Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation

Flavio Cunha

James Heckman*

Abstract

This paper estimates models of the evolution of cognitive and noncognitive skills and explores the role of family environments in shaping these skills at different stages of the life cycle of the child. Central to this analysis is identification of the technology of 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 generally more effective in raising noncognitive skills. Noncognitive skills promote the formation of cognitive skills but, in most specifications of our model, cognitive skills do not promote the formation of noncognitive skills. 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.

I. Introduction

The importance of cognitive skills in explaining socioeconomic success is now firmly established. An emerging body of empirical research documents the importance of noncognitive skills for predicting wages, schooling and participation in risky behaviors.[1] 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 outcomes.

There is a substantial body of empirical research on the determinants of cognitive test scores and their growth.[2] There is no previous research on the determinants of the evolution of noncognitive skills. This paper identifies and estimates models of the technology of skill formation. Building on the theoretical analyses of Cunha and Heckman (2007) and Cunha, Heckman, Lochner, and Masterov (2006), we estimate the joint evolution of cognitive and noncognitive skills over the life cycle of children.

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.[3] It also formalizes the notion that cognitive skills can 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.

Psychologists who study child development 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; Zigler and Butterfield, 1968). 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 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, are first steps toward providing a unified treatment of the early intervention and family influence literatures.

The conventional approach to estimating cognitive production functions is best exemplified by the research of Todd and Wolpin (2003; 2005). A central problem with the production function approach is accounting for the endogeneity of inputs. Another problem is the wealth of candidate parental input measures available in many data sets. 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. The constructed indices often have an *ad hoc* quality about them and may be poor proxies for the true combination of inputs that enter the technology.

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 analyses 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 the model of Cunha and Heckman (2007). 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 recurring problem in the literature on cognitive production functions. Studies in this tradition typically use a test score as a measure of output (see, for example, 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 also be used. Thus, we anchor the scale of the latent factors that generate test scores by determining how the factors predict adult outcomes.[4] This 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 mothers’ ability affect the formation of cognitive skills but not noncognitive skills. (3) Parental inputs appear to affect cognitive skill formation more strongly at earlier ages. They affect noncognitive skill formation more strongly at later ages. Ages where parental inputs have higher marginal productivity, holding all inputs constant, are called "sensitive" periods. The sensitive periods for cognitive skills occur earlier in the life cycle of the child than do sensitive periods for noncognitive skills. Our evidence is consistent with the evidence presented in Carneiro and Heckman (2003) that noncognitive skills are more malleable at later ages than cognitive skills. See also the evidence in Heckman (2007) and in Borghans et al. (2007). We also find that (4) measurement error in inputs is substantial and that correcting for measurement error greatly affects our estimates.

The plan of this paper is as follows. Section II briefly summarizes our previous research on models of skill formation. Section III presents our analysis of identification using dynamic factor models. Section IV discusses our empirical findings. Section V concludes. We use a technical appendix to present our likelihood function. A website provides supporting material.[5]

II. A Model of Cognitive and Noncognitive Skill Formation

Cunha and Heckman (2007) analyze multiperiod models of childhood skill formation followed by a period of adulthood.[6] They extend the model of Becker and Tomes (1986), who assume childhood lasts one period, and that investment inputs at different stages of the lifecycle of a child are perfect substitutes and are equally productive. Becker and Tomes do not distinguish cognitive from noncognitive skills. Cunha and Heckman (2007) analyze models with two kinds of skills: θ^C and θ^N, where θ^C is cognitive skill and θ^N is noncognitive skill.

Let $\theta^I_{k,t}$ denote parental investments in child skill k in period t, $k \in \{C, N\}$ and $t \in \{1, \ldots, T\}$, where T is the number of periods of childhood. Let h be the level of human capital as the child starts adulthood which depends on both θ_{T+1}^{C} and θ_{T+1}^{N}. The parents fully control the investment in the child. A better model would incorporate 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 leave the development of that model for another occasion.

Assume that each agent is born with initial conditions $\theta_1 = \left(\theta_1^C, \theta_1^N\right)$. 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 1×2 vector of skill or ability stocks. The technology of production of skill k in period t is

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

for $k \in \{C, N\}$ and $t \in \{1, \ldots, T\}$.[7] 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 is an argument of (1). Cunha and Heckman (2007) summarize the evidence in economics and psychology about the interaction between cognitive and noncognitive skills in the production of human capital.

Adult human capital h is a combination of period T+1 skills accumulated by the end of childhood:

(2) $$h = g\left(\theta_{T+1}^{C}, \theta_{T+1}^{N}\right).$$

The function g is assumed to be continuously differentiable and increasing in θ_{T+1}^{C} and θ_{T+1}^{N}. This specification of human capital assumes that there is no comparative advantage in the labor market or in other areas of social performance.[8]

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 Child-Parent Center programs promote greater efficiency in learning in schools and reduce problem behaviors. See Blau and Currie (2006), Cunha and Heckman (2007), Cunha et al. (2006), and 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 et al. (2006) for a review of the literature. Emotionally nurturing early environments producing motivation and self-discipline 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 dynamic complementarity.

Using technology (1), Cunha and Heckman (2007) define critical and sensitive periods for investment. At some ages, and for certain skills, parental investment may be more productive than in other periods. Such periods are “sensitive” periods. If investment is productive only in a single period, it is a “critical” period for that investment.

Cunha and Heckman (2007) discuss the role of complementarity in investments. If early investments are complementary with later investments, then low early investments, associated with disadvantaged childhoods, make later investments less productive. High early investments have a multiplier effect in making later investments more productive. If investment inputs are not perfect substitutes but instead are complements, government investment in the early years for disadvantaged children promotes investment in the later years.

Cunha and Heckman (2007) show that 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 later period interventions for a child from a disadvantaged environment depend critically on the nature of the technology of skill production. If early and late investments are perfect complements, on efficiency grounds a low early investment should be followed up by low later investments.

If inputs are perfect substitutes, later interventions can, in principle, eliminate initial skill deficits. At a sufficiently high level of later-period investment, it is technically possible to offset low initial investments. However, it may not be cost effective to do so. Cunha and Heckman (2007) give exact conditions for no investment to be an efficient outcome in this case. Under those conditions, it would be more efficient to give children bonds that earn interest, rather than invest in their human capital in order to raise their incomes.

The key to understanding optimal investment in children is to understand the technology and market environment in which agents operate. This paper focuses on identifying and estimating the technology of skill formation, which is a vital ingredient for designing skill formation policies, and evaluating their performance.

III. 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, and measurement error is likely to be a serious problem. In addition, the inputs are endogenous because parents choose them.

General nonlinear specifications of technology (1) raise additional problems regarding measurement error in latent variables in nonlinear systems (see Schennach, 2004). This paper estimates linear specifications of (1). A more general nonlinear analysis requires addressing additional econometric and computational considerations, which are addressed in Cunha, Heckman, and Schennach (2007).

A. *Identifying A Linear Technology*

Using a 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 cross effects in a simpler setting, it is unlikely that a more general nonlinear model will overturn these results. Identifying a 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) survey this literature. 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.[9] In their paper, they use a scalar measure of cognitive ability $\left(\theta_{t+1}^{C}\right)$ in period t+1 that depends on period t cognitive ability $\left(\theta_{t}^{C}\right)$ and investment. We denote investment by θ_{t}^{I} in this and remaining sections, rather than $\theta_{k,t}^{I}$, as in the preceding section. This notation reflects the fact that we cannot empirically distinguish between investment in cognitive skills and investment in noncognitive skills. Todd and Wolpin assume a linear-in-parameters technology

$$\theta_{t+1}^{C} = a_t \theta_t^{C} + b_t \theta_t^{I} + \eta_t, \tag{3}$$

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 to proxy parental investment.[10] We use a version of the inputs into the home score as well, but in a different way than they do.

Todd and Wolpin (2003; 2005) discuss problems arising from endogenous inputs $\left(\theta_t^C, \theta_t^I\right)$ that depend on unobservable η_t. In their 2005 paper, they use IV methods coupled with fixed effect methods.[11] Reliance on IV is problematic because of the ever-present controversy about the validity of exclusion restrictions. 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.[12] The CNLSY data used by Todd and Wolpin (2005) and in this paper have a multiplicity of investment measures subsumed in a "home score" measure which combines many diverse parental input measures into a score that weights all components equally.[13] As we note below, use of arbitrary aggregates calls into question the validity of instrumental variable estimation strategies for inputs.

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 their analysis. Test scores are arbitrarily normed. Any monotonic function of a 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, for example, Carneiro and Heckman, 2003; and 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 in this paper by using adult outcomes to anchor the scale of the test score. Cunha, Heckman, and Schennach (2007) address this problem in a more general way for arbitrary monotonically increasing transformations of the factors. In this paper, we develop an interpretable scale for θ_t^C, θ_t^N that is robust to all affine transformations of the units in which factors $\left(\theta_t^C, \theta_t^N\right)$ are measured. For example, using adult earnings Y as the anchor, we write

(4) $$\ln Y = \mu + \delta^C \theta_{T+1}^C + \delta^N \theta_{T+1}^N + \varepsilon,$$

where the scales of θ_{T+1}^C and θ_{T+1}^C are unknown. For any affine transformation of θ_{T+1}^k, corresponding to different units of measuring the factors, the value of δ^k and the intercept adjust and we can uniquely identify the left hand side of

(5) $$\frac{\partial \ln Y}{\partial \theta_t^I} = \delta^k \left(\frac{\partial \theta_{T+1}^k}{\partial \theta_t^I}\right) \text{for } k \in \{C, N\};\ t \in \{1, \ldots, T\}$$

for any scale of θ_t^I. Thus, although the scale of δ^k is not uniquely determined, nor is the scale of θ_{T+1}^k, the scale of $\delta^k \theta_{T+1}^k$ is uniquely determined by its effect on log earnings and we can define the effects of all inputs on ln$Y$ relative to their effects on earnings.

The scale for measuring investment θ_t^I is also arbitrary. We report results for alternative normalizations of the units of investment. Natural scales are in dollars or log dollars. Using elasticities,

$$\left(\frac{\partial \ln Y}{\partial \theta_t^I}\right)\theta_t^I = \left(\delta^k \frac{\partial \theta_{T+1}^k}{\partial \theta_t^I}\right)\theta_t^I$$

produces parameters that are invariant to linear transformations of the units in which investment is measured. This approach generalizes to multiple factors and multiple anchors and we apply it in this paper. We now develop our empirical approach to identifying and estimating the technology of skill formation.

B. 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

$$\begin{pmatrix} \theta_{t+1}^{N} \\ \theta_{t+1}^{C} \end{pmatrix} = \begin{pmatrix} \gamma_1^N & \gamma_2^N \\ \gamma_1^C & \gamma_2^C \end{pmatrix} \begin{pmatrix} \theta_t^N \\ \theta_t^C \end{pmatrix} + \begin{pmatrix} \gamma_3^N \\ \gamma_3^C \end{pmatrix} \theta_t^I + \begin{pmatrix} \eta_t^N \\ \eta_t^C \end{pmatrix}, \tag{6}$$

where θ_t^I can be a vector. (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 $\theta_t' = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ by vectors of measurements on skills which can include test scores as well as outcome measures.[14] In our analysis, test scores and parental inputs are indicators of the latent skills and latent investments. We account for measurement errors in output and input variables. We find substantial measurement errors in the proxies for parental investment and in the proxies for cognitive and noncognitive skills. (4) Instead of imposing a particular index of parental input based on components of the home score, we estimate an index that best fits the data. (5) Instead of relying solely on exclusion restrictions to generate instruments to correct for measurement error in the proxies for θ_t, and for endogeneity, we use covariance restrictions that exploit a feature of our data that there are many more measurements on θ_{t+1} and θ_t than the number of latent factors. This allows us to secure identification from cross equation restrictions using multiple indicator-multiple cause (MIMIC) (Jöreskog and Goldberger, 1975) and linear structural relationship (LISREL) (Jöreskog, Sörbom, and Magidson, 1979) models.[15] When instruments are needed, they arise from the internal logic of the model developed in Cunha et al. (2006) and Cunha and Heckman (2007), using methods developed by Madansky (1964) and Pudney (1982). (6) Instead of relying on test scores as measures of output and change in output due to parental investments, we anchor the scale of the test scores using adult outcome measures: log earnings and the probability of high school graduation. We thus estimate the effect of parental investments on the adult earnings of the child and on the probability of high school graduation.

C. A Model for the Measurements

We assume access to measurement systems that can be represented by a dynamic factor structure:

$$Y_{j,t}^k = \mu_{j,t}^k + \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k, \text{ for } j \in \{1, \ldots, m_t^k\}, \; k \in \{C, N, I\},$$

where m_t^k is the number of measurements on cognitive skills, noncognitive skills, and investments in period *t*; and where θ_t^k is a dynamic factor for component *k*, $k \in \{C, N, I\}$. We account for latent initial conditions of the process, $\left(\theta_1^C, \theta_1^N\right)$, which correspond to endowment of abilities. Because we have multiple measurements of abilities in the first period of our data, we can also identify the distribution of the latent initial conditions. We also identify the distribution of each $\theta_t = \left(\theta_t^C, \theta_t^N, \theta_t^I\right)$, as well as the dependence across θ_t and $\theta_{t'}$, $t \neq t'$.

As above, let θ_t^C denote the stock of cognitive skill of the agent in period *t*. We do not observe θ_t^C directly. Instead, we observe a vector of measurements, such as test scores, $Y_{j,t}^C$, for $j \in \{1, 2, \ldots, m_t^C\}$. Assume that:

(7) $$Y_{j,t}^C = \mu_{j,t}^C + \alpha_{j,t}^C \theta_t^C + \varepsilon_{j,t}^C \text{ for } j \in \{1, 2, \ldots, m_t^C\}$$

and set $\alpha_{1,t}^C = 1$. Some normalization is needed to set the scale of the factors. The $\mu_{j,t}^C$ may depend on regressors.

