeta\cdot\omega_2}\right]} = \frac{E\left[i\theta e^{i\zeta\cdot\theta}\right]}{E\left[e^{i\zeta\cdot\theta}\right]} \\
&= \frac{\nabla_\zeta E\left[e^{i\zeta\cdot\theta}\right]}{E\left[e^{i\zeta\cdot\theta}\right]} = \nabla_\zeta \ln\left(E\left[e^{i\zeta\cdot\theta}\right]\right)
\end{aligned}$$

Substituting this expression into Equation (3.10), we obtain:

$$\begin{aligned}
& (2\pi)^{-L}\int e^{-i\chi\cdot\theta}\exp\left(\int_0^\chi \nabla_\zeta \ln\left(E\left[e^{i\zeta\cdot\theta}\right]\right)\cdot d\zeta\right)d\chi \\
=\; & (2\pi)^{-L}\int e^{-i\chi\cdot\theta}\exp\left(\ln\left(E\left[e^{i\chi\cdot\theta}\right]\right)-\ln\left(E\left[e^{i0\cdot\theta}\right]\right)\right)d\chi \\
=\; & (2\pi)^{-L}\int e^{-i\chi\cdot\theta}\exp\left(\ln\left(E\left[e^{i\chi\cdot\theta}\right]\right)\right)d\chi \\
=\; & (2\pi)^{-L}\int e^{-i\chi\cdot\theta}E\left[e^{i\chi\cdot\theta}\right]d\chi,
\end{aligned}$$

where we have used the fact that the path integral of the gradient of a scalar field gives the scalar field itself and that $\ln\left(E\left[e^{i0\cdot\theta}\right]\right)=\ln\left(E\left[1\right]\right)=0$. Note that the integral obtained is invariant to the continuous path that is selected. The last integral is equal to $p_\theta(\theta)$ since the inverse Fourier transform of the characteristic function $E\left[e^{i\zeta\cdot\theta}\right]$ yields the density of θ. ■

A.2 Proof of Theorem 3

Since $m \geq 3$, we can use Theorem 1 in Hu and Schennach (2006) to prove that the distribution of θ is identified, after setting $x=Z_1$, $y=Z_2$, $z=Z_3$, and $x^*=\theta$ in the notation of that paper. Assumption 1 implies that Assumption 1 in Hu and Schennach (2006) holds. All our other assumptions are taken from their paper. Theorem 1 in Hu and Schennach (2006) then implies that the joint density of Z_2 and θ is identified, which provides the density of θ, after integration over Z_2. ■

B Computational Methods

B.1 Sequential Monte Carlo Methods

The text shows how to write the likelihood recursively. We now discuss how to evaluate it. We have to compute integrals of the type:

$$Q = \int p\left(z \mid \theta\right) \pi\left(\theta\right) d\theta. \tag{B.1}$$

To motivate our approach, abstract from the recursive nature of the problem and assume for expositional convenience that it is a static model. If the dimension of θ is large, deterministic integration methods are computationally infeasible. An alternative strategy is to use Monte Carlo integration. First, one generates a random sample $\{\theta\left(l\right)\}_{l=1}^{L}$ from $\pi\left(\theta\right)$ Second, one evaluates $p\left(z \mid \theta\left(l\right)\right)$. Finally, one computes the sample moment:

$$Q_L = \frac{1}{L}\sum_{l=1}^{L} p\left(z \mid \theta\left(l\right)\right).$$

One can show under our assumptions that $\text{plim}_{L\rightarrow\infty} Q_L = Q$. The key is to be able to sample $\theta\left(l\right)$ from $\pi\left(\theta\right)$.

When one is unable to sample directly from $\pi\left(\theta\right)$, one can still perform Monte Carlo integration using an Importance Sampling algorithm (Geweke, 1989). In this case, one samples from a known density $q\left(\theta\right)$, with the property that for every θ such that $\pi\left(\theta\right) > 0$ then $q\left(\theta\right) > 0$. One can rewrite the integral (B.1) as

$$Q = \int p\left(z \mid \theta\right) \frac{\pi\left(\theta\right)}{q\left(\theta\right)} q\left(\theta\right) d\theta.$$

Monte Carlo integration can be implemented by generating a random sample $\{\theta\left(l\right)\}_{l=1}^{L}$ from $q\left(\theta\right)$ and computing the sample moment:

$$Q_L = \frac{1}{L}\sum_{l=1}^{L} p\left(z \mid \theta\left(l\right)\right) w\left(l\right)$$

where

$$w\left(l\right) = \frac{\pi\left(\theta\left(l\right)\right)}{q\left(\theta\left(l\right)\right)}.$$

If we don't know $\pi\left(\theta\right)$, it may be difficult to compute $w\left(l\right)$. We return to this point below.

Nonlinear filtering uses this same basic principle of importance sampling, but in a sequential fashion. The literature terms this method sequential importance sampling (*SIS*)[23]. The basic concepts we use were introduced by Hammersley and Morton

[23] For details on Sequential importance Sampling and Sequential Monte Carlo Methods we refer the reader to Doucet, de Freitas, and Gordon (2001).

(1954). Here we follow the discussion in Doucet, de Freitas, and Gordon (2001). To fix ideas, let $\theta^t = \{\theta_1, \theta_2, \ldots, \theta_t\}, w^t = \{w_1, \ldots, w_t\}$. Let $p\left(\theta^t \middle| z^t\right)$ denote the joint posterior density at period t and let the marginal density of $\theta_j$be represented by $p\left(\theta_j \middle| z^t\right)$. Let $\left\{\theta^t(l), w^t(l)\right\}$ denote the random sample and corresponding weights that characterize the joint posterior $p\left(\theta^t \middle| z^t\right)$. By this we mean that $\left\{\theta^t(l), w^t(l)\right\}$ can approximate $p\left(\theta^t \middle| z^t\right)$ in a discrete fashion:

$$p\left(\theta^t \middle| z^t\right) \approx \sum_{l=1}^{L} w^t(l)\,\delta\left(\theta^t - \theta^t(l)\right)$$

where $\delta(.)$ is the delta function. The weights $w^t(l)$ are chosen using importance sampling:

$$w^t(l) \propto \frac{p\left(\theta^t(l) \middle| z^t\right)}{q\left(\theta^t(l) \middle| z^t\right)}. \tag{B.2}$$

