Formulating, Identifying and 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

First draft, October 2003
This draft, May 29, 2006

*This research was supported by NIH R01-HD32058-3, NSF SES-024158, a grant to the Committee on Economic Development PAES Group from the PEW Foundation, a grant from the Wellspring Foundation, a grant from the "Report to the Nation" of America's Promise and support from the J.B. Pritzker Consortium on Early Childhood Development at the Harris School of Public Policy, University of Chicago. Cunha aknowledges support from the Claudio Haddad dissertation fund at the University of Chicago. The first draft of this paper was presented at a conference at the Minneapolis Federal Reserve, October 2003. We received helpful comments from Robert Pollak at a seminar at Washington University, February 2004, from Susanne Schennach who is the coauthor of successor paper, Petra Todd, and Kenneth Wolpin. A website contains supplementary material is available at http://jenni.uchicago.edu/idest-tech.

Abstract

This paper formulates and estimates models of the evolution of cognitive and noncognitive skills over the life cycle of children and explores the role of family environments in shaping these skills at different stages of the life cycle. Central to this analysis is the identification of the technology of human skill formation. We estimate a dynamic factor model to solve the problem of endogeneity of inputs and multiplicity of inputs relative to instruments. We identify the scale of the factors by estimating their effects on adult outcomes. In this fashion we avoid reliance on test scores and changes in test scores that have no natural metric. Parental investments are more effective in raising noncognitive skills. Noncognitive skills promote the formation of cognitive skills but not vice versa. Parental inputs have different effects at different stages of the child's life cycle with cognitive skills affected more at early ages and noncognitive skills affected more at later ages.

JEL code C31

Keywords: cognitive skills; noncognitive skills; dynamic factor analysis; parental influence

Flavio Cunha	James J. Heckman
Department of Economics	Department of Economics
University of Chicago	University of Chicago
1126 E. 59th Street	1126 E. 59th Street
Chicago, IL 60637	Chicago, IL 60637
flavio@uchicago.edu	jjh@uchicago.edu
phone: 773-256-6216	phone: 773-702-0634
	fax: 773-702-8490

1 Introduction

The importance of cognitive skills in explaining socioeconomic success is now firmly established (see Cawley, Heckman, and Vytlacil, 2001; Herrnstein and Murray, 1994; Murnane, Willett, and Levy, 1995; Neal and Johnson, 1996). An emerging body of empirical research documents the importance of noncognitive skills for wages, schooling and participation in risky behaviors (See Bowles, Gintis, and Osborne, 2001, Heckman and Rubinstein, 2001, and Heckman, Stixrud, and Urzua, 2006). Heckman, Stixrud, and Urzua (2006) demonstrate that cognitive and noncognitive skills are equally important in explaining a variety of aspects of social and economic life in the sense that movements from the bottom to the top of the noncognitive and cognitive distributions have comparable effects on many outcome measures.

There is a substantial body of empirical research on the determinants of cognitive test scores and their growth.[1] There is little work on the determinants of the evolution of noncognitive skills. This paper develops, identifies and estimates models of the technology of skill formation. Building on the theoretical analyses of Cunha and Heckman (2003) and Cunha, Heckman, Lochner, and Masterov (2006), we analyze the joint evolution of cognitive and noncognitive skills over the life cycle of the child.

We model the self productivity of skills as well as their dynamic complementarity. Our technology formalizes the notion that noncognitive skills foster acquisition of cognitive skills by making children more adventuresome and open to learning.[2] It also formalizes the notion that cognitive skills promote the formation of noncognitive skills. With our estimated technology, it is possible to define and measure critical and sensitive periods in the life cycle of child development, and to determine at which ages inputs most affect the evolution of skills.

Child development psychologists have long advocated the importance of understanding the formation of noncognitive skills for interpreting the effects of early childhood intervention programs (see Raver and Zigler, 1997). Heckman, Stixrud, and Urzua (2006) note that the Perry Preschool program did not raise IQ but promoted success among its participants in a variety of aspects of

[1] Todd and Wolpin (2003) survey the vast educational production function literature as well as the child development literature.

[2] Cameron (2004) reports evidence for such effects in her experimental studies of macaque monkeys, and Meaney (2001) reports similar results for rodents. See the evidence in Knudsen, Heckman, Cameron, and Shonkoff (2006).

social and economic life. Our analysis of noncognitive skills, their role in shaping cognitive skills, our investigation of the role of cognitive skills in shaping noncognitive skills, and our determination of the effectiveness of parental inputs on the formation of both types of skill over the life cycle, is a first step toward providing a unified treatment of the early intervention and family influence literatures.

The conventional approach to the estimation of cognitive production functions, best exemplified by the research of Todd and Wolpin (2003, 2005), is to estimate a production function for cognitive test scores relating inputs to outputs. As they emphasize, a central problem with this approach is accounting for the endogeneity of inputs. Another problem is the wealth of candidate parental input measures available in the Children of NLSY (CNLSY) data described below that Todd and Wolpin (2005) analyze, and that we utilize in this paper. The confluence of these two problems–endogeneity and the multiplicity of input measures–places great demands on standard instrumental variable (IV) and fixed effect procedures, such as those used by Todd and Wolpin. It is common in studies of cognitive production functions for analysts to have more inputs than instruments. Indices of inputs are used to circumvent this problem and reduce the parental input data to more manageable dimensions. As emphasized by Todd and Wolpin (2005), fixed effects methods invoke strong assumptions about separability of the technology of skill formation in observables and unobservables, and the way unobservables enter the model.

Our approach to the identification of the technology of skill formation bypasses these problems. We estimate a dynamic factor model that exploits cross equation restrictions (covariance restrictions in linear systems) to secure identification using a version of dynamic state space models (Shumway and Stoffer, 1982; Watson and Engle, 1983). The idea underlying our approach is to model cognitive and noncognitive skills, as well as parental investments as low dimensional latent variables. Building on the analysis of Jöreskog and Goldberger (1975), Jöreskog, Sörbom, and Magidson (1979), Bollen (1989) and Carneiro, Hansen, and Heckman (2003), we 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, we can identify the latent state space dynamics generating the evolution of skills through cross-equation restrictions. When instruments are required, they are internally justified by our model. We economize on the instruments required to secure identification, which are often scarce. We solve the problem of the multiplicity of measures of parental investments by using all of them as proxies for low dimensional latent investments. Instead of creating an arbitrary index of parental inputs, we estimate an index that best predicts latent skill dynamics.

We also address a recurrent problem in the literature on cognitive production functions. Studies in this tradition typically use a test score as a measure of output (see, e.g. Hanushek, 2003). Yet test scores are arbitrarily normalized. Any monotonic transformation of a test score is also a valid test score. Value added—the change in test scores over stages (or grades)—is not invariant to monotonic transformations.

We solve the problem of defining a scale for output by anchoring our test scores using the adult earnings of the child, which have a well defined cardinal scale. Other anchors such as high school graduation, college enrollment and the like could be used. We normalize the latent factors that generate test scores by determining how the factors predict adult outcomes.[3] This approach sets the scale of the test scores and factors in an interpretable metric.

Applying our methodology to CNLSY data we find that: (1) Both cognitive and noncognitive skills change over the life cycle of the child. (2) Parental inputs affect the formation of both noncognitive skills and cognitive skills. Direct measures of mother's ability affect cognitive skills but not noncognitive skills. (3) We find evidence for sensitive periods for parental inputs in the acquisition of cognitive skills and noncognitive skills. The sensitive periods for cognitive skills occur earlier in the life cycle of the child than do sensitive periods for noncognitive skills. In other words, parental inputs appear to affect cognitive skill formation more strongly at earlier ages. They affect noncognitive skill formation more strongly at later ages. This finding is consistent with the evidence presented in Carneiro and Heckman (2003) that noncognitive skills are more malleable at later ages than cognitive skills.

The plan of this paper is as follows. Section 2 presents our model of skill formation. Section 3 presents our analysis of identification using dynamic factor models. Section 4 discusses our empirical findings. Section 5 concludes. We use a technical appendix to develop the details of our sample

[3] See Cawley, Heckman, and Vytlacil (1999) for an analysis that anchors test scores in earnings outcomes. We substantially extend their analysis by allowing for investment at different life cycle stages to affect the evolution of test scores.

likelihood. A website provides supporting material.[4]

2 A Model of Cognitive and Noncognitive Skill Formation

The influential analysis of Becker and Tomes (1979, 1986) assumes only one period of childhood. In Cunha, Heckman, Lochner, and Masterov (2006), we analyze a model where there are two periods of childhood, "1" and "2", followed by adult working life. This paper summarizes and extends the analysis and estimates the technology derived in it. We assume that there are two kinds of skills: θ^C and θ^N. θ^C is cognitive skill and θ^N is noncognitive skill. Becker and Tomes consider one output associated with "human capital" that can be interpreted as a composite of these skills.

Let I_t^k denote parental investments in child skill k in period t, $k = C, N$ and $t = 1, 2$. Let h be the level of human capital as the child starts adulthood which depends on both θ_2^C and θ_2^N. The parents fully control the investment in the child. A richer model which we leave for another occasion would incorporate, among other features, investment decisions of the child as influenced by the parent through the process of preference formation, and through parental incentives for influencing child behavior.

We first describe how skills evolve over time. Assume that 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 let $\theta_t = \left(\theta_t^C, \theta_t^N\right)$ denote the vector of skill or ability stocks. The technology of production of skill k in period t is:

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

for $k = C, N$ and $t = 1, 2$. We assume that f_t^k is twice continuously differentiable, increasing and concave in I_t^k.[5] In this model, stocks of both skills and abilities produce next period skills and influence the productivity of investments. Cognitive skills can promote the formation of noncognitive skills and vice versa because θ_{t-1} is an argument of (1).

[4] See http://jenni.uchicago.edu/idest-tech.

[5] Twice continuous differentiability is only a convenience.

Adult human capital h is a combination of different period 2 skills:

$$h = g\left(\theta_2^C, \theta_2^N\right). \tag{2}$$

The function g is assumed to be continuously differentiable and increasing in $\left(\theta_2^C, \theta_2^N\right)$. This specification of human capital assumes that there is no comparative advantage in the labor market or in other areas of social performance.[6]

Period 1 is a *critical period* for θ^C if investments in θ^C are productive in period 1 but not in period 2:

$$\frac{\partial \theta_2^C}{\partial I_2^C} = \frac{\partial f_2^C\left(\theta_1, I_2^C\right)}{\partial I_2^C} \equiv 0 \text{ for all } \theta_1, I_2^C,$$

but

$$\frac{\partial \theta_1^C}{\partial I_1^C} = \frac{\partial f_1^C\left(\theta_0, I_1^C\right)}{\partial I_1^C} > 0 \text{ for some } \theta_0, I_1^C.$$

Cunha, Heckman, Lochner, and Masterov (2006) survey a literature that establishes that the early periods (before age 8) are critical for intelligence as measured by IQ but not for noncognitive skills.

Period 1 is a *sensitive period* for θ^C if at the same level of inputs, investment is more productive in stage 1 than in stage 2:

$$\left.\frac{\partial \theta_2^C}{\partial I_2^C}\right|_{\theta_1 = s, I_2^C = i} < \left.\frac{\partial \theta_1^C}{\partial I_1^C}\right|_{\theta_0 = s, I_1^C = i}.$$

The evidence summarized in Cunha, Heckman, Lochner, and Masterov (2006) suggests that early investments in both cognitive and noncognitive abilities and skills are more productive than later investments.[7]

Direct complementarity of skill l acquired in period 1 on the output of investment k in producing skill k in period 2 is defined as:

$$\frac{\partial^2 \theta_2^k}{\partial I_2^k \partial \theta_1^l} > 0, \quad l = C, N, \quad k = C, N, \quad l \neq k.$$

[6]Thus we 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.

[7]For a discussion of critical and sensitive periods in animals and humans, see Knudsen, Heckman, Cameron, and Shonkoff (2006).

Early stocks of abilities and skills promote later skill acquisition by making later 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. See Currie and Blau (2006), the survey by Cunha, Heckman, Lochner, and Masterov (2006), and the discussion in Heckman, Stixrud, and Urzua (2006).

Technology (1) is sufficiently rich to capture the evidence on learning in rodents and macaque monkeys documented by Meaney (2001) and Cameron (2004) respectively. See Knudsen, Heckman, Cameron, and Shonkoff (2006) for a review of the literature. Emotionally nurturing early environments producing θ^N create preconditions for later cognitive learning. More emotionally secure young animals explore their environments more actively and learn more quickly. This is an instance of complementarity.

