ESTIMATING THE TECHNOLOGY OF COGNITIVE AND NONCOGNITIVE SKILL FORMATION

FLAVIO CUNHA
University of Pennsylvania, Philadelphia, PA 19102, U.S.A.

JAMES J. HECKMAN
University of Chicago, Chicago, IL 60637, U.S.A. and Geary Institute,
University College Dublin, Dublin 4, Ireland and American Bar
Foundation, Chicago, IL 60611, U.S.A.

SUSANNE M. SCHENNACH
University of Chicago, Chicago, IL 60637, U.S.A.
ESTIMATING THE TECHNOLOGY OF COGNITIVE AND NONCOGNITIVE SKILL FORMATION

BY FLAVIO CUNHA, JAMES J. HECKMAN, AND SUSANNE M. SCHENNACH

This paper formulates and estimates multistage production functions for children’s cognitive and noncognitive skills. Skills are determined by parental environments and investments at different stages of childhood. We estimate the elasticity of substitution between investments in one period and stocks of skills in that period to assess the benefits of early investment in children compared to later remediation. We establish nonparametric identification of a general class of production technologies based on nonlinear factor models with endogenous inputs. A by-product of our approach is a framework for evaluating childhood and schooling interventions that does not rely on arbitrarily scaled test scores as outputs and recognizes the differential effects of the same bundle of skills in different tasks. Using the estimated technology, we determine optimal targeting of interventions to children with different parental and personal birth endowments. Substitutability decreases in later stages of the life cycle in the production of cognitive skills. It is roughly constant across stages of the life cycle in the production of noncognitive skills. This finding has important implications for the design of policies that target the disadvantaged. For most configurations of disadvantage it is optimal to invest relatively more in the early stages of childhood than in later stages.

KEYWORDS: Cognitive skills, noncognitive skills, dynamic factor analysis, endogeneity of inputs, anchoring test scores, parental influence.

© 2010 The Econometric Society DOI: 10.3982/ECTA6551
A large body of research documents the importance of cognitive skills in producing social and economic success. An emerging body of research establishes the parallel importance of noncognitive skills, that is, personality, social, and emotional traits. Understanding the factors that affect the evolution of cognitive and noncognitive skills is important for understanding how to promote successful lives.

This paper estimates the technology governing the formation of cognitive and noncognitive skills in childhood. We establish identification of general nonlinear factor models that enable us to determine the technology of skill formation. Our multistage technology captures different developmental phases in the life cycle of a child. We identify and estimate substitution parameters that determine the importance of early parental investment for subsequent lifetime achievement, and the costliness of later remediation if early investment is not undertaken.

Cunha and Heckman (2007) present a theoretical framework that organizes and interprets a large body of empirical evidence on child and animal development. Cunha and Heckman (2008) estimate a linear dynamic factor model that exploits cross-equation restrictions (covariance restrictions) to secure identification of a multistage technology for child investment. With enough measurements relative to the number of latent skills and types of investment, it is possible to identify the latent state space dynamics that generate the evolution of skills.

The linear technology used by Cunha and Heckman (2008) imposes the assumption that early and late investments are perfect substitutes over the feasible set of inputs. This paper identifies a more general nonlinear technology by extending linear state space and factor analysis to a nonlinear setting. This extension allows us to identify crucial elasticity of substitution parameters that govern the trade-off between early and late investments in producing adult skills.

---


3 See Heckman, Stixrud, and Urzua (2006), Borghans, Duckworth, Heckman, and ter Weel (2008), and the references they cite. See also the special issue of the Journal of Human Resources, 43, Fall 2008 (Kniesner and ter Weel (2008)) on noncognitive skills.


5 This evidence is summarized in Knudsen, Heckman, Cameron, and Shonkoff (2006a) and Cunha and Heckman (2008).

Drawing on the analyses of Schennach (2004a) and Hu and Schennach (2008), we establish identification of the technology of skill formation. We relax the strong independence assumptions for error terms in the measurement equations that are maintained in Cunha and Heckman (2008) and Carneiro, Hansen, and Heckman (2003). The assumption of linearity of the technology in inputs that is used by Cunha and Heckman (2008) and Todd and Wolpin (2003, 2005) is not required because we allow inputs to interact in producing outputs. We generalize the factor-analytic index function models used by Carneiro, Hansen, and Heckman (2003) to allow for more general functional forms for measurement equations. We solve the problem of defining a scale for the output of childhood investments by anchoring test scores using adult outcomes of the child, which have a well defined cardinal scale. We determine the latent variables that generate test scores by estimating how these latent variables predict adult outcomes.7 Our approach sets the scale of test scores and latent variables in an interpretable metric. Using this metric, analysts can meaningfully interpret changes in output and conduct interpretable value-added analyses.8 We also solve the problem of missing inputs in estimating technologies in a way that is much more general than the widely used framework of Olley and Pakes (1996) that assumes perfect proxies for latent factors. We allow for imperfect proxies and establish that measurement error is substantial in the data analyzed in this paper.

The plan of this paper is as follows. Section 2 briefly summarizes the previous literature to motivate our contribution to it. Section 3 presents our identification analysis. Section 4 discusses the data used to estimate the model, our estimation strategy, and the model estimates. Section 5 concludes. Extensive appendices comprise the Supplemental Material (Cunha, Heckman, and Schennach (2010)).

2. A MODEL OF COGNITIVE AND NONCOGNITIVE SKILL FORMATION

We analyze a model with multiple periods of childhood, \( t \in \{1, 2, \ldots, T\} \), \( T \geq 2 \), followed by \( A \) periods of adult working life, \( t \in \{T + 1, T + 2, \ldots, T + A\} \). The \( T \) childhood periods are divided into \( S \) stages of development, \( s \in \{1, \ldots, S\} \) with \( S \leq T \). Adult outcomes are produced by cognitive skills, \( \theta_{C,T+1} \), and noncognitive skills, \( \theta_{N,T+1} \), at the beginning of the adult years.9 Denote parental investments at age \( t \) in child skill \( k \) by \( I_{k,t} \), \( k \in \{C, N\} \).