Suppose that at time t we have $\left\{\theta^t(l), w^t(l)\right\}$ which can discretely approximate $p\left(\theta^t \middle| z^t\right)$. Given observation z_{t+1} we wish to approximate $p\left(\theta^{t+1} \middle| z^{t+1}\right)$. If the importance density is chosen to factorize such that:

$$q\left(\theta^{t+1} \middle| z^{t+1}\right) \equiv q\left(\theta_{t+1} \middle| \theta^t, z^{t+1}\right) q\left(\theta^t \middle| z^t\right) \tag{B.3}$$

then one can obtain a random sample $\left\{\theta^{t+1}(l)\right\}$ from $q\left(\theta^{t+1} \middle| z^{t+1}\right)$ by augmenting the existing sample $\left\{\theta^t(l)\right\}$ from $q\left(\theta^t \middle| z^t\right)$ with the new sample $\left\{\theta^{t+1}(l)\right\}$ from $q\left(\theta^{t+1} \middle| \theta^t, z^{t+1}\right)$. Next, we need to augment the existing weights $w^t(l)$ with the new weights $w^{t+1}(i)$, which we derive next. First, note that from Bayes' theorem:

$$\begin{aligned} p\left(\theta^{t+1} \middle| z^{t+1}\right) &= \frac{p\left(z_{t+1} \middle| \theta^{t+1}, z^t\right) p\left(\theta^{t+1} \middle| z^t\right)}{p\left(z_{t+1} \middle| z^t\right)} \\ &= \frac{p\left(z_{t+1} \middle| \theta^{t+1}, z^t\right) p\left(\theta_{t+1} \middle| \theta^t, z^t\right) p\left(\theta^t \middle| z^t\right)}{p\left(z_{t+1} \middle| z^t\right)}. \end{aligned}$$

Using the measurement equation (4.2) and the technology equations (4.3):

$$\frac{p\left(z_{t+1} \middle| \theta^{t+1}, z^t\right) p\left(\theta_{t+1} \middle| \theta^t, z^t\right) p\left(\theta^t \middle| z^t\right)}{p\left(z_{t+1} \middle| z^t\right)} = \frac{p\left(z_{t+1} \middle| \theta_{t+1}\right) p\left(\theta_{t+1} \middle| \theta_t\right) p\left(\theta^t \middle| z^t\right)}{p\left(z_{t+1} \middle| z^t\right)}$$

and note that:

$$p\left(\theta^{t+1} \middle| z^{t+1}\right) \propto p\left(z_{t+1} \middle| \theta_{t+1}\right) p\left(\theta_{t+1} \middle| \theta_t\right) p\left(\theta^t \middle| z^t\right). \tag{B.4}$$

If we substitute (B.3) and (B.4) into (B.2) we obtain:

$$w_{t+1}(i) \propto w_t(l) \frac{p\left(z_{t+1} \middle| \theta_{t+1}(l)\right) p\left(\theta_{t+1}(l) \middle| \theta_t(l)\right)}{q\left(\theta_{t+1}(l) \middle| \theta^t(l), z^{t+1}\right)}$$

and note that once we calculate the weights $w^{t+1}(l)$ we can estimate the updated density:

$$p\left(\theta_{t+1}\middle| Z^{t+1}\right) \approx \sum_{l=1}^{L} w_{t+1}(l)\,\delta\left(\theta_{t+1}-\theta_{t+1}(l)\right).$$

It is possible to show that this estimator is consistent under the conditions assumed in this paper. Sequential Monte Carlo methods (SMC) generate an algorithm that updates weights in sequential importance sampling which can be used to approximate updated densities.

B.2 Implementation of Nonlinear Filtering

One would like to use $p\left(\theta^{t}\middle| Z^{t}\right)$ as the importance density. But because we don't know this density, we use $q\left(\theta^{t+1}\middle| z^{t+1}\right)$ instead. However, this choice is not without its problems. As shown by Doucet, Godsill, and Andrieu (2000), if we use the factorization in (B.3), it turns out that the variance of the importance weights increases over time, this implies that after a certain number of recursive steps, all but one particle (i.e., one realization of the sample $\{\theta_{t+1}(l)\}$) will have negligible weights.

This degeneracy problem can be overcome by a resampling strategy. This step consists of eliminating particles with low importance weights and multiplying particles with high importance weights. That is, resampling is a mapping from the pair $\{\theta_{t+1}(l), w_{t+1}(l)\}$ to a pair $\left\{\theta^{*}_{t+1}(l), w^{*}_{t+1}(l)\right\}$ where $w^{*}_{t+1}(l)=\frac{1}{L}$ for all l. Because of the elimination of the low-probability particles and multiplication of the high-probability particles, the particle filter has also been known as "survival-of-the-fittest" filtering.

There are many versions of the particle filter in the literature of nonlinear filtering. The different filters have in common the sequential importance sampling scheme shown above. The filters differ in at least two dimensions: the calculation of the importance weights and the resampling scheme used.

In our application we use what is known as the Sampling Importance Resampling (SIR) filter, introduced by Gordon, Salmond, and Smith (1993). The implementation of the SIR filter is as follows:

1. Fix a large integer L
2. Draw a sample size of L from $\theta_t(l) \sim p\left(\theta_t\middle| \theta^{*}_{t-1}(l)\right)$.
3. Calculate $\tilde{w}_t(l) = p\left(z_t\middle| \theta_t(l)\right)$
4. Calculate the survival probability: $w_t(l) = \frac{\tilde{w}_t(l)}{\sum_{j=1}^{L}\tilde{w}_t(j)}$
5. Resample with replacement L values of $\theta_t(l)$ with relative weights $w_t(l)$. Denote by $\theta^{*}_t(l)$, $k=1,\ldots,L$, the ones that are resampled. Reset weight $w_t(l)=\frac{1}{L}$.
6. Go back to step 2 and iterate until you reach T.

C Data Appendix

C.1 Survey Measures[24]

The measures of quality of a child's home environment that are included in the CNLSY/79 survey are the components of the Home Observation Measurement of the Environment - Short Form (HOME-SF). They are a subset of the measures used to construct the HOME scale designed by Bradley and Caldwell (1980, 1984) to assess the emotional support and cognitive stimulation children receive through their home environment, planned events and family surroundings. These measurements have been used extensively as inputs to explain child characteristics and behaviors (see e.g. Todd and Wolpin, 2005). As discussed in Linver, Brooks-Gunn, and Cabrera (2004), some of these items are not useful because they do not vary much among families (i.e., more than 90% to 95% of all families make the same response). Web appendix tables 1-8 show the raw correlations of the home score items with a variety of cognitive and noncognitive outcomes at different ages of the child.[25] Our empirical study uses measurements on the following parental investments: the number of books available to the child, a dummy variable indicating whether the child has a musical instrument, a dummy variable indicating whether the family receives a daily newspaper, a dummy variable indicating whether the child receives special lessons, a variable indicating how often the child goes to museums, and a variable indicating how often the child goes to the theater. We also report results from some specifications that use family income as a proxy for parental inputs, but none of our empirical conclusions rely on this particular measure.