To fix these notions and provide a framework for interpreting the evidence presented in Section 4, consider the following specialization of the model. Suppose initial conditions are the same for everyone and assume that first period skills are just due to first period investment. This assumption is made for analytical convenience and is relaxed in our empirical work reported below. Thus

$$\theta_1^C = f_1^C\left(\theta_0, I_1^C\right) = I_1^C$$

and

$$\theta_1^N = f_1^C\left(\theta_0, I_1^C\right) = I_1^N,$$

where we assume that I_1^C and I_1^N are scalars. Assume a CES model for the second period technologies producing cognitive and noncognitive skill:

$$\begin{aligned} \theta_2^C &= f_2^C\left(\theta_1, I_2^C\right) \\ &= \left\{\gamma_1\left(\theta_1^C\right)^\alpha + \gamma_2\left(\theta_1^N\right)^\alpha + (1-\gamma_1-\gamma_2)\left(I_2^C\right)^\alpha\right\}^{\frac{1}{\alpha}}, \quad \text{where} \quad \begin{array}{l} 1 \geq \gamma_1 \geq 0, \\ 1 \geq \gamma_2 \geq 0, \\ 1 \geq 1-\gamma_1-\gamma_2 \geq 0, \end{array} \end{aligned} \tag{3}$$

and

$$
\begin{aligned}
\theta_2^N &= f_2^N\left(\theta_1, I_2^N\right) \\
&= \left\{\eta_1\left(\theta_1^C\right)^\sigma + \eta_2\left(\theta_1^N\right)^\sigma + (1-\eta_1-\eta_2)\left(I_2^N\right)^\sigma\right\}^{\frac{1}{\sigma}}, \quad \text{where } \begin{array}{l} 1 \geq \eta_1 \geq 0, \\ 1 \geq \eta_2 \geq 0, \\ 1 \geq 1-\eta_1-\eta_2 \geq 0, \end{array}
\end{aligned} \tag{4}
$$

where $\frac{1}{1-\alpha}$ is the elasticity of substitution in the inputs producing θ_2^C and $\frac{1}{1-\sigma}$ is the elasticity of substitution of inputs in producing θ_2^N where $\alpha \in (-\infty, 1]$ and $\sigma \in (-\infty, 1]$.

The CES technology is well known and has convenient properties. Notice that I_2^N and I_2^C are direct complements with (θ_1^C, θ_1^N) irrespective of the substitution parameters α and σ, except in limiting cases. Focusing on the technology for producing θ_2^C, when $\alpha = 1$, the inputs are perfect substitutes in the intuitive use of that term (the elasticity of substitution is infinite). The inputs θ_1^C, θ_1^N and I_2^C can be ordered by their relative productivity in producing θ_2^C. The higher γ_1 and γ_2, the higher the productivity of θ_1^C and θ_1^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 $\theta_2^C = \min\left\{\theta_1^C, \theta_1^N, I_2^C\right\}$. In this technology, early investments are a *bottleneck* for later investment. Compensation for adverse early environments through late investments is impossible.

The evidence from numerous studies surveyed in Cunha, Heckman, Lochner, and Masterov (2006) shows that IQ is no longer malleable after age eight. Taken at face value, this implies that if θ_1^C is IQ, and period 1 stops at age eight, for all values of I_2^C, $\theta_2^C = \theta_1^C$. Period 1 is a critical period for IQ but not necessarily for other skills and abilities. More generally, period 1 is a critical period if

$$\frac{\partial \theta_t^C}{\partial I_t^C} = 0 \text{ for } t > 1.$$

For technology (3), this condition is satisfied if $\gamma_1 + \gamma_2 = 1$.

The evidence surveyed in Cunha, Heckman, Lochner, and Masterov (2006) shows substantial positive results for adolescent interventions in producing noncognitive skills (θ_2^N) and at best modest gains for cognitive skills $\left(\theta_2^C\right)$ from such interventions. Technologies (3) and (4) rationalize this pattern. Since the populations targeted by adolescent intervention studies tend to come from families with poor backgrounds, we would expect I_1^C and I_1^N for such families to be below average. Thus, θ_1^C and θ_1^N will be below average. Interventions make I_2^C and I_2^N relatively large for the treatment group in comparison to the control group in the adolescent intervention experiments. At stage 2, θ_2^C (cognitive ability) is essentially the same in the control and treatment groups, while θ_2^N (noncognitive ability) is higher for the treated group. Large values of $(\gamma_1 + \gamma_2)$ which produce small coefficients on I_2^C, or small values of $(\eta_1 + \eta_2)$ (so the coefficient on I_2^N is large) and high values of α and σ can produce this pattern. Another case that rationalizes their evidence is the parameter configuration $\alpha \to -\infty$ and $\sigma = 1$. Under these conditions:

$$\theta_2^C = \min\left\{\theta_1^C, \theta_1^N, I_2^C\right\}, \tag{5}$$

while

$$\theta_2^N = \eta_1 \theta_1^C + \eta_2 \theta_1^N + (1 - \eta_1 - \eta_2) I_2^N. \tag{6}$$

In this case, the attainable period 2 stock of cognitive skill $\left(\theta_2^C\right)$ is limited by the minimum value of $\theta_1^C, \theta_1^N, I_2^C$. Any level of investment in period 2 such that $I_2^C > \min\left\{\theta_1^C, \theta_1^N\right\}$ is ineffective in incrementing the stock of cognitive skills. Period 1 is a bottleneck period. Unless sufficient skill investments are made in θ^C in period 1, it is not possible to raise skill θ^C in period 2. This phenomenon does not appear in the production of the noncognitive skill, provided that $(1 - \eta_1 - \eta_2) > 0$. More generally, the higher σ and the larger $(1 - \eta_1 - \eta_2)$, the more productive is investment I_2^N in producing θ_2^N.

To complete the CES example, assume that adult human capital h is a CES function of the two skills accumulated at stage two:

$$h = \left\{\tau \left(\theta_2^C\right)^{\phi} + (1 - \tau) \left(\theta_2^N\right)^{\phi}\right\}^{\frac{\rho}{\phi}} = g\left(I_1, I_2\right), \tag{7}$$

where $\rho \in (0, 1)$, $\tau \in [0, 1]$, and $\phi \in (-\infty, 1]$. In this parameterization, $\frac{1}{1-\phi}$ is the elasticity of substitution across different skills in the production of adult human capital. Equation (7) reminds us that the market, or life in general, requires use of multiple skills. Heckman, Stixrud, and

Urzua (2006) show the importance of both cognitive and noncognitive skills in producing market effectiveness and estimate the trade-off among the skills. In general, different tasks require cognitive and noncognitive skills in different proportions. One way to remedy early skill deficits is to make compensatory investments. Another way is to motivate people from disadvantaged environments to pursue tasks that do not require the skill that deprived early environments do not produce. A richer theory would account for this choice of tasks and its implications for remediation. For the sake of simplifying the argument, we work with equation (7) that captures the notion that skills can trade off against each other in producing effective people. Highly motivated, but not very bright, people may be just as effective as bright but unmotivated people. That is one of the lessons from the GED program (See Heckman and LaFontaine, 2006).

Our analysis is simplified by assuming that investments are general or public goods investments: $I_1^C = I_1^N = I_1$, $I_2^C = I_2^N = I_2$.[8] Cunha and Heckman (2003) develop the more general case of skill-specific investments which requires substantially more notational complexity.

With common investment goods, we can solve out for θ_2^C and θ_2^N in terms of I_1 to simplify (3) and (4) to reach

$$\theta_2^C = \left\{(\gamma_1 + \gamma_2)(I_1)^{\alpha} + (1 - \gamma_1 - \gamma_2)(I_2)^{\alpha}\right\}^{\frac{1}{\alpha}} \tag{8}$$

and

$$\theta_2^N = \left\{(\eta_1 + \eta_2)(I_1)^{\sigma} + (1 - \eta_1 - \eta_2)(I_2)^{\sigma}\right\}^{\frac{1}{\sigma}}. \tag{9}$$

If we substitute these expressions into the production function for adult human capital (7), we obtain

$$h = \left\{\tau\left[\tilde{\gamma}(I_1)^{\alpha} + (1 - \tilde{\gamma})(I_2)^{\alpha}\right]^{\frac{\phi}{\alpha}} + (1 - \tau)\left[\tilde{\phi}(I_1)^{\sigma} + \left(1 - \tilde{\phi}\right)(I_2)^{\sigma}\right]^{\frac{\phi}{\sigma}}\right\}^{\frac{\rho}{\phi}}, \tag{10}$$

where $\tilde{\gamma} = \gamma_1 + \gamma_2$, $\tilde{\phi} = \eta_1 + \eta_2$. Equation (10) expresses adult human capital as a function of the entire sequence of childhood investments in human capital. Current investments in human capital are combined with the past stock of skills in order to increment the stock of current skills.

A conveniently simple formulation of the problem arises if we assume that $\alpha = \sigma = \phi$ so that

[8] Thus when a parent reads to the child in the first period of childhood, such reading may be an investment in all kinds of skills. It is an investment in cognitive skills, as it helps the child get exposure to language and new words. It can also be an investment in noncognitive skills, if reading nurtures the self confidence and persistence of the child.

CES substitution among inputs in producing outputs and CES substitution among skill in producing human capital are the same. This produces the familiar CES expression for adult human capital stocks:

$$h = \left\{ \gamma I_1^{\phi} + (1-\gamma) I_2^{\phi} \right\}^{\frac{\rho}{\phi}}, \tag{11}$$

where $\gamma = \tau\tilde{\gamma} + (1-\tau)\tilde{\phi}$ and $\phi = \alpha = \sigma$. The parameter γ is *a skill multiplier*. It arises because I_1 affects the accumulation of θ_1^C and θ_1^N. These stocks of skills in turn affect the productivity of I_2 in forming θ_2^C and θ_2^N. Thus γ captures the net effect of I_1 on h through both self-productivity and direct complementarity.[9] The parameter γ also illustrates the concepts of sensitive and critical periods. For example, if $\gamma = 1$, then period one is critical for h. On the other hand, if $\gamma = 0$ period two is critical for h. Whenever $\frac{1}{2} < \gamma < 1$, period one is sensitive for h, but not critical. Finally, if $0 < \gamma \leq \frac{1}{2}$, then period two is sensitive for h.

The quantity $\frac{1}{1-\phi}$ is a measure of how easy it is to substitute between I_1 and I_2 where the substitution arises from both the task performance (human capital) function in equation (7) and the technology of skill formation. Within the CES technology, ϕ is a measure of the ease of substitution of inputs. In this analytically convenient case, the parameter ϕ plays a dual role. First, it informs us how easily one can substitute across different skills in order to produce one unit of adult human capital h. Second, it also represents the degree of complementarity (or substitutability) between early and late investments in producing skills. In this second role, the parameter ϕ dictates how easy it is to compensate for low levels of stage 1 skills in producing late skills.[10]

When ϕ is small, low early investments I_1 are not easily remediated by late investments I_2 in producing human capital. The other face of CES complementarity is that when ϕ 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 $\phi \to -\infty$, (11) converges to $h = (\min\{I_1, I_2\})^{\rho}$. The Leontief case contrasts with the case of perfect CES substitutes, which arises when $\phi = 1$: $h =$

[9] Direct complementarity between I_1 and I_2 is $\frac{\partial^2 h}{\partial I_1 \partial I_2} > 0$. As long as $\rho > \phi$, I_1 and I_2 are direct complements, because $\text{sign}\left(\frac{\partial^2 h}{\partial I_1 \partial I_2}\right) = \text{sign}\,(\rho - \phi)$. This definition of complementarity is to be distinguished from the notion based on the elasticity of substitution between I_1 and I_2, which is $\frac{1}{1-\phi}$. When $\phi < 0$, I_1 and I_2 are sometimes described as complements. When $\phi > 0$, I_1 and I_2 are sometimes described as substitutes. When $\rho = 1$, I_1 and I_2 are always direct complements, but if $1 > \phi > 0$, they are CES substitutes.

[10] In principle, compensation can come through two channels: (i) through skill investment or (ii) through choice of market activities, substituting deficits in one skill by the relative abundance in the other through choice of tasks.

$[\gamma I_1 + (1-\gamma) I_2]^\rho$. When we impose the further restriction that $\gamma = \frac{1}{2}$, we generate the model that is implicitly assumed in the literature of human capital investments that collapse childhood into a single period. In this special case, the output of adult human capital is determined by the total amount of human capital investments, regardless of how investment is distributed across childhood periods. In the case of perfect CES 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.

To analyze the optimal timing of investment, it is convenient to work with the technology embodied in (11). We now show how the ratio of early to late investments varies as a function of ϕ, γ, and ρ. Consider the following model in which parents maximize the present value of net wealth of their children.[11] In order to do that, 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, b, given total parental resources M. Period "1" could include *in utero* investments. Parents cannot extract resources from children, so $b \geq 0$. From period "3" to period T, the age of retirement from the workforce, persons are assumed to work full time. Let r denote the time-invariant interest rate, set exogenously and assumed to be constant for all periods, and let q denote the present value of future earnings per efficiency unit of human capital $\{w_t\}_{t=3}^{T}$:

$$q = \sum_{t=3}^{T} \left(\frac{1}{1+r} \right)^{t-3} w_t.$$[12]

Lifetime earnings of children when they start working at period "3" are given by $qg(I_1, I_2)$, where g is the function determining the adult stock of human capital. Discounted to period 1, the present value of lifetime earnings is $\frac{q}{(1+r)^2} g(I_1, I_2)$. The problem of the parents is to maximize the present value of the child's net wealth:

$$\max_{I_1, I_2, b} \left\{ \frac{1}{(1+r)^2} \left[qg(I_1, I_2) + b \right] \right\},$$

[11] This setup is overly simplistic but allows us to focus on the important points. See Cunha (2004) and Cunha and Heckman (2003) for more general models.