We have a similar equation for noncognitive skills at age *t*, relating θ_t^N to proxies for it:

(8) $$Y_{j,t}^N = \mu_{j,t}^N + \alpha_{j,t}^N \theta_t^N + \varepsilon_{j,t}^N \text{ for } j \in \{1,\ldots,m_t^N\}$$

and we normalize $\alpha_{1,t}^N = 1$. Finally, we model the measurement equations for investments, θ_t^I:

(9) $$Y_{j,t}^I = \mu_{j,t}^I + \alpha_{j,t}^I \theta_t^I + \varepsilon_{j,t}^I \text{ for } j \in \{1,\ldots,m_t^I\}$$

and the factor loading $\alpha_{1,t}^I = 1$. The ε's are measurement errors that account for the fallibility of our measures of latent skills and investments.[16]

We analyze a linear law of motion for skills:

(10) $$\theta_{t+1}^k = \gamma_0^k + \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \eta_t^k \text{ for } k \in \{C,N\} \text{ and } t \in \{1,\ldots,T\},$$

where the error term 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 assume that the η_t^k, $k \in \{C,N\}$, are independent of $\left(\theta_1^C, \theta_1^N\right)$. Below, we show how to relax the independence assumption and allow for unobserved inputs.

We allow the components of θ_t to be freely correlated for any *t* and with any vector $\theta_{t'}$, $t' \neq t$, and we can identify this dependence. We assume that any variables in the $\mu_{j,t}^k$ are independent of θ_t for $k \in \{C,N,I\}$ and $t \in \{1,\ldots,T\}$. We now establish conditions under which the technology parameters are identified.

D. Semiparametric Identification

The goal of the analysis is to recover the joint distribution of $\left\{\theta_t^C, \theta_t^N, \theta_t^I\right\}_{t=1}^T$, the distributions of $\left\{\eta_t^k\right\}_{t=1}^T$ and $\left\{\varepsilon_{j,t}^k\right\}_{j=1}^{m_t^k}$ nonparametrically, as well as the parameters $\left\{\alpha_{j,t}^k\right\}_{j=1,t=1}^{m_t^k}, \left\{\gamma_{j,t}^k\right\}_{j=1}^3$ for

$k \in \{C, N, I\}$, and for $t \in \{1, \ldots, T\}$. Identification of the means of the measurements is straightforward under our assumptions.[17]

1. Classical Measurement Error for the Case of Two Measurements Per Latent Factor: $m_t^C = m_t^N = m_t^I = 2$

We make the following assumptions about the $\varepsilon_{j,t}^k$:

Assumption 1 $\varepsilon_{j,t}^k$ *is mean zero and independent across agents and over time for* $t \in \{1,...,T\}$; $j \in \{1,2\}$; and $k \in \{C, N, I\}$;

Assumption 2 $\varepsilon_{j,t}^k$ *is mean zero and independent of* $\left(\theta_\tau^C, \theta_\tau^N, \theta_\tau^I\right)$ *for all* $t, \tau \in \{1,...,T\}$; $j \in \{1,2\}$; and $k \in \{C, N, I\}$;

Assumption 3 $\varepsilon_{j,t}^k$ *is mean zero and independent from* $\varepsilon_{j,t}^l$ *for* $i, j \in \{1,2\}$ *and i≠j;* $k, l \in \{C, N, I\}$ $t \in \{1,...,T\}$.

Identification of the Factor Loadings.

Since we observe $\left\{\left[Y_{j,t}^k\right]_{j=1}^2\right\}_{t=1}^T$ for every person, we can compute $\text{Cov}\left(Y_{1,t}^k, Y_{2,\tau}^l\right)$ from the data for all *t*,τ and *k,l* pairs, where $t, \tau \in \{1,...,T\}$; $k, l \in \{C, N, I\}$. Consider, for example, measurements on cognitive skills. Recall that $\alpha_{1,t}^C = 1$. We know the left hand side of each of the following equations:

$$\text{Cov}\left(Y_{1,t}^C, Y_{1,t+1}^C\right) = \text{Cov}\left(\theta_t^C, \theta_{t+1}^C\right), \tag{11}$$

$$\text{Cov}\left(Y_{2,t}^C, Y_{1,t}^C\right) = \alpha_{2,t}^C \text{Cov}\left(\theta_t^C, \theta_{t+1}^C\right), \tag{12}$$

(13) $$\operatorname{Cov}\left(Y_{1,t}^{C}, Y_{2,t+1}^{C}\right) = \alpha_{2,t+1}^{C} \operatorname{Cov}\left(\theta_{t}^{C}, \theta_{t+1}^{C}\right).$$

We can identify $\alpha_{2,t}^{C}$ by taking the ratio of (12) to (11) and $\alpha_{2,t+1}^{C}$ from the ratio (13) to (11).

Proceeding in the same fashion, we can identify $\alpha_{j,t}^{k}$ for $t \in \{1,\ldots,T\}$ and $j \in \{1,2\}$, up to the normalizations $\alpha_{1,t}^{k} = 1$, $k \in \{C,N,I\}$.

The Identification of the Joint Distribution of $\left\{\left(\theta_{t}^{C}, \theta_{t}^{N}, \theta_{t}^{I}\right)\right\}_{t=1}^{T}$.

Once the parameters $\alpha_{1,t}^{k}$ and $\alpha_{2,t}^{k}$ are identified (up to the normalization $\alpha_{1,t}^{k} = 1$), we can rewrite (7), (8), and (9) as

$$\frac{Y_{j,t}^{k}}{\alpha_{j,t}^{k}} = \frac{\mu_{j,t}^{k}}{\alpha_{j,t}^{k}} + \theta_{t}^{k} + \frac{\varepsilon_{j,t}^{k}}{\alpha_{j,t}^{k}},\ j \in \{1,2\} \text{ for } \alpha_{j,t}^{k} \neq 0,\ k \in \{C,N,I\};\ t \in \{1,\ldots,T\}.$$ [18]

Now, define

$$Y_{j} = \left\{\left(\frac{Y_{j,t}^{C}}{\alpha_{j,t}^{C}}, \frac{Y_{j,t}^{N}}{\alpha_{j,t}^{N}}, \frac{Y_{j,t}^{I}}{\alpha_{j,t}^{I}}\right)\right\}_{t=1}^{T} \quad \text{for } j = 1,2.$$

Similarly, define

$$\varepsilon_{j} = \left\{\left(\frac{\varepsilon_{j,t}^{C}}{\alpha_{j,t}^{C}}, \frac{\varepsilon_{j,t}^{N}}{\alpha_{j,t}^{N}}, \frac{\varepsilon_{j,t}^{I}}{\alpha_{j,t}^{I}}\right)\right\}_{t=1}^{T} \quad \text{for } j = 1,2,$$

and

$$\mu_{j} = \left\{\left(\frac{\mu_{j,t}^{C}}{\alpha_{j,t}^{C}}, \frac{\mu_{j,t}^{N}}{\alpha_{j,t}^{N}}, \frac{\mu_{j,t}^{I}}{\alpha_{j,t}^{I}}\right)\right\}_{t=1}^{T} \quad \text{for } j = 1,2.$$

Let θ denote the vector of all factors in all time periods:

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

We rewrite the measurement equations as

$$Y_1 = \mu_1 + \theta + \varepsilon_1,$$

$$Y_2 = \mu_2 + \theta + \varepsilon_2.$$

Under the assumption that measurement error is classical, we can apply Kotlarski's Theorem (Kotlarski, 1967) and identify the joint distribution of θ as well as the distributions of ε_1 and ε_2. Since $\alpha_{j,t}^k$ is identified, it is possible to recover the distribution of $\varepsilon_{j,t}^k$ for

$j \in \{1, 2, \ldots, m_t^k\};\ k \in \{C, N, I\}$ and $t \in \{1, 2, \ldots, T\}$.

Example 1 *Suppose that* $\theta \sim N\left(0, \Sigma\right),\ \varepsilon_{j,t}^k \sim \left(0, \sigma_{k,j,t}^2\right)$. *We observe the vectors Y_1 and Y_2, μ_1 and μ_2 are identified and the Y_1 and Y_2 can be adjusted accordingly. As previously established, we can identify the factor loadings* $\alpha_{j,t}^k$ *by taking the ratio of covariances such as* (12) *to* (11). *To identify the distribution of the factors, we need to identify the variance-covariance matrix* Σ. *We can compute the variance of the factor* θ_t^k *from the covariance between* $Y_{1,t}^k$ *and* $Y_{1,t}^k$*:*

$$\text{Cov}\left(Y_{1,t}^2, Y_{2,t}^k\right) = \alpha_{2,t}^k \text{Var}\left(\theta_t^k\right)\ for\ k \in \{C, N, I\}\ .$$

Recall that $\alpha_{2,t}^k$ is identified and the covariance on the left hand side can be formed from the data. The covariance of any two elements of θ can be computed from the corresponding moments:

(14) $$\text{Cov}\left(Y_{1,t}^k, Y_{1,\tau}^l\right) = \text{Cov}\left(\theta_t^k, \theta_\tau^l\right) \text{ for } k, l \in \{C, N, I\} \text{ and } t, \tau \in \{1, \ldots, T\},$$

and

(15) $$\text{Cov}\left(Y_{j,t}^k, Y_{k,\tau}^l\right) = \alpha_{j,t}^k \alpha_{k,\tau}^l \text{Cov}\left(\theta_t^k, \theta_\tau^l\right),$$

where the coefficients $\alpha^{k}_{j,t}, \alpha^{l}_{k,\tau}$ *are known by the previous argument. Since we know* $\operatorname{Var}\left(Y^{k}_{j,t}\right), \left(\alpha^{k}_{j,t}\right)$ *and* $\operatorname{Var}\left(\theta^{k}_{j,t}\right)$*, we can identify* $\sigma^{2}_{k,j,t}$ *from these ingredients:*

$$\operatorname{Var}\left(Y^{k}_{j,t}\right) - \left(\alpha^{k}_{j,t}\right)^{2} \operatorname{Var}\left(\theta^{k}_{j,t}\right) = \sigma^{2}_{k,j,t},\ k \in \{C, N, I\},\ t \in \{1, \ldots, T\}.$$

The Identification of the Technology Parameters Assuming Independence of η.

Assume that η^{k}_{t} is independent of $\left(\theta^{C}_{t}, \theta^{N}_{t}, \theta^{I}_{t}\right)$. Consider, for example, the law of motion for noncognitive skills,

$$\theta^{N}_{t+1} = \gamma^{N}_{0} + \gamma^{N}_{1}\theta^{N}_{t} + \gamma^{N}_{2}\theta^{C}_{t} + \gamma^{N}_{3}\theta^{I}_{t} + \eta^{N}_{t} \text{ for } t \in \{1, \ldots, T\}. \tag{16}$$

Assume that η^{N}_{t} is serially independent but possibly correlated with η^{C}_{t}. We substitute the measurement equations $Y^{N}_{1,t+1}, Y^{N}_{1,t}, Y^{C}_{1,t}, Y^{I}_{t}$ for $\theta^{N}_{t+1}, \theta^{N}_{t}, \theta^{C}_{t}, \theta^{I}_{t}$, respectively:

$$Y^{N}_{1,t+1} = \gamma^{N}_{0} + \gamma^{N}_{1}Y^{N}_{1,t} + \gamma^{N}_{2}Y^{C}_{1,t} + \gamma^{N}_{3}Y^{I}_{1,t} + \left(\varepsilon^{N}_{1,t+1} - \gamma^{N}_{1,t}\varepsilon^{N}_{1,t} - \gamma^{N}_{2,t}\varepsilon^{C}_{1,t} - \gamma^{N}_{3,t}\varepsilon^{I}_{1,t} + \eta^{N}_{t+1}\right). \tag{17}$$

If we estimate (17) by least squares, we do not obtain consistent estimators of γ^{N}_{k} for $k \in \{1, 2, 3\}$ because the regressors $Y^{N}_{1,t}, Y^{C}_{1,t}, Y^{I}_{1,t}$ are correlated with the error term ω_{t+1}, where

$$\omega_{t+1} = \varepsilon^{N}_{1,t+1} - \gamma^{N}_{1,t}\varepsilon^{N}_{1,t} - \gamma^{N}_{2,t}\varepsilon^{C}_{1,t} - \gamma^{N}_{3,t}\varepsilon^{I}_{1,t} + \eta^{N}_{t+1}.$$

However, we can instrument $Y^{N}_{1,t}, Y^{C}_{1,t}, Y^{I}_{1,t}$, using $Y^{N}_{2,t}, Y^{C}_{2,t}, Y^{I}_{2,t}$ as instruments by applying two-stage-least squares to recover the parameters γ^{N}_{k} for k=1,2,3. See Madansky (1964) or Pudney (1982) for the precise conditions on the factor loadings. Our instruments are “internal instruments” justified by the model. The justification follows from the investment equation in Cunha and Heckman (2007) that relates investment in one period to investment in other periods.[19] The suggested instruments are also independent of η^{N}_{t} because of the assumed lack of serial correlation in η_t^N.[20] We can repeat the argument for different time periods. In this way, we can identify stage-specific technologies for each stage of the child's life cycle. We can perform a parallel analysis for the cognitive skill equation.

2. Non-classical Measurement Error

We can replace Assumption 3 with the following assumption and still obtain full identification of the model.