As measurements of noncognitive skills we use components of the Behavior Problem Index (BPI), created by Peterson and Zill (1986), and designed to measure the frequency, range, and type of childhood behavior problems for children age four and over, although in our empirical analysis we only use children age six to thirteen. The Behavior Problem score is based on responses from the mothers to 28 questions about specific behaviors that children age four and over may have exhibited in the previous three months. Three response categories are used in the questionnaire: often true, sometimes true, and not true. In our empirical analysis we use the following subscores of the behavioral problems index: (1) antisocial, (2) anxious/depressed, (3) headstrong, (4) hyperactive, (5) peer problems. Among other characteristics, a child who scores low on the antisocial subscore is a child who often cheats or tell lies, is cruel or mean to others, and does not feel sorry for misbehaving. A child who displays a low score on the anxious/depressed measurement is a child who experiences sudden changes in mood, feels no one loves him/her, is fearful, or feels worthless or inferior. A child with low scores on the headstrong measurement is tense, nervous, argues too much, and is disobedient at home, for example. Children will score low on the hyperactivity subscale if they have difficulty concentrating, act without thinking, and are restless or overly active. Finally, a child will be assigned a low score on the peer problem subscore if they have problems getting along with others, are not liked

[24] Additional information can be found on the website http://jenni.uchicago.edu/elast-sub.

[25] See http://jenni.uchicago.edu/elast-sub.

by other children, and are not involved with others.

For measurements of cognitive skills we use the Peabody Individual Achievement Test (PIAT), which is a wide-ranging measure of academic achievement of children aged five and over. It is widely used in developmental research. Todd and Wolpin (2005) use the raw PIAT test score as their measure of cognitive outcomes. The CNLSY/79 includes two subtests from the full PIAT battery: PIAT Mathematics and PIAT Reading Recognition[26]. The PIAT Mathematics measures a child's attainment in mathematics as taught in mainstream education. It consists of 84 multiple-choice items of increasing difficulty. It begins with basic skills such as recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The PIAT Reading Recognition subtest measures word recognition and pronunciation ability. Children read a word silently, then say it aloud. The test contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include the ability to match letters, name names, and read single words aloud.

[26]We do not use the PIAT Reading Comprehension battery since it is not administered to the children who score low in the PIAT Reading Recognition.

D Implementation of the Anchoring

D.1 Nonlinear Anchoring

Let the random variable $D = 1$ if agent i has graduated from high school and zero otherwise. Let $\ln\theta^N_{T+1}$ and $\ln\theta^C_{T+1}$ denote the stocks of noncognitive and cognitive skills of the agent upon completion of childhood. We assume

$$\Pr\left(D = 1 \middle| \mu_D, \ln\theta^N_{T+1}, \ln\theta^C_{T+1}\right) = \Phi\left(\mu_D + \alpha^N_D \ln\theta^N_T + \alpha^C_D \ln\theta^C_T\right),$$

For any $\ln\theta^k$ we can define the anchoring functions g^k for $k = C, N$ as:

$$g^C\left(\ln\theta^C\right) = \frac{\int \Phi\left(\mu_D + \alpha^N_D \ln\theta^N_T + \alpha^C_D \ln\theta^C\right) p\left(\ln\theta^N_T\right) d\ln\theta^N_T}{1 - \int \Phi\left(\alpha^N_D \ln\theta^N_T + \alpha^C_D \ln\theta^C\right) p\left(\ln\theta^N_T\right) d\ln\theta^N_T}$$

and

$$g^N\left(\ln\theta^N\right) = \frac{\int \Phi\left(\mu_D + \alpha^N_D \ln\theta^N + \alpha^C_D \ln\theta^C_T\right) p\left(\ln\theta^C_T\right) d\ln\theta^C_T}{1 - \int \Phi\left(\alpha^N_D \ln\theta^N + \alpha^C_D \ln\theta^C_T\right) p\left(\ln\theta^C_T\right) d\ln\theta^C_T},$$

where μ_D can depend on covariates on which we condition. We seek to estimate the technologies $\tilde{f}^k$ in terms of the anchored factors:

$$\tilde{f}^k(g^N\left(\ln\theta^N_t\right), g^C(\ln\theta^C_t), \ln I_t, \ln\theta^C_M, \ln\theta^N_M, \eta^k_t)$$
$$= g^k\left[f^k(\ln\theta^N_t, \ln\theta^C_t, \ln I_t, \ln\theta^C_M, \ln\theta^N_M, \eta^k_t)\right], \quad k = C, N.$$

So,

$$g(\ln\theta_{t+1}) = \tilde{f}^k(g^N\left(\ln\theta^N_t\right), g^C(\ln\theta^C_t), \ln I_t, \ln\theta^C_M, \ln\theta^N_M, \eta^k_t)) \tag{D.1}$$

and

$$\ln\theta_{t+1} = g^{-1}\left[\tilde{f}^k(g^N\left(\ln\theta^N_t\right), g^C(\ln\theta^C_t), \ln I_t, \ln\theta^C_M, \ln\theta^N_M, \eta^k_t))\right]$$

Let Z_t denote the measurement variables for all period-t factors. Let ε_t denote the measurement error. We can rewrite the estimation system as:

$$Z_t = \alpha_t \ln\theta_t + \varepsilon_t \tag{D.2}$$

$$\ln\theta^k_{t+1} = \left(g^k\right)^{-1}\left\{\tilde{f}^k_t\left[g^C\left(\ln\theta^C_t\right), g^N\left(\ln\theta^N_t\right), \ln I_t, \ln\theta^C_M, \ln\theta^N_M, \eta^k_t\right]\right\} \text{ for } k = C, N \tag{D.3}$$

$$\Pr\left(D = 1 \middle| \mu_D, \ln\theta^N_T, \ln\theta^C_T\right) = \Phi\left(\mu_D + \alpha^N_D \ln\theta^N_T + \alpha^C_D \ln\theta^C_T\right) \tag{D.4}$$

D.2 Linear Anchoring

Suppose that the we model high school graduation according to a linear probability model:

$$\Pr\left(D = 1 \middle| \mu_D, \ln\theta^N_T, \ln\theta^C_T\right) = \mu_D + \alpha^N_D \ln\theta^N_T + \alpha^C_D \ln\theta^C_T$$

We can define linear anchoring functions:

$$g^N\left(\ln\theta^N\right) = \alpha_D^N \ln\theta^N$$

$$g^C\left(\ln\theta^C\right) = \alpha_D^C \ln\theta^C$$

Estimation of the system on this anchored system uses measurement equation (D.2), but replaces (D.3) with:

$$\ln\theta_{t+1}^k = \frac{1}{\alpha_D^k}\left\{\tilde{f}_t^k\left[\alpha_D^C \ln\theta_t^C, \alpha_D^N \ln\theta_t^N, \ln I_t, \ln\theta_M^C, \ln\theta_M^N, \eta_t^k\right]\right\} \text{ for } k = C, N$$

and, instead of (D.4), we estimate:

$$\Pr\left(D_i = 1 \middle| \mu_D, \ln\theta_T^N, \ln\theta_T^C\right) = \mu_D + \alpha_D^N \ln\theta_T^N + \alpha_D^C \ln\theta_T^C.$$

References

Becker, G. S. and N. Tomes (1979, December). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy 87*(6), 1153–1189.