[12] We abstract from endogenously determined on-the-job training, learning-by-doing, and assume that agents supply labor inelastically.

subject to the standard budget constraint

$$I_1 + \frac{1}{1+r} I_2 + \frac{1}{(1+r)^2} b = M, \tag{12}$$

and the constraint that parents cannot leave negative bequests to their children

$$b \geq 0, \tag{13}$$

where $g(I_1, I_2)$ is defined in equation (11) and is concave in I_1 and I_2.

When $\phi = 1$, early and late investments are perfect CES substitutes. This is a CES version of Becker and Tomes (1979, 1986). In this case, the optimal investment strategy is straightforward. The price of early investment is \$1. The price of the late investment is $\$\frac{1}{(1+r)}$. Thus the parents can purchase $(1+r)$ units of I_2 for every unit of I_1 foregone. The amount of human capital produced from one unit of I_1 is γ, while $\$(1+r)$ of I_2 produces $(1+r)(1-\gamma)$ units of human capital. Therefore, the parent invests early if $\gamma > (1-\gamma)(1+r)$ and late otherwise. Two forces act in opposite directions. High productivity of initial investment (the skill multiplier) drives the agent toward making early investments. Intertemporal prices (the interest rate) drive the agent to invest late. It is optimal to invest early if $\gamma > \frac{1+r}{2+r}$.

As $\phi \to -\infty$, the CES production function converges to the Leontief case and the optimal investment strategy is to set $I_1 = I_2$. CES complementarity dominates and the profile of investments is such that $\frac{I_1}{I_2}$ converges to one. In this extreme case, CES complementarity has a dual face. Investments in the young are essential. At the same time, later investments are needed to harvest early investments. On efficiency grounds, early disadvantages should be perpetuated, and compensatory investments at later ages are economically inefficient.

For $-\infty < \phi < 1$, the first-order conditions are necessary and sufficient given concavity of the technology in terms of I_1 and I_2. Let μ, λ denote the Lagrange multipliers associated with constraints (12) and (13), respectively. The first-order conditions for I_1, I_2, and b are

$$\frac{q}{(1+r)^2} \rho\gamma \left\{ \gamma I_1^{\phi} + (1-\gamma) I_2^{\phi} \right\}^{\frac{\rho-\phi}{\phi}} I_1^{\phi-1} = \mu, \tag{14}$$

$$\frac{q}{(1+r)}\rho(1-\gamma)\left\{\gamma I_1^{\phi}+(1-\gamma)I_2^{\phi}\right\}^{\frac{\rho-\phi}{\phi}}I_2^{\phi-1}=\mu, \tag{15}$$

$$\mu-1=\lambda(1+r)^2. \tag{16}$$

Notice that if restriction (13) is not binding, then $\lambda = 0$, $\mu = 1$ and optimal early and late investments are only functions of (q, r). In this case, unconstrained families that make bequests will all invest the same in their children. The only difference is in the transfers of assets to their children. If $M_A > M_B$ then $b_A > b_B$.

For an interior solution, if we take the ratio of (14) to (15) and rearrange terms we obtain

$$\frac{I_1}{I_2}=\left[\frac{\gamma}{(1-\gamma)(1+r)}\right]^{\frac{1}{1-\phi}}. \tag{17}$$

Figure 1 plots the ratio of early to late investments as a function of the skill multiplier γ, under different values of the complementarity parameter ϕ. When $\phi \to -\infty$, we obtain the Leontief technology and there is high CES-complementarity between early and late investments. In this case, the ratio is not sensitive to variations in γ. CES-complementarity dominates, and the optimal investment profile distributes investments equally across different periods. When $\phi = 0$, technology g is given by the Cobb-Douglas function:

$$h=(I_1)^{\rho\gamma}(I_2)^{\rho(1-\gamma)}.$$

In this case, from equation (17), $\frac{I_1}{I_2}$ is close to zero for low values of γ, but explodes to infinity as γ approaches one.

Taking logs of (17) yields the expression

$$\log\left(\frac{I_1}{I_2}\right)=\left(\frac{1}{1-\phi}\right)\log\left(\frac{\gamma}{1-\gamma}\right)-\left(\frac{1}{1-\phi}\right)\log(1+r). \tag{18}$$

This expression does not depend on ρ. In the special case $\gamma = \frac{1+r}{2+r}$, investment will be the same in both periods regardless of the value assumed by ϕ. More generally, the ratio of early to late investments varies with the complementarity between early and late investments, ϕ, with the skill multiplier for human capital, γ, and with the interest rate. *Ceteris paribus*, if $\frac{\gamma}{(1-\gamma)(1+r)} > 1$, the greater the CES complementarity, (i.e., the lower ϕ), 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 the skill multiplier, γ, 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. Finally, 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. The higher the interest rate today, the cheaper it is to postpone investments. Expression (18) is the basis for the internal instruments we develop in section 3.

We summarize the lessons from this analysis in Table 1. When CES complementarity is high, the skill multiplier γ plays a limited role in shaping the ratio of early to late investments. High early investments should be followed by high late investments. As the degree of CES complementarity decreases, the role of the skill multiplier increases, and the higher the multiplier, the more investments should be concentrated in early ages.

This simple model also has implications for the optimal timing of interventions. If $M_A > M_B$ and family A is unconstrained while family B is constrained, then for family B, $\lambda_B > 0$, $\mu_B = \left[1 + \lambda_B (1+r)^2\right]$. Consequently, in equilibrium, the marginal return to one dollar invested in the poor child from family B is above the marginal return to the same dollar invested in the rich child from family A, so family B underinvests compared to the less constrained family A.

There is no trade-off between equity and efficiency in *early* childhood investments. Government policies to promote early accumulation of human capital should be targeted to the children of poor families. However, the optimal second period intervention for a child from a disadvantaged environment depends critically on the nature of human capital aggregation function (11) and the technology of skill production. If I_1 and I_2 are perfect CES complements, then a low level of I_1 cannot be compensated at any level of investment by a high I_2.

On the other hand, suppose that $\phi = 1$, so inputs are perfect substitutes. The technology in this case is:

$$h = \left[\gamma I_1 + (1-\gamma) I_2\right]^{\rho}, \ 0 \leq \gamma \leq 1. \tag{19}$$

In this case, a second-period intervention can, in principle, eliminate initial skill deficits (low values of I_1). At a sufficiently high level of second-period investment, it is technically possible to offset low first period investments. However, it may not be cost effective to do so. For example, if $\rho = 1$ and $q(1-\gamma) < 1+r$, then the gains from future earnings do not justify the costs of investment. It would be more efficient to give the child a bond that earns interest rather than to invest in human capital in order to put the child at a certain level of income. Carneiro and Heckman (2003) show that at current levels of spending, reduced classroom size interventions in U.S. public schools have this feature and are economically inefficient.

We previously discussed the concepts of critical and sensitive periods in terms of the technical possibilities of remediation. These concepts were defined in terms of the technology of skill formation. Here, we consider the net effects of interventions operating through investment and market substitution. The higher ϕ, the greater are the possibilities for alleviating early disadvantage. When $\phi = 1$, as in the example just discussed, it is always technically possible to remediate early disadvantage. But it may not be economically efficient to do so. From an economic point of view, critical and sensitive periods should be defined in terms of the costs and returns of remediation, and not solely in terms of technical possibilities. We now turn to identifying and estimating key aspects of the technology.

3 Identifying the Technology using Dynamic Factor Models

Identifying and estimating technology (1) is a challenging task. Both the inputs and outputs can only be proxied, producing the problem of measurement error. In addition, the inputs are likely to be endogenous because parents choose them in the light of their preferences and early outcomes of children. General nonlinear specifications of technology (1) raise additional problems regarding measurement error in latent variables for nonlinear systems. (See Schennach, 2004).

In this paper we estimate log linear specifications of (1) and therefore assume Cobb Douglas technologies. Log linear specifications are a traditional starting point. A more general nonlinear analysis would entail additional econometric and computational complications although it would identify key substitution (α, σ) parameters. We leave that task for the future (See Cunha, Heckman, and Schennach, 2006).[13]

3.1 Identifying A Log Linear Technology

Using a log linear specification we can identify critical and sensitive periods for inputs. We can also identify cross effects, as well as self productivity of the stocks of skills. If we find little evidence of self productivity, sensitive or critical periods, or little evidence of cross effects in a simpler setting, it is unlikely that a more general nonlinear model will overturn these results. Identifying the log linear technology raises many challenges that we address in this paper.

There is a large body of research that estimates the determinants of the evolution of cognitive skills. Todd and Wolpin (2003) present a valuable survey of this literature that is a guide for interpretation and estimation. To our knowledge, there is no previous research on estimating the evolution of noncognitive skills.

The empirical analysis reported in Todd and Wolpin (2005) represents the state of the art in modeling the determinants of the evolution of cognitive skills.[14] In their paper, they use a scalar measure of cognitive ability (θ_{t+1}^{C}) in period $t+1$ that depends on period t cognitive ability (θ_{t}^{C}) and investment (I_t). They assume a linear-in-parameters technology

$$\theta_{t+t}^{C} = a_t \theta_t^{C} + b_t I_t + \eta_t \tag{20}$$

where η_t represents unobserved inputs, measurement error, or both. They allow inputs to have different effects at different stages of the child's life cycle. They use the components of the "home score" measure as measures of parental investment.[15] We use a version of the inputs into the home score as well, but in a different way than they do.

Using logs of latent skills and investments, one can interpret (20) as a Cobb Douglas repre-

[13] We could have equally well adopted a linear specification assuming perfect substitution among inputs. As an approximation, this would also be valid. The results reported in this paper are silent about the magnitudes of the key substitution parameters (α, σ), introduced in the last section, but allow us to approximate other features of the technology and distinguish the effects early versus late investment.

[14] Todd and Wolpin (2005) discuss a paper by Fryer and Levitt (2004) that uses inappropriate static methods to estimate a dynamic model of investment. Fryer and Levitt assume that parental inputs do not cumulate. Alternatively, they assume 100% depreciation of investment in each period. They also do not account for endogeneity of inputs.

[15] This measure originates in the work of Bradley and Caldwell (1980, 1984) and is discussed further in section 4.

sentation of technology (1). The Cobb Douglas case is intermediate between the case of perfect substitutes and perfect complements. It sets $\alpha = \sigma = 0$. The effects of investment and lagged skill can depend on the stage of the investment (a and b have t subscripts) so it is possible to identify critical and sensitive periods. Todd and Wolpin (2003, 2005) provide an extensive discussion of problems arising from endogenous inputs (θ_t^C, I_t dependent on unobservable η_t). In their 2005 paper, they use instrumental variable (IV) methods coupled with fixed effect methods (see Hsiao, 1986, Baltagi, 1995, and Arellano, 2003, for descriptions of these procedures). Reliance on IV is problematic because of the ever present controversy about the validity of exclusion restrictions.[16] As stressed by Todd and Wolpin, fixed effect methods require very special assumptions about the nature of the unobservables, their persistence over time and the structure of agent decision rules. The CNLSY data that Todd and Wolpin (2005) (and we) use has a multiplicity of investment measures subsumed in a "home score" measure which, as they discuss, combines many diverse parental input measures into a score that weights all components equally.[17]

Todd and Wolpin (2005) and the large literature they cite use a cognitive test score as a measure of output. This imparts a certain arbitrariness to the analysis. Test scores are arbitrarily normed. Any monotonic function of the test score is a perfectly good alternative test score. A test score is only a relative rank. While Todd and Wolpin use raw scores and others use ranks (see, e.g. Carneiro and Heckman, 2003; Carneiro, Heckman, and Masterov, 2005), none of these measures is intrinsically satisfactory because there is no meaningful cardinal scale for test scores. We address this problem by using adult outcomes to anchor the scale of the test score.

In Cunha, Heckman, and Schennach (2006), we address this problem in a general way for arbitrary monotonic transformations of the factors. In this paper, we develop an interpretable scale for θ that is robust to all affine transformations.[18] We assume that investment I_t is measured in the scale of log dollars. We report results for alternative normalizations of the units of investment.

[16] Fixed effect methods do not easily generalize to the nonlinear frameworks that are suggested by our analysis of the technology of skill formation presented in Section 2 but that concern is not relevant to this paper. See however the analysis of Altonji and Matzkin (2005) for one approach to fixed effects in nonlinear systems.

[17] There are many other papers that use this score. See e.g. Baydar and Brooks-Gunn, 1991, and the papers cited by Todd and Wolpin.