Assumption 4 *$\varepsilon_{1,t}^k$ is independent of $\varepsilon_{j,\tau}^l$ for $j \in \{2,\dots,m_t^k\}$; $k,l \in \{C,N,I\}$ and $t,\tau \in \{1,2,\dots,T\}$, $m_t^k \geq 2$. Otherwise the $\varepsilon_{j,\tau}^l$, for $j \in \{2,\dots,m_t^k\}$; $k,l \in \{C,N,I\}$ and $t,\tau \in \{1,2,\dots,T\}$ can be arbitrarily dependent.*

The proof of identification is as follows. Let $Y_{j,t}^k = \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k$, for $j \in \{1,\dots,m_t^k\}$; $t \in \{1,\dots,T\}$ and $k \in \{C,N,I\}$. Normalize $\alpha_{1,t}^k = 1$ for all $k \in \{C,N,I\}$ and $t \in \{1,\dots,T\}$. Within a k system, for a fixed t, we can compute $\operatorname{Cov}\left(Y_{j,t}^k, Y_{1,t}^k\right) = \alpha_{j,t}^k \operatorname{Var}\left(\theta_t^k\right)$, $j \in \{1,\dots,m_t^k\}$. For temporally adjacent systems, we can compute

$$
\begin{aligned}
&\operatorname{Cov}\left(Y_{1,t-1}^k, Y_{1,t}^k\right) = \operatorname{Cov}\left(\theta_{t-1}^k, \theta_t^k\right), \\
&\operatorname{Cov}\left(Y_{1,t-1}^k, Y_{j,t}^k\right) = \alpha_{j,t}^k \operatorname{Cov}\left(\theta_{t-1}^k, \theta_t^k\right),\ j \in \{2,\dots,m_t^k\}.
\end{aligned} \tag{18}
$$

Hence we can identify $\alpha_{j,t}^k$, $j \in \{1,\dots,m_t^k\}$; $t \in \{1,\dots,T\}$; and $k \in \{C,N,I\}$ and thus $\operatorname{Var}\left(\theta_t^k\right)$, $t \in \{1,\dots,T\}$; $k \in \{C,N,I\}$. With these ingredients in hand, we can identify $\operatorname{Var}\left(\varepsilon_{j,t}^k\right)$, $t \in \{1,\dots,T\}$, as well as

$$\operatorname{Cov}\left(\varepsilon_{j,t}^k, \varepsilon_{j',t}^k\right) = \operatorname{Cov}\left(Y_{j,t}^k, Y_{j',t}^k\right) - \alpha_{j,t}^k \alpha_{j',t}^k \operatorname{Var}\left(\theta_t^k\right),$$

since we know every ingredient on the right hand side of the preceding equation. By a similar argument, we can identify

(19) $$\mathrm{Cov}\left(\varepsilon_{j,t}^{k},\varepsilon_{j',\tau}^{l}\right)=\mathrm{Cov}\left(Y_{j,t}^{k},Y_{j',\tau}^{l}\right)-\alpha_{j,t}^{k}\alpha_{j',\tau}^{l}\mathrm{Cov}\left(\theta_{t}^{k},\theta_{\tau}^{l}\right).$$

We can rewrite the measurement equations as a system:

$$\frac{Y_{j,t}^{k}}{\alpha_{j,t}^{k}}=\frac{\mu_{j,t}^{k}}{\alpha_{j,t}^{k}}+\theta_{t}^{k}+\frac{\varepsilon_{j,t}^{k}}{\alpha_{j,t}^{k}},\ j\in\{1,\ldots,m_{t}^{k}\};\ t\in\{1,\ldots,T\};\ k\in\{C,N,I\}.$$

Applying Schennach (2004), we can identify the joint distribution of

$(\theta_1^C,\ldots,\theta_T^C,\theta_1^N\ldots,\theta_T^N,\theta_1^I,\ldots,\theta_T^I)$ as well as the joint distribution of $\{\varepsilon_{j,t}^k\}$, $j\in\{1,\ldots,m_t^k\}$;

$t\in\{1,\ldots,T\}$ and $k\in\{C,N,I\}$ using multivariate deconvolution.

Example 2 *Assume access to three measurements for cognitive, non-cognitive, and investment factors, respectively. Suppose that* $\theta=\left(\theta^C,\theta^N,\theta^I\right)\sim N\left(0,\Sigma\right),\varepsilon_{1,t}^k\sim N\left(0,\sigma_{k,1,t}^2\right)$*, but* $(\varepsilon_2,\varepsilon_3)\sim N(0,\Omega)$*, where* Ω *need not be diagonal. As discussed above, we identify* $\mathrm{Var}(\theta_t^k)$*,* $k\in\{C,N,I\}$*. Again, any element of the variance-covariance matrix* Σ *is obtained from* (14)*. Furthermore, any element of the matrix* Ω *can be obtained from* (19)*. Finally, we can identify* $\sigma_{k,l,t}^2$ *from* $\mathrm{Var}(Y_{l,t}^k)$.

For this more general measurement error system, we can identify stage-specific technologies using the same proof structure as was used for the case with classical measurement error.

3. The Identification of the Technology with Correlated Omitted Inputs.

It is unrealistic to assume that omitted inputs are serially independent. Fortunately, we can relax this assumption. Assume now that η_t^k is *not* independent of $\theta_t=(\theta_t^C,\theta_t^N,\theta_t^I)$ Consider a model in which η_t^k can be decomposed into two parts:

$$\eta_t^N = \lambda + \nu_{t+1}^N \text{ and } \eta_t^C = \gamma_4^C \lambda + \nu_{t+1}^C$$

so that the equations of motion can be written as

$$\theta_{t+1}^N + \gamma_0^N + \gamma_1^N \theta_t^N + \gamma_2^N \theta_t^C + \gamma_3^N \theta_t^I + \lambda + \nu_{t+1}^N, \tag{20}$$

$$\theta_{t+1}^C + \gamma_0^C + \gamma_1^C \theta_t^N + \gamma_2^C \theta_t^C + \gamma_3^C \theta_t^I + \gamma_4^C \lambda + \nu_{t+1}^C. \tag{21}$$

The term λ is a time-invariant input permitted to be freely correlated with θ_t . We allow λ to have a different impact on cognitive and non-cognitive skill accumulation. Let $\nu_t = (\nu_t^N, \nu_t^C)$. We make the following assumption.

Assumption 5 *The error term* ν_t *is independent of* θ_t , λ, ν_τ *for any* $\tau \neq t$.

Under this assumption, we can identify both a stage-invariant technology and a stage-varying technology. We first analyze the stage-invariant case. Consider, for example, the law of motion for noncognitive skills. For any periods *t*, *t*+1 we can compute the difference

$$\theta_{t+1}^N - \theta_t^N = \gamma_1^N \left(\theta_t^N - \theta_{t-1}^N\right) + \gamma_2^N \left(\theta_t^C - \theta_{t-1}^C\right) + \gamma_3^N \left(\theta_t^I - \theta_{t-1}^I\right) + \nu_{t+1}^N - \nu_t^N. \tag{22}$$

We use the measurement equations for $Y_{1,t+1}^k$ and $Y_{1,t}^k$ to proxy θ_{t+1} and θ_t :

$$\begin{aligned} Y_{1,t+1}^N - Y_{1,t}^N &= \gamma_1^N \left(Y_{1,t}^N - Y_{1,t-1}^N\right) + \gamma_2^N \left(Y_{1,t}^C - Y_{1,t-1}^C\right) + \gamma_3^N \left(Y_{1,t}^I - Y_{1,t-1}^I\right) + \nu_{t+1}^N - \nu_t^N \\ &+ \left\{\left(\varepsilon_{1,t+1}^N - \varepsilon_{1,t}^N\right) - \gamma_1^N \left(\varepsilon_{1,t}^N - \varepsilon_{1,t-1}^N\right) - \gamma_2^N \left(\varepsilon_{1,t}^C - \varepsilon_{1,t-1}^C\right) - \gamma_3^N \left(\varepsilon_{1,t}^I - \varepsilon_{1,t-1}^I\right)\right\}. \end{aligned} \tag{23}$$

OLS applied to (23) does not produce consistent estimates of γ_1^N, γ_2^N and γ_3^N because the regressors $(Y_{1,t}^k - Y_{1,t-1}^k)$ are correlated with the error term ω, where

$$\omega = \left(\varepsilon_{1,t+1}^N - \varepsilon_{1,t}^N\right) - \gamma_1^N \left(\varepsilon_{1,t}^N - \varepsilon_{1,t-1}^N\right) - \gamma_2^N \left(\varepsilon_{1,t}^C - \varepsilon_{1,t-1}^C\right) - \gamma_3^N \left(\varepsilon_{1,t}^I - \varepsilon_{1,t-1}^I\right).$$

However, we can instrument $(Y_{1,t}^{k} - Y_{1,t-1}^{k})$ using $\left\{\left(Y_{j,t-1}^{k} - Y_{j,t-2}^{k}\right)\right\}_{j=2}^{m_t^k}$ as the instruments. These instruments are valid because of the generalization of investment equation (9) in Cunha and Heckman (2007) to a T period model.[21] Using a two-stage least squares regression with these instruments allows us to recover the parameters γ_1^N, γ_2^N and γ_3^N. We can identify γ_0^N if we assume that $\mathrm{E}(\lambda) = 0$ Following a parallel argument, we can identify $\gamma_0^N, \gamma_1^N, \gamma_2^N$ and γ_3^N using the data on the evolution of cognitive test scores.

Next, define

$$\psi_{t+1}^{k} = \theta_{t+1}^{k} - \left(\gamma_0^k + \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I\right).$$

From the measurement equations, we know the joint distribution of $\left(\theta_{t+1}^{k}, \theta_t^N, \theta_t^C, \theta_t^I\right)$ for $k \in \{C, N\}$. We have established how to obtain the parameter values $\gamma_0^N, \gamma_1^N, \gamma_2^N$ and γ_3^N. Consequently, we know the distribution of ψ_t^k for $k \in \{C, N\}$ and $t \in \{1, \ldots, T\}$. We have $2T$ equations:

$$\begin{array}{ll}
\psi_T^N = \lambda + \nu_T^N & \psi_T^C = \gamma_4^C \lambda + \nu_T^C \\
\psi_{T-1}^N = \lambda + \nu_{T-1}^N & \psi_{T-1}^C = \gamma_4^C \lambda + \nu_{T-1}^C \\
\vdots & \vdots \\
\psi_1^N = \lambda + \nu_1^N & \psi_1^C = \gamma_4^C \lambda + \nu_1^C
\end{array}.$$

Under Assumption 5 we can apply Kotlarski's Theorem to this system and obtain the distribution of λ and ν_t for any t. Note that we can identify the parameter γ_4^C from the covariance:

$$\mathrm{Cov}\left(\psi_t^N, \psi_\tau^C\right) = \gamma_4^C \mathrm{Var}\left(\lambda\right)$$

since the variance of λ is known. This approach solves the problem raised by correlated omitted inputs for stage-invariant technologies.

For the stage-varying case, a similar argument applies. In place of equation (22), we can write

$$(24)\quad \theta_{t+1}^N - \theta_t^N = \gamma_{0,t}^N - \gamma_{0,t-1}^N + \gamma_{1,t}^N \theta_t^N - \gamma_{1,t-1}^N \theta_{t-1}^N + \gamma_{1,t}^N \theta_t^C - \gamma_{2,t-1}^N \theta_{t-1}^C + \gamma_{3,t}^N \theta_t^I - \gamma_{3,t-1}^N \theta_{t-1}^I + \nu_{t+1}^N - \nu_t^N .$$

Using the measurement equations for $Y_{1,t+1}^k$ and $Y_{1,t}^k$ to proxy θ_{t+1} and θ_t, we obtain

$$\begin{aligned} Y_{1,t+1}^N - Y_{1,t}^N &= \gamma_{0,t}^N - \gamma_{0,t-1}^N + \gamma_{1,t}^N Y_{1,t}^N - \gamma_{1,t-1}^N Y_{1,t-1}^N + \gamma_{2,t}^N Y_{1,t}^C - \gamma_{2,t-1}^N Y_{1,t-1}^C + \gamma_{3,t}^N Y_{1,t}^I + \gamma_{3,t-1}^N Y_{1,t-1}^I + \nu_{t+1}^N - \nu_t^N \\ &\quad + \left\{ \left(\varepsilon_{1,t+1}^N - \varepsilon_{1,t}^N \right) - \left(\gamma_{1,t}^N \varepsilon_{1,t}^N - \gamma_{1,t-1}^N \varepsilon_{1,t-1}^N \right) - \left(\gamma_{2,t}^N \varepsilon_{1,t}^C - \gamma_{2,t-1}^N \varepsilon_{1,t-1}^C \right) - \left(\gamma_{3,t}^N \varepsilon_{1,t}^I - \gamma_{3,t-1}^N \varepsilon_{1,t-1}^I \right) \right\}. \end{aligned}$$

We can instrument $Y^k_{1,t}$, $Y^k_{1,t-1}$, $k \in \{C, N, I\}$, using $\left\{ Y_{j,t-l}^k \right\}_{j=2}^{m_t^k}$, $k \in \{C, N, I\}$ and $l \geq 2$, as instruments. The validity of the instruments is based on the generalization of investment equation (9) in Cunha and Heckman (2007), discussed in our analysis of stage-invariant technologies. Thus we can identify the coefficients of (24) except for the intercepts. We can identify relative intercepts $(\gamma_{0,t}^N - \gamma_{0,t-1}^N)$, $t \in \{2, ..., T\}$. With these intercepts in hand, we can identify the remaining parameters by the preceding proof.

E. Anchoring the Factors in the Metric of Earnings

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

$$(25)\qquad \ln Y = \mu_T + \delta_N \theta_T^N + \delta_C \theta_T^C + \varepsilon.$$

Define

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

Assume $\delta_N = 0$ and $\delta_C = 0$.[22] For any given normalization of the test scores we can transform the θ_t to an earnings metric by multiplying equation system (6) by *D*:

$$D\theta'_{t+1} = \left(DA_t D^{-1}\right)\left(D\theta'_t\right) + \left(DB\right)\theta_t^I + \left(D\eta_t\right), \tag{26}$$

and work with $D\theta'_{t+1}$ and $D\theta'_t$ in place of θ_{t+1} and θ_t. The cross terms in $\left(DA_t D^{-1}\right)$ are affected by this change of units but not the self-productivity terms. The relative magnitude of θ_t^I on the outcomes can be affected by this change in scale. We can use other anchors besides earnings. 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.

IV. Estimating the Technology of Skill Formation

We use a sample of 1053 white males from the Children of the National Longitudinal Survey of Youth, 1979 (CNLSY/79) data set. Starting in 1986, the children of the NLSY/79 female respondents have been assessed every two years. The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-confidence 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 1 presents summary statistics of the measures of skill and investment used in this paper. The web appendix presents a more complete description of our data set.

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 outcomes (see for example, Todd and Wolpin, 2005).[23] 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 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 aged 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. We standardize these variables so that among other characteristics, a child who scores low on the antisocial subscore is a child who often cheats or tell lies, is cruel or mean to others, and does not feel sorry for misbehaving. A child who displays a low score on the anxious/depressed measurement is a child who experiences sudden changes in mood, feels no one loves him/her, is fearful, or feels worthless or inferior. A child with low scores on the headstrong measurement is tense, nervous, argues too much, and is disobedient at home. Children will score low on the hyperactivity subscale if they have difficulty concentrating, act without thinking, and are restless or overly active. Finally, a child will be assigned a low score on the peer problem subscore if they have problems getting along with others, are not liked 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 commonly used in research on child development. 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 test 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. They 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 instead of relying on an arbitrary index. Measured inputs that are not very informative on family investment decisions will have estimated factor loadings that are close to zero. Covariance restrictions in our model substitute for the missing instruments to secure identification.

Second, our 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 for 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. Appendix 2 presents our sample likelihood. We now present and discuss our empirical results using the CNLSY data.