Becker, G. S. and N. Tomes (1986, July). Human capital and the rise and fall of families. *Journal of Labor Economics 4*(3, Part 2), S1–S39.

Bradley, R. H. and B. M. Caldwell (1980, December). The relation of home environment, cognitive competence, and iq among males and females. *Child Development 51*(4), 1140–1148.

Bradley, R. H. and B. M. Caldwell (1984, June). The relation of infants' home environments to achievement test performance in first grade: A follow-up study. *Child Development 55*(3), 803–809.

Cameron, J. (2004). Evidence for an early sensitive period for the development of brain systems underlying social affiliative behavior. Unpublished manuscript, Oregon National Primate Research Center.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422. 2001 Lawrence R. Klein Lecture.

Cawley, J., J. J. Heckman, and E. J. Vytlacil (1999, November). On policies to reward the value added by educators. *Review of Economics and Statistics 81*(4), 720–727.

Cawley, J., J. J. Heckman, and E. J. Vytlacil (2001, September). Three observations on wages and measured cognitive ability. *Labour Economics 8*(4), 419–442.

Center for Human Resource Research (Ed.) (2004). *NLSY79 Child and Young Adult Data User's Guide*. Ohio State University, Columbus, Ohio.

Cunha, F. and J. J. Heckman (2003). The technology of skill formation. Unpublished manuscript, University of Chicago, presented at AEA meetings, January, 2003, San Diego, CA and Federal Reserve Bank of Minneapolis, October, 2004. Revised May, 2005 for presentation at the Society for Economic Dynamics and Control.

Cunha, F. and J. J. Heckman (2006). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Unpublished manuscript, University of Chicago, Department of Economics.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*. Amsterdam: North-Holland. forthcoming.

Currie, J. and D. Blau (2006). Who's minding the kids? preschool, day care, and after school care. In F. Welch and E. Hanushek (Eds.), *Handbook of the Economics of Education*. Amsterdam: North-Holland. forthcoming.

Darolles, S., J.-P. Florens, and E. Renault (2002). Nonparametric instrumental regression. Working Paper 05-2002, Centre interuniversitaire de recherche en économie quantitative, CIREQ.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. New York: Springer-Verlag.

Doucet, A., S. Godsill, and C. Andrieu (2000, July). On sequential monte carlo sampling methods for Bayesian filtering. *Statistics and Computing 10*(3), 197–208.

Geweke, J. (1989, November). Bayesian inference in econometric models using monte carlo integration. *Econometrica 57*(6), 1317–1339.

Gordon, N., D. Salmond, and A. Smith (1993, April). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F 140*(2), 107–113.

Greenough, W. T., J. E. Black, and C. S. Wallace (1987, June). Experience and brain development. *Child Development 58*(3), 539–559.

Hammersley, J. M. and K. W. Morton (1954). Poor man's monte carlo. *Journal of the Royal Statistical Society. Series B (Methodological) 16*(1), 23–38.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3). In press.

Herrnstein, R. J. and C. A. Murray (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hu, Y. and S. M. Schennach (2006). Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions. Working Paper, University of Chicago.

Jöreskog, K. G. and A. S. Goldberger (1975, September). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association 70*(351), 631–639.

Knudsen, E. I., J. J. Heckman, J. Cameron, and J. P. Shonkoff (2006). Building america's future workforce: Economic, neurobiological and behavioral perspectives on investment in human skill development. *Proceedings of the National Academy of Sciences*. Forthcoming.

Levitt, P. (2003, October). Structural and functional maturation of the developing primate brain. *Journal of Pediatrics 143*(4, Supplement), S35–S45.

Linver, M. R., J. Brooks-Gunn, and N. Cabrera (2004, April-September). The home observation for measurement of the environment (HOME) inventory: The derivation of conceptually designed subscales. *Parenting: Science & Practice 4*(2/3), 99–114.

Matzkin, R. L. (2003, September). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Matzkin, R. L. (2007). Nonparametric identification. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6. Amsterdam: Elsevier.

Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience 24*(1), 1161–1192.

Murnane, R. J., J. B. Willett, and F. Levy (1995, May). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*(2), 251–266.

Newey, W. K. and J. L. Powell (2003, September). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Olds, D. L. (2002, September). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science 3*(2), 153–172.

Olley, G. S. and A. Pakes (1996, November). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Peterson, J. L. and N. Zill (1986, May). Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family 48*(2), 295–307.

Schennach, S. M. (2004a, January). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Schennach, S. M. (2004b). Nonparametric estimation in the presence of measurement error. *Econometric Theory 20*, 1046–1093.