---

7Cawley, Heckman, and Vytlacil (1999) anchor test scores in earnings outcomes.
8Cunha and Heckman (2008) develop a class of anchoring functions invariant to affine transformations. This paper develops a more general class of monotonic transformations and presents a new analysis of joint identification of the anchoring equations and the technology of skill formation.
9This model generalizes the model of Becker and Tomes (1986), who assume only one period of childhood (\( T = 1 \)) and consider one output associated with “human capital” that can be interpreted as a composite of cognitive (\( C \)) and noncognitive (\( N \)) skills. We do not model post-childhood investment.
Skills evolve in the following way. Each agent is born with initial conditions \( \theta_1 = (\theta_{C,1}, \theta_{N,1}) \). Family environments and genetic factors may influence these initial conditions (see Olds (2002) and Levitt (2003)). We denote by \( \theta_p = (\theta_{C,p}, \theta_{N,p}) \) parental cognitive and noncognitive skills, respectively. \( \theta_t = (\theta_{C,t}, \theta_{N,t}) \) denotes the vector of skill stocks in period \( t \). Let \( \eta_t = (\eta_{C,t}, \eta_{N,t}) \) denote shocks and/or unobserved inputs that affect the accumulation of cognitive and noncognitive skills, respectively. The technology of production of skill \( k \) in period \( t \) and developmental stage \( s \) depends on the stock of skills in period \( t \), investment at \( t \), \( I_{k,t} \), parental skills, \( \theta_p \), shocks in period \( t \), \( \eta_{k,t} \), and the production function at stage \( s \),

\[
\theta_{k,t+1} = f_{k,s}(\theta_t, I_{k,t}, \theta_p, \eta_{k,t})
\]

for \( k \in \{C, N\}, \ t \in \{1, 2, \ldots, T\}, \ \text{and} \ s \in \{1, \ldots, S\} \). We assume that \( f_{k,s} \) is monotone increasing in its arguments, twice continuously differentiable, and concave in \( I_{k,t} \). In this model, stocks of current period skills produce next period skills and affect the current period productivity of investments. Stocks of cognitive skills can promote the formation of noncognitive skills and vice versa because \( \theta_t \) is an argument of (2.1).

Direct complementarity between the stock of skill \( l \) and the productivity of investment \( I_{k,t} \) in producing skill \( k \) in period \( t \) arises if

\[
\frac{\partial^2 f_{k,s}(\cdot)}{\partial I_{k,t} \partial \theta_{l,t}} > 0, \quad t \in \{1, \ldots, T\}, \ l, k \in \{C, N\}.
\]

Period \( t \) stocks of abilities and skills promote the acquisition of skills by making investment more productive. Students with greater early cognitive and noncognitive abilities are more efficient in later learning of both cognitive and noncognitive skills. The evidence from the early intervention literature suggests that the enriched early environments of the Abecedarian, Perry, and Chicago Child–Parent Center (CPC) programs promoted greater efficiency in learning in schools and reduced problem behaviors.\(^{10}\)

Adult outcome \( j, Q_j \), is produced by a combination of different skills at the beginning of period \( T + 1 \):

\[
Q_j = g_j(\theta_{C,T+1}, \theta_{N,T+1}), \quad j \in \{1, \ldots, J\}.
\]

\(^{10}\)See, for example, Cunha, Heckman, Lochner, and Masterov (2006), Heckman, Malofeeva, Pinto, and Savelyev (2009), Heckman, Moon, Pinto, Savelyev, and Yavitz (2010a, 2010b), and Reynolds and Temple (2009).

\(^{11}\)To focus on the main contribution of this paper, we focus on investment in children. Thus we assume that \( \theta_{T+1} \) is the adult stock of skills for the rest of life, contrary to the evidence reported in Borghans, Duckworth, Heckman, and ter Weel (2008). The technology could be extended to accommodate adult investment as in Ben-Porath (1967) or its generalization Heckman, Lochner, and Taber (1998).
These outcome equations capture the twin concepts that both cognitive and noncognitive skills matter for performance in most tasks in life, and have different effects in different tasks in the labor market and in other areas of social performance. Outcomes include test scores, schooling, wages, occupational attainment, hours worked, criminal activity, and teenage pregnancy.

In this paper, we identify and estimate a constant elasticity of substitution (CES) version of technology (2.1) where we assume that $\theta_{C,t}$, $\theta_{N,t}$, $I_{C,t}$, $I_{N,t}$, $\theta_{C,P}$, and $\theta_{N,P}$ are scalars. Outputs of skills at stage $s$ are governed by

\begin{equation}
\theta_{C,t+1} = \left[ \gamma_{s,C,1}\theta_{\phi_{s,C}} + \gamma_{s,C,2}\theta_{\phi_{s,N}} + \gamma_{s,C,3}\theta_{\phi_{s,C}} + \gamma_{s,C,4}\theta_{\phi_{s,P}} + \gamma_{s,C,5}\theta_{\phi_{s,N}} \right]^{1/\phi_{s,C}}
\end{equation}

and

\begin{equation}
\theta_{N,t+1} = \left[ \gamma_{s,N,1}\theta_{\phi_{s,N}} + \gamma_{s,N,2}\theta_{\phi_{s,N}} + \gamma_{s,N,3}\theta_{\phi_{s,N}} + \gamma_{s,N,4}\theta_{\phi_{s,P}} + \gamma_{s,N,5}\theta_{\phi_{s,N}} \right]^{1/\phi_{s,N}},
\end{equation}

where $\gamma_{s,k,l} \in [0,1]$, $\sum_{l} \gamma_{s,k,l} = 1$ for $k \in \{C, N\}$, $l \in \{1, \ldots, 5\}$, $t \in \{1, \ldots, T\}$, and $s \in \{1, \ldots, S\}$. $1/(1 - \phi_{s,k})$ is the elasticity of substitution in the inputs producing $\theta_{k,t+1}$, where $\phi_{s,k} \in (\infty, 1]$ for $k \in \{C, N\}$. It is a measure of how easy it is to compensate for low levels of stocks $\theta_{C,t}$ and $\theta_{N,t}$ inherited from the previous period with current levels of investment $I_{C,t}$ and $I_{N,t}$. For the moment, we ignore the shocks $\eta_{k,t}$ in (2.1), although they play an important role in our empirical analysis.

A CES specification of adult outcomes is

\begin{equation}
Q_{j} = \{ \rho_{j}(\theta_{C,T+1})^{\phi_{Q,j}} + (1 - \rho_{j})(\theta_{N,T+1})^{\phi_{Q,j}} \}^{1/\phi_{Q,j}},
\end{equation}

where $\rho_{j} \in [0,1]$ and $\phi_{Q,j} \in (\infty, 1]$ for $j = 1, \ldots, J$. $1/(1 - \phi_{Q,j})$ is the elasticity of substitution between different skills in the production of outcome $j$. The ability of noncognitive skills to compensate for cognitive deficits in producing adult outcomes is governed by $\phi_{Q,j}$. The importance of cognition in producing output in task $j$ is governed by the share parameter $\rho_{j}$.

To gain some insight into this model, consider a special case investigated in Cunha and Heckman (2007), where childhood lasts two periods ($T = 2$), there is one adult outcome ("human capital") so $J = 1$, and the elasticities of substitution are the same across technologies (2.3) and (2.4) and in the outcome (2.5), so $\phi_{s,C} = \phi_{s,N} = \phi_{Q} = \phi$ for all $s \in \{1, \ldots, S\}$. Assume that there is one investment good in each period that increases both cognitive and noncognitive skills, though not necessarily by the same amount ($I_{k,t} \equiv I_{t}$, $k \in \{C, N\}$). In this case, the adult outcome is a function of investments, initial endowments, and parental characteristics, and can be written as

\begin{equation}
Q = [\tau_{1}I_{1}^{\phi} + \tau_{2}I_{2}^{\phi} + \tau_{3}\theta_{C,1}^{\phi} + \tau_{4}\theta_{N,1}^{\phi} + \tau_{5}\theta_{C,P}^{\phi} + \tau_{6}\theta_{N,P}^{\phi}]^{1/\phi},
\end{equation}
where $\tau_i$ for $i = 1, \ldots, 6$ depend on the parameters of equations (2.3)–(2.5). Cunha and Heckman (2007) analyze the optimal timing of investment using a special version of the technology embodied in (2.6).