A. 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 IV.A.5.

1. Estimates of Time-Invariant Technology Parameters

Using the CNLSY data, we 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, there are no omitted inputs correlated with θ_t, and the measurement error is classical. In Table 2 we report results in the scale of standardized test scores. We normalize the scale of the investment factor θ_t^I on different measures. Columns (1) and (4) show the estimated noncognitive and cognitive skill technologies, respectively, when we normalize the investment factor on family income. Columns (2) and (5) show the estimated parameters when we normalize the investment factor in "trips to the museum". Finally, in columns (3) and (6) we show the results when we normalize the factor loading in "trips to the theater". The estimated technology is robust to different normalization assumptions.[26]

Table 2 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 in one period affect the accumulation of next period cognitive skills, but cognitive skills in one period do not affect the accumulation of next period noncognitive skills; (3) the estimated parental investment factor affects noncognitive skills slightly more strongly than cognitive skills, but 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 in affecting the evolution of ability after controlling for parental investments, and mother's ability. We contrast the OLS estimates of this model (presented in Table 16) with our measurement-error corrected versions in Section IV.1.6.

The dynamic factors are statistically dependent. Table 3 shows the evolution of the correlation patterns across the dynamic factors. The correlation between cognitive and noncognitive skills is 0.18 at ages six and seven, and grows to around 0.28 at ages 12 and 13. There is a strong contemporaneous correlation among noncognitive skill and the home investment. The correlation starts off at 0.40 at ages six and seven and grows to 0.55 by ages 12 and 13. The same pattern is true for the correlation between cognitive skills and home investments. The correlation between these two variables goes from 0.38 at ages six and seven to 0.61 at ages 12 and 13.

2. Allowing for Non-Classical Measurement Error

We check the robustness of our findings by relaxing the assumption that the error terms in the measurement equations are classical. We allow the measurement errors to be freely correlated and estimate their dependence. Table 4 shows the estimated technologies for noncognitive and cognitive skills estimated under these more general conditions.[27] The main conclusions based on Table 2 are robust to the assumption that measurement error is classical.[28] In Table 5 we show the estimated contemporaneous correlation across the measurement errors in our measures of noncognitive skills. Most of the correlations across the error terms are low. In fact, no correlation in any period exceeds, in absolute value, 0.2, and most are well below it.

Table 6 reports the contemporaneous correlation of the error terms in the measurement equations for investment. We assume that the error term in family income is independent of the remaining error terms. Virtually all correlations are well below 0.04 in absolute value. The only exceptions are the correlations between “trips to the museum” and “trips to the theater” in periods 1 and 2. In sum, these findings suggest that the assumption that the measurement error is classical is not at odds with the data we analyze, and allowing for correlation in errors does not change the main conclusions obtained from the simpler technology assuming classical measurement error.

3. Allowing for Correlated Omitted Inputs

We next investigate the assumption that the error term in the technology equations η_t is independent of the vector θ_t, by allowing for the presence of a time-invariant omitted input λ, as discussed in section III.4.3.[29] The results, displayed in Table 7, are consistent with the results shown in Table 2. Accounting for correlated omitted inputs does not reverse any major conclusion. Note that for purposes of identification, we normalize the coefficient on λ in the cognitive technology equation to one, $\gamma_4^C = 1$, and we estimate the coefficient on the noncognitive technology equation, $\gamma_4^N = .2835$.

4. Anchoring our estimates of the factor scale using adult outcomes

Table 8 reports estimates of the time-invariant technology that use the earnings data for persons age 23–28 to anchor the output of the production function in a log dollar metric.[30] We initially assume that η_t is serially uncorrelated and that measurement error is classical. We relax these assumptions below, when we report estimates of more general specifications. Our fitted earnings function is linear in age, and depends on the final level of the factors θ_{T+1}^C and θ_{T+1}^N. The coefficient on cognitive skills in the log earnings equations is estimated to be 0.052 (standard error is 0.0109). For noncognitive skills, we estimate a loading of 0.14 (with a standard error of 0.054). These estimates are consistent with estimates reported in Heckman, Stixrud, and Urzua (2006). From equation (25), 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 θ_T^I on outcomes.

Columns (1) and (3) in Table 8 transform the estimates in Table 2 by *D* into a log earnings metric. The two cross effects are ordered in the same direction as in the model reported in Table 2 where we use the metric of test scores. The effect of noncognitive skills on cognitive skills is precisely estimated.

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 these conclusions with regard to the log earnings anchor, 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 equation. It is interesting to note that in the metric of the probability of graduating from high school, the estimated parental investment factor affects *cognitive* skills more strongly than noncognitive skills. This is because cognitive skills receive higher weight in the high school graduation equation than in the log earnings equation. The relative strength of these effects is reversed across the two metrics. The choice of a metric is not innocuous.

5. Evidence of Sensitive Periods of Investment in Skills

We now report evidence on critical and sensitive periods. Our analysis in this section 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 or not there are sensitive periods in the development of skills provided that we normalize the investment factor on an input that is used at all stages of the child's life cycle. Results for an unanchored stage-specific technology, not correcting for nonclassical measurement error and serially correlated omitted inputs are presented in Table 9. Using several alternative measures, including family income, trips to the museum, and trips to the theater, we estimate the same qualitative ordering on the sensitivity of parental investments at different stages of the life cycle.[31] Using a likelihood ratio test, we test and reject the hypothesis that the parameters describing the technologies are invariant over stages of the lifecycle.[32]

Although we use test scores as a measure of output, transformation of output units 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 parental investment in both cognitive and noncognitive skills. Sensitive periods for parental investments in cognitive skills occur at earlier ages than sensitive periods 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 six and seven to ages eight and nine) is around 0.11 (with a standard error of 0.032). For the transition from period two to period three (ages eight and nine to ten and eleven) the corresponding coefficient decreases rather sharply to 0.0364 (with a standard error of 0.014). For the final transition (ages ten and eleven to ages 12 and 13), the estimate is about the same: 0.0379, with a standard error of 0.014. The difference between the early coefficient and the two later coefficients is statistically significant. This finding is consistent with periods 1 and 2 being sensitive periods for cognitive skills.[33]

For noncognitive skills in period one, the coefficient on investments is only 0.0533, with a standard error of 0.013. Then, it increases to 0.1067 in period two. It decreases to 0.0457 in the final transition. This evidence suggests that sensitive periods for the development of noncognitive skills occur at later ages in comparison to sensitive periods for cognitive skills.[34]

For the sake of completeness, in Table 10 we show the estimated technologies for cognitive and noncognitive skills when we allow the error term in the measurement equations for noncognitive skills and investments to be correlated. Again, the estimates in Tables 9 and 10 are very similar, suggesting that the assumption of independence across measurement errors does not substantially affect our estimates. Table 11 shows that the qualitative evidence on sensitive periods reported in Table 9 is robust to anchoring. Period 1 is the sensitive period for cognitive skills. Period 2 is the sensitive period for noncognitive skills in all of these specifications. The effects of parental investment on noncognitive skill remain strong at all stages, and are stronger and more precisely determined than in the case where we impose a stage-invariant technology.[35]

6. Estimating the Components of the Home Investment Dynamic Factor

The CNLSY/1979 reports an aggregate HOME score by taking a simple mean of the variables presented in Table 12 which assigns each component of the score the same weight. For expositional purposes we call these ad hoc weights.

The logic of the factor model speaks against uniform weighting. From equation (9), it follows that different components of the HOME score, $H_{i,t}$, weight latent θ_T^I differently. A uniformly weighted average of the mean adjusted components of the scores for person *i* is

(27) $$\tilde{H}_{i,t} = \frac{1}{m_t^I}\sum_{j=1}^{m_t^I}\left(Y_{i,j,t}^I - \mu_{j,t}^I\right) = \left(\frac{1}{m_t^I}\sum_{j=1}^{m_t^I}\alpha_{i,j,t}^I\right)\theta_{i,t}^I + \frac{1}{m_t^I}\sum_{j=1}^{m_t^I}\varepsilon_{i,j,t}^I.$$

There is no guarantee that the term in front of $\theta_{i,t}^I$ in equation (27) is 1. Thus the mean-adjusted HOME scores may be biased for $\theta_{i,t}^I$ for each person even if the $\varepsilon_{i,j,t}^I$ are mutually independent and m_t^I gets large so the second term converges to zero.

The measurement error of the standard mean adjusted home score $\tilde{H}_{i,t}$ for $\theta_{i,t}^I$ for person *i* is thus

$$\tilde{H}_i - \theta^I_{i,t} = \left(\frac{1}{m^I_t} \sum_{j=1}^{m^I_t} \alpha^I_{j,t} - 1 \right) \theta^I_{i,t} + \frac{1}{m^I_t} \sum_{j=1}^{m^I_t} \varepsilon^I_{i,j,t}.$$

Unless the term in parentheses on the right hand side equals zero, the measurement error is correlated with the true score. Instrumenting $\tilde{H}_{i,t}$ by a variable $Z_{i,t}$ correlated with $\theta^I_{i,t}$ but uncorrelated with $\varepsilon^I_{i,j,t}$ and $j \in \{1, \ldots, m^I_t\}$, will not produce consistent estimates of the skill technology. Trivially, it would produce consistent estimates of the technology parameter for $\theta^I_{i,t}$ divided by $\frac{1}{m^I_t} \sum_{j=1}^{m^I_t} \alpha^I_{j,t}$.[36]

Rewriting (9) and removing the means $(\tilde{Y}^I_{j,t} = Y^I_{j,t} - \mu^I_{j,t})$, we obtain

$$\frac{\tilde{Y}^I_{i,j,t}}{\alpha^I_{j,t}} = \theta^I_{i,t} + \frac{\varepsilon^I_{i,j,t}}{\alpha^I_{j,t}}.$$

An unweighted average of the inverse-factor-weighted mean adjusted scores is unbiased for θ^I_t for each person. The minimum variance unbiased combination of the inverse-factor loading weighted $\tilde{Y}^I_{i,j,t}$ in the case of uncorrelated $\frac{\varepsilon^I_{i,j,t}}{\alpha^I_{j,t}}$ assigns weight

$$\omega_{j,t} = \frac{\left(\alpha^I_{j,t}\right)^2}{\operatorname{Var}\left(\varepsilon^I_{j,t}\right)} \left[\sum_{k=1}^{m^I_t} \frac{\left(\alpha^I_{k,t}\right)^2}{\operatorname{Var}\left(\varepsilon^I_{k,t}\right)} \right]^{-1}$$

to $\frac{\tilde{Y}^I_{j,t}}{\alpha^I_{j,t}}$ where $\sum_{j=1}^{m^I_{k,t}} \omega_{j,t} = 1$. If an observation on θ is weighted more heavily, the smaller is its variance $\left(\frac{\operatorname{Var}\left(\varepsilon^I_{j,t}\right)}{\left(\alpha^I_{j,t}\right)^2} \right)$.

Dropping the i subscript to simplify the notation, and arraying $\frac{\tilde{Y}_{j,t}^{I}}{\alpha_{j,t}^{I}}$ into a vector,

$$\tilde{Y}_t^I = \left(\frac{\tilde{Y}_{1,t}^I}{\alpha_{1,t}^I}, \ldots, \frac{\tilde{Y}_{m_t^I,t}^I}{\alpha_{m_t^I,t}^I} \right),$$

and the $\frac{\varepsilon_{j,t}^{I}}{\alpha_{j,t}^{I}}$ into a vector,

$$\varepsilon_t^I = \left(\frac{\varepsilon_{1,t}^I}{\alpha_{1,t}^I}, \ldots, \frac{\varepsilon_{m_t^I,t}^I}{\alpha_{m_t^I,t}^I} \right),$$

and defining $V = \mathrm{E}\left(\varepsilon_t^I, \varepsilon_t^{I'}\right),$ where in this expression "E" denotes expectation, we can produce optimal weights ω_t as the solution to

$$\min \omega_t' V \omega_t \text{ subject to } \iota' \omega_t = 1,$$

where ι is a 1 by m_t^I vector of ones. The solution in the general case is

$$\omega_t = \frac{1}{\left(\iota' V_t^{-1} \iota\right)} \left(V_t^{-1} \iota\right),$$

which specializes to the weight previously given when V_t is diagonal.[37] These weights are optimal normalized home scores in the sense that they produce a minimum variance unbiased estimator of θ_t^I that will produce less bias for the true coefficient of θ_t^I in a least squares regression using θ_t^I as a regressor.

The importance of these weights depends on the importance of the measurement error in the components of these scores. 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. 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 modeled as

$R_t = \alpha^I_{R,t}\theta^I_t + \varepsilon^I_{R,t}$ so that Var(R_t) = $(\alpha^I_{R,t})^2$ Var(θ^I_t) + Var($\varepsilon^I_{R,t}$), because of the independence between θ^I_t and $\varepsilon^I_{R,t}$. We can decompose the total unobserved variance into two terms: one that is due to the parental input, the other that is orthogonal to it. The latter arises from measurement error. The relative importance of the two sources of error can be computed as:

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

and

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

Table 12 reports that $s_{I,R,1}$ = 0.1359 (for the first stage, corresponding to ages six and seven), while $s_{I,\varepsilon,1}$ = 0.8641. Most of the unobservable variance in "the number of books a child has" is actually not informative on the unobserved parental input θ^I_t. We report the same measures for the other input variables in Table 12. Over stages of the life cycle, all of the input measures tend to become relatively more error laden as proxies for θ^I_t.

Table 12 also displays the estimated optimal weights $\omega_{j,t}$ for each measurement j at each period t. The weights are far from uniform across inputs, as is assumed in constructing the traditional home score. Note further that the weights change over the life cycle reflecting the differential importance of measurement error variance at different ages. The change in the error variance reflects in part the change in $\alpha^{I}_{j,t}$ with t. Our estimates show that whether the child has special lessons has high weight early on (ages six and seven to eight and nine), but the weight declines considerably in the later periods (ages ten and eleven to 12/13). The variable that indicates the number of books at home, on the other hand, exhibits the opposite behavior. It starts small in early ages, but becomes more important at later ages. It is interesting to note that variables that describe the number of books at home and whether the family takes a newspaper, although informative about home investments, receive lower weights in our method than other components of the home score. The optimal weighting differs greatly from the uniform weighting traditionally used in constructing home scores.

In our sample, the covariance between the measurement error $\left(\widetilde{H}_h - \theta^I_t\right)$ and the true score θ^I_t is relatively weak (see Table 13). This happens because $\frac{1}{m^I_t}\sum_{k=1}^{m^I_t}\alpha^I_{t,k}$ is close to 1. See Table 14 for the factor loadings and normalized factor loadings for noncognitive, cognitive, and parental investment (HOME score) components. The fact that $\frac{1}{m^I_t}\sum_{k=1}^{m^I_t}\alpha^I_{t,k} = 1$ implies that in our sample, standard IV methods designed to protect against classical measurement error in the standard HOME score are likely to be effective.