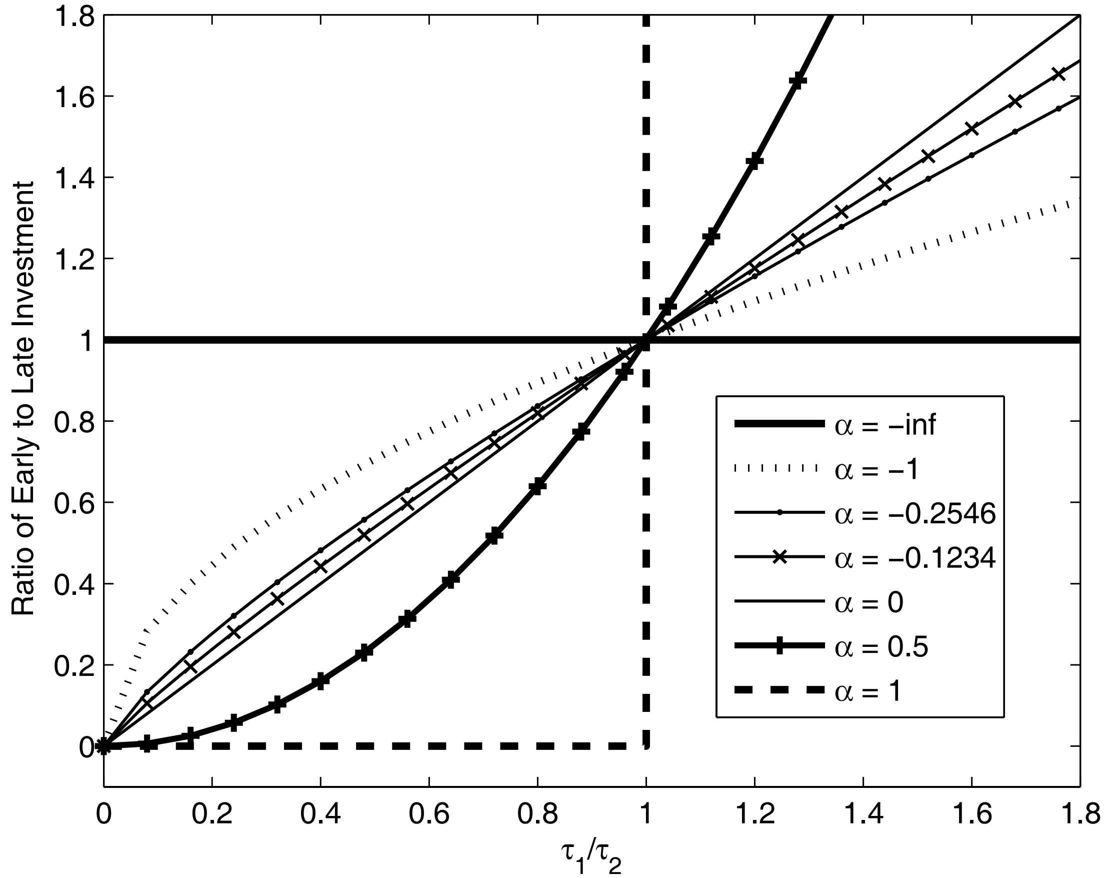
Shumway, R. H. and D. S. Stoffer (1982, May). An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis 3*(3), 253–264.

Todd, P. E. and K. I. Wolpin (2003, February). On the specification and estimation of the production function for cognitive achievement. *Economic Journal 113*(485), F3–33.

Todd, P. E. and K. I. Wolpin (2005). The production of cognitive achievement in children: Home, school and racial test score gaps. Unpublished manuscript. Under revision for publication.

Watson, M. W. and R. F. Engle (1983, December). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics 23*(3), 385–400.

Figure 1
Ratio of Early to Late Investment in Human Capital



Parents maximize the net present value of child wealth investing in period "1", I_1 and "2", I_2, and how much to transfer in risk-free assets. Assuming a CES function, single skill ($\theta_{Ct} = \theta_{Nt}$) and common investments in all skills, the human capital equation reduces to

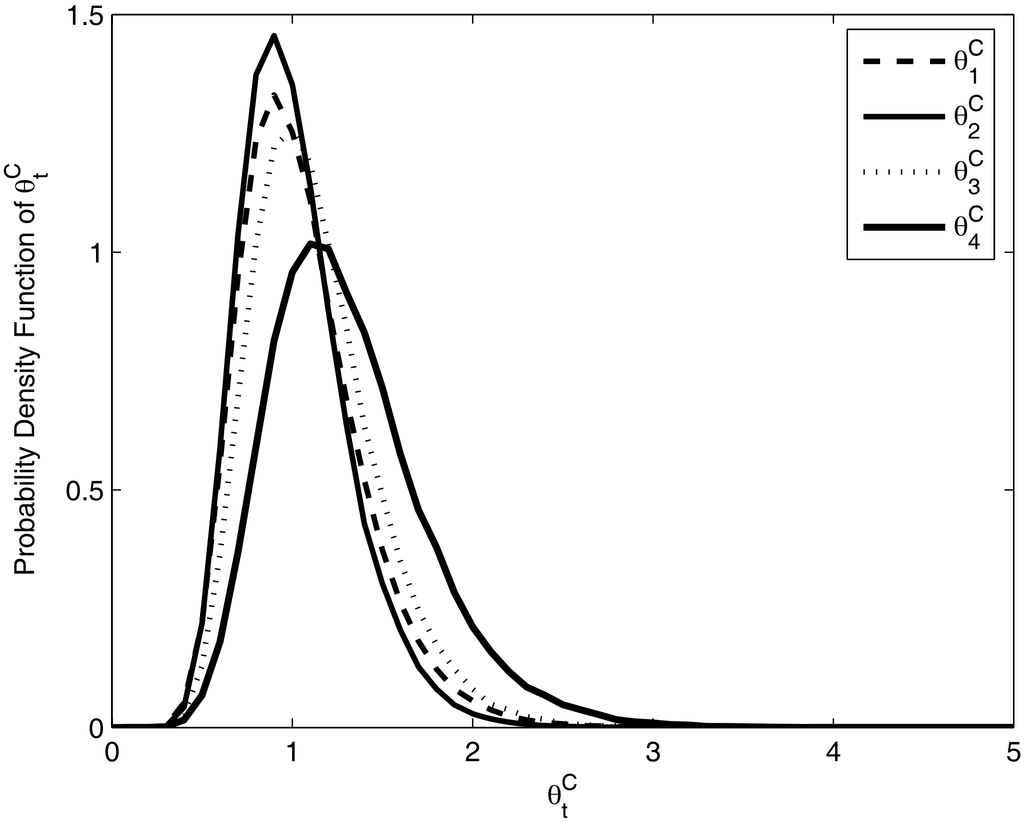
$$h = \left\{\tau_2 \left(I_2\right)^{\alpha} + \tau_1 \left(I_1\right)^{\alpha} + \tau_0 \left(\theta_0\right)^{\alpha}\right\}^{\frac{1}{\alpha}},$$

where τ_2, τ_1 and τ_0 are parameters discussed in Section 2. Assuming that the prices of investments are the same in both periods and the interest rate is r, an interior solution for the ratio of investment is given by

$$\frac{I_1}{I_2} = \left[\frac{\tau_1}{\tau_2\left(1+r\right)}\right]^{\frac{1}{1-\alpha}}.$$

The Euler equation between early and late investments. We assume an interest rate $r = 0$. The elasticity of substitution of the inputs in producing skill θ is obtained by $\frac{1}{1-\alpha}$. The values of $\alpha = -0.2546$ and $\alpha = -0.1234$ are estimated parameters for the elasticity in the case of two skills, cognitive and noncognitive respectively.

Figure 2a
Probability Density Function of Stock of Cognitive Skills by Age



These marginal densities of cognitive skills for four periods were obtained by 100,000 Monte Carlo simulations of the following model with estimates reported in Table 3.