[18] An affine transformation is e.g. $\chi_1 + \chi_2\theta$, $\chi_2 \neq 0$, so it includes linear transformations but also allows for intercepts.

For example, using the earnings as the anchor in period T, we write

$$\ln E_T = \mu_T + \delta\theta_T + \varepsilon_T, \tag{21}$$

where the scale of θ_T is unknown. For any affine transformation of θ_T, corresponding to different units of measuring investment, the value of δ adjusts and we can identify the left hand side of

$$\frac{\partial \ln E_T}{\partial I_{T-1}} = \delta \left(\frac{\partial \theta_T}{\partial I_{T-1}} \right). \tag{22}$$

Thus, although the scale of δ is not uniquely determined, nor is the scale of θ_T, the scale of $\delta\theta_T$ is uniquely determined by its effect on log earnings and define all factors relative to their effects on earnings. This approach generalizes to multiple factors and we apply it in this paper. We now develop our empirical approach to identifying and estimating the technology of skill formation.[19]

3.2 Estimating the Technology of Production of Cognitive and Noncognitive Skills

Our analysis departs from that of Todd and Wolpin (2005) in six ways. (1) We analyze the evolution of both cognitive and noncognitive outcomes using the equation system

$$\begin{pmatrix} \theta_{t+1}^N \\ \theta_{t+1}^C \end{pmatrix} = A_t \begin{pmatrix} \theta_t^N \\ \theta_t^C \end{pmatrix} + B_t I_t + \begin{pmatrix} \eta_t^N \\ \eta_t^C \end{pmatrix} \tag{23}$$

where I_t can be a vector and B_t a suitably dimensioned coefficient matrix. (2) We determine how stocks of cognitive and noncognitive skills at date t affect the stocks 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. (3) We develop a 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

[19] To check the robustness of our results, we could use another anchor - such as the probability of high school completion - using the methods developed in Carneiro, Hansen, and Heckman (2003) to incorporate the factors in the (nonlinear) outcome equations. See the discussion below in section 4.1.2. In this paper we use a linear probability model for high school choices.

indicators of the latent skills and latent investments. In some specifications we also use earnings as well as test scores as indicators of latent skills. We account for measurement errors in output and input measures. We find substantial measurement errors in the proxies for parental investment. (4) Instead of imposing a particular index of parental input based on components of the home score, we implicitly estimate an index. (5) Instead of relying solely on exclusion restrictions to generate instruments to correct for measurement error in the proxies for θ_t and I_t, and for endogeneity, we use covariance restrictions that exploit the feature of our data that we have many more measurements on θ_{t+1}, θ_t and I_t than the number of unobserved factors. This allows us to secure identification from cross equation restrictions using *MIMIC* (Jöreskog and Goldberger, 1975) and *LISREL* (Jöreskog, Sörbom, and Magidson, 1979) models.[20] When instruments are needed, they arise from the internal logic of the model presented in section 2, equation (18), using methods developed by Madansky (1964) and Pudney (1982). (6) Instead of relying on test scores as a measure of output and change in output due to parental investments, we anchor the scale of the test scores using adult outcome measures as log earnings and probability of high school graduation. Thus we estimate the effect of parental investments on the adult earnings of the child.

In terms of the model of Todd and Wolpin (equation (20)), we assume access to measurement systems for θ^C_{t+1}, θ^C_t, I_t and assume that we can represent the measurements by a dynamic factor structure:

$$\begin{aligned} Y^C_{j,t+1} &= \mu^C_{j,t+1} + \alpha^C_{j,t+1}\theta^C_{t+1} + \varepsilon^C_{j,t+1}, \quad \text{for} \quad j = 1, \ldots, m^C_{t+1}, \\ Y^C_{j,t} &= \mu^C_{j,t} + \alpha^C_{j,t}\theta^C_t + \varepsilon^C_{j,t}, \quad \text{for} \quad j = 1, \ldots, m^C_t, \\ X_{k,t} &= \mu^X_{k,t} + \beta_{k,t} I_t + \varepsilon^I_{k,t}, \quad \text{for} \quad k = 1, \ldots, m^I_t, t = 1, \ldots, T, \end{aligned} \tag{24}$$

where m^C_t, m^I_t are, respectively, the number of measurements on cognitive skills and investments in period t. Assuming that the components of $(\varepsilon^C_{j,t+1},\ \varepsilon^C_{j,t},\ \varepsilon^I_{k,t})$ are mutually independent and are independent of $(\varepsilon^C_{j,t'+1},\ \varepsilon^C_{j,t'},\ \varepsilon^I_{k,t'})$ $t \neq t'$, and all pairs are independent of θ^C_{t+1}, θ^C_t, I_t, we can, under the conditions we present in the next section, identify the parameters of (20). In a linear system, our approach to identification is based on the covariance restrictions used in *LISREL* (see,

[20] See Carneiro, Hansen, and Heckman (2003) and Hansen, Heckman, and Mullen (2004) for some recent extensions.

e.g. Bollen, 1989) and in *MIMIC* models (Jöreskog and Goldberger, 1975). Our approach can be extended to general nonlinear systems so we can use it to estimate more general technologies. See Cunha, Heckman, and Schennach (2006). However, in this paper, we confine our analysis to (log) linear technologies.

In our model, a "fixed effect" is an unmeasured factor that enters each stage of the production function. Exploiting the rich nature of the CNLSY data, we assume that fixed effects are factors that can be proxied by measurement equations. Since we allow for and identify correlations among the unobserved factors, we account for dependence among the latent inputs. We account for latent initial conditions of the process, θ_0, which correspond to endowment abilities. Because we have many different measures of abilities in the first period of our data, we identify the distribution of the latent initial conditions. We also identify the distribution of each θ_t as well as the dependence across θ_t and $\theta_{t'}$, $t \neq t'$.

Todd and Wolpin (2005) adopt a more agnostic posture and assume no measurements are available to proxy unobserved fixed effects. In this respect, their approach is more general than ours. They refer to "mental ability" as one candidate fixed effect. We have multiple measurements of mental (and other abilities) over the life cycle of the child and use these as proxies for "true" mental ability. We now present a formal statement of our identification strategy for the parameters of the model of equation (23).

3.3 A Model for the Measurements

Let θ_t^C, θ_t^N denote, respectively, the stock of cognitive and noncognitive skills of the agent in period t. We do not observe θ_t^C or θ_t^N directly. 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. Assume that:

$$Y_{j,t}^k = \mu_{j,t}^k + \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k \quad \text{for} \quad j = 1, \ldots, m_t^k \quad \text{and} \quad k = C, N \tag{25}$$

and set $\alpha_{1,t}^C = \alpha_{1,t}^N = 1$. Some normalization is needed to set the scale of the factors. The $\mu_{j,t}^k$ may depend on regressors. We normalize the factors on adult earnings instead of test scores as described in section 3.1 using a log linear earnings function. We make the following assumptions about the

$\varepsilon_{j,t}^{k}$ for $k = C, N$:

1. $\varepsilon_{j,t}^{k}$ is independent across agents and over time for $j = 1, 2, \ldots, m_t^k$, $k = C, N$, $t = 1, \ldots, T$;

2. $\varepsilon_{i,t}^{k}$ is independent from $\varepsilon_{j,t}^{l}$ for $i, j = 1, 2, \ldots, m_t^k$ and $i \neq j$, $k = C, N$, $l = C, N$, $t = 1, \ldots, T$;

3. $\varepsilon_{j,t}^{k}$ is independent of θ_{τ}^{k} for all $t, \tau = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, m_t^k$, $k = C, N$.

Let I_t denote parental investments in the skills of the child. We do not observe investments I_t directly. We observe a vector of measurements $X_{k,t}$. We represent these by

$$X_{k,t} = \mu_{k,t}^{X} + \beta_{k,t} I_t + \varepsilon_{k,t}^{I} \quad \text{for} \quad k = 1, \ldots, m_t^I, t = 1, \ldots, T, \tag{26}$$

and we set $\beta_{1,t} = 1$, a normalization that sets the scale of I_t. Again, the $\mu_{k,t}^{X}$ may depend on regressors. The following assumptions are made about $\varepsilon_{k,t}^{I}$:

1. $\varepsilon_{k,t}^{I}$ is independent across agents and over time for $k = 1, 2, \ldots, m_t^I$, $t = 1, \ldots, T$;

2. $\varepsilon_{j,t}^{I}$ is independent from $\varepsilon_{k,t}^{I}$ for $j, k = 1, 2, \ldots, m_t^I$ and $k \neq j$, $t = 1, \ldots, T$;

3. $\varepsilon_{k,t}^{I}$ is independent from I_{τ} for all $t, \tau = 1, 2, \ldots, T$ and $k = 1, 2, \ldots, m_t^I$, $t = 1, \ldots, T$.

4. $\varepsilon_{k,t}^{I}$ is independent of θ_{τ}^{k} and $\varepsilon_{j,t}^{l}$ for all $t, \tau = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, m_t^k$, $k = C, N$, $l = C, N$.

The ε's are components of measurement error accounting for the fallible measure of the latent skills and investments. Let S, A denote mother's education and latent cognitive ability, respectively. We have a vector of measurements on mother's cognitive ability, M_k for $k = 1, \ldots, m_A$:

$$M_k = \mu_k^M + \delta_k A + \varepsilon_k^A \tag{27}$$

where the error term ε_k^A satisfies:

1. ε_k^A is independent from ε_j^A for $j \neq k$.

2. ε_k^A is independent from A, all θ's and I_t.

3. ε_k^A is independent of all $\varepsilon_{l,t}^I, \varepsilon_{j,t}^k$: $l = 1, \ldots, m_T^I$; $t = 1, \ldots, T$; $j = 1, \ldots, m_t^k$; $k = C, N$.

We assume that the mother's education is not measured with error.

We analyze a simple log linear law of motion for skills. Let $\Gamma_t = (A_t, B_t)$

$$\theta_{t+1}^k = \gamma_{1,t}^k \theta_t^N + \gamma_{2,t}^k \theta_t^C + \gamma_{3,t}^k I_t + \gamma_{4,t}^k S + \gamma_{5,t}^k A + \eta_t^k \quad \text{for } k = C, N \text{ and } t = 1, \ldots, T, \tag{28}$$

where the error term η_t^k satisfies the property that η_t^k is independent across agents and over time for the same agents, but η_t^C and η_t^N are freely correlated. We show how to relax the independence assumption at our website appendix and allow for unobserved inputs, but in this paper we assume independence.[21]

We allow the components of $(\theta_{t+1}, \theta_t, I_t, A)$ to be freely correlated for any t and with any vector $(\theta_{t'+1}, \theta_{t'}, I_{t'}, A)$, $t' \neq t$, and we can identify this dependence. We assume that m_t^k is independent of θ_t, I_t, S, and A for $k = C, N$ and $t = 1, \ldots, T$. When the means associated with each measurement system contain regressors, we assume that the regressors are independent of the errors in all measurement systems. We now establish conditions under which the technology parameters are identified.

3.4 Semiparametric Identification

The goal of the analysis is to recover the joint distributions of $\left\{\theta_t^C, \theta_t^N\right\}_{t=1}^T$, A, $\{I_t\}_{t=1}^T$, $\left\{\eta_t^k\right\}_{t=1,k=C,N}^T$, $\left\{\varepsilon_{j,t}^k\right\}_{j=1}^{m_t^k}$, $\left\{\varepsilon_{k,t}^I\right\}_{k=1}^{m_t^I}$ and $\left\{\varepsilon_{k,t}^A\right\}_{k=1}^{m_t^A}$ nonparametrically, as well as the parameters $\left\{\alpha_{j,t}^k\right\}_{j=1,t=1}^{m_t^k}$, $\left\{\beta_{j,t}\right\}_{j=2}^{m_t^I}$, $\left\{\gamma_{j,t}^k\right\}_{j=1}^5$ for $k = C, N$. For simplicity we do not discuss identification of the means of the measurements because under our assumptions, this is straightforward.

3.4.1 The identification of $\left\{\alpha_j^C\right\}_{j=2}^{m_t^C}$, $\left\{\alpha_j^N\right\}_{j=2}^{m_t^N}$, $\left\{\beta_{k,t}\right\}_{k=1}^{m_t^I}$ and the distributions of $\left\{\theta_t^C, \theta_t^N\right\}_{t=1}^T$, $\{I_t\}_{t=1}^T$

Suppose that $m_t^k \geq 3$, $k = C, N$. Since we observe $\left\{Y_{j,t}^C\right\}_{j=1}^{m_t^C}$ for every person, we can compute from the data $Cov\left(Y_{i,t}^C, Y_{j,t}^C\right)$ for all i, j pairs. Let $Var\left(\theta_t^C\right) = \sigma_{\theta_t^C}^2$ denote the variance of θ_t^C,

[21] When we apply the econometric framework that relaxes independence, our empirical conclusions do not change in any important way.