Let $R(Q) = \sum_{t=2}^{4+2} (\frac{1}{1+r})^t wQ$ denote the net present value of the child’s future income computed with respect to the date of birth, and $w$ is the return per unit $Q$. Parents have resources $M$ that they use to invest in period 1, $I_1$, and period 2, $I_2$. The objective of the parent is to maximize the net present value of the child’s future income given parental resource constraints. Assuming an interior solution, that the price of investment in period 1 is one, the relative price of investment in period 2 is $1/(1+r)$, the optimal ratio of period 1 investment to period 2 investment is

$$\log \left( \frac{I_1}{I_2} \right) = \left( \frac{1}{1-\phi} \right) \left[ \log \left( \frac{\tau_1}{\tau_2} \right) - \log(1+r) \right]. \tag{2.7}$$

Figure 1 plots the ratio of early to late investment as a function of $\tau_1/\tau_2$ for different values of $\phi$ assuming $r = 0$. Ceteris paribus, the higher $\tau_1$ relative to $\tau_2$, the higher first period investment should be relative to second period investment. The parameters $\tau_1$ and $\tau_2$ are determined in part by the productivity of investments in producing skills, which are generated by the technology parameters $\gamma_{s,k,3}$ for $s \in \{1, 2\}$ and $k \in \{C, N\}$. They also depend on the relative importance of cognitive skills, $\rho$, versus noncognitive skills, $1-\rho$, in producing the adult outcome $Q$. Ceteris paribus, if $\tau_1/\tau_2 > (1+r)$, the higher the CES complementarity (i.e., the lower $\phi$), the greater is the ratio of optimal early to late investment. The greater $r$, the smaller should be the optimal ratio of early to late investment. In the limit, if investments complement each other strongly, optimality implies that they should be equal in both periods.

This example builds intuition about the importance of the elasticity of substitution in determining the optimal timing of life-cycle investments. However, it oversimplifies the analysis of skill formation. It is implausible that the elasticity of substitution between skills in producing adult outcomes $(1/(1-\phi Q))$ is the same as the elasticity of substitution between inputs in producing skills, and that a common elasticity of substitution governs the productivity of inputs in producing both cognitive and noncognitive skills.

Our analysis allows for multiple adult outcomes and multiple skills. We allow the elasticities of substitution governing the technologies for producing cognitive and noncognitive skills to differ at different stages of the life cycle, and for both to be different from the elasticities of substitution for cognitive and noncognitive skills in producing adult outcomes. We test and reject the assumption that $\phi_{s,C} = \phi_{s,N}$ for $s \in \{1, \ldots, S\}$.}

---

\(^{12}\)See Appendix A1 for the derivation of this expression in terms of the parameters of equations (2.3)–(2.5).
3. IDENTIFYING THE TECHNOLOGY USING DYNAMIC FACTOR MODELS

Identifying and estimating technology (2.1) is challenging. Both inputs and outputs can only be proxied. Measurement error in general nonlinear specifications of technology (2.1) raises serious econometric challenges. Inputs may be endogenous and the unobservables in the input equations may be correlated with unobservables in the technology equations.

This paper addresses these challenges. Specifically, we execute the following tasks: (i) Determine how stocks of cognitive and noncognitive skills at date $t$ affect the stocks of skills at date $t + 1$, identifying both self-productivity (the effects of $\theta_{N,t}$ on $\theta_{N,t+1}$ and of $\theta_{C,t}$ on $\theta_{C,t+1}$) and cross-productivity (the effects of $\theta_{C,t}$ on $\theta_{N,t+1}$ and of $\theta_{N,t}$ on $\theta_{C,t+1}$) at each stage of the life cycle. (ii) Develop a nonlinear dynamic factor model where $(\theta_t, I_t, \theta_P)$ is proxied by vectors of measurements which include test scores and input measures as well as outcome measures. In our analysis, test scores and personality evaluations are indicators of latent skills. Parental inputs are indicators of latent
investment. We account for measurement error in these proxies. (iii) Estimate the elasticities of substitution for the technologies governing the production of cognitive and noncognitive skills. (iv) Anchor the scale of test scores using adult outcome measures instead of relying on test scores as measures of output. This allows us to avoid relying on arbitrary test scores as measurements of output. Any monotonic function of a test score is a valid test score. (v) Account for the endogeneity of parental investments when parents make child investment decisions in response to the characteristics of the child that may change over time as the child develops and as new information about the child is revealed.

Our analysis of identification proceeds in the following way. We start with a model where measurements are linear and separable in the latent variables, as in Cunha and Heckman (2008). We establish identification of the joint distribution of the latent variables without imposing conventional independence assumptions about measurement errors. With the joint distribution of latent variables in hand, we nonparametrically identify technology (2.1) given alternative assumptions about \( \eta_{k,t} \). We then extend this analysis to identify nonparametric measurement and production models. We anchor the latent variables in adult outcomes to make their scales interpretable. Finally, we account for endogeneity of inputs in the technology equations and model investment behavior.

### 3.1. Identifying the Distribution of the Latent Variables

We use a general notation for all measurements to simplify the econometric analysis. Let \( Z_{a,k,t,j} \) be the \( j \)th measurement at time \( t \) on measure of type \( a \) for factor \( k \). We have measurements on test scores and parental and teacher assessments of skills (\( a = 1 \)) on investment (\( a = 2 \)) and on parental endowments (\( a = 3 \)). Each measurement has a cognitive and noncognitive component, so \( k \in \{ C, N \} \). We initially assume that measurements are additively separable functions of the latent factors \( \theta_{k,t} \) and \( I_{k,t} \):

\[
Z_{1,k,t,j} = \mu_{1,k,t,j} + \alpha_{1,k,t,j} \theta_{k,t} + \varepsilon_{1,k,t,j},
\]

\[
Z_{2,k,t,j} = \mu_{2,k,t,j} + \alpha_{2,k,t,j} I_{k,t} + \varepsilon_{2,k,t,j},
\]

where \( E(\varepsilon_{a,k,t,j}) = 0, j \in \{1, \ldots, M_{a,k,t}\}, t \in \{1, \ldots, T\}, k \in \{C, N\}, a \in \{1, 2\} \) and where \( \varepsilon_{a,k,t,j} \) are uncorrelated across the \( j \). Assuming that parental endowments are measured only once in period \( t = 1 \), we write

\[
Z_{3,k,1,j} = \mu_{3,k,1,j} + \alpha_{3,k,1,j} \theta_{k,P} + \varepsilon_{3,k,1,j},
\]