Table 15 displays the reliability in the test scores for cognitive and noncognitive skills in a manner comparable to the estimates of the share of measurement error for the components of the HOME score in Table 12. The components of both cognitive and noncognitive tests are measured with substantial error. Simple unweighted averages of the components of these tests are biased for θ^C_t and θ^N_t for each person. We display the proportion of the variance due to measurement error in each of these scores for each test in the second column. The share of measurement error is roughly stable across ages but fluctuates for some components (for example, hyperactivity).

Our evidence of substantial measurement error in all of the measures of inputs and outputs suggests that simple OLS estimates of the technology of skill formation are likely to be considerably biased. Table 16 presents an OLS version of the model with estimates reported in columns 1 and 4 of Table 2 that use income as the investment anchor.[38] The contrast between the estimates reported in the first and fourth columns of Table 2 and the least squares estimates in Table 16 is striking. Generally, OLS coefficients are downward-biased, showing much smaller self productivity, cross productivity and investment productivity effects. The estimated effect of the HOME score on the Math score is perverse.

V. Conclusion

This paper identifies and estimates a model of investment in child cognitive and noncognitive skills using dynamic factor models. The model is based on the analysis of Cunha and Heckman (2007) and Cunha, Heckman, Lochner, and Masterov (2006).

Our empirical methodology accounts for the proxy nature of the measurements of 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 model of Cunha and Heckman (2007). To avoid the arbitrariness that arises in using test scores to measure the output of parental investments, we anchor estimated effects of investment in the metric of adult earnings and in the metric of the probability of high school graduation. The choice of the metric affects our conclusions about the relative productivity of parental investment on cognitive and noncognitive skills. We report results for alternative normalizations of the scale of parental investment and generally find agreement among alternative specifications.

We reach the following major conclusions. (1) We find high levels of self productivity in the production of cognitive and noncognitive skills. (2) We find 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) We also find substantial evidence of measurement error in the home input proxies and corollary evidence of attenuation bias in the OLS estimates of the technology of skill formation. (4) The estimated relative effect of parental input on cognitive and noncognitive skills depends on the metric in which we measure input.

Different adult outcomes are affected differently by cognitive and noncognitive skills. Sensitive periods occur at different stages for cognitive and noncognitive skills. Therefore, different stages of the child's life cycle are sensitive periods for investment to achieve different adult outcomes.

To show this, we simulate the effect of a ten percent increase in investment at different stages of the life cycle of the child on log earnings at 23 (Table 17A) and on high school graduation (Table 17B). These estimates include the cross effects of each skill on the other, self productivity, and direct investment effects.[39]

For the log earnings outcome, the strongest effect is for investment in period 2. This operates primarily through its effect on noncognitive skills which then percolate into the next period and raise both cognitive and noncognitive skills. The strongest effect of investment on earnings operating through effects on cognitive skills is in stage 1. Even at stage 1, however, the effects of investment on cognitive skills and noncognitive skills are equally strong.

For high school graduation, the strongest effect of investment comes in period 1 and it operates primarily through its effects on cognitive skills. Even though period 1 is important, the effects of investment in later periods are substantial.

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. These problems are addressed in Cunha, Heckman, and Schennach (2007), who also present a more general nonlinear approach to anchoring the test scores in an outcome measure.

Appendix 1. Interpretation of the Measurement Equations as Derived Demands

Write a production function

$$\theta_t^I = \varphi_t^I \left(X_{1,t}, \ldots, X_{m_t^I, t} \right),$$

which is the output of the investment sector. Let $X_t = \left(X_{1,t}, \ldots, X_{m_t^I, t} \right)$. $P_t = \left(P_{1,t}, \ldots, P_{m_t^I, t} \right)$ is the price vector.

The problem of the family is to minimize costs,

$$\min \left[P_t' X_t + \lambda \left(\theta_t^I - \varphi_t^I \left(X_t \right) \right) \right].$$

The first order condition for this problem is

$$P_{i,t} - \lambda \frac{\partial \varphi_t^I \left(X_{i,t} \right)}{\partial X_{i,t}} = 0, \; i \in \{1, \ldots, m_t^I\}$$

for an interior solution. We can derive input demands as a function of prices and output levels

$$X_{i,t} = h_{j,t} \left(P_t, \theta_t^I \right), \; j \in \{1, \ldots, m_t^I\},$$

which implicitly define the measurement equations.

For the Cobb-Douglas case, the technology is

$$\theta_t^I = A_t \prod_{i-1}^{m_t^I} X_{i,t}^{\alpha_i}.$$

The input demand function for input *i* is

$$\ln X_{i,t} = \frac{1}{\sum_{j=1}^{m_t^I} \alpha_j} \ln \theta_t^I - \frac{\ln A_t}{\sum_{j=1}^{m_t^I} \alpha_j} - \sum_{j=1}^{m_t^I} \ln\left(\frac{\alpha_j P_i}{\alpha_i P_j}\right),\ i \in \{1, \ldots, m_t^I\}.$$

Accounting for measurement error,

$$Y_{i,t}^I = \ln X_{i,t} + \varepsilon_{i,t}^I.$$

In the Cobb-Douglas case, all inputs (measurements) have the same factor loading on $\ln \theta_t^I$. Only the intercepts which depend on the share parameters and the prices are different. In the Cobb-Douglas case, one would use logs of the factors.[40]

In the Leontief case,

$$\theta_t^I = \min\left\{\frac{X_{1,t}}{\alpha_1}, \ldots, \frac{X_{m_t^I,t}}{\alpha_{m_t^I}}\right\}.$$

The input demand equations are

$$X_{i,t} = \alpha_i \theta_t^I,\ i \in \{1, \ldots, m_t^I\},$$

and in logs,

$$\ln X_{i,t} = \ln \alpha_i \ln \theta_t^I,$$

so

$$Y_{i,t}^{I} = \ln X_{i,t} + \varepsilon_{i,t}^{I},\ i \in \{1,\ldots,m_t^I\}.$$

Thus all factor loadings on $\ln\theta_t^I$ are unity.

A generalized Leontief function writes

$$\theta_t^I = \min\left\{\frac{X_{1,t}^{\tau_1}}{\alpha_1},\ldots,\frac{X_{m_t^I,t}^{\tau_{m_t^I}}}{\alpha_{m_t^I}}\right\}.$$

Thus,

$$\ln X_{i,t} = \frac{1}{\tau_i}\ln\alpha_i + \frac{1}{\tau_i}\ln\theta_t^I,\ i \in \{1,\ldots,m_t^I\}.$$

In this case, the factor loadings are input-specific.

Appendix 2. Sample Likelihood for the Basic Estimation Strategy

We derive the likelihood and describe the basic estimation strategy for the model with classical measurement error and without serially correlated η_t. The likelihood for the more general models we estimate follows from a straightforward modification of the analysis in this appendix. 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 and 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

$$Y_t' = \left(Y_{1,t}^N,\ldots,Y_{m_t^N,t}^N, Y_{1,t}^C,\ldots,Y_{m_t^C,t}^C, Y_{1,t}^I,\ldots,Y_{m_t^I,t}^I\right).$$

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

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

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

(B.1) $$Y_t = \alpha_t \theta_t + \varepsilon_t.$$

Recall that we use S,A to denote the 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 $Q_t = \mathrm{Var}\left(\eta_t\right)$.

We assume that $\theta_1 \left| S, A \sim N\left(a_1, P_1\right)\right.$. In the text, we establish the conditions for identification of a_1 and P_1. We also assume that $\varepsilon_t \sim N\left(0, H_t\right)$ and $\eta_t \sim N\left(0, Q_t\right)$. Then, given a normality assumption, together with linearity, it follows that $Y_t \sim N\left(\mu_1, F_t\right)$ 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 work underway, we relax this assumption. To proceed in the normal case, we apply the Kalman filtering procedure (for details on the derivations see, for example, Harvey, 1989 or Durbin and Koopman, 2001). If we define $Y^t = \left(Y_1, \ldots, Y_t\right), a_{t+1} = \mathrm{E}\left(\theta_{t+1} \left| S, A, Y^t\right.\right),$ and $P_{t+1} = \mathrm{Var}\left(\theta_{t+1} \left| S, A, Y^t\right.\right)$, it is straightforward to establish that:

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

and

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

Consequently, using (B.1) we obtain $Y_{t+1} \left| S, \lambda, Y^t \sim N\left(\mu_t, F_t\right)\right.$ 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:

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

where Y_i^{t-1} is the history of Y_i up to time period *t*-1. 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:

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

Assuming that observations are i.i.d. over children, 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. Extensions to the other cases are straightforward and for the sake of brevity are deleted.

References

Altonji, Joseph G. and Rosa L. Matzkin. 2005. "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors." *Econometrica* 73(4):1053–1102.

Arellano, Manuel. 2003. *Panel Data Econometrics*. New York: Oxford University Press.

Baltagi, Badi H. 1995. *Econometric Analysis of Panel Data*. New York: Wiley.

Baydar, Nazli and Jeanne Brooks-Gunn. 1991. "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, Gary S. and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4(3, Part 2):S1–S39.

Blau, David and Janet Currie. 2006. "Preschool, Daycare, and Afterschool Care: Who's Minding the Kids? " In Eric Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, *Handbooks in Economics*, vol. 2, chap. 20. Amsterdam: North-Holland, 1163–1278.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Borghans, Lex, Angela L. Duckworth, James J. Heckman, and Bas ter Weel. 2007. "The Economics and Psychology of Personality and Motivation." Unpublished manuscript, University of Chicago, Department of Economics. Forthcoming, *Journal of Human Resources*.

Borghans, Lex, Bas ter Weel, and Bruce A. Weinberg. 2007. "Interpersonal Styles and Labor Market Outcomes." Working Paper 12846, NBER.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The Determinants of Earnings: A Behavioral Approach." *Journal of Economic Literature* 39(4):1137–1176.

Bradley, Robert H. and Bettye M. Caldwell. 1980. "The Relation of Home Environment, Cognitive Competence, and IQ among Males and Females." *Child Development* 51(4):1140–1148.

__________. 1984. "The Relation of Infants' Home Environments to Achievement Test Performance in First Grade: A Follow-up Study." *Child Development* 55(3):803–809.

Cameron, Judy. 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, Pedro, Karsten Hansen, and James J. Heckman. 2003. "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, Pedro and James J. Heckman. 2003. "Human Capital Policy." In James J. Heckman, Alan B. Krueger, and Benjamin M. Friedman, eds., *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.

Carneiro, Pedro, James J. Heckman, and Dimitriy V. Masterov. 2005. "Labor Market Discrimination and Racial Differences in Pre-Market Factors." *Journal of Law and Economics* 48(1):1–39.

Cawley, John, James J. Heckman, and Edward J. Vytlacil. 1999. "On Policies to Reward the Value Added by Educators." *Review of Economics and Statistics* 81(4):720–727.

__________. 2001. “Three Observations on Wages and Measured Cognitive Ability.” *Labour Economics* 8(4):419–442.

Center for Human Resource Research, ed. 2004. *NLSY79 Child and Young Adult Data User’s Guide*. Ohio State University, Columbus, Ohio.

Cunha, Flavio and James J. Heckman. 2007. “The Technology of Skill Formation.” *American Economic Review* 97(2):31–47.

Cunha, Flavio, James J. Heckman, Lance J. Lochner, and Dimitriy V. Masterov. 2006. “Interpreting the Evidence on Life Cycle Skill Formation.” In Eric A. Hanushek and Frank Welch, eds., *Handbook of the Economics of Education*, chap. 12. Amsterdam: North-Holland, 697–812.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2007. “Estimating the Technology of Cognitive and Noncognitive Skill Formation.” Unpublished manuscript, University of Chicago, Department of Economics. Presented at the Yale Conference on Macro and Labor Economics, May 5–7, 2006. Under revision, *Econometrica*.

Durbin, James and Siem Jan Koopman, 2001. *Time Series Analysis by State Space Methods*. New York: Oxford University Press.

Fryer, Roland and Steven Levitt. 2004. “Understanding the Black-White Test Score Gap in the First Two Years of School.” *Review of Economics and Statistics* 86(2):447–464.

Hansen, Karsten T., James J. Heckman, and Kathleen J. Mullen. 2004. “The Effect of Schooling and Ability on Achievement Test Scores.” *Journal of Econometrics* 121(1-2):39–98.

Hanushek, Eric. 2003. “The Failure of Input-Based Schooling Policies.” *Economic Journal* 113(485):F64–F98.

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

Heckman, James J. 2007. “The Technology and Neuroscience of Capacity Formation.” *Proceedings of the National Academy of Sciences* Forthcoming, presented at Economic Causes and Consequences of Population Aging, Robert Fogel 80th Birthday Celebration, November 18, 2006.

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

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior.” *Journal of Labor Economics* 24(3):411–482.

Herrnstein, Richard J. and Charles A. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.

Jöreskog, Karl G. and Arthur S. Goldberger. 1975. “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, Karl G., Dag Sörbom, and Jay Magidson. 1979. *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books.

Knudsen, Eric I., James J. Heckman, Judy Cameron, and Jack P. Shonkoff. 2006. “Economic, neurobiological, and behavioral perspectives on building America’s future workforce.” *Proceedings of the National Academy of Sciences* 103(27):10155–10162.

Kotlarski, Ignacy I. 1967. "On Characterizing the Gamma and Normal Distribution." *Pacific Journal of Mathematics* 20:69–76.

Levitt, Pat. 2003. "Structural and Functional Maturation of the Developing Primate Brain." *Journal of Pediatrics* 143(4, Supplement):S35–S45.

Linver, Miriam R., Jeanne Brooks-Gunn, and Natasha Cabrera. 2004. "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, Albert. 1964. "Instrumental Variables in Factor Analysis." *Psychometrika* 29(2):105–113.

Meaney, Michael 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, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 77(2):251–266.

Olds, David L. 2002. "Prenatal and Infancy Home Visiting by Nurses: From Randomized Trials to Community Replication." *Prevention Science* 3(2):153–172.

Peterson, James L. and Nicholas Zill. 1986. "Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children." *Journal of Marriage and the Family* 48(2):295–307.

Pudney, S. E. 1982. "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. Cybele and Edward F. Zigler. 1997. “Social Competence: An Untapped Dimension in Evaluating Head Start’s Success.” *Early Childhood Research Quarterly* 12(4):363–385.

Schennach, Susanne M. 2004. “Estimation of Nonlinear Models with Measurement Error.” *Econometrica* 72(1):33–75.