$t = 1, 2, \ldots, T$, and note that it may change over time. Recall that $\alpha_{1,t}^C = 1$. Consequently, we obtain:

$$Cov\left(Y_{1,t}^C, Y_{2,t}^C\right) = \alpha_{2,t}^C \sigma_{\theta_t}^2 \tag{29}$$

$$Cov\left(Y_{1,t}^C, Y_{3,t}^C\right) = \alpha_{3,t}^C \sigma_{\theta_t}^2 \tag{30}$$

$$Cov\left(Y_{2,t}^C, Y_{3,t}^C\right) = \alpha_{2,t}^C \alpha_{3,t}^C \sigma_{\theta_t}^2. \tag{31}$$

Taking ratios:

$$\frac{Cov\left(Y_{2,t}^C, Y_{3,t}^C\right)}{Cov\left(Y_{1,t}^C, Y_{3,t}^C\right)} = \alpha_{2,t}^C$$

$$\frac{Cov\left(Y_{2,t}^C, Y_{3,t}^C\right)}{Cov\left(Y_{1,t}^C, Y_{2,t}^C\right)} = \alpha_{3,t}^C.$$

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, (29) 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 (25) as:

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

We can apply Kotlarski's Theorem (Kotlarski, 1967) and identify the distributions of θ_t^C and $\frac{\varepsilon_{j,t}^C}{\alpha_{j,t}^C}$. Since $\alpha_{j,t}^C$ is identified, it is possible to recover the distribution of $\varepsilon_{j,t}^C$ for $j = 1, 2, \ldots, m_t^C$, and $t = 1, 2, \ldots, T$.

The preceding analysis can also be applied to the system of measurements of noncognitive skills. Consequently, it is possible to identify the loadings $\alpha_{j,t}^N$ for $j = 1, 2, \ldots, m_t^N$ up to the normalization $\alpha_{j,t}^N = 1$. Once factor loadings are identified, we can apply Kotlarski's Theorem and obtain identification of the distributions of θ_t^N and $\varepsilon_{j,t}^N$ for $t = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, m_t^N$.

Proceeding in a similar fashion for the measurements on investments, we conclude that the parameters $\beta_{k,t}$, $k = 1, 2, \ldots, m_t^I$ are identified up to the normalization $\beta_{1,t} = 1$. Also, we can apply Kotlarski's Theorem and conclude that the distributions of I_t and $\varepsilon_{k,t}^I$ are identified for $t = 1, 2, \ldots, T$ and $k = 1, 2, \ldots, m_t^I$. Suppose that we have a vector of M_k test scores for parents,

$k = 1, \ldots, m_A$, where $m_A \geq 3$. We can use the same principles to identify the distribution of A for parents.

Finally, we can identify the covariance between θ_t^C, θ_t^N by computing:

$$Cov\left(Y_{1,t}^C, Y_{1,t}^N\right) = Cov\left(\theta_t^C, \theta_t^N\right)$$

and we can proceed analogously to identify the covariances $Cov\left(\theta_t^C, I_t\right)$ and $Cov\left(\theta_t^N, I_t\right)$, and the joint covariances among all factors associated with all measurement systems.[22]

3.4.2 The Identification of the Technology Parameters

Consider, for example, the law of motion for noncognitive skills:

$$\theta_{t+1}^N = \gamma_{1,t}^N \theta_t^N + \gamma_{2,t}^N \theta_t^C + \gamma_{3,t}^N I_t + \gamma_{4,t}^N S + \gamma_{5,t}^N A + \eta_t^N \text{ for } t = 1, \ldots, T. \tag{32}$$

A parallel analysis can be performed for cognitive skills. We assume that η_t^N is serially independent but possibly correlated with all of the latent factors. We substitute the measurement equations $Y_{1,t+1}^N, Y_{1,t}^N, Y_{1,t}^C, X_{1,t}$, and M_1, for $\theta_{t+1}^N, \theta_t^N, \theta_t^C, I_t$, and A respectively:

$$Y_{1,t+1}^N = \gamma_{1,t}^N Y_{1,t}^N + \gamma_{2,t}^N Y_{1,t}^C + \gamma_{3,t}^N X_{1,t} + \gamma_{4,t}^N S + \gamma_{5,t}^N M_1 + \left(\varepsilon_{1,t+1}^N - \gamma_{1,t}^N \varepsilon_{1,t}^N - \gamma_{2,t}^N \varepsilon_{1,t}^C - \gamma_{3,t}^N \varepsilon_{1,t}^I - \gamma_{5,t}^N \varepsilon_1^A + \eta_{t+1}^N\right). \tag{33}$$

If we estimate (33) by least squares, we do not obtain consistent estimators of $\gamma_{k,t}^N$ for $k = 1, \ldots, 5$ because the regressors $Y_{1,t}^N$, $Y_{1,t}^C$, $X_{1,t}$, and M_1 are correlated with the error term ω_{t+1} where:

$$\omega_{t+1} = \varepsilon_{1,t+1}^N - \gamma_{1,t}^N \varepsilon_{1,t}^N - \gamma_{2,t}^N \varepsilon_{1,t}^C - \gamma_{3,t}^N \varepsilon_{1,t}^I - \gamma_{5,t}^N \varepsilon_1^A + \eta_{t+1}^N.$$

However, we can instrument $Y_{1,t}^N, Y_{1,t}^C, X_{1,t}$, and M_1 using $\left(Y_{j,t}^N\right)_{j=2}^{m_t^N}, \left(Y_{j,t}^C\right)_{j=2}^{m_t^C}, \left(X_{j,t}\right)_{k=2}^{m_t^I}, \left(M_k\right)_{k=2}^{m_A}$. The $\left(Y_{j,t}^N\right)_{j=1}^{m_t^N}, \left(Y_{j,t}^C\right)_{j=2}^{m_t^C}$ are valid instruments as long as in equation (23) $A_t \neq (0)$, so the factors are correlated over time. The $\left(X_{j,t}\right)_{k=2}^{m_t^I}$ are valid instruments for $X_{1,t}$ as a consequence of equation (18)

[22] We can recover the joint distributions from a standard multivariate Fourier inversion argument. See Cunha, Heckman, and Schennach (2006).

connecting investments in different periods. The $(M_k)_{k=2}^{m_A}$ are valid for M_1 because of the common factor generating them. Using two stage least squares with these instruments allows us to recover the parameters $\gamma_{k,t}^N$ for $k = 1,2,3,4,5$. See Pudney (1982). Our instruments are the internal instruments justified by the model. We can perform a parallel analysis for the cognitive factor. The suggested instruments are also independent of η_t^N because of the lack of serial correlation in η_t^N.[23] We can repeat the argument for different time subscripts and the same results apply. We now turn to our estimates.

3.5 Anchoring the Factors in the Metric of Earnings

We set the scale of the factors by estimating their effects on log earnings for children when they become adults. Let E_T be adult earnings. We write

$$\ln E_T = \mu_T + \delta_N \theta_T^N + \delta_C \theta_T^C + \varepsilon_T \tag{34}$$

Define

$$D = \begin{pmatrix} \delta_N & 0 \\ 0 & \delta_C \end{pmatrix}.$$

We assume $\delta_N \neq 0$ and $\delta_C \neq 0$.

For any given normalization of the test scores we can transform the θ_t to an earnings metric by multiplying (23) by D:

$$D\theta_{t+1} = \left(DA_tD^{-1}\right)(D\theta_t) + (DB)\,I_t + (D\eta_t)\,, \tag{35}$$

and working with $D\theta_{t+1}$ and $D\theta_t$ in place of θ_{t+1} and θ_t. The cross terms in (DA_tD^{-1}) are affected by this change of units but not the self-productivity terms. The relative magnitude of I_t on the outcomes can be affected by this change in scale. We report results from two anchors in this paper: (a) log earnings and (b) the probability of graduating from high school. For the latter, we use a linear probability model to represent outcomes.

[23] See our website for an analysis of the case in which n_t^k are serially correlated for $k = C, N$.

4 Estimating the Technology of Skill Formation

We use a sample of the 1053 white males from the Children of the NLSY/79 (CNLSY/79) data set. 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. Table 2 presents summary statistics of our data.

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.[24] 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

[24] See http://jenni.uchicago.edu/idest-tech.

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 high 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 high 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 high scores on the headstrong measurement is tense, nervous, argues too much, and is disobedient at home, for example. Children will score high 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 high score on the peer problem subscore if they have problems getting along with others, are not liked 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[25]. 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.

Our dynamic factor models allow us to exploit the wealth of measures available in these data.

[25] 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.

The Appendix presents the sample likelihood. The dynamic factor models enable us to solve several problems. First, there are many proxies for parental investments in children's cognitive and noncognitive development. Even if all parents provided responses to all of the measures of family input, we would still face the problem of selecting which variables to use and how to find enough instruments for so many endogenous variables. 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 the 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. It often happens that mothers do not provide responses to all items of the HOME-SF score. Similarly, some children may take the PIAT Reading Recognition exam, but not the PIAT mathematics test. Another missing data problem that arises is that the mothers may provide information about whether the child has peer problems or not, but may refuse to issue statements regarding the child's hyperactivity level. For such cases, some researchers drop the observations on the parents who do not respond to certain items, or do not analyze the items that are not responded to by many parents, even though these items may be very informative. With our setup, we do not need to drop the parents or entire items in our analysis. Assuming that the data are missing randomly, we integrate out the missing items from the sample likelihood. We now present and discuss our empirical results using the CNLSY data.

4.1 Empirical Results

We first present our estimates of an age-invariant version of the technology where we assume no critical and sensitive periods. We report estimates of a model with critical and sensitive periods in section 4.1.3.

4.1.1 Estimates of Time-Invariant Technology Parameters

Using the CNLSY data, we first estimate the simplest version of the model that imposes the restriction that the coefficients on the technology equations do not vary over periods of the child's life cycle. We first report results in the scale of standardized test scores. We discuss estimates in the scales of log earnings and the probability of graduating from high school below. We normalize the scale of the investment factor I_t by using “trips to the theater”. We report the effect of alternative normalizations on our estimates below. Recall that we use adult outcome measures (log earnings and high school graduation) of the child as outcomes generated, in part, by the terminal cognitive and noncognitive factors $\left(\theta_T^N, \theta_T^C\right)$ respectively. Dropping this information has negligible effects on the estimates.

Table 3 shows the estimated parameter values and their standard errors. From this table, we see that: (1) both cognitive and noncognitive skills show strong persistence over time; (2) noncognitive skills affect the accumulation of next period cognitive skills, but cognitive skills do not affect the accumulation of next period noncognitive skills; (3) the estimated parental investment factor affects noncognitive skills somewhat more strongly than cognitive skills, although the differences are not statistically significant; (4) the mother's ability affects the child's cognitive ability but not noncognitive ability; (5) the mother's education plays no role after controlling for parental investments. These results are robust to alternative normalizations of the factor loadings on the measurements associated with family inputs that set the scale of the parental investment factor as we discuss below.

The dynamic factors are estimated to be statistically dependent. Table 4 shows the evolution of the correlation patterns across the dynamic factors. Early in the life cycle, the correlation between cognitive and noncognitive skills is strong. The correlation is 0.21 as early as ages 6 and 7, and it grows to around 0.29 at ages 12 and 13. There is also strong contemporaneous correlation among noncognitive skill and the home investment. The correlation starts off at 0.32 at ages 6 and 7 and grows to 0.53 by ages 12 and 13. The same pattern is true for the correlation between cognitive skills and home investments. In fact, the correlation between these two variables actually doubles from 0.27 at ages 6 and 7 to 0.57 at ages 12 and 13.

We also report results adding family income at age t to the list of parental inputs. One advantage of this proxy is that it provides a natural scale for I_t. However, as Todd and Wolpin (2003, 2005) note, family income is *not* a valid input measure and using it as an input can make the interpretation of estimated relationships as production functions problematic.[26] However, when we add this "input," and normalize off of it, our estimates are barely affected. Compare the estimates in table 5 with those in table 3. The same conclusion holds for the contemporaneous correlation matrices, which we do not show for the sake of briefness.

We compare results from these estimations with those obtained from OLS regression. Table 6 reports the OLS estimates of the technology parameters with roughly comparable specifications. In this table, instead of using all measures of noncognitive and cognitive skills available in the data, we use only the antisocial measure for noncognitive skills and the PIAT mathematics for cognitive skills. Qualitatively, the results reported in Table 6 are similar to those reported in Table 3. There is strong persistence of noncognitive and cognitive skills. Noncognitive skills affect the accumulation of cognitive skills but not vice versa. The estimates suggest that noncognitive skills respond more strongly to investment than cognitive skills. However, the estimated parameter values tend to be consistently lower in the OLS regressions in comparison to the corresponding estimates from our dynamic factor approach. One possible explanation is attenuation bias due to measurement error. We report evidence on the severity of the measurement error in section 4.1.4.[27]

4.1.2 Anchoring our estimates of the factor scale using adult outcomes

We now report estimates that use the earnings data for persons age 23–28 to anchor the output of the production function in a dollar metric. Our fitted earnings function is linear and quadratic in age, and depends on the final level of the factors θ_T^C and θ_T^N. The coefficient on cognitive skills in the log earnings equations is estimated to be 0.14 (standard error is 0.054). For noncognitive skills, we estimate a loading of 0.052 (with a standard error of 0.0109). These estimates are consistent with estimates reported in Heckman, Stixrud, and Urzua (2006). From equation (34) it is clear that anchoring does not affect the estimates of self productivity but can affect the estimates of cross productivity. It can also affect the magnitude of the estimated effect of I_t on outcomes.