\[
E(\varepsilon_{3,k,1,j}) = 0, \quad j \in \{1, \ldots, M_{3,k,1}\} \text{ and } k \in \{C, N\}.
\]

\[13\] An economic model that rationalizes the investment measurement equations in terms of family inputs is presented in Appendix A2. See also Cunha and Heckman (2008).
The $\alpha_{a,k,t,j}$ are factor loadings. The parameters and variables are defined conditional on $X$. To reduce the notational burden, we keep $X$ implicit. Following standard conventions in factor analysis, we set the scale of the factors by assuming $\alpha_{a,k,t,j} = 1$ and normalize $E(\theta_{k,t}) = 0$ and $E(I_{k,t}) = 0$ for all $k \in \{C,N\}$, $t = 1, \ldots, T$. Separability makes the identification analysis transparent. We consider a more general nonseparable model below. Given measurements $Z_{a,k,t,j}$, we can identify the mean functions $\mu_{a,k,t,j}$, $a \in \{1, 2, 3\}$, $t \in \{1, \ldots, T\}$, $k \in \{C,N\}$, which may depend on the $X$.

### 3.2. Identification of the Factor Loadings and of the Joint Distributions of the Latent Variables

We first establish identification of the factor loadings under the assumption that the $\varepsilon_{a,k,t,j}$ are uncorrelated across $t$ and that the analyst has at least two measures of each type of child skills and investments in each period $t$, where $T \geq 2$. Without loss of generality, we focus on $\alpha_{1,C,t,j}$ and note that similar expressions can be derived for the loadings of the other latent factors.

Since $Z_{1,C,t,1}$ and $Z_{1,C,t+1,1}$ are observed, we can compute $\operatorname{Cov}(Z_{1,C,t,1}, Z_{1,C,t+1,1})$ from the data. Because of the normalization $\alpha_{1,C,t,1} = 1$ for all $t$, we obtain

\[ \operatorname{Cov}(Z_{1,C,t,1}, Z_{1,C,t+1,1}) = \operatorname{Cov}(\theta_{C,t}, \theta_{C,t+1}). \]

In addition, we can compute the covariance of the second measurement on cognitive skills at period $t$ with the first measurement on cognitive skills at period $t+1$:

\[ \operatorname{Cov}(Z_{1,C,t,2}, Z_{1,C,t+1,1}) = \alpha_{1,C,t,2} \operatorname{Cov}(\theta_{C,t}, \theta_{C,t+1}). \]

If $\operatorname{Cov}(\theta_{C,t}, \theta_{C,t+1}) \neq 0$, we can identify the loading $\alpha_{1,C,t,2}$ from the ratio of covariances

\[ \frac{\operatorname{Cov}(Z_{1,C,t,2}, Z_{1,C,t+1,1})}{\operatorname{Cov}(Z_{1,C,t,1}, Z_{1,C,t+1,1})} = \alpha_{1,C,t,2}. \]

\[ 14 \text{This formulation assumes that measurements } a \in \{1, 2, 3\} \text{ proxy only one factor. This is not strictly required for identification. One can identify the correlated factor model if there is one measurement for each factor that depends solely on the one factor, and standard normalizations and rank conditions are imposed. The other measurements can be generated by multiple factors. This follows from the analysis of Anderson and Rubin (1956), who give precise conditions for identification in factor models. Carneiro, Hansen, and Heckman (2003) consider alternative specifications. The key idea in classical factor approaches is one normalization of the factor loading for each factor in one measurement equation to set the scale of the factor and at least one measurement dedicated to each factor.} \]

\[ 15 \text{In our framework, parental skills are assumed to be constant over time as a practical matter because we only observe parental skills once.} \]
If there are more than two measures of cognitive skill in each period \( t \), we can identify \( \alpha_{1,t,j} \) for \( j \in \{2, 3, \ldots, M_{1,t}\} \), \( t \in \{1, \ldots, T\} \) up to the normalization \( \alpha_{1,C,t,1} = 1 \). The assumption that the \( \epsilon_{a,k,t,j} \) are uncorrelated across \( t \) is then no longer necessary. Replacing \( Z_{1,C,t-1,1} \) by \( Z_{a',k',t,3} \) for some \((a', k', t')\) which may or may not be equal to \((1, C, t)\), we may proceed in the same fashion.\(^{16}\) Note that the same third measurement \( Z_{a',k',t,3} \) can be reused for all \( a, t \), and \( k \), implying that in the presence of serial correlation, the total number of measurements needed for identification of the factor loadings is \( 2L + 1 \) if there are \( L \) factors.

Once the parameters \( \alpha_{1,C,t,j} \) are identified, we can rewrite (3.1), assuming \( \alpha_{1,C,t,j} \neq 0 \), as

\[
\begin{align*}
\frac{Z_{1,C,t,j}}{\alpha_{1,C,t,j}} &= \frac{\mu_{1,C,t,j}}{\alpha_{1,C,t,j}} + \theta_{C,t} + \frac{\epsilon_{1,C,t,j}}{\alpha_{1,C,t,j}}, \\
&\quad j \in \{1, 2, \ldots, M_{1,C,t}\}.
\end{align*}
\]

In this form, it is clear that the known quantities \( Z_{1,C,t,j}/\alpha_{1,C,t,j} \) play the role of repeated error-contaminated measurements of \( \theta_{C,t} \). Collecting results for all \( t = 1, \ldots, T \), we can identify the joint distribution of \( \{\theta_{C,t}\}_t \). Proceeding in a similar fashion for all types of measurements, \( a \in \{1, 2, 3\} \), on abilities \( k \in \{C, N\} \), using the analysis in Schennach (2004a, 2004b), we can identify the joint distribution of all the latent variables. Define the matrix of latent variables by \( \theta \), where

\[
\theta = \{\theta_{C,t}\}_t, \{\theta_{N,t}\}_t, \{I_{C,t}\}_t, \{I_{N,t}\}_t, \theta_{C,P}, \theta_{N,P}.
\]

Thus, we can identify the joint distribution of \( \theta \), \( p(\theta) \).