Figure 2b

Probability Density Function of Stock of Noncognitive Skills by Age

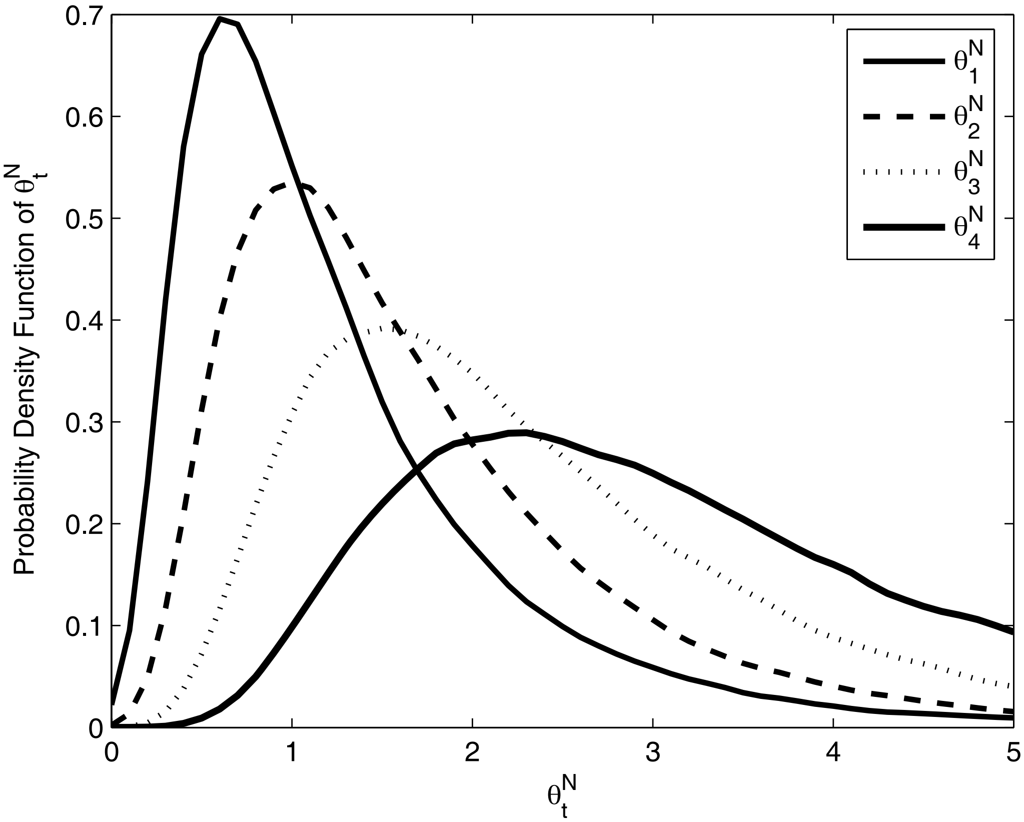


Table 1
The Technology Equations
Unanchored Model

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Constant	0.6932	0.0374	1.0541	0.0834
Current Period Noncognitive Skills	0.7912	0.0297	0.0213	0.0103
Current Period Cognitive Skills	0.0372	0.0178	0.8673	0.0423
Current Period Investments	0.0828	0.0269	0.0599	0.0217
Mother's Cognitive Skills	0.0250	0.0105	0.0314	0.0139
Current Period Noncognitive Skills	0.0639	0.0207	0.0201	0.0102
Parameter of the Elasticity of Substitution	-0.1710	0.0322	-0.8961	0.0763
Variance of Shocks	0.2921	0.0221	0.0585	0.0131

Z_t denotes the measurements on skills and investments at age t. $\theta = \left(\theta_t^N, \theta_t^C, I_t, \theta_M^C, \theta_M^N\right)$ denotes the child's noncognitive ability, child's cognitive ability, parental investment, mother's cognitive ability, and mother's noncognitive ability, respectively. ε_t denotes the measurement error vector. The measurement equations are

$$Z_t = \alpha_t \ln \theta_t + \varepsilon_t,$$

and the technology equations are

$$\ln \theta_{t+1}^k = \ln f^k \left(e^{\ln \theta_t^N}, e^{\ln \theta_t^C}, e^{\ln I_t}, e^{\ln \theta_M^C}, e^{\ln \theta_M^N}\right) + \eta_t^k,$$

where f^k is the technology for skill θ_{t+1}^k and η_t^k are the period omitted inputs for $k = C, N$. We model the functions f^k as CES production functions in the following way:

$$f^k \left(\theta_t^N, \theta_t^C, I_t, \theta_M^C, \theta_M^N\right) = B_k \left[\psi_N^k \left(\theta_t^N\right)^{\phi_k} + \psi_C^k \left(\theta_t^C\right)^{\phi_k} + \psi_I^k \left(I_t\right)^{\phi_k} + \psi_{M_C}^k \left(\theta_M^C\right)^{\phi_k} + \psi_{M_N}^k \left(\theta_M^N\right)^{\phi_k}\right]^{\frac{1}{\phi_k}},$$

subject to

$$\sum_{s \in \{N,C,I,M_C,M_N\}} \psi_s^k = 1.$$

Table 1 shows the estimated parameter values and standard errors of B_k, ψ_s^k for $s \in \{N, C, I, M_C, M_N\}$, and ϕ_k as well as the $Var\left(\eta_{t+1}^k\right)$.

Table 2
Contemporaneous Correlation Matrices
Unanchored Model

Period 1 - Children ages 6 and 7

	Noncognitive	Cognitive	Investment	Mother's Cognitive	Mother's Noncognitive
Noncognitive	1.0000	0.1778	0.2580	0.1260	0.2541
Cognitive	0.1778	1.0000	0.0003	0.2038	0.1255
Investment	0.2580	0.0003	1.0000	0.2794	0.1930
Mother's Cognitive	0.1260	0.2038	0.2794	1.0000	0.1822
Mother's Noncognitive	0.2541	0.1255	0.1930	0.1822	1.0000

Period 2 - Children ages 8 and 9

	Noncognitive	Cognitive	Investment	Mother's Cognitive	Mother's Noncognitive
Noncognitive	1.0000	0.2225	0.2965	0.1319	0.2028
Cognitive	0.2225	1.0000	0.0806	0.2925	0.0936
Investment	0.2965	0.0806	1.0000	0.2814	0.2444
Mother's Cognitive	0.1319	0.2925	0.2814	1.0000	0.1822
Mother's Noncognitive	0.2028	0.0936	0.2444	0.1822	1.0000