[26] They stress the point that controlling for family income, changing one input requires changing other inputs if the budget is exhausted on other inputs. We use family income as a proxy for I_t and do not directly condition on family income.

[27] We experimented with alternative specifications of the OLS regression using different test scores and input measures and obtain the same approximate results. One specification was based on indices constructed from the first principal components of the test scores and family investments. It shows severe attenuation bias as would be predicted from our analysis of measurement error in section 4.1.4.

Table 7, which transforms the estimates in Table 3 by D into a log earnings metric, shows that some of our conclusions are altered when we anchor outcomes in adult earnings. The two cross effects are ordered in the same direction as in the model where we use the metric of test scores and the statistical inference on the coefficients is unchanged. In the metric of earnings the effect of cognitive skills on noncognitive skill is weaker and the effect of noncognitive skills on cognitive skills is stronger than in the test score metric. In the metric of earnings, the estimated parental investment factor affects *cognitive* skills more strongly than noncognitive skills. The relative strength of these effects is reversed across the two metrics. The other qualitative conclusions from Table 3 stand.

One problem that might arise in using log earnings as an anchor for this sample is that log earnings are observed for the children who are born to very young mothers, making it a very selected sample. To check the robustness of our results we also use high school graduation for a person at least 19 years-old to anchor the parameters of the technology equations. We model the probability of high school graduation as a linear probability model and find that the coefficient on noncognitive skills is 0.098 (with a standard deviation of 0.031) and the coefficient on cognitive skills is 0.26 (with a standard deviation of 0.07). On table 8, we show the empirical results from using high-school graduation and only note that the same conclusions from the comparison of tables 3 and 7 are also found comparing table 8 to table 3.

4.1.3 Evidence of Sensitive Periods of Investment in Skills

We now report evidence on sensitive periods. Our analysis in section 3 presents conditions under which we can identify the parameters of the technology when they are allowed to vary over stages of the life cycle. We can identify whether there are sensitive periods in the development of skills provided that we normalize our investment factor on an input that is used at all stages of the life cycle. Using several alternative measures including trips to the theater, the number of books, as well as family income as a "proxy," we obtain the same qualitative ordering in terms of critical and sensitive periods.[28] All of our estimated models include an equation for the adult earnings of the child based on the period T value of the factors but "output" is reported in test score units.

[28] In the text we report the results for the normalization relative to "trips to the theater and musical performances". In our website appendix, we report estimates of alternative normalizations using family income and the number of books. Family income does not appear among the family inputs unless it is used as a normalization.

Using a likelihood ratio test, we test and reject the hypothesis that the parameters describing the technologies are invariant over stages of the lifecycle. Specifically, we use a likelihood ratio test. Under the restricted model, we estimate 277 parameters and the value of the log likelihood at the maximum is -53877. Under the unrestricted model, we estimate 305 parameters and the log likelihood attains the maximum value of -53800. The statistic $\lambda = -2(\ln L_R - \ln L_U)$ is distributed as chi-square with 28 $(= 305 - 277)$ degrees of freedom. We find that λ is 155, significantly above the critical value of 41.337 at a 5% significance level.

Our estimates are reported in table 9. Although we use test scores as a measure of output, transformation by D will not affect our inference about sensitive periods because D is time invariant. When we allow the coefficients of the technology to vary over time we find evidence of sensitive periods for both cognitive and noncognitive skills. A sensitive period for parental investments in cognitive skills occurs at an earlier age than the sensitive period for parental investments in noncognitive skills. The coefficient on investments in the technology for cognitive skills for the transition from period one to period two (ages 6 and 7 to ages 8 and 9) is around 0.11 (with a standard error of 0.0152). For the transition from period two to period three (ages 8 and 9 to 10 and 11) this same coefficient decreases rather sharply to 0.0397 (with a standard error of 0.0131). For the final transition (ages 10 and 11 to ages 12 and 13), this coefficient is about the same: 0.0294, with a standard error of 0.0103. The difference between the early coefficient and the later two is statistically significant. This finding is consistent with periods 1 and 2 being sensitive periods for cognitive skills.[29]

For noncognitive skills in period one, the coefficient on investments is only 0.0462, with a standard error of 0.0131. Then, it increases to 0.1119 in period two. It reduces to 0.0375 in the final transition. This evidence suggests that the sensitive periods for the development of noncognitive skills tend to take place at later ages in comparison to sensitive periods for cognitive skills.[30] Table 10 shows that the correlation among the dynamic factors does not change even after we allow for the technology parameters to vary over time. Compare it to table 4. When we use alternative

[29] For the coefficients on cognitive skills, the lower bound for the t statistic for the hypothesis $\gamma_{I,2}^{C} = \gamma_{I,1}^{C}$ is 2.73. For the hypothesis $\gamma_{I,2}^{C} = \gamma_{I,3}^{C}$ it is 3.43.

[30] For the coefficients of investments on noncognitive skills, the lower bound for the t statistic for the hypothesis $\gamma_{I,2}^{N} = \gamma_{I,1}^{N}$ is 2.16 and for the hypothesis $\gamma_{I,2}^{N} = \gamma_{I,3}^{N}$ it is 2.34.

normalizations of parental input measures and add family income to the list, we obtain the same results. See tables 9 and 10 in our web appendix.

4.1.4 Estimating the Components of the Home Investment Dynamic Factor

In table 11 we show how our method constructs an implicit home score by estimating factor loadings on the inputs used to form the conventional home score. We use the estimates generating the parameters reported in tables 9 and 10. Thus we normalize the scale of the investment factor by "trips to the theater". The CNLSY/1979 reports an aggregate HOME score by adding these variables and assigning each one of them the same weight. For expositional purposes we call these ad-hoc weights. The advantage of working with (dynamic) factor models is that the relative weights on the components of the home score are estimated rather than imposed. We set the absolute scale of the factor by the normalizations discussed in the preceding section.

For example, consider the number of books available to the child. This variable is correlated with parental inputs because parents who invest more in the development of their children will tend to spend more resources on books. But the number of books is unlikely to be a perfect indicator of total parental input. Our method allows for imperfect proxies. Under our method, the number of books a child has at age t (R_t) is modelled as $R_t = \alpha^I_{R,t} I_t + \varepsilon^I_{R,t}$ so that $Var\left(R_t\right) = \left(\alpha^I_{R,t}\right)^2 Var\left(I_t\right) + Var\left(\varepsilon^I_{R,t}\right)$, because of the independence between I_t and $\varepsilon^I_{R,t}$. We can decompose the total unobserved variance in two terms: one that is due to the parental input, the other that is orthogonal to it. The relative importance of the two measures can be computed as:

$$s_{I,R,t} = \frac{\left(\alpha^I_{R,t}\right)^2 Var\left(I_t\right)}{\left(\alpha^I_{R,t}\right)^2 Var\left(I_t\right) + Var\left(\varepsilon^I_{R,t}\right)}$$

and

$$s_{I,\varepsilon,t} = \frac{Var\left(\varepsilon^I_{B,t}\right)}{\left(\alpha^I_{R,t}\right)^2 Var\left(I_t\right) + Var\left(\varepsilon^I_{R,t}\right)}.$$

Table 11 reports that $s_{I,R,1} = 0.1359$ (corresponding to 6 and 7), while $s_{I,\varepsilon,1} = 0.8641$. So, most of the unobservable variance in "the number of books a child has" is actually not informative on the parental input unobserved variable I_t. We report the same measures for the other input variables in table 11. Over stages of the life cycle, all of the input measures tend to become more error laden as a proxy for I_t.

One can interpret the inverse of the factor loadings on the investment inputs as a measure of the strength of the relationship between the measure and the latent factor. Consequently, for every measurement $X_{k,t}$ we obtain the relationship:

$$\frac{1}{\alpha_{k,t}^{I}} E\left(X_{k,t}\right) = E\left(I_t\right).$$

If we have m_t^I measurements for investments, we can construct the implicit relative weights on input in predicting I_t, $w_{k,t}$, $k = 1, 2, ..., m_t^I$, as:

$$w_{i,t} = \frac{\frac{1}{\alpha_{k,t}^{I}}}{\sum_{i=1}^{m_t^I} \frac{1}{\alpha_{i,t}^{I}}}.$$

This is a weighted average of the inputs that proxies I_t, which is a measure of the true home score.

Table 11 displays the estimated weights $w_{i,t}$ for each measurement i at each period t. Note that the weights are not stable over stages of the life cycle. Our estimates show that the number of books receives high weight early on (ages 6/7 and 8/9), but the weight declines considerably in the later periods (ages 10/11 and 12/13). The variable that indicates whether the child receives special lessons, on the other hand, exhibits the opposite behavior. It starts small in early ages, but it becomes more important at later ages. It is interesting to remark that variables that describe how often children attend theater or visit museums, although informative about the home investments, receive lower weights in our method than the weights that weight all items equally strongly.

5 Conclusion

This paper formulates, identifies and estimates a model of investment in child cognitive and noncognitive skills using dynamic factor models. We present a simple model of human skill development that generates critical and sensitive periods of investment and that can explain the body of evidence surveyed in Cunha, Heckman, Lochner, and Masterov (2006).

Our empirical methodology accounts for the proxy nature of the measurements on parental investments and outcomes and for the endogeneity of inputs. It allows us to utilize the large number of potentially endogenous proxy variables available in our data set without exhausting the available instruments. Our instruments are justified by the internal logic of the model. To avoid the arbitrariness in using test scores to measure output of parental investments, we normalize estimated effects of investment in the metric of adult earnings. The choice of the metric affects conclusions on cross effects but not our results on self productivity or sensitive periods. We report results for alternative normalizations of the scale of parental investment and find agreement in the conclusions from alternative specifications.

Our estimated technology displays the following features. (1) High levels of self productivity in the production of cognitive and noncognitive skills. (2) Evidence of sensitive periods for parental investments in both types of skills with the sensitive period for cognitive skill investments occurring earlier in the life cycle than the sensitive period for investments in noncognitive skills. (3) Substantial evidence of measurement error in the home input proxies and corollary evidence of attenuation bias in the OLS estimates of the model. The estimated strength of cross effects is sensitive to the metric in which we measure output.

Missing from this paper is an estimate of the key substitution parameters that determine the cost of later remediation relative to early investment. To recover these crucial parameters requires a more general specification of the technology and more advanced econometric methods. This task is underway in Cunha, Heckman, and Schennach (2006), who also present a more general nonlinear approach to anchoring the test scores in an outcome measure.

A Sample Likelihood

We derive the likelihood and describe the estimation strategy. In period t, let $m_t = m_t^N + m_t^C + m_t^I$ where m_t^N is the number of measurements on the noncognitive factor, and m_t^C, m_t^I are defined accordingly for the cognitive and investment factors. Here we explicitly allow for the number of measurements to be period specific. Let Y_t denote the $(m_t \times 1)$ vector

$$Y_t' = \begin{pmatrix} Y_{1,t}^N & \dots & Y_{m_t^N,t}^N & Y_{1,t}^C & \dots & Y_{m_t^C,t}^C & X_{1,t} & \dots & X_{m_t^I,t} \end{pmatrix}.$$

At each period t, let $\theta_t' = \left(\theta_t^N, \theta_{t,}^C, I_t\right)$. We use α_t to denote the $(m_t \times 3)$ matrix containing the factor loadings.

$$\alpha_t = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ \alpha_{m_t^N,t}^N & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & \alpha_{m_t^C,t}^C & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \alpha_{m_t^I,t}^I \end{bmatrix}$$

Let ε_t denote the $(m_t \times 1)$ vector of uniquenesses and $H_t = Var\left(\varepsilon_t\right)$, where H_t is $(m_t \times m_t)$ matrix. With this notation, we can write the observation equations as:

$$Y_t = \alpha_t \theta_t + \varepsilon_t \tag{A.1}$$

We remind the reader that we use S, A to denote mother's education and cognitive ability. Let G_t be a (3×3) matrix of coefficients. Let ψ_1 and ψ_2 denote (3×1) vectors. The G_t matrix and the vectors ψ_1 and ψ_2 contain the technology parameters for both the cognitive and noncognitive factors:

$$\theta_{t+1} = G_t \theta_t + \psi_1 S + \psi_2 A + \eta_{t+1}$$

where η_{t+1} is a (3×1) vector of error terms in the technology equations. Define $Q_t = Var(\eta_t)$.