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

\[
W_i = \left( \begin{array}{c}
Z_{1,C,i,1} \\
Z_{1,C,i,2} \\
Z_{3,C,i,1,1} \\
Z_{3,C,i,1,2}
\end{array} \right), \quad i \in \{1, 2\}.
\]

\(^{16}\) The idea is to write

\[
\frac{\text{Cov}(Z_{1,C,t,2}, Z_{a',k',t,3})}{\text{Cov}(Z_{1,C,t,1}, Z_{a',k',t,3})} = \frac{\alpha_{1,C,t,2, a',k',t,3}}{\alpha_{1,C,t,1, a',k',t,3}} \frac{\text{Cov}(\theta_{C,t}, \theta_{k',t'})}{\text{Cov}(\theta_{C,t}, \theta_{k',t'})} = \frac{\alpha_{1,C,t,2}}{\alpha_{1,C,t,1}} = \alpha_{1,C,t,2}.
\]

This only requires uncorrelatedness across different \( j \) but not across \( t \).
These vectors consist of the first and the second measurements for each factor, respectively. The corresponding measurement errors are

$$\omega_i = \left( \begin{array}{c} \varepsilon_{1,C,i,t} \\ \alpha_{1,C,i,t} \\ \varepsilon_{2,C,i,t} \\ \alpha_{2,C,i,t} \\ \varepsilon_{3,C,i,t} \\ \alpha_{3,C,i,t} \end{array} \right)_{t=1}^{T},$$

$$\omega_i = \left( \begin{array}{c} \varepsilon_{1,N,i,t} \\ \alpha_{1,N,i,t} \\ \varepsilon_{2,N,i,t} \\ \alpha_{2,N,i,t} \end{array} \right)_{t=1}^{T},$$

$$\varepsilon_{3,N,1,i} / \alpha_{3,N,1,i} , \quad i \in \{1, 2\}.$$

Identification of the distribution of $\theta$ is obtained from the following theorem. Let $L$ denote the total number of latent factors, which in our case is $4T + 2$.

**Theorem 1:** Let $W_1, W_2, \theta, \omega_1,$ and $\omega_2$ be random vectors taking values in $\mathbb{R}^L$ and related through

$$W_1 = \theta + \omega_1,$$

$$W_2 = \theta + \omega_2.$$

If (i) $E[\omega_1|\theta, \omega_2] = 0$ and (ii) $\omega_2$ is independent from $\theta$, then the density of $\theta$ can be expressed in terms of observable quantities as:

$$p_\theta(\theta) = (2\pi)^{-L} \int e^{-i\chi \cdot \theta} \exp \left( \int_0^\chi \frac{E[iW_1e^{i\zeta W_2}]}{E[e^{i\zeta W_2}]} d\zeta \right) d\chi,$$

where in this expression $i = \sqrt{-1}$, provided that all the requisite expectations exist and $E[e^{i\zeta W_2}]$ is nonvanishing. Note that the innermost integral is the integral of a vector-valued field along a piecewise smooth path joining the origin and the point $\chi \in \mathbb{R}^L$, while the outermost integral is over the whole $\mathbb{R}^L$ space. If $\theta$ does not admit a density with respect to the Lebesgue measure, $p_\theta(\theta)$ can be interpreted within the context of the theory of distributions. If some elements of $\theta$ are perfectly measured, one may simply set the corresponding elements of $W_1$ and $W_2$ to be equal. In this way, the joint distribution of mismeasured and perfectly measured variables is identified.

For the proof, see Appendix A3.1.17

The striking improvement in this analysis over the analysis of Cunha and Heckman (2008) is that identification can be achieved under much weaker conditions regarding measurement errors—far fewer independence assumptions are needed. The asymmetry in the analysis of $\omega_1$ and $\omega_2$ generalizes previous analysis which treats these terms symmetrically. It gives the analyst a more flexible toolkit for the analysis of factor models. For example, our analysis allows analysts to accommodate heteroscedasticity in the distribution of $\omega_1$ that may

17The results of Theorem 1 are sketched informally in Schennach (2004a, footnote 11).
depend on $\omega_2$ and $\theta$. It also allows for potential correlation of components within the vectors $\omega_1$ and $\omega_2$, thus permitting serial correlation within a given set of measurements.

The intuition for identification in this paper, as in all factor analyses, is that the signal is common to multiple measurements, but the noise is not. To extract the noise from the signal, the disturbances have to satisfy some form of orthogonality with respect to the signal and with respect to each other. These conditions are various uncorrelatedness assumptions, conditional mean assumptions, or conditional independence assumptions. They are used in various combinations in Theorem 1, in Theorem 2 below, and in other results in this paper.

### 3.3. The Identification of a General Measurement Error Model

In this section, we extend the previous analysis for linear factor models to consider a measurement model of the general form

$$Z_j = a_j(\theta, \epsilon_j) \quad \text{for} \quad j \in \{1, \ldots, M\},$$

where $M \geq 3$ and where the indicator $Z_j$ is observed while the latent factor $\theta$ and the disturbance $\epsilon_j$ are not. The variables $Z_j$, $\theta$, and $\epsilon_j$ are assumed to be vectors of the same dimension. In our application, the vector of observed indicators and corresponding disturbances is

$$Z_j = (Z_{1,C,t,j})_{t=1}^T, (Z_{1,N,t,j})_{t=1}^T, (Z_{2,C,t,j})_{t=1}^T, (Z_{2,N,t,j})_{t=1}^T, Z_{3,C,1,j}, Z_{3,N,1,j})',$$

$$\epsilon_j = (\epsilon_{1,C,t,j})_{t=1}^T, (\epsilon_{1,N,t,j})_{t=1}^T, (\epsilon_{2,C,t,j})_{t=1}^T, (\epsilon_{2,N,t,j})_{t=1}^T, \epsilon_{3,C,1,j}, \epsilon_{3,C,N,1,j})',$$

while the vector of unobserved latent factors is

$$\theta = (\theta_{C,t})_{t=1}^T, (\theta_{N,t})_{t=1}^T, (I_{C,t})_{t=1}^T, (I_{N,t})_{t=1}^T, (\theta_{C,P}, \theta_{N,P})'.$$

The functions $a_j(\cdot, \cdot)$ for $j \in \{1, \ldots, M\}$ in equations (3.7) are unknown. It is necessary to normalize one of them (e.g., $a_1(\cdot, \cdot)$) in some way to achieve identification, as established in the following theorem.

**THEOREM 2:** The distribution of $\theta$ in equations (3.7) is identified under the following conditions:

(i) The joint density of $\theta$, $Z_1$, $Z_2$, and $Z_3$ is bounded and so are all their marginal and conditional densities.\(^{18}\)

\(^{18}\)This is a density with respect to the product measure of the Lebesgue measure on $\mathbb{R}^L \times \mathbb{R}^L \times \mathbb{R}^L$ and some dominating measure $\mu$. Hence $\theta$, $Z_1$, and $Z_2$ must be continuously distributed while $Z_3$ may be continuous or discrete.
(ii) \(Z_1, Z_2,\) and \(Z_3\) are mutually independent conditional on \(\theta.\)

(iii) \(p_{Z_1|Z_2}(Z_1 | Z_2)\) and \(p_{\theta|Z_1}(\theta | Z_1)\) form a bounded complete family of distributions indexed by \(Z_2\) and \(Z_1,\) respectively.

(iv) Whenever \(\theta \neq \tilde{\theta}, p_{Z_3|\theta}(Z_3 | \theta)\) and \(p_{Z_3|\tilde{\theta}}(Z_3 | \tilde{\theta})\) differ over a set of strictly positive probability.

(v) There exists a known functional \(\Psi,\) mapping a density to a vector, that has the property that \(\Psi[p_{Z_1|\theta}(\cdot | \theta)] = \theta.\)

See Appendix A3.2 for the proof\(^{19}\)

The proof of Theorem 2 proceeds by casting the analysis of identification as a linear algebra problem analogous to matrix diagonalization. In contrast to the standard matrix diagonalization used in linear factor analyses, we do not work with random vectors. Instead, we work with their densities. This approach offers the advantage that the problem remains linear even when the random vectors are related nonlinearly.