Period 3 - Children ages 10 and 11

	Noncognitive	Cognitive	Investment	Mother's Cognitive	Mother's Noncognitive
Noncognitive	1.0000	0.2508	0.3338	0.1475	0.2265
Cognitive	0.2508	1.0000	0.1532	0.3253	0.1389
Investment	0.3338	0.1532	1.0000	0.3406	0.2167
Mother's Cognitive	0.1475	0.3253	0.3406	1.0000	0.1822
Mother's Noncognitive	0.2265	0.1389	0.2167	0.1822	1.0000

Period 4 - Children ages 12 and 13

	Noncognitive	Cognitive	Investment	Mother's Cognitive	Mother's Noncognitive
Noncognitive	1.0000	0.2684	0.3651	0.1101	0.2185
Cognitive	0.2684	1.0000	0.2181	0.3572	0.0906
Investment	0.3651	0.2181	1.0000	0.3160	0.1930
Mother's Cognitive	0.1101	0.3572	0.3160	1.0000	0.1822
Mother's Noncognitive	0.2185	0.0906	0.1930	0.1822	1.0000

Table 3
The Technology Equations
Anchoring on the Probability of Graduating from High School using a Probit Model

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Constant	1.4226	0.0484	0.9842	0.0932
Current Period Noncognitive Skills	0.7403	0.0359	0.0455	0.0133
Current Period Cognitive Skills	0.0516	0.0234	0.7206	0.0581
Current Period Investment	0.1262	0.0302	0.1168	0.0384
Mother's Cognitive Skills	0.0151	0.0178	0.0724	0.0265
Mother's Noncognitive Skills	0.0668	0.0269	0.0446	0.0153
Substitution	-0.1234	0.0419	-0.2543	0.0839
Variance of Shocks	0.2761	0.0628	0.1106	0.0326

$\theta_t = \left(\theta_t^N, \theta_t^C, I_t, \theta_M^C, \theta_M^N\right)$ denotes child's noncognitive ability, child's cognitive ability, parental investment, mother's cognitive ability, and mother's noncognitive ability, respectively. For each skill k, we define an anchor g^k such that

$$\ln A^k = g^k\left(\ln \theta^k\right), \ k = C, N,$$

such that $\ln A^k$ is the factor according to anchor g^k.

To obtain an anchor function g^k we estimate the probability of graduating from high school. The metrics of the factors are, as a result, defined in terms of the probability of graduating from high school. Let the random variable D take the value $d_h = 1$ if person h graduates from high school and $d_h = 0$ otherwise. We model

$$\Pr\left(D = 1 \mid \mu_D, \theta_T^C, \theta_T^N\right) = \Phi\left(\mu_D + \alpha_D^C \ln \theta_T^C + \alpha_D^N \ln \theta_T^N\right).$$

The anchoring functions are:

$$\ln A^N = g^N\left(\ln \theta^N\right) = \frac{\int \Phi\left(\mu_D + \alpha_D^C \ln \theta_T^C + \alpha_D^N \ln \theta^N\right) p\left(\ln \theta_T^C\right) d\ln \theta_T^C}{1 - \int \Phi\left(\mu_D + \alpha_D^C \ln \theta_T^C + \alpha_D^N \ln \theta^N\right) p\left(\ln \theta_T^C\right) d\ln \theta_T^C},$$

$$\ln A^C = g^C\left(\ln \theta^C\right) = \frac{\int \Phi\left(\mu_D + \alpha_D^C \ln \theta_T^C + \alpha_D^N \ln \theta^N\right) p\left(\ln \theta_T^N\right) d\ln \theta_T^N}{1 - \int \Phi\left(\mu_D + \alpha_D^C \ln \theta_T^C + \alpha_D^N \ln \theta^N\right) p\left(\ln \theta_T^N\right) d\ln \theta_T^N}.$$

Z_t denotes the measurements on skills and investments at age t. We estimate the model

$$Z_t = \alpha_t \ln \theta_t + \varepsilon_t,$$

$$\ln \theta_{t+1}^k = g^{-1}\left[\tilde{f}^k\left(g^N\left(\ln \theta_t^N\right), g^C\left(\ln \theta_t^C\right), \ln I_t, \ln \theta_M^C, \ln \theta_M^N\right)\right] + \tilde{\eta}_t^k,$$

and

$$\Pr\left(D = 1 \mid \mu_D, \theta_T^C, \theta_T^N\right) = \Phi\left(\mu_D + \alpha_D^C \ln \theta_T^C + \alpha_D^N \ln \theta_D^N\right).$$

Regressors enter through the μ_D.

Table 4
The Weights in the Construction of the Investment Factor - Anchored Model

	Ages 6 and 7				Ages 8 and 9			
	Estimated Weights[1]	Ad Hoc Weights[2]	Share of Total Residual Variance due to Factors[3]	Share of Total Residual Variance due to Uniqueness[4]	Estimated Weights[1]	Ad Hoc Weights[2]	Share of Total Residual Variance due to Factors[3]	Share of Total Residual Variance due to Uniqueness[4]
Number of Books	0.3177	0.1667	0.1133	0.8867	0.3650	0.1667	0.0694	0.9306
Musical Instrument	0.2011	0.1667	0.1294	0.8706	0.1531	0.1667	0.1255	0.8745
Newspaper	0.1860	0.1667	0.1522	0.8478	0.1859	0.1667	0.0827	0.9173
Child has special lessons	0.1448	0.1667	0.2559	0.7441	0.1579	0.1667	0.1468	0.8532
Child goes to museums	0.0716	0.1667	0.3116	0.6884	0.0701	0.1667	0.2242	0.7758
Child goes to theater	0.0788	0.1667	0.3159	0.6841	0.0680	0.1667	0.2859	0.7141
	Ages 10 and 11				**Ages 12 and 13**			
	Estimated Weights[1]	Ad Hoc Weights[2]	Share of Total Residual Variance due to Factors[3]	Share of Total Residual Variance due to Uniqueness[4]	Estimated Weights[1]	Ad Hoc Weights[2]	Share of Total Residual Variance due to Factors[3]	Share of Total Residual Variance due to Uniqueness[4]
Number of Books	0.1662	0.1667	0.0875	0.9125	0.1399	0.1667	0.0550	0.9450
Musical Instrument	0.1920	0.1667	0.1027	0.8973	0.2078	0.1667	0.0449	0.9551
Newspaper	0.2186	0.1667	0.0749	0.9251	0.2208	0.1667	0.0378	0.9622
Child has special lessons	0.2347	0.1667	0.0844	0.9156	0.2583	0.1667	0.0399	0.9601
Child goes to museums	0.0975	0.1667	0.1485	0.8515	0.0913	0.1667	0.0967	0.9033
Child goes to theater	0.0911	0.1667	0.1842	0.8158	0.0820	0.1667	0.1423	0.8577