We assume that $\theta_1 | S, A \sim N(a_1, P_1)$. In the text, we establish the conditions for identification of a_1 and P_1. We also assume that $\varepsilon_t \sim N(0, H_t)$ and $\eta_t \sim N(0, Q_t)$. Then, given a normality assumption, together with linearity, it follows that $Y_1 \sim N(\mu_1, F_1)$ where:

$$\mu_1 = \alpha_1 a_1 \text{ and } F_1 = \alpha_1 P_1 \alpha_1' + H_1.$$

Normality is not required for identification but it facilitates computation. In future work we plan to relax this assumption. To proceed, we apply the Kalman filtering procedure (for details on the derivations see, for example, Harvey, 1989 or Durbin and Koopman, 2001). If we define $Y^t = (Y_1, \ldots, Y_t)$, $a_{t+1} = E(\theta_{t+1} | S, A, Y^t)$, and $P_{t+1} = Var(\theta_{t+1} | S, A, Y^t)$, it is straightforward to establish that:

$$a_{t+1} = G_t a_t + G_t P_t \alpha_t' (\alpha_t P_t \alpha_t' + H_t)^{-1} + \psi_1 S + \psi_2 A,$$

and

$$P_{t+1} = G_t P_t G_t' - G_t P_t \alpha_t' \alpha_t P_t (\alpha_t P_t \alpha_t' + H_t)^{-1} G_t'.$$

Consequently, using the relationship in (A.1) we obtain that $Y_{t+1} | S, \lambda, Y^t \sim N(\mu_t, F_t)$ where:

$$\mu_t = \alpha_t a_t \text{ and } F_t = \alpha_t P_t \alpha_t' + H_t.$$

Assuming that we observe mother's schooling, S, and mother's education, A, we can decompose the contribution of individual i to the likelihood as:

$$f(y_{i,T}, y_{i,T-1}, \ldots, y_{i,1} | S_i, A) = f(y_{i,1} | S_i, A) \prod_{t=2}^{T} f\left(y_{i,t} | S_i, A, Y_i^{t-1}\right).$$

In general we observe S but not A. However, we have shown that we can identify the distribution of A if we have a set of cognitive test scores for the mother, M. Consequently, we can integrate A out:

$$f(y_{i,T}, y_{i,T-1}, \ldots, y_{i,1} | S_i) = \int f(y_{i,1} | S_i, A) \prod_{t=2}^{T} f\left(y_{i,t} | S_i, A, Y_i^{t-1}\right) f_A(A)\, dA.$$

Assuming that observations are i.i.d. over children, we get that the likelihood of the data is:

$$\prod_{i=1}^{n} f\left(y_{i,T}, y_{i,T-1}, \ldots, y_{i,1} \middle| S_i\right) = \prod_{i=1}^{n} \int f\left(y_{i,1} \middle| S_i, A\right) \prod_{t=2}^{T} f\left(y_{i,t} \middle| S_i, A, Y_i^{t-1}\right) f_A\left(A\right) dA.$$

Missing data can be integrated out and so all cases can be used even in the presence of missing data.

References

Altonji, J. G. and R. L. Matzkin (2005, July). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica 73*(4), 1053–1102.

Arellano, M. (2003). *Panel Data Econometrics*. New York: Oxford University Press.

Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. New York: Wiley.

Baydar, N. and J. Brooks-Gunn (1991, November). Effects of maternal employment and child-care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from the children of the national longitudinal survey of youth. *Developmental Psychology 27*(6), 932–945.

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.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bowles, S., H. Gintis, and M. Osborne (2001, December). The determinants of earnings: A behavioral approach. *Journal of Economic Literature 39*(4), 1137–1176.

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.

Carneiro, P. and J. J. Heckman (2003). Human capital policy. In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.

Carneiro, P., J. J. Heckman, and D. V. Masterov (2005, April). Labor market discrimination and racial differences in pre-market factors. *Journal of Law and Economics 48*(1), 1–39.

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.

Cunha, F. (2004). Skill formation in a laitner economy. Ph.D. thesis, University of Chicago.

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., 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.

Cunha, F., J. J. Heckman, and S. M. Schennach (2006). A nonseparable model for the evolution of skills: Estimating substitution in latent skills over the life cycle. Unpublished manuscript, University of Chicago, Department of Economics.

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.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. New York: Oxford University Press.

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

Fryer, R. and S. Levitt (2004, May). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics 86*(2), 447–464.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004, July-August). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*(1-2), 39–98.

Hanushek, E. (2003, February). The failure of input-based schooling policies. *Economic Journal 113*(485), F64–F98.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.

Heckman, J. J. and P. LaFontaine (2006). *America's Dropout Problem: The GED and the Importance of Social and Emotional Skills*. Chicago: University of Chicago Press. Forthcoming.

Heckman, J. J. and Y. Rubinstein (2001, May). The importance of noncognitive skills: Lessons from the ged testing program. *American Economic Review 91*(2), 145–149.

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*. Forthcoming.

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

Hsiao, C. (1986). *Analysis of Panel Data*. New York: Cambridge University Press.

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.

Jöreskog, K. G., D. Sörbom, and J. Magidson (1979). *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books.

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.

Kotlarski, I. I. (1967). On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics 20*, 69–76.

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.

Madansky, A. (1964, June). Instrumental variables in factor analysis. *Psychometrika 29*(2), 105–113.

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.

Neal, D. A. and W. R. Johnson (1996, October). The role of premarket factors in black-white wage differences. *Journal of Political Economy 104*(5), 869–895.

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

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.

Pudney, S. E. (1982, December). Estimating latent variable systems when specification is uncertain: Generalized component analysis and the eliminant method. *Journal of the American Statistical Association 77*(380), 883–889.

Raver, C. C. and E. F. Zigler (1997). Social competence: An untapped dimension in evaluating head start's success. *Early Childhood Research Quarterly 12*(4), 363–385.

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

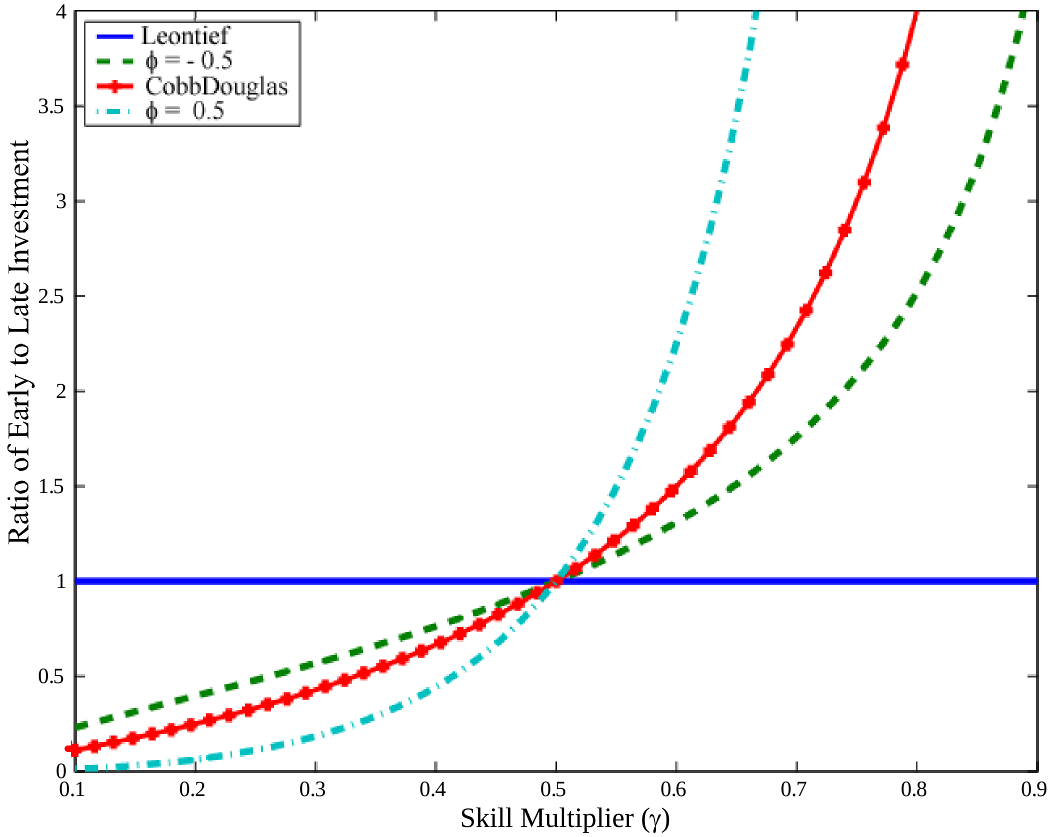
Shumway, R. H. and D. S. Stoffer (1982, May). An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis 3*(3), 253–264.

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
The Ratio of Early to Late Investment in Human Capital
As a function of the Skill Multiplier for Different Values of Complementarity



This figure shows the optimal ratio of early to late investments, $\frac{I_1}{I_2}$, as a function of the skill multiplier parameter γ, for different values of the complementarity parameter ϕ, assuming that the interest rate r is zero. The optimal ratio $\frac{I_1}{I_2}$ is the solution of the parental problem of maximizing the present value of the child's wealth through investments in human capital, h, and transfers of risk-free bonds, b. In order to do that, parents have to decide how to allocate a total of M dollars into early and late investments in human capital, I_1 and I_2, respectively, and risk-free bonds. Let q denote the present value as of period "3" of the future prices of one efficiency unit of human capital: $q = \sum_{t=3}^{T} \frac{w_t}{(1+r)^{t-3}}$. The parents solve

$$\max \left(\frac{1}{1+r}\right)^2 [qh + b]$$

subject to the budget constraint

$$I_1 + \frac{I_2}{(1+r)} + \frac{b}{(1+r)^2} = M$$

and the technology of skill formation:

$$h = \left[\gamma I_1^{\phi} + (1-\gamma)\, I_2^{\phi}\right]^{\frac{\rho}{\phi}}$$

for $0 < \rho < 1$, $0 \leq \gamma \leq 1$, and $\phi \leq 1$. From the first-order conditions it follows that $\frac{I_1}{I_2} = \left[\frac{\gamma}{(1-\gamma)(1+r)}\right]^{\frac{1}{1-\phi}}$. This ratio is plotted in this figure when $\phi \to -\infty$ (Leontief), $\phi = -0.5$, $\phi = 0$ (Cobb-Douglas) and $\phi = 0.5$ and for values of the skill multiplier γ between 0.1 and 0.9.

Table 1. The Ratio of Optimal Early and Late Investments $\frac{I_1}{I_2}$ Under Different Assumptions About the Skill Formation Technology

	Low Self-Productivity: $\gamma < \frac{(1+r)}{(2+r)}$	High Self-Productivity: $\gamma > \frac{(1+r)}{(2+r)}$
High Degree of Complementarity: $\phi < 0$	$\frac{I_1}{I_2} \rightarrow 1$ as $\phi \rightarrow -\infty$	$\frac{I_1}{I_2} \rightarrow 1$ as $\phi \rightarrow -\infty$
Low Degree of Complementarity: $0 \leq \phi \leq 1$	$\frac{I_1}{I_2} \rightarrow 0$ as $\phi \rightarrow 1$	$\frac{I_1}{I_2} \rightarrow \infty$ as $\phi \rightarrow 1$

Note: This table summarizes the behavior of the ratio of optimal early to late investments according to four cases: I_1 and I_2 have high complementarity, but self-productivity is low; I_1 and I_2 have both high complementarity and self-productivity; I_1 and I_2 have low complementarity and self-productivity; and I_1 and I_2 have low complementarity, but high self-productivity. When I_1 and I_2 exhibit high complementary, complementarity dominates and is a force towards equal distribution of investments between early and late periods. Consequently, self-productivity plays a limited role in determining the ratio $\frac{I_1}{I_2}$ (row 1). On the other hand, when $I_1$and I_2 exhibit a low degree of complementarity, self-productivity tends to concentrate investments in the *late* period if self-productivity is low, but in the *early* period if it is high (row 2).

Table 2: 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.

Table 3
The Technology Equations
Measurement Variables are Standardize with Mean Zero and Variance One

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.8835	0.0215	0.0282	0.0131
Current Period Cognitive Skills	0.0181	0.0130	0.9814	0.0388
Current Period Investment	0.0601	0.0206	0.0566	0.0137
Mother's Education	0.0067	0.0088	0.0047	0.0062
Mother's Ability	-0.0063	0.0069	0.0290	0.0104
Variance of Shocks	0.1357	0.0219	0.0340	0.0091

[1]Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. In table 3 we show the estimated parameter values and standard errors of γ, ψ_1, and ψ_2 as well as the $Var\left(\eta_{t+1}^N\right)$ and $Var\left(\eta_{t+1}^C\right)$.