The conditional independence requirement of assumption (ii) is weaker than the full independence assumption traditionally made in standard linear factor models as it allows for heteroscedasticity. Assumption (iii) requires \(\theta, Z_1,\) and \(Z_2\) to be vectors of the same dimensions, while assumption (iv) can be satisfied even if \(Z_3\) is a scalar. The minimum number of measurements needed for identification is therefore \(2L + 1,\) which is exactly the same number of measurements as in the linear, classical measurement error case.

Versions of assumption (iii) appear in the nonparametric instrumental variable literature (e.g., Newey and Powell (2003), Darolles, Florens, and Renault (2002)). Intuitively, the requirement that \(p_{Z_1|Z_2}(Z_1 | Z_2)\) forms a bounded complete family requires that the density of \(Z_1\) vary sufficiently as \(Z_2\) varies (and similarly for \(p_{\theta|Z_1}(\theta | Z_1)\)).\(^{20}\)

Assumption (iv) is automatically satisfied, for instance, if \(\theta\) is univariate and \(a_3(\theta, \varepsilon_3)\) is strictly increasing in \(\theta.\) However, it holds much more generally. Since \(a_3(\theta, \varepsilon_3)\) is nonseparable, the distribution of \(Z_3\) conditional on \(\theta\) can change with \(\theta,\) thus making it possible for assumption (iv) to be satisfied even if \(a_3(\theta, \varepsilon_3)\) is not strictly increasing in \(\theta.\)

Assumption (v) specifies how the observed \(Z_1\) is used to determine the scale of the unobserved \(\theta.\) The most common choices of the functional \(\Psi\) would be

\(^{19}\)A vector of correctly measured variables \(C\) can trivially be added to the model by including \(C\) in the list of conditioning variables for all densities in the statement of the theorem. Theorem 2 then implies that \(p_{\theta|C}(\theta | C)\) is identified. Since \(p_C(C)\) is identified, it follows that \(p_{\theta,C}(\theta, C) = p_{\theta|C}(\theta | C)p_C(C)\) is also identified.

\(^{20}\)In the case of classical measurement error, bounded completeness assumptions can be phrased in terms of primitive conditions that require nonvanishing characteristic functions of the distributions of the measurement errors as in Mattner (1993). However, apart from this special case, very little is known about primitive conditions for bounded completeness, and research is still ongoing on this topic. See d’Haultfoeuille (2006).
the mean, the mode, the median, or any other well defined measure of location. This specification allows for nonclassical measurement error. One way to satisfy this assumption is to normalize \( a_1(\theta, \varepsilon_1) \) to be equal to \( \theta + \varepsilon_1 \), where \( \varepsilon_1 \) has zero mean, median, or mode. The zero mode assumption is particularly plausible for surveys where respondents face many possible wrong answers, but only one correct answer. Moving the mode of the answers away from zero would therefore require a majority of respondents to misreport in exactly the same way—an unlikely scenario. Many other nonseparable functions can also satisfy this assumption. With the distribution of \( p_\theta(\theta) \) in hand, we can identify the technology using the analysis presented below in Section 3.4.

Note that Theorem 2 does not claim that the distributions of the errors \( \varepsilon_j \) or that the functions \( a_j(\cdot, \cdot) \) are identified. In fact, it is always possible to alter the distribution of \( \varepsilon_j \) and the dependence of the function \( a_j(\cdot, \cdot) \) on its second argument in ways that cancel each other out, as noted in the literature on nonseparable models. However, lack of identifiability of these features of the model does not prevent identification of the distribution of \( \theta \).

Nevertheless, various normalizations that ensure that the functions \( a_j(\theta, \varepsilon_j) \) are fully identified are available. For example, if each element of \( \varepsilon_j \) is normalized to be uniform (or any other known distribution), the \( a_j(\theta, \varepsilon_j) \) are fully identified. Other normalizations discussed in Matzkin (2003, 2007) are also possible. Alternatively, one may assume that the \( a_j(\theta, \varepsilon_j) \) are separable in \( \varepsilon_j \) with zero conditional mean of \( \varepsilon_j \) given \( \theta \). We invoke these assumptions when we identify the policy function for investments in Section 3.6.2 below.

The conditions that justify Theorems 1 and 2 are not nested within each other. Their different assumptions represent different trade-offs best suited for different applications. While Theorem 1 would suffice for the empirical analysis of this paper, the general result established in Theorem 2 will likely be quite useful as larger sample sizes become available.

Carneiro, Hansen, and Heckman (2003) present an analysis for nonseparable measurement equations based on a separable latent index structure, but invoke strong independence and “identification-at-infinity” assumptions. Our approach for identifying the distribution of \( \theta \) from general nonseparable measurement equations does not require these strong assumptions. Note that it also allows the \( \theta \) to determine all measurements and for the \( \theta \) to be freely correlated.

### 3.4. Nonparametric Identification of the Technology Function

Suppose that the shocks \( \eta_{k,t} \) are independent over time. Below, we analyze a more general case that allows for serial dependence. Once the density of

---

22Observe that Theorem 2 covers the identifiability of the outcome \( Q_j \) functions (2.2) even if we supplement the model with errors \( \varepsilon_j, j \in \{1, \ldots, J\} \), that satisfy the conditions of the theorem.
θ is known, one can identify nonseparable technology function (2.1) for \( t \in \{1, \ldots, T\} \), \( k \in \{C, N\} \), and \( s \in \{1, \ldots, S\} \). Even if \((\theta_t, I_t, \theta_p)\) were perfectly observed, one could not separately identify the distribution of \( \eta_{k,t} \) and the function \( f_{k,s} \), because, without further normalizations, a change in the density of \( \eta_{k,t} \) can be undone by a change in the function \( f_{k,s} \).

One solution to this problem is to assume that (2.1) is additively separable in \( \eta_{k,t} \). Another way to avoid this ambiguity is to normalize \( \eta_{k,t} \) to have a uniform density on \([0, 1]\). Any of the normalizations suggested by Matzkin (2003, 2007) could be used. Assuming \( \eta_{k,t} \) is uniform \([0, 1]\), we establish that \( f_{k,s} \) is non-parametrically identified, by noting that, from the knowledge of \( p_\theta \), we can calculate, for any \( \tilde{\theta} \in \mathbb{R} \),

\[
\Pr[\theta_{k,t+1} \leq \tilde{\theta} \mid \theta_t, I_k, \theta_P] = G(\tilde{\theta} \mid \theta_t, I_k, \theta_P).
\]

We identify technology (2.1) using the relationship

\[
f_{k,s}(\theta_t, I_k, \theta_P, \eta_{k,t}) = G^{-1}(\eta_{k,t} \mid \theta_t, I_k, \theta_P),
\]

where \( G^{-1}(\eta_{k,t} \mid \theta_t, I_k, \theta_P) \) denotes the inverse of \( G(\tilde{\theta} \mid \theta_t, I_k, \theta_P) \) with respect to its first argument (assuming it exists), that is, the value \( \tilde{\theta} \) such that \( \eta_{k,t} = G(\tilde{\theta} \mid \theta_t, I_k, \theta_P) \). By construction, this operation produces a function \( f_{k,s} \) that generates outcomes \( \theta_{k,t+1} \) with the appropriate distribution, because a continuously distributed random variable is mapped into a uniformly distributed variable under the mapping defined by its own cumulative distribution function (c.d.f.).