[1] Z_t^I denotes investments at age t. ε_t^I denotes the measurement errors associated with the investment measurements. α_t^I denotes the matrix containing the factor loadings associated with the investment factor I. The investment measurement equations are

$$Z_t^I = \alpha_t^I \ln I_t + \varepsilon_t^I.$$

To construct our estimated weights we take the corresponding element in the factor-loading matrix. To fix ideas, let $\alpha_t^I(i,j)$ denote the element in row i and column j of the matrix α_t^I. For measurement $Z_{i,t}^I$ we have $E(\ln I_t) = \frac{1}{\alpha_t^I(i,3)} E\left(Z_{i,t}^I\right)$ for $i = 1, \ldots, T$. We say that $\frac{1}{\alpha_t^I(i,3)}$ is the contribution of measurement $Z_{i,t}^I$ for investment factor I_t. We define the estimated weight $w_{i,t}$ of measurement $Z_{i,t}^I$ in the construction of the investment factor by

$$w_{i,t} = \frac{\frac{1}{\alpha_t^I(i,3)}}{\sum_{j=1}^{m_t^I} \frac{1}{\alpha_t^I(j,3)}}.$$

[2] *Ad hoc* weighting is uniform weighting. If there are m_t^I measures, each measure has weight $\frac{1}{m_t^I}$.

[3] Let $\sigma_{I_t}^2$ denote the variance of the investment factor at period t. For each measurement on parental investment k, the total residual variance is $\sigma_{k,t}^2 = \left(\alpha_{k,t}^I\right)^2 \sigma_{I_t}^2 + \sigma_{\varepsilon_{k,t}^I}^2$, where $\sigma_{\varepsilon_{k,t}^I}^2$ is the variance of the uniqueness $\varepsilon_{k,t}^I$. The share of the total residual variance that is due to the factor is $s_{I_t} = \frac{\left(\alpha_{k,t}^I\right)^2 \sigma_{I_t}^2}{\sigma_{k,t}^2}$.

[4] Analogously, the share of the total residual that is due to the uniqueness is $s_{\varepsilon_{k,t}} = \frac{\sigma_{\varepsilon_{k,t}^I}^2}{\sigma_{k,t}^2}$.

Table 5

Probability of Graduating from High School

As A Function of Endowments At School Entry ("Initial Conditions")

And Parental Investments at 8-9 (Period 1), 10-11 (Period 2), and 12-13 (Period 3)*

	Low Initial Conditions	High Initial Conditions
Low I_1, Low I_2, Low I_3	0.291	0.534
Low I_1, Low I_2, High I_3	0.423	0.685
Low I_1, High I_2, High I_3	0.537	0.789
High I_1, High I_2, High I_3	0.627	0.855

*"Low" refers to bottom 10^{th} percentile of the relevant distribution.

"High" refers to the top 10^{th} percentile of the relevant distribution.

I_1 is investment in period 1 (children are 8-9 years-old), I_2 is investment in period 2 (children are 10-11 years-old), I_3 is investment in period 3 (children are 12-13 years-old).

Mother's cognitive and noncognitive skills are fixed at the bottom 10^{th} percentile level throughout.

Appendix Table C-1: Summary Dynamic Measurements - White Children NLSY/1979

	Ages 6 and 7			Ages 8 and 9			Ages 10 and 11			Ages 12 and 13		
	Obs	Mean	Std Error	Obs	Mean	Std Error	Obs	Mean	Std Error	Obs	Mean	Std Error
Piat Math[1]	753	-1.0376	0.5110	799	0.0423	0.6205	787	0.7851	0.6101	690	1.2451	0.5783
Piat Reading Recognition[1]	751	-1.0654	0.4303	795	-0.0932	0.6543	783	0.6179	0.7334	688	1.1442	0.7852
Antisocial Score[1]	753	0.0732	0.9774	801	-0.0843	1.0641	787	-0.0841	1.0990	717	-0.0658	1.0119
Anxious Score[1]	778	0.1596	1.0016	813	-0.0539	1.0187	813	-0.0753	1.0771	730	-0.0664	1.0561
Headstrong Score[1]	780	0.0192	0.9882	813	-0.2127	1.0000	812	-0.2146	1.0416	729	-0.2123	1.0572
Hyperactive Score[1]	780	-0.0907	0.9673	815	-0.1213	1.0148	813	-0.0983	0.9902	729	-0.0349	0.9910
Conflict Score[1]	779	0.0177	0.9977	815	-0.0057	0.9935	814	-0.0441	1.0304	731	-0.0472	1.0420
Number of Books[2]	629	3.9173	0.3562	821	3.9220	0.3104	676	3.6746	0.6422	730	3.6315	0.6768
Musical Instrument[3]	628	0.4650	0.4992	821	0.4896	0.5002	674	0.5504	0.4978	728	0.5907	0.4921
Newspaper[3]	629	0.5326	0.4993	821	0.5043	0.5003	674	0.4985	0.5004	728	0.5000	0.5003
Child has special lessons[3]	627	0.5470	0.4982	820	0.7049	0.4564	672	0.7247	0.4470	727	0.7717	0.4200
Child goes to museums[4]	628	2.2596	0.9095	821	2.3082	0.8286	672	2.2604	0.8239	729	2.2195	0.8178
Child goes to theater[4]	630	1.8111	0.8312	820	1.8012	0.7532	674	1.8309	0.8000	728	1.8475	0.7920

[1]The variables are standardized with mean zero and variance one across the entire CNLSY/79 sample.

[2]The variable takes the value 1 if the child has no books, 2 if the child has 1 or 2 books, 3 if the child has 3 to 9 books and 4 if the child has 10 or more books.

[3]For example, for musical instrument, the variable takes value 1 if the child has a musical instrument at home and 0 otherwise. Other variables are defined accordingly.

[4]For example, for "museums", the variable takes the value 1 if the child never went to the museum in the last calendar year, 1 if the child went to the museum once or twice in the last calendar year, 3 if the child went to the museum several times in the past calendar year, 4 if the child went to the museum about once a month in the last calendar year, and 5 if the child went to a museum once a week in the last calendar year.

[5]For example, for "Child spends time with father indoors", the variable takes the value 1 if the child never spends time with the father indoors, 2 if the child spends time with the father indoors a few times in a year, 3 if the child spend time with the father indoors about once a month, 4 if the child spends time with the father indoors about once a week, 5 if the child spends time with the father indoors at least four times a week, and 6 if the child spends time with the father once a day or more often.