Table 4
Contemporaneous Correlation Matrices

Period 1 - Children ages 6 and 7

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2089	0.3217
Cognitive	0.2089	1.0000	0.2712
Investment (Home)	0.3217	0.2712	1.0000

Period 2 - Children ages 8 and 9

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2406	0.3797
Cognitive	0.2406	1.0000	0.3478
Investment (Home)	0.3797	0.3478	1.0000

Period 3 - Children ages 10 and 11

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2668	0.4500
Cognitive	0.2668	1.0000	0.4459
Investment (Home)	0.4500	0.4459	1.0000

Period 4 - Children ages 12 and 13

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2868	0.5318
Cognitive	0.2868	1.0000	0.5714
Investment (Home)	0.5318	0.5714	1.0000

[1]Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. Suppose that the initial distribution of the dynamic factors is $\theta_1 \sim N(a_1, P_1)$. In table 4, we show the covariance matrix P_t for differente ages. To see how this matrix is updated, let G contain the matrix of the technology parameters γ. Let H_t denote the covariance-variance matrix of the uniquenesses ε_t. Let Q_t denote the variance-covariance matrix of the error terms in the technology equations, η_{t+1}. This matrix is updated by the standard Kalman rule:

$$P_{t+1} = G_t P_t G_t' - G_t P_t \alpha_t' \alpha_t P_t \left(\alpha_t P_t \alpha_t' + H_t\right)^{-1} G_t' P_t$$

Table 5
The Technology Equations

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.8848	0.0212	0.0285	0.0132
Current Period Cognitive Skills	0.0041	0.0264	1.0779	0.0250
Current Period Investment	0.0705	0.0302	0.0414	0.0187
Mother's Education	0.0039	0.0056	0.0079	0.0028
Mother's Ability	-0.0168	0.0444	0.0268	0.0127
Variance of Shocks	0.1381	0.0296	0.0249	0.0034

[1] Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score). Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. In table 5 we show the estimated parameter values and standard errors of γ, ψ_1, and ψ_2 as well as the $Var\left(\eta_{t+1}^N\right)$ and $Var\left(\eta_{t+1}^C\right)$. The difference between table 3 and table 5 is that in the former we normalized the investment factor in "trips to the theater" while in the latter we normalized it in "log family income".

Table 6
OLS Regressions

	Dependent Variables			
	Standardized Social Score[1]		Standardized Math Score[1]	
	Coefficient	T-Statistic	Coefficient	T-Statistic
Lagged Standardized Antisocial Score	0.6271**	28.24	0.0269**	2.14
Lagged Standardized Math Score	0.0409	1.00	0.5703**	24.67
Lagged Number of Books[2]	0.0710	1.27	0.0154	0.48
Lagged Musical Instruments[3]	0.1223**	2.57	0.04784*	1.78
Lagged Newspaper[3]	-0.0080	-0.17	0.0402	1.53
Lagged Special Lessons[3]	0.0824	1.57	0.0512*	1.72
Lagged Museum[4]	0.0501	1.68	0.0018	0.11
Lagged Theater[4]	-0.0466	-1.43	0.0036	0.20
Mother's Education	0.0287**	2.46	0.0276**	4.18
Mother's ASVAB (AR)[5]	-0.0626	-1.53	0.0271	1.17
Mother's ASVAB (WK)[6]	-0.0695	-1.59	-0.0039	-0.16
Mother's ASVAB (PC)[7]	0.0077	0.17	-0.0257	-0.99
Mother's ASVAB (NO)[8]	0.0852**	2.09	0.0321	1.39
Mother's ASVAB (CS)[9]	0.0431	1.19	-0.0277	-1.35
Mother's ASVAB (MK)[10]	-0.0062	-0.16	-0.0008	-0.04
Constant	-0.9703**	-2.48	-0.0985	-0.45
Adjusted R Squared	0.4392		0.6348	
Observations[11]	1367		1367	

[1]The variables are standardized with mean zero and variance one across the entire CNLSY/79 sample. Let m(t) denote the dependent variable at age t. Let X(t) denote the inputs at age t. For each dependent variable we run a OLS regression: m(t) = bm(t-1) + aX(t-1) + e(t).

[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]AR stands for Arithmetic Reasoning. It is standardized with mean zero and variance one across the entire NLSY/1979 sample

[6]WK stands for Word Knowledge. It is standardized with mean zero and variance one across the entire NLSY/1979 sample

[7]PC stands for Paragraph Composition. It is standardized with mean zero and variance one across the entire NLSY/1979 sample

[8]NO stands for Numerical Operations. It is standardized with mean zero and variance one across the entire NLSY/1979 sample

[9]CS stands for Coding Speed. The variable is standardized with mean zero and variance one across the entire NLSY/1979 sample

[10]MK stands for Mathematics Knowledge. It is standardized with mean zero and variance one across the entire NLSY/1979 sample

[11]The total number of observations is given by the number of children times the number of periods we have nonmissing data for the children. On average, we have around 650 children per period. For each child, we have at most 4 periods of data.

*Statistically significant at 10%. ** Statistically significant at 5%.

Table 7
The Technology Equations - Anchored in Adult Earnings of the Child
Measurement Variables are Standardize with Mean Zero and Variance One

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.8835	0.0215	0.0790	0.0367
Current Period Cognitive Skills	0.0065	0.0046	0.9814	0.0388
Current Period Investment	0.0030	0.0010	0.0079	0.0019
Mother's Education	0.0003	0.0004	0.0007	0.0009
Mother's Ability	-0.0003	0.0003	0.0041	0.0015
Variance of Shocks	0.0003	0.0001	0.0007	0.0002

[1]Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. In table 7 we show the estimated parameter values and standard errors of γ, ψ_1, and ψ_2 as well as the $Var\left(\eta_{t+1}^N\right)$ and $Var\left(\eta_{t+1}^C\right)$.

Table 8
The Technology Equations -
We anchor the parameters on the probability of graduating from High School

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.8848	0.0231	0.0745	0.0125
Current Period Cognitive Skills	0.0007	0.0365	1.0744	0.0250
Current Period Investment	0.0073	0.0031	0.0120	0.0043
Mother's Education	0.0004	0.0016	0.0022	0.0029
Mother's Ability	-0.0016	0.0119	0.0058	0.0167

[1] Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. In table 8 we show the estimated parameter values and standard errors of γ, ψ_1, and ψ_2 as well as the $Var\left(\eta_{t+1}^N\right)$ and $Var\left(\eta_{t+1}^C\right)$. Note that while in table 7 we anchor on adult log earnings of the children, here we anchor on the probability of graduating from high school.

Table 9

The Technology Equations[1]

Estimated Parameter Values - Technology from Period 1 to Period 2

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.9862	0.0141	0.0624	0.0242
Current Period Cognitive Skills	0.0508	0.0712	0.9235	0.0912
Current Period Investment	0.0462	0.0131	0.1170	0.0152
Mother's Education	0.0035	0.0032	-0.0080	0.0059
Mother's Ability	0.0074	0.0410	0.0567	0.0413
Variance of Shocks	0.1425	0.0141	0.0646	0.0152

Estimated Parameter Values - Technology from Period 2 to Period 3

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.9405	0.0132	0.0220	0.0072
Current Period Cognitive Skills	-0.0417	0.0814	0.8871	0.0213
Current Period Investment	0.1119	0.0173	0.0397	0.0131
Mother's Education	-0.0025	0.0079	0.0124	0.0083
Mother's Ability	-0.0048	0.0132	0.0067	0.0069
Variance of Shocks	0.1286	0.0152	0.0232	0.0064

Estimated Parameter Values - Technology from Period 3 to Period 4

	Next Period Noncognitive Skills		Next Period Cognitive Skills	
	Mean	Standard Error	Mean	Standard Error
Current Period Noncognitive Skills	0.7597	0.0351	0.0021	0.0041
Current Period Cognitive Skills	0.0511	0.0254	0.9124	0.0409
Current Period Investment	0.0375	0.0144	0.0294	0.0103
Mother's Education	0.0082	0.0125	0.0036	0.0041
Mother's Ability	-0.0073	0.0183	0.0193	0.0413
Variance of Shocks	0.1574	0.0172	0.0103	0.0029

[1]The parental investment factor is normalized on "trips to the theater". Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. In table 9 we show the estimated parameter values and standard errors of γ_t, $\psi_{1,t}$, and $\psi_{2,t}$ as well as the $Var\left(\eta_{t+1}^N\right)$ and $Var\left(\eta_{t+1}^C\right)$.

Table 10
Per Period Correlation Matrices
Technology Parameters are allowed to vary over time

Period 1 - Children ages 6 and 7

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.1822	0.3263
Cognitive	0.1822	1.0000	0.2704
Investment (Home)	0.3263	0.2704	1.0000

Period 2 - Children ages 8 and 9

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2687	0.3698
Cognitive	0.2687	1.0000	0.3684
Investment (Home)	0.3698	0.3684	1.0000

Period 3 - Children ages 10 and 11

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2832	0.4094
Cognitive	0.2832	1.0000	0.4279
Investment (Home)	0.4094	0.4279	1.0000

Period 4 - Children ages 12 and 13

	Noncognitive	Cognitive	Investment (Home)
Noncognitive	1.0000	0.2858	0.4647
Cognitive	0.2858	1.0000	0.5653
Investment (Home)	0.4647	0.5653	1.0000

[1]Let $Y_t^N = \left(Y_{1,t}^N, ..., Y_{m_t^N,t}^N\right)'$ denote the measurements of noncognitive skills. Let $Y_t^C = \left(Y_{1,t}^C, ..., Y_{m_t^C,t}^C\right)'$ denote the measurements of cognitive skills. Let $X_t = \left(X_{1,t}, ..., X_{m_t^I,t}\right)'$ denote the measurements of parental investment (from the HOME-SF score. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. The measurement equations are:

$$Y_t = \alpha_t \theta_t + \varepsilon_t$$

and the technology equations are:

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations and the error terms in the technology equations. Suppose that the initial distribution of the dynamic factors is $\theta_1 \sim N\left(a_1, P_1\right)$. In table 10, we show the covariance matrix P_t for differente ages. To see how this matrix is updated, let G contain the matrix of the technology parameters γ. Let H_t denote the covariance-variance matrix of the uniquenesses ε_t. Let Q_t denote the variance-covariance matrix of the error terms in the technology equations, η_{t+1}. This matrix is updated by the standard Kalman rule:

$$P_{t+1} = G_t P_t G_t' - G_t P_t \alpha_t' \alpha_t P_t \left(\alpha_t P_t \alpha_t' + H_t\right)^{-1} G_t' P_t$$

Table 11
The Weights in the Construction of the Investment Factor

	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	0.3079	0.1667	0.1242	0.8758	0.2971	0.1667	0.0943	0.9057
Musical Instrument	0.1997	0.1667	0.1417	0.8583	0.1667	0.1667	0.1312	0.8688
Newspaper	0.1932	0.1667	0.1517	0.8483	0.2148	0.1667	0.0766	0.9234
Child has special lessons	0.1431	0.1667	0.2808	0.7192	0.1560	0.1667	0.1801	0.8199
Child goes to museums	0.0740	0.1667	0.3063	0.6937	0.0768	0.1667	0.2158	0.7842
Child goes to theater	0.0821	0.1667	0.3068	0.6932	0.0821	0.1667	0.2473	0.7527
	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	0.1396	0.1667	0.1056	0.8944	0.1380	0.1667	0.0568	0.9432
Musical Instrument	0.1707	0.1667	0.1196	0.8804	0.2001	0.1667	0.0500	0.9500
Newspaper	0.2082	0.1667	0.0754	0.9246	0.2253	0.1667	0.0370	0.9630
Child has special lessons	0.1971	0.1667	0.1074	0.8926	0.2396	0.1667	0.0457	0.9543
Child goes to museums	0.0866	0.1667	0.1693	0.8307	0.0910	0.1667	0.1011	0.8989
Child goes to theater	0.0821	0.1667	0.2074	0.7926	0.0821	0.1667	0.1412	0.8588

[1]Let Y_t^N, Y_t^C, and X_t denote the measurements on noncognitive skills, cognitive skills and parental investment (from the HOME-SF score), respectively. Let $Y_t = \left(Y_t^N, Y_t^C, X_t\right)$. Let $\theta = \left(\theta_t^N, \theta_t^C, I_t\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let S denote mother's education and A denote mother's cognitive ability. We estimate the model

$$Y_t = \alpha_t \theta_t + \varepsilon_t \tag{1}$$

$$\theta_{t+1} = \gamma_t \theta_t + \psi_{1,t} S + \psi_{2,t} A + \eta_{t+1} \tag{2}$$

where α_t is the factor-loading matrix, γ_t is the technology-parameters matrix, $\psi_{k,t}$ are parameter vectors. The vectors $\varepsilon_t, \eta_{t+1}$ contain the uniquenesses of the measurement equations (1) and the error terms in the technology equations (2).

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

[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 $\sigma_{k,t}^2 = \left(\alpha_{k,t}^I\right)^2 \sigma_{I_t}^2 + \sigma_{\varepsilon_{k,t}}^2$, where $\sigma_{\varepsilon_{k,t}}^2$ is the variance of the uniqueness in measurement k at period t. 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 $s_{\varepsilon_{k,t}} = \frac{\sigma_{\varepsilon_{k,t}}^2}{\sigma_{k,t}^2}$.