The more traditional separable technology with zero mean disturbance, \( \theta_{k,t+1} = f_{k,s}(\theta_t, I_k, \theta_P) + \eta_{k,t} \), is covered by our analysis if we define

\[
f_{k,s}(\theta_t, I_k, \theta_P) = E[\theta_{k,t+1} \mid \theta_t, I_k, \theta_P],
\]

where the expectation is taken under the density \( p_{\theta_{k,t+1} \mid \theta_t, I_k, \theta_P} \), which can be calculated from \( p_\theta \). The density of \( \eta_{k,t} \) conditional on all variables is identified from

\[
p_{\eta_{k,t} \mid \theta_t, I_k, \theta_P}(\eta_{k,t} \mid \theta_t, I_k, \theta_P) = p_{\theta_{k,t+1} \mid \theta_t, I_k, \theta_P}(\eta_{k,t} + E[\theta_{k,t+1} \mid \theta_t, I_k, \theta_P] \mid \theta_t, I_k, \theta_P),
\]

since \( p_{\theta_{k,t+1} \mid \theta_t, I_k, \theta_P} \) is known once \( p_\theta \) is known. We now show how to anchor the scales of \( \theta_{C,t+1} \) and \( \theta_{N,t+1} \) using measures of adult outcomes.

\footnote{See, for example, Matzkin (2003, 2007).}
3.5. Anchoring Skills in an Interpretable Metric

It is common in the empirical literature on child schooling and investment to measure outcomes by test scores. However, test scores are arbitrarily scaled. To gain a better understanding of the relative importance of cognitive and noncognitive skills and their interactions, and the relative importance of investments at different stages of the life cycle, it is desirable to anchor skills in a common scale. In what follows, we continue to keep the conditioning on the regressors implicit.

We model the effect of period $T + 1$ cognitive and noncognitive skills on adult outcomes $Z_{4,j}$ for $j \in \{1, \ldots, J\}$. Suppose that there are $J_1$ observed outcomes that are linear functions of cognitive and noncognitive skills at the end of childhood, that is, in period $T$,

$$Z_{4,j} = \mu_{4,j} + \alpha_{4,C,j} \theta_{C,T+1} + \alpha_{4,N,j} \theta_{N,T+1} + \epsilon_{4,j} \quad \text{for} \quad j \in \{1, \ldots, J_1\}.$$  

When adult outcomes are linear and separable functions of skills, we define the anchoring functions to be

$$g_{C,j}(\theta_{C,T+1}) = \mu_{4,j} + \alpha_{4,C,j} \theta_{C,T+1},$$

$$g_{N,j}(\theta_{N,T+1}) = \mu_{4,j} + \alpha_{4,N,j} \theta_{N,T+1}.$$  

We can also anchor using nonlinear functions. One example would be an outcome produced by a latent variable $Z^*_j$, for $j \in \{J_1 + 1, \ldots, J\}$:

$$Z^*_j = \tilde{g}_j(\theta_{C,T+1}, \theta_{N,T+1}) - \epsilon_{4,j}.$$  

Note that we do not observe $Z^*_j$, but we observe the variable $Z_{4,j}$ which is defined as

$$Z_{4,j} = \begin{cases} 1, & \text{if } \tilde{g}_j(\theta_{C,T+1}, \theta_{N,T+1}) - \epsilon_{4,j} \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

In this notation,

$$\Pr(Z_{4,j} = 1 \mid \theta_{C,T+1}, \theta_{N,T+1})$$

$$= \Pr[\epsilon_{4,j} \leq \tilde{g}_j(\theta_{C,T+1}, \theta_{N,T+1}) \mid \theta_{C,T+1}, \theta_{N,T+1}]$$

$$= F_{\epsilon_{4,j}}[\tilde{g}_j(\theta_{C,T+1}, \theta_{N,T+1}) \mid \theta_{C,T+1}, \theta_{N,T+1}]$$

$$= g_j(\theta_{C,T+1}, \theta_{N,T+1}).$$

Adult outcomes such as high school graduation, criminal activity, drug use, and teenage pregnancy may be represented in this fashion.

$^{24}$The $Z_{4,j}$ correspond to the $Q_j$ of Section 2.
To establish identification of \( g_j(\theta_{C,T+1}, \theta_{N,T+1}) \) for \( j \in \{ J_1 + 1, \ldots, J \} \), we include the dummy \( Z_{4,j} \) in the vector \( \theta \). Assuming that the dummy \( Z_{4,j} \) is measured without error, the corresponding element of the two repeated measurement vectors \( W_1 \) and \( W_2 \) are identical and equal to \( Z_{4,j} \). Theorem 1 implies that the joint density of \( Z_{4,j} \), \( \theta_c,t \), and \( \theta_{N,t} \) is identified. Thus, it is possible to identify \( \Pr[Z_{4,j} = 1 \mid \theta_{C,T+1}, \theta_{N,T+1}] \).

We can extract two separate “anchors” \( g_{C,j}(\theta_{C,T+1}) \) and \( g_{N,j}(\theta_{N,T+1}) \) from the function \( g_j(\theta_{C,T+1}, \theta_{N,T+1}) \), by integrating out the other variable, for example,

\[
(3.9) \quad g_{C,j}(\theta_{C,T+1}) = \int g_j(\theta_{C,T+1}, \theta_{N,T+1}) p_{\theta_{N,T+1}}(\theta_{N,T+1}) d\theta_{N,T+1},
\]

\[
g_{N,j}(\theta_{N,T+1}) = \int g_j(\theta_{C,T+1}, \theta_{N,T+1}) p_{\theta_{C,T+1}}(\theta_{C,T+1}) d\theta_{C,T+1},
\]

where the marginal densities, \( p_{\theta_{j,t}}(\theta_{N,T+1}) \), \( j \in \{ C, N \} \), are identified by applying the preceding analysis. Both \( g_{C,j}(\theta_{C,T+1}) \) and \( g_{N,j}(\theta_{N,T+1}) \) are assumed to be strictly monotonic in their arguments.

The “anchored” skills, denoted by \( \tilde{\theta}_{j,k,t} \), are defined as

\[
\tilde{\theta}_{j,k,t} = g_{k,j}(\theta_{k,t}), \quad k \in \{ C, N \}, \quad t \in \{ 1, \ldots, T \}.
\]

The anchored skills inherit the subscript \( j \) because different anchors generally scale the same latent variables differently.