Shumway, Robert H. and David S. Stoffer. 1982. “An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm.” *Journal of Time Series Analysis* 3(3):253–264.

Todd, Petra E. and Kenneth I. Wolpin. 2003. “On the Specification and Estimation of the Production Function for Cognitive Achievement.” *Economic Journal* 113(485):F3–33.

__________. 2005. “The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps.” Unpublished manuscript. Under revision for publication.

Watson, Mark W. and Robert F. Engle. 1983. “Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models.” *Journal of Econometrics* 23(3):385–400.

Zigler, Edward F. and Earl C. Butterfield. 1968. “Motivational Aspects of Changes in IQ Test Performance of Culturally Deprived Nursery School Children.” *Child Development* 39(1):1–14.

* Cunha: University of Pennsylvania. Heckman: University of Chicago, American Bar Foundation and University College Dublin. This research was supported by NIH R01-HD043411, NSF SES-024158, the Committee for Economic Development with a grant from The Pew Charitable Trusts and the Partnership for America's Economic Success, and the J.B. Pritzker Consortium on Early Childhood Development at the Harris School of Public Policy, University of Chicago. Flavio Cunha also acknowledges support from the Claudio Haddad dissertation fund at the University of Chicago. The views expressed in this paper are those of the authors and not necessarily those of the funders listed here. 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 a coauthor of a successor paper, Petra Todd, and Kenneth Wolpin. We also received helpful comments from the editor and three anonymous references. A website containing supplementary material is available at http://jenni.uchicago.edu/idest-tech. The data used in this article can be obtained beginning [six months after publication] through [three years hence] from James J. Heckman, University of Chicago, Department of Economics, 1126 E. 59th Street, Chicago IL 60637; j-hec kman@uchicago.edu.

1 See Bowles, Gintis, and Osborne (2001), Heckman and Rubinstein (2001), and Heckman, Stixrud, and Urzua (2006).

2 Todd and Wolpin (2003) survey the educational production function literature as well as the child development literature.

3 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) and the evidence summarized in Cunha and Heckman (2007).

4 Cawley, Heckman, and Vytlacil (1999) anchor 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.

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

6 See also their web appendix, where more general models of skill formation are analyzed.

7 We assume that f_t^k is twice continuously differentiable, increasing and concave in $\theta_{k,t}^I$. Twice continuous differentiability is only a convenience.

8 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. Borghans, ter Weel, and Weinberg (2007b) discuss evidence against this assumption. 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. If we assume that the output (reward) in adult task j is $g_j\left(\theta_{T+1}^C, \theta_{T+1}^N; \eta\right)$, where η is a person-specific parameter and there are J distinct tasks, we can define $g_j\left(\theta_{T+1}^C, \theta_{T+1}^N, \eta\right) = \arg\max_j \left\{g_j\left(\theta_{T+1}^C, \theta_{T+1}^N, \eta\right)\right\}_{j=1}^J$ and capture the operation of comparative advantage in the labor market.

9 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 percent depreciation of investment in each period. They also do not account for endogeneity of inputs or measurement error in inputs which we find to be substantial.

10 This measure originates in the work of Bradley and Caldwell (1980; 1984) and is discussed further in section IV.

11 See Hsiao (1986); Baltagi (1995); and Arellano (2003) for descriptions of these procedures.

12 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 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.

13 There are many other papers that use this score. See for example, Baydar and Brooks-Gunn, 1991, and the papers cited by Todd and Wolpin.

14 In this and later sections, θ_t includes the investment factor, whereas in section II it only includes stocks of skills at date *t*.

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

16 Measurement equations (7), (8), and (9) can be interpreted as output-constant demand equations arising from the following two-stage maximization problem. Families use inputs $X_{j,t}$ with prices $P_{j,t}$, $j \in \{1,\ldots,m_t^I\}$, to produce family investment $\theta_t^I = \phi_t^I\left(X_{1,t},\ldots,X_{m_t,t}^I\right)$. For the

problem of minimizing the cost of achieving a given output, one can derive demand functions $X_{j,t} = h_{j,t}\left(P_{1,t},\ldots,P^{I}_{m_t,t},\theta^{I}_{t}\right)$ under general conditions. Specifications (7) - (9) are consistent with Cobb-Douglas and Leontief technologies, when θ^{I}_{t} is measured in logs. Prices appear in the intercepts. These technologies impose restrictions on the factor loadings of the inputs. See Appendix 1 which develops this point further.

17 Obviously, we can not identify the mean of the factor, $\mathrm{E}\left(\theta^{k}_{t}\right)$, and the intercepts $\mu^{k}_{j,t}$ at the same. It is necessary either to normalize the intercept in one equation $\mu^{k}_{1,t} = 0$ and identify $\mathrm{E}\left(\theta^{k}_{t}\right)$, or to normalize $\mathrm{E}\left(\theta^{k}_{t}\right) = 0$ and identify all of the intercepts $\mu^{k}_{j,t}$.

18 Obviously, we can not identify the mean of the factor, $\mathrm{E}\left(\theta^{k}_{t}\right)$, and the intercepts $\mu^{k}_{j,t}$ at the same. It is necessary either to normalize the intercept in one equation $\mu^{k}_{1,t} = 0$ and identify $\mathrm{E}\left(\theta^{k}_{t}\right)$, or to normalize $\mathrm{E}\left(\theta^{k}_{t}\right) = 0$ and identify all of the intercepts $\mu^{k}_{j,t}$.

19 See equation (9) in Cunha and Heckman (2007).

20 See our website for an analysis of the case in which η^{k}_{t} are serially correlated for $k \in \{C, N\}$.

21 See their web appendix.

22 Note that we can identify the loadings δ_N and δ_C from:

$$\mathrm{Cov}\left(\ln Y, Y^{N}_{1,T}\right) = \delta_N \mathrm{Var}\left(\theta^{\mathrm{N}}_{\mathrm{T}}\right) + \delta_C \mathrm{Cov}\left(\theta^{\mathrm{N}}_{\mathrm{T}}, \theta^{\mathrm{C}}_{\mathrm{T}}\right)$$

and

$$\mathrm{Cov}\left(\ln Y, Y^{C}_{1,T}\right) = \delta_N \mathrm{Cov}\left(\theta^{\mathrm{N}}_{\mathrm{T}}, \theta^{\mathrm{C}}_{\mathrm{T}}\right) + \delta_C \mathrm{Var}\left(\theta^{\mathrm{C}}_{\mathrm{T}}\right)$$

which gives us two linearly independent equations in two unknowns (δ_N, δ_C) The solution is:

$$\begin{pmatrix} \delta_N \\ \delta_C \end{pmatrix} = \frac{1}{\mathrm{Var}\left(\theta_T^N\right)\mathrm{Var}\left(\theta_T^C\right) - \mathrm{Cov}\left(\theta_T^N, \theta_T^C\right)^2} \begin{bmatrix} \mathrm{Var}\left(\theta_T^C\right) & -\mathrm{Cov}\left(\theta_T^N, \theta_T^C\right) \\ -\mathrm{Cov}\left(\theta_T^N, \theta_T^C\right) & \mathrm{Var}\left(\theta_T^N\right) \end{bmatrix}.$$

23 As discussed in Linver, Brooks-Gunn, and Cabrera (2004), some of these items are not useful because they do not vary much among families (that is, more than 90 percent to 95 percent of all families make the same response).

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

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.

26 The magnitude of the estimated parental investment effect clearly depends on the scale in which investments are measured.

27 We use family income to normalize investment.

28 As discussed in Section 2, to generalize our results to allow for nonclassical measurement error, we need to assume that the error term in one of the measures is independent of all measurement errors. For the measurements for noncognitive skills, we impose this assumption on the error term in the anti-social score equation.

29 We normalize investment on family income.

30 Investment is normalized on family income.

31 In the text we report the results for the normalization of investment relative to family income. In our website appendix, we report estimates of alternative normalizations using “trips to the theater” and “trips to the museum”.

32 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 $\Lambda = -2\left(\ln L_R - \ln L_U\right)$, where "*R*" denotes restricted and "*U*" denotes unrestricted, 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 five percent significance level.

33 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.

34 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.

35 When we anchor on high school graduation instead of log earnings, we find that parental effects on the cognitive factor are stronger than on the noncognitive factor. This is also found in the stage-invariant technology. See Web Appendix Table 11.

36 The proof is straightforward. Divide both sides of (27) by $\frac{1}{m_t^I}\sum_{j=1}^{m_t^I}\alpha_{j,t}^I$, substitute into technology (6) and apply standard IV.

37 In terms of $H_t = \mathrm{E}\left(\varepsilon_t^I \varepsilon_t^{I\prime}\right)$, $\omega_t = \frac{1}{\left(\alpha_t^{I\prime} H_t^{-1} \alpha_t^I\right)}\left(\mathrm{Diag}\ \alpha_t^I\right) H_t^{-1} \alpha_t^I$ where $\mathrm{Diag}\ \alpha_t^I$ arrays the elements of α_t^I on a diagonal matrix.

38 We get similar results for other anchors.

39 From technology (6), we obtain the effect of investment in a period *k* stages before the terminal period on adult abilities as

$$\theta_{\mathrm{T+1}} = \left(\prod_{j=0}^{k} A_{T-j} \right) B_{T-k} \theta_{T-k}^{I}.$$

The effects of variations in components of θ_{T-k}^{I} operating through cognitive and noncognitive skills are reported in Tables 17A and 17B. (The top element of B_{T-k} corresponds to the noncognitive effect of investment in the period; the bottom element corresponds to the cognitive effect of investment in the period.)

40 One can always write $\tilde{\theta}_t^I = \log \theta_t^I$ and work with $\tilde{\theta}_t^I$ everywhere.

Table 1
Summary Dynamic Measurements
White Children NLSY/1979

	Ages 6 and 7			Ages 8 and 9			Ages 10 and 11			Ages 12 and 13		
	Observation	Mean	Standard Error	Observations	Mean	Standard Error	Observations	Mean	Standard Error	Observations	Mean	Standard Error
Piat Math[a]	753	-1.0376	0.5110	799	0.0423	0.6205	787	0.7851	0.6101	690	1.2451	0.5783
Piat Reading Recognition[a]	751	-1.0654	0.4303	795	-0.0932	0.6543	783	0.6179	0.7334	688	1.1442	0.7852
Antisocial Score[a]	753	0.0732	0.9774	801	-0.0843	1.0641	787	-0.0841	1.0990	717	-0.0658	1.0119
Anxious Score[a]	778	0.1596	1.0016	813	-0.0539	1.0187	813	-0.0753	1.0771	730	-0.0664	1.0561
Headstrong Score[a]	780	0.0192	0.9882	813	-0.2127	1.0000	812	-0.2146	1.0416	729	-0.2123	1.0572
Hyperactive Score[a]	780	-0.0907	0.9673	815	-0.1213	1.0148	813	-0.0983	0.9902	729	-0.0349	0.9910
Conflict Score[a]	779	0.0177	0.9977	815	-0.0057	0.9935	814	-0.0441	1.0304	731	-0.0472	1.0420
Number of Books[b]	629	3.9173	0.3562	821	3.9220	0.3104	676	3.6746	0.6422	730	3.6315	0.6768
Musical Instrument[c]	628	0.4650	0.4992	821	0.4896	0.5002	674	0.5504	0.4978	728	0.5907	0.4921
Newspaper[c]	629	0.5326	0.4993	821	0.5043	0.5003	674	0.4985	0.5004	728	0.5000	0.5003
Child has special lessons[c]	627	0.5470	0.4982	820	0.7049	0.4564	672	0.7247	0.4470	727	0.7717	0.4200
Child goes to museums[d]	628	2.2596	0.9095	821	2.3082	0.8286	672	2.2604	0.8239	729	2.2195	0.8178
Child goes to theater[d]	630	1.8111	0.8312	820	1.8012	0.7532	674	1.8309	0.8000	728	1.8475	0.7920
Natural Logarithm of Family Income[e]	865	10.4915	1.3647	936	10.4494	1.5689	881	10.5454	1.3168	795	10.6169	1.1877
Mother's Highest Grade Completed[f]	1053	12.9620	2.2015	1053	12.9620	2.2015	1053	12.9620	2.2015	1053	12.9620	2.2015
Mother's	776	0.6050	1.0132	776	0.6050	1.0132	776	0.6050	1.0132	776	0.6050	1.0132

Arithmetic Reasoning Score[g]												
Mother's Word Knowledge Score[g]	776	0.5894	0.7666	776	0.5894	0.7666	776	0.5894	0.7666	776	0.5894	0.7666
Mother's Paragraph Composition Score[g]	776	0.5464	0.7311	776	0.5464	0.7311	776	0.5464	0.7311	776	0.5464	0.7311
Mother's Numerical Operations Score[g]	776	0.4945	0.8189	776	0.4945	0.8189	776	0.4945	0.8189	776	0.4945	0.8189
Mother's Coding Speed Score[g]	776	0.4554	0.8084	776	0.4554	0.8084	776	0.4554	0.8084	776	0.4554	0.8084
Mother's Mathematical Knowledge Score[g]	776	0.5297	1.0259	776	0.5297	1.0259	776	0.5297	1.0259	776	0.5297	1.0259

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

b. 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.

c. 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.

d. 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.

e. Family Income is CPI adjusted. Base year is 2000.

f. Mother's Highest Grade Completed by Age 28.

g. Components of the ASVAB Battery. The variables are standardized with mean zero and variance one across the entire NLSY/79 sample.

Table 2
Unanchored Technology Equations[a]
Measurement Error is Classical, Absence of Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill $\left(\theta_{t+1}^{N}\right)$			Cognitive Skill $\left(\theta_{t+1}^{C}\right)$		
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.884	0.884	0.884	0.028	0.028	0.028
	(0.021)	(0.021)	(0.021)	(0.013)	(0.013)	(0.013)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.003	0.003	0.003	0.977	0.977	0.977
	(0.013)	(0.012)	(0.013)	(0.038)	(0.038)	(0.038)
Parental Investment, $\left(\theta_t^I\right)$	0.072	0.078	0.080	0.064	0.069	0.071
	(0.020)	(0.021)	(0.024)	(0.013)	(0.014)	(0.015)
Maternal Education, *S*	0.004	0.004	0.004	0.003	0.003	0.003
	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)
Maternal Cognitive Skill, *A*	-0.006	-0.006	-0.006	0.025	0.025	0.025
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)

Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let *S* denote mother's education and *A* denote mother's cognitive ability. The technology equations are:

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_{1,t}^k S + \psi_{2,t}^k A + \eta_{t+1}^k.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_1^k$, and ψ_2^k in columns (1) through (6). In columns (1) and (4), the parental investment factor is normalized in the log-family income equation. In columns (2) and (5), the parental investment factor is normalized in trips to the museum. In columns (3) and (6), we normalize the parental investment factor in trips to the theater.

Table 3
Contemporaneous Correlation Matrices
Measurement Error is Classical, Absence of Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

	Noncognitive	Cognitive	Investments
Period 1 – Children ages 6 and 7			
Noncognitive	1.0000	0.1892	0.3426
Cognitive	0.1892	1.0000	0.2921
Investments	0.3426	0.2921	1.0000
Period 2 - Children ages 8 and 9			
Noncognitive	1.0000	0.2334	0.4065
Cognitive	0.2334	1.0000	0.3835
Investments	0.4065	0.3835	1.0000
Period 3 - Children ages 10 and 11			
Noncognitive	1.0000	0.2643	0.4785
Cognitive	0.2643	1.0000	0.4892
Investments	0.4785	0.4892	1.0000
Period 4 - Children ages 12 and 13			
Noncognitive	1.0000	0.2845	0.5511
Cognitive	0.2845	1.0000	0.6111
Investments	0.5511	0.6111	1.0000

Table 4
Unanchored Technology Equations[a]
Measurement Error is Non-Classical, Absence of Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill $\left(\theta_{t+1}^{N}\right)$	Cognitive Skill $\left(\theta_{t+1}^{C}\right)$
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.8672	0.0264
	(0.024)	(0.011)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.0045	0.9739
	(0.014)	(0.038)
Parental Investment, $\left(\theta_t^I\right)$	0.0801	0.0647
	(0.018)	(0.012)
Maternal Education, *S*	0.0041	0.0026
	(0.008)	(0.010)
Maternal Cognitive Skill, *A*	-0.0092	0.0252
	(0.006)	(0.009)

a. Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let *S* denote mother's education and *A* denote mother's cognitive ability. The technology equations are:

$$\theta_{t+1}^{k} = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_{1,t}^k S + \psi_{2,t}^k A + \eta_{t+1}^k.$$

In this table we show the estimated parameter values and standard errors (in parenthesis) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_1^k,$ and ψ_2^k for noncognitive (k = N) and cognitive (k = C) skills. Investment is normalized in family income.

Table 5
Contemporaneous Correlation Matrices in Measurement Error
Measurements for Noncognitive Skills
White Males, CNLSY/79

	Anti-Social	Anxious	Headstrong	Hyperactive	Peer Conflict
Period 1 - Children ages 6 and 7					
Anti-Social	1.0000	0.0000	0.0000	0.0000	0.0000
Anxious	0.0000	1.0000	-0.0054	-0.0083	0.0479
Headstrong	0.0000	-0.0054	1.0000	0.0193	-0.1113
Hyperactive	0.0000	-0.0083	0.0193	1.0000	-0.1721
Peer Conflict	0.0000	0.0479	-0.1113	-0.1721	1.0000
Period 2 - Children ages 8 and 9					
Anti-Social	1.0000	0.0000	0.0000	0.0000	0.0000
Anxious	0.0000	1.0000	-0.0023	-0.0020	0.0117
Headstrong	0.0000	-0.0023	1.0000	0.0328	-0.1941
Hyperactive	0.0000	-0.0020	0.0328	1.0000	-0.1652
Peer Conflict	0.0000	0.0117	-0.1941	-0.1652	1.0000
Period 3 – Children ages 10 and 11					
Anti-Social	1.0000	0.0000	0.0000	0.0000	0.0000
Anxious	0.0000	1.0000	0.0196	0.0001	-0.0007
Headstrong	0.0000	0.0196	1.0000	0.0067	-0.0312
Hyperactive	0.0000	0.0001	0.0067	1.0000	-0.0002
Peer Conflict	0.0000	-0.0007	-0.0312	-0.0002	1.0000
Period 4 - Children ages 12 and 13					
Anti-Social	1.0000	0.0000	0.0000	0.0000	0.0000
Anxious	0.0000	1.0000	-0.0797	-0.1495	-0.0105
Headstrong	0.0000	-0.0797	1.0000	0.0692	0.0049
Hyperactive	0.0000	-0.1495	0.0692	1.0000	0.0092
Peer Conflict	0.0000	-0.0105	0.0049	0.0092	1.0000

Table 6
Contemporaneous Correlation Matrices in Measurement Error
Measurements for Parental Investments
White Males, CNLSY/79

	Income	Books	Musical	Newspaper	Lessons	Museum	Theater
Period 1 - Children ages 6 and 7							
Family Income	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of Books	0.0000	1.0000	-0.0044	0.0050	-0.0029	-0.0269	-0.0678
Musical Instruments	0.0000	-0.0044	1.0000	-0.0047	0.0027	0.0257	0.0647
Newspaper Subscriptions	0.0000	0.0050	-0.0047	1.0000	-0.0031	-0.0290	-0.0731
Number of Special Lessons	0.0000	-0.0029	0.0027	-0.0031	1.0000	0.0168	0.0423
Trips to Museum	0.0000	-0.0269	0.0257	-0.0290	0.0168	1.0000	0.3960
Trips to Theater	0.0000	-0.0678	0.0647	-0.0731	0.0423	0.3960	1.0000
Period 2 - Children ages 8 and 9							
Family Income	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of Books	0.0000	1.0000	-0.0008	0.0052	0.0018	-0.0160	-0.0484
Musical Instruments	0.0000	-0.0008	1.0000	-0.0019	-0.0006	0.0058	0.0175
Newspaper Subscriptions	0.0000	0.0052	-0.0019	1.0000	0.0039	-0.0355	-0.1076
Number of Special Lessons	0.0000	0.0018	-0.0006	0.0039	1.0000	-0.0121	-0.0366
Trips to Museum	0.0000	-0.0160	0.0058	-0.0355	-0.0121	1.0000	0.3291
Trips to Theater	0.0000	-0.0484	0.0175	-0.1076	-0.0366	0.3291	1.0000
Period 3 - Children ages 10 and 11							
Family Income	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of Books	0.0000	1.0000	-0.0001	-0.0001	-0.0001	0.0052	0.0007
Musical Instruments	0.0000	-0.0001	1.0000	0.0002	0.0003	-0.0137	-0.0017
Newspaper Subscriptions	0.0000	-0.0001	0.0002	1.0000	0.0002	-0.0083	-0.0010
Number of Special Lessons	0.0000	-0.0001	0.0003	0.0002	1.0000	-0.0130	-0.0016
Trips to Museum	0.0000	0.0052	-0.0137	-0.0083	-0.0130	1.0000	0.0693
Trips to Theater	0.0000	0.0007	-0.0017	-0.0010	-0.0016	0.0693	1.0000
Period 4 - Children ages 12 and 13							
Family Income	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of Books	0.0000	1.0000	0.0003	-0.0007	0.0000	0.0017	0.0158
Musical Instruments	0.0000	0.0003	1.0000	-0.0006	0.0000	0.0016	0.0151
Newspaper Subscriptions	0.0000	-0.0007	-0.0006	1.0000	0.0000	-0.0034	-0.0313
Number of Special Lessons	0.0000	0.0000	0.0000	0.0000	1.0000	0.0001	0.0010
Trips to Museum	0.0000	0.0017	0.0016	-0.0034	0.0001	1.0000	0.0803
Trips to Theater	0.0000	0.0158	0.0151	-0.0313	0.0010	0.0803	1.0000

Table 7
Unanchored Technology Equations[a]
Measurement Error is Classical, Allows for Omitted Input λ Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill $\left(\theta_{t+1}^{N}\right)$	Cognitive Skill $\left(\theta_{t+1}^{C}\right)$
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.8848	0.0276
	(0.021)	(0.013)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.0022	0.9891
	(0.013)	(0.039)
Parental Investment, $\left(\theta_t^I\right)$	0.0797	0.0844
	(0.020)	(0.017)
Omitted Correlated Inputs, λ	0.2835	1.0000
	(0.134)	(normalized)

a. Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let λ denote omitted inputs that are potentially correlated with θ_t. The technology equations are:

$$\theta_{t+1}^{k} = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \gamma_4^k \lambda + \nu_{t+1}^{k}.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k$ and γ_4^k. Note that for identification purposes we normalize $\gamma_4^C = 1$. Investment is normalized on family income.

Table 8
Anchored Technology Equations[a]
Anchoring on Log Earnings and Graduation from High School
Measurement Error is Non-Classical, No Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill		Cognitive Skill	
	(1)	(2)	(4)	(5)
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.8844	0.8843	0.0100	0.0687
	(0.0210)	(0.0210)	(0.0046)	(0.0319)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.0084	0.0012	0.9777	0.9771
	(0.0364)	(0.0053)	(0.0380)	(0.0380)
Parental Investment, $\left(\theta_t^I\right)$	0.0101	0.0079	0.0032	0.0173
	(0.0028)	(0.0022)	(0.0007)	(0.0035)
Maternal Education, *S*	0.0006	0.0004	0.0002	0.0008
	(0.0011)	(0.0009)	(0.0005)	(0.0027)
Maternal Cognitive Skill, *A*	-0.0008	-0.0007	0.0013	0.0068
	(0.0008)	(0.0007)	(0.0005)	(0.0024)

Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let *S* denote mother's education and *A* denote mother's cognitive ability. The technology equations are:

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_1^k S + \psi_2^k A + \eta_{t+1}^k.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_1^k,$ and ψ_2^k in columns (1) through (4). In columns (1) and (3), we anchor the skill factors in log earnings of the child when adult. In columns (2) and (4), we anchor the skill factors in the probability of graduating from high-school using a linear probability model. The investment factor is normalized in family income.

Table 9
Unanchored Stage Specific Technology Equations[a]
Measurement Error is Classical, No Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill $\left(\theta_{t+1}^N\right)$			Cognitive Skill $\left(\theta_{t+1}^C\right)$		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.9849	0.9383	0.7570	0.0605	0.0212	0.0014
	(0.014)	(0.015)	(0.010)	(0.012)	(0.008)	(0.008)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.0508	-0.0415	0.0412	0.9197	0.8845	0.9099
	(0.043)	(0.041)	(0.041)	(0.023)	(0.021)	(0.019)
Parental Investment, $\left(\theta_t^I\right)$	0.0533	0.1067	0.0457	0.1125	0.0364	0.0379
	(0.013)	(0.022)	(0.019)	(0.032)	(0.014)	(0.014)
Maternal Education, *S*	0.0034	-0.0028	0.0138	0.0050	0.0131	0.0021
	(0.007)	(0.007)	(0.008)	(0.010)	(0.012)	(0.014)
Maternal Cognitive Skill, *A*	0.0007	-0.0077	-0.0134	0.0506	0.0044	0.0194
	(0.001)	(0.001)	(0.002)	(0.013)	(0.008)	(0.007)

a. Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let *S* denote mother's education and *A* denote mother's cognitive ability. The technology equations are:

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_{1,t}^k S + \psi_{2,t}^k A + \eta_{t+1}^k.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_{1,t}^k,$ and $\psi_{2,t}^k$. Stage 1 consists of the transition from ages 6-7 to ages 8-9. Stage 2 refers to the transition from ages 8-9 to 10-11. Stage 3 is the transition from ages 10-11 to 12-13.

Table 10
Unanchored Stage Specific Technology Equations[a]
Measurement Error is Non-Classical, No Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill $\left(\theta_{t+1}^N\right)$			Cognitive Skill $\left(\theta_{t+1}^C\right)$		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.9884	0.9427	0.7568	0.0597	0.0211	0.0014
	(0.016)	(0.018)	(0.012)	(0.012)	(0.008)	(0.008)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.0497	-0.0463	0.0418	0.9192	0.8846	0.9101
	(0.046)	(0.042)	(0.042)	(0.024)	(0.023)	(0.021)
Parental Investment, $\left(\theta_t^I\right)$	0.0532	0.1002	0.0435	0.1116	0.0367	0.0378
	(0.013)	(0.022)	(0.019)	(0.033)	(0.002)	(0.001)
Maternal Education, *S*	0.0032	-0.0029	0.0138	-0.0050	0.0131	0.0021
	(0.007)	(0.007)	(0.008)	(0.010)	(0.010)	(0.010)
Maternal Cognitive Skill, *A*	-0.0008	-0.0062	-0.0119	0.0510	0.0045	0.0194
	(0.001)	(0.001)	(0.002)	(0.021)	(0.010)	(0.004)

a. Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let *S* denote mother's education and *A* denote mother's cognitive ability. The technology equations are:

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_{1,t}^k S + \psi_{2,t}^k A + \eta_{t+1}^k.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_{1,t}^k,$ and $\psi_{2,t}^k$. Stage 1 consists of the transition from ages 6-7 to ages 8-9. Stage 2 refers to the transition from ages 8-9 to 10-11. Stage 3 is the transition from ages 10-11 to 12-13.

Table 11
Anchored Stage Specific Technology Equations[a]
Anchor: Log Earnings of the Child Between Ages 23-28
Measurement Error is Classical, No Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Noncognitive Skill $\left(\theta_{t+1}^N\right)$			Cognitive Skill $\left(\theta_{t+1}^C\right)$		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Lagged Noncognitive Skill, $\left(\theta_t^N\right)$	0.9849	0.9383	0.7570	0.0216	0.0076	0.0005
	(0.014)	(0.015)	(0.010)	(0.004)	(0.003)	(0.003)
Lagged Cognitive Skill, $\left(\theta_t^C\right)$	0.1442	-0.1259	0.1171	0.9197	0.8845	0.9099
	(0.120)	(0.115)	(0.115)	(0.023)	(0.021)	(0.019)
Parental Investment, $\left(\theta_t^I\right)$	0.0075	0.0149	0.0064	0.0056	0.0018	0.0019
	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)
Maternal Education, *S*	0.0005	-0.0004	0.0019	-0.0003	0.0007	0.0001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Maternal Cognitive Skill, *A*	0.0001	-0.0011	-0.0019	0.0025	0.0002	0.0010
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)

a. Let $\theta_t = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let *S* denote mother's education and *A* denote mother's cognitive ability. The technology equations are:

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_{1,t}^k S + \psi_{2,t}^k A + \eta_{t+1}^k.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_{1,t}^k,$ and $\psi_{2,t}^k$. Stage 1 consists of the transition from ages 6-7 to ages 8-9. Stage 2 refers to the transition from ages 8-9 to 10-11. Stage 3 is the transition from ages 10-11 to 12-13.

Table 12
The Weights in the Construction of the Investment Factor

	Estimated Weights[a]	Ad Hoc Weights[b]	Share of Total Residual Variance due to Factors[c]	Share of Total Residual Variance due to Uniqueness[d]
Ages 6 and 7				
Log Family Income	0.0787	--	0.1188	0.8812
Number of Books	0.0919	0.1667	0.1359	0.8641
Musical Instrument	0.0917	0.1667	0.1358	0.8642
Newspaper	0.1083	0.1667	0.1564	0.8436
Child has special lessons	0.2251	0.1667	0.2783	0.7217
Child goes to museums	0.2019	0.1667	0.2569	0.7431
Child goes to theater	0.2025	0.1667	0.2575	0.7425
Ages 8 and 9				
Log Family Income	0.0646	--	0.0686	0.9314
Number of Books	0.0987	0.1667	0.1011	0.8989
Musical Instrument	0.1338	0.1667	0.1323	0.8677
Newspaper	0.0828	0.1667	0.0862	0.9138
Child has special lessons	0.1990	0.1667	0.1848	0.8152
Child goes to museums	0.1912	0.1667	0.1789	0.8211
Child goes to theater	0.2299	0.1667	0.2076	0.7924
Ages 10 and 11				
Log Family Income	0.0721	--	0.0537	0.9463
Number of Books	0.1310	0.1667	0.0934	0.9066
Musical Instrument	0.1566	0.1667	0.1097	0.8903
Newspaper	0.0973	0.1667	0.0711	0.9289
Child has special lessons	0.1386	0.1667	0.0983	0.9017
Child goes to museums	0.1785	0.1667	0.1232	0.8768
Child goes to theater	0.2260	0.1667	0.1510	0.8490
Ages 12 and 13				
Log Family Income	0.0862	--	0.0349	0.9651
Number of Books	0.1314	0.1667	0.0523	0.9477
Musical Instrument	0.1109	0.1667	0.0445	0.9555
Newspaper	0.0968	0.1667	0.0390	0.9610
Child has special lessons	0.1036	0.1667	0.0417	0.9583
Child goes to museums	0.1890	0.1667	0.0735	0.9265
Child goes to theater	0.2821	0.1667	0.1059	0.8941

a. Let $\tilde{Y}^I_{k,t}$ denote the k^{th} measurement on parental investment (from the HOME-SF score) with the mean removed, $k = 1,\ldots,m^I_t$. Let θ^I_t denote the investment dynamic factor. Note that $\frac{\tilde{Y}^I_{k,t}}{\alpha^I_{k,t}} = \theta^I_t + \frac{\varepsilon^I_{k,t}}{\alpha^I_{k,t}}$, where $\alpha^I_{k,t}$ is the factor loading and $\varepsilon_{k,t}$ is the measurement error with variance $\sigma^2_{\varepsilon_{k,t}}$. Let $\hat{\alpha}^I_{k,t}$ and $\hat{\sigma}^2_{\varepsilon_{k,t}}$ denote their maximum likelihood estimators, respectively. To construct our estimated weights, we calculate $\hat{\theta}^I_t$ which is the solution of

$$\hat{\theta}^I_t = \arg\min \left\{ \sum_{k=1}^{m^I_t} \left[\frac{\left(\frac{\tilde{Y}^I_{k,t}}{\hat{\alpha}^I_{k,t}} - \theta^I_t \right)^2}{\frac{\left(\hat{\alpha}^I_{k,t}\right)^2}{\hat{\sigma}^2_{\varepsilon_{k,t}}}} \right] \right\}.$$

The solution is

$$\hat{\theta}^I_t = \sum \omega^I_{k,t} \frac{\tilde{Y}^I_{k,t}}{\hat{\alpha}^I_{k,t}},$$

Where the weight $\omega^I_{k,t}$ satisfies $\omega^I_{k,t} = \frac{\left(\hat{\alpha}^I_{k,t}\right)^2 / \hat{\sigma}^2_{\varepsilon_{k,t}}}{\sum_l \left(\hat{\alpha}^I_{l,t}\right)^2 / \hat{\sigma}^2_{\varepsilon_{k,t}}}$.

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

c. Let $\sigma^2_{I_t}$ denote the variance of the investment factor at period t. For each measurement on parental investment k, the total residual variance is $\sigma^2_{k,t} = \left(\alpha^I_{k,t}\right)^2 \sigma^2_{I_t} + \sigma^2_{\varepsilon_{k,t}}$, where $\sigma^2_{\varepsilon_{k,t}}$ is the variance of the uniquenesses in measurement *k* at period *t*. The share of the total residual variance that is due to the factor is $s_{I_t} = \frac{\left(\alpha_{k,t}^{I}\right)^2 \sigma_{I_t}^2}{\sigma_{k,t}^2}$.

d. Analogously, the share of total residual that is due to the uniquenesses is $s_{\varepsilon_{k,t}} = \frac{\sigma_{\varepsilon_{k,t}}^2}{\sigma_{k,t}^2}$.

Table 13
Covariance between Measurement Error and the Dynamic Factors
White Males, CNLSY/79

	Period 1	Period 2	Period 3	Period 4
$Cov\left(\frac{1}{m_t^N}\sum_{k=1}^{m_t^N}\left(\alpha_{t,k}^I-1\right)\theta_t^I+\frac{1}{m_t^N}\sum_{k=1}^{m_t^N}\varepsilon_{t,k}^I,\theta_t^I\right)$	-0.0271	-0.0593	-0.0498	-0.0099
$Cov\left(\frac{1}{m_t^C}\sum_{k=1}^{m_t^C}\left(\alpha_{t,k}^I-1\right)\theta_t^I+\frac{1}{m_t^C}\sum_{k=1}^{m_t^C}\varepsilon_{t,k}^I,\theta_t^I\right)$	0.0113	0.0346	0.0421	0.0441
$Cov\left(\frac{1}{m_t^I}\sum_{k=1}^{m_t^I}\left(\alpha_{t,k}^I-1\right)\theta_t^I+\frac{1}{m_t^I}\sum_{k=1}^{m_t^I}\varepsilon_{t,k}^I,\theta_t^I\right)$	0.0237	0.0216	0.0066	0.0029

Table 14
Estimated Factor Loadings and Standard Errors
White Males, CNLSY/1979

	Period 1	Period 2	Period 3	Period 4
Noncognitive Skills (Normalization: Anti-Social Score)				
Anxiety Score	0.9006	0.8910	0.9364	1.0122
	(0.0231)	(0.0236)	(0.0233)	(0.0234)
Headstrong Score	1.0671	0.9692	0.9590	1.1071
	(0.0366)	(0.0368)	(0.0368)	(0.0372)
Hyperactivity Score	1.0028	0.8980	0.8673	0.9208
	(0.0329)	(0.0331)	(0.0332)	(0.0337)
Peer Conflict Score	0.7647	0.7252	0.7974	0.8472
	(0.0188)	(0.0182)	(0.0189)	(0.0194)
Cognitive Skills (Normalization: PIAT-Math Score)				
Reading Recognition Score	1.2995	1.4878	1.6533	1.8307
	(0.0244)	(0.0274)	(0.0322)	(0.0411)
Parental Investments (Normalization: Family Income)				
Number of Books	0.2710	0.2514	0.6244	0.6797
	(0.0098)	(0.0097)	(0.0163)	(0.0167)
Number of Musical Instruments	0.3908	0.4346	0.5180	0.4622
	(0.0103)	(0.0151)	(0.0158)	(0.0153)
Newspaper Subscriptions	0.4191	0.3559	0.4318	0.4430
	(0.0311)	(0.0308)	(0.0317)	(0.0318)
Special Lessons	0.5546	0.4678	0.4487	0.3899
	(0.0373)	(0.0351)	(0.0350)	(0.0346)
Trips to the Museum	0.9874	0.8619	0.9247	0.9384
	(0.0412)	(0.0427)	(0.0433)	(0.0434)
Trips to the Theater	0.8895	0.8113	0.9764	1.0578
	(0.0277)	(0.0257)	(0.0299)	(0.0331)

Table 15
Fraction of Total Variance Explained by Skill Factor versus Uniqueness
White Males, CNLSY/79

	Share of Total Variance Explained by Factor	Share of Total Variance Explained by Uniqueness
Ages 6-7		
Non-Cognitive Measurements		
Anti-Social Score	0.5321	0.4679
Anxiety Score	0.3673	0.6327
Headstrong Score	0.6289	0.3711
Hyperactivity Score	0.5500	0.4500
Peer Conflict Score	0.2073	0.7927
Cognitive Measurements		
PIAT-Math	0.3512	0.6488
PIAT-Reading Recognition	0.9473	0.0527
Ages 8-9		
Non-Cognitive Measurements		
Anti-Social Score	0.5409	0.4591
Anxiety Score	0.3983	0.6017
Headstrong Score	0.5620	0.4380
Hyperactivity Score	0.4371	0.5629
Peer Conflict Score	0.2005	0.7995
Cognitive Measurements		
PIAT-Math	0.3938	0.6062
PIAT-Reading Recognition	0.9119	0.0881
Ages 10-11		
Non-Cognitive Measurements		
Anti-Social Score	0.5266	0.4734
Anxiety Score	0.4460	0.5540
Headstrong Score	0.5368	0.4632
Hyperactivity Score	0.4286	0.5714
Peer Conflict Score	0.2738	0.7262
Cognitive Measurements		
PIAT-Math	0.3835	0.6165
PIAT-Reading Recognition	0.9181	0.0819
Ages 12-13		
Non-Cognitive Measurements		
Anti-Social Score	0.5040	0.4960
Anxiety Score	0.4803	0.5197
Headstrong Score	0.6324	0.3676

Hyperactivity Score	0.4064	0.5936
Peer Conflict Score	0.2613	0.7387
Cognitive Measurements		
PIAT-Math	0.3561	0.6439
PIAT-Reading Recognition	0.9149	0.0851

Note: let $Y_{j,t}^{k}$ denote the j^{th} measurement on skill k (k = Noncognitive/Cognitive). Let θ_t^k denote the skill k dynamic factor. Note that $Y_{j,t}^{k} = \alpha_{j,t}^{k}\theta_t^I + \varepsilon_{j,t}^I$, where $\alpha_{j,t}^{k}$ is the factor loading and $\varepsilon_{k,t}$ is the uniqueness with variance $\sigma_{\varepsilon_{k,t}}^2$. The total residual variance is $\sigma_{k,t}^2 = \left(\alpha_{j,t}^k\right)^2 Var\left(\theta_t^k\right) + \sigma_{\varepsilon_{k,t}}^2$.

The share of the total residual variance that is due to the factor is $s_{\theta_t^k} = \dfrac{\left(\alpha_{j,t}^k\right)^2 Var\left(\theta_t^k\right)}{\sigma_{k,t}^2}$.

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

Table 16
OLS Estimation of the Technology Equations
Measurement Error is Classical, Absence of Omitted Inputs Correlated with θ_t
White Males, CNLSY/79

Dependent Variable	Antisocial Score (*t* + 1)	PIAT Math (*t*+1)
Antisocial Score, *t*	0.6431	0.0333
	(0.0165)	(0.0096)
PIAT Math, *t*	0.0933	0.5909
	(0.0317)	(0.0184)
HOME Score, *t*	0.0147	-0.0137
	(0.0059)	(0.0034)
Maternal Education	0.0358	0.0208
	(0.0091)	(0.0053)
Maternal ASVAB Arithmetics	-0.0254	0.0658
	(0.0190)	(0.0110)

Table 17A
The Percentage Impact on Log Earnings at Age 23 of an Exogenous Increase by Ten Percent in Investments at Different Periods
White Males, CNLSY/79

Total Percentage Impact on Earnings	Percentage Impact on Log Earnings Exclusively through Cognitive Skills	Percentage Impact on Log Earnings Exclusively through Noncognitive Skills
	Period 1	
0.2487	0.1247	0.1240
(0.0302)	(0.0151)	(0.0150)
	Period 2	
0.3065	0.0445	0.2620
(0.0358)	(0.0052)	(0.0306)
	Period 3	
0.2090	0.0540	0.1550
(0.0230)	(0.0059)	(0.0170)

Note: Let $\tilde{Y}^{I}_{j,t}$ denote the j^{th} measurement on the parental investment dynamic factor θ^{I}_{t} with the mean removed. We obtain the predicted parental investment $\hat{\theta}^{I}_{t}$ by applying the weights reported in Table 12 and measurements in the following way:

$$\hat{\theta}^{I}_{t} = \sum_{j=1}^{m^{I}_{t}} \omega_{j,t} \tilde{Y}^{I}_{j,t}.$$

We then simulate the model and obtain the adult level of cognitive and non-cognitive skills. Using the anchoring equation, we then predict log earnings, log E. We then perform a counterfactual simulation. We investigate the level of adult skills if investments at different periods were increased by ten percent and we check the impact on log earnings, log E_{τ}, where E_{τ} is the counterfactual earnings if investment in period τ were ten percent higher, $\tau = 1, 2, 3$. In this table, we report the percentage change in earnings, that is log $E_{\tau} - \log E$.

Table 17B
The Percentage Impact on the Probability of Graduating from High School of an Exogenous Increase by Ten Percent in Investments at Different Periods
White Males, CNLSY/79

Total Percentage Impact	Percentage Impact through Cognitive Skills	Percentage Impact Exclusively through Noncognitive Skills
	Period 1	
0.6441	0.5480	0.0961
(0.0.789)	(0.0672)	(0.0118)
	Period 2	
0.3980	0.1951	0.2029
(0.0466)	(0.0229)	(0.0238)
	Period 3	
0.3565	0.2366	0.1198
(0.0389)	(0.0258)	(0.0131)

Note: Let $\tilde{Y}_{j,t}^{I}$ denote the j^{th} measurement on the parental investment dynamic factor θ_t^I with the mean removed. We obtain the predicted parental investment $\hat{\theta}_t^I$ by applying the weights reported in Table 12 and measurements in the following way:

$$\hat{\theta}_t^I = \sum_{j=1}^{m_t^I} \omega_{j,t} \tilde{Y}_{j,t}^{I}.$$

We then simulate the model and obtain the adult level of cognitive and non-cognitive skills. Using the anchoring equation, we then predict the probability of graduating from high school, p. We then perform a counterfactual simulation. We investigate the level of adult skills if investments at different periods were increased by ten percent and we check the impact on the probability of graduating from high school, p_τ, where p_τ is the counterfactual graduation probability if investment in period τ were ten percent higher. In this table, we report the percentage change in probability of graduating, that is $\log p_\tau - \log p$. Standard errors in parentheses.