We combine the identification of the anchoring functions with the identification of the technology function \( \tilde{f}_{k,s,j}(\theta, I_{k,t}, \theta_P, \eta_{k,t}) \) established in the previous section to prove that the technology function expressed in terms of the anchored skills—denoted by \( \tilde{f}_{k,s,j}(\tilde{\theta}_{j,t}, I_{k,t}, \theta_P, \eta_{k,t}) \)—is also identified. To do so, redefine the technology function to be

\[
\tilde{f}_{k,s,j}(\tilde{\theta}_{j,C,t}, \tilde{\theta}_{j,N,t}, I_{k,t}, \theta_C, \theta_P, \eta_{k,t})
\]

\[
\equiv g_{k,j}(f_{k,s}(g_{C,j}^{-1}(\tilde{\theta}_{j,C,t}), g_{N,j}^{-1}(\tilde{\theta}_{j,N,t}), I_{k,t}, \theta_C, \theta_P, \eta_{k,t})),
\]

\[
k \in \{ C, N \},
\]

where \( g_{k,j}^{-1}(\cdot) \) denotes the inverse of the function \( g_{k,j}(\cdot) \). Invertibility follows from the assumed monotonicity. It is straightforward to show that

\[
\tilde{f}_{k,s,j}(\tilde{\theta}_{j,C,t}, \tilde{\theta}_{j,N,t}, I_{k,t}, \theta_C, \theta_P, \eta_{k,t})
\]

\[
= f_{k,s,j}(g_{C,j}(\theta_{C,t}), g_{N,j}(\theta_{N,t}), I_{k,t}, \theta_C, \theta_P, \theta_{N,P}, \eta_{k,t})
\]

\[
= g_{k,j}(f_{k,s}(g_{C,j}(g_{C,j}(\theta_{C,t})), g_{N,j}(g_{N,j}(\theta_{N,t})), I_{k,t}, \theta_C, \theta_P, \theta_{N,P}, \eta_{k,t}))
\]

\[
= g_{k,j}(f_{k,s}(\theta_{C,t}, \theta_{N,t}, I_{k,t}, \theta_C, \theta_{N,P}, \eta_{k,t})))
\]

\[
= g_{k,j}(\theta_{k,t+1}) = \tilde{\theta}_{k,j,t+1},
\]
as desired. Hence, $\tilde{f}_{k,s,j}$ is the equation of motion for the anchored skills $\tilde{\theta}_{k,i,t+1}$ that is consistent with the equation of motion $f_{k,s}$ for the original skills $\theta_{k,i,t}$.

### 3.6. Accounting for Endogeneity of Parental Investment

#### 3.6.1. Allowing for Unobserved Time-Invariant Heterogeneity

Thus far, we have maintained the assumption that the error term $\eta_{k,i,t}$ in the technology (2.1) is independent of all the other inputs $(\theta_t, I_{k,t}, \theta_P)$ as well as $\eta_{\ell,t}, k \neq \ell$. This implies that variables not observed by the econometrician are not used by parents to make their decisions regarding investments $I_{k,t}$. This is a very strong assumption. The availability of data on adult outcomes can be exploited to relax this assumption and allow for endogeneity of the inputs. This subsection develops an approach for a nonlinear model based on time-invariant heterogeneity.

To see how this can be done, suppose that we observe at least three adult outcomes, so that $J \geq 3$. We can then write outcomes as functions of $T + 1$ skills as well as unobserved (by the economist) time-invariant heterogeneity component, $\pi$, on which parents make their investment decisions:

$$Z_{4,j} = \alpha_{4,C,j} \theta_{C,T+1} + \alpha_{4,N,j} \theta_{N,T+1} + \alpha_{4,\pi,j} \pi + \varepsilon_{4,j}$$

for $j \in \{1, 2, \ldots, J\}$.

We can use the analysis of Section 3.2, suitably extended to allow for measurements $Z_{4,j}$, to secure identification of the factor loadings $\alpha_{4,C,j}, \alpha_{4,N,j},$ and $\alpha_{4,\pi,j}$. We can apply the argument of Section 3.4 to secure identification of the joint distribution of $(\theta_t, I_t, \theta_P, \pi)$.\(^{25}\) Write $\eta_{k,i,t} = (\pi, \nu_{k,i,t})$. Extending the preceding analysis, we can identify a more general version of the technology:

$$\theta_{k,t+1} = f_{k,s}(\theta_t, I_{k,t}, \theta_P, \pi, \nu_{k,i,t}).$$

$\pi$ is permitted to be correlated with the inputs $(\theta_t, I_t, \theta_P)$, and $\nu_{k,i,t}$ is assumed to be independent from the vector $(\theta_t, I_t, \theta_P, \pi)$ as well as $\nu_{l,t}$ for $l \neq k$. The next subsection develops a more general approach that allows $\pi$ to vary over time.

#### 3.6.2. More General Forms of Endogeneity

This subsection relaxes the invariant heterogeneity assumption by using exclusion restrictions based on economic theory to identify the technology under more general conditions. $\pi_t$ evolves over time and agents make investment

---

\(^{25}\)We discuss the identification of the factor loadings in this case in Appendix A4.
decisions based on it. Define $y_t$ as family resources in period $t$ (e.g., income, assets, constraints). As in Sections 3.2 and 3.3, we assume that suitable multiple measurements of $(\theta_P, (\theta_t, I_{C,t}, I_{N,t}, y_t)_{t=1}^T)$ are available to identify their (joint) distribution. In our application, we assume that $y_t$ is measured without error.\(^{26}\) We further assume that the error term $\eta_{k,t}$ can be decomposed into two components, $(\pi_t, \nu_{k,t})$, so that we may write the technology as

$$\theta_{k,t+1} = f_{k,t}(\theta_t, I_{k,t}, \theta_P, \pi_t, \nu_{k,t}).$$

(3.10)

$\pi_t$ is assumed to be a scalar shock independent over people, but not over time. It is a common shock that affects all technologies, but its effect may differ across technologies. The component $\nu_{k,t}$ is independent of $\theta_t, I_{k,t}, \theta_P$, and $y_t$, and independent of $\nu_{k,t'}$ for $t' \neq t$. Its realization takes place at the end of period $t$, after investment choices have already been made and implemented. The shock $\pi_t$ is realized before parents make investment choices, so we expect $I_{k,t}$ to respond to it.

We analyze a model of investment of the form

$$I_{k,t} = q_{k,t}(\theta_t, \theta_P, y_t, \pi_t), \quad k \in \{C, N\}, t \in \{1, \ldots, T\}.$$  

(3.11)

Equation (3.11) is the investment policy function that maps state variables for the parents, $(\theta_t, \theta_P, y_t, \pi_t)$, to the control variables $I_{k,t}$ for $k \in \{C, N\}$.\(^{27}\)

Our analysis relies on the assumption that the disturbances $\pi_t$ and $\nu_{k,t}$ in equation (3.10) are both scalar, although all other variables may be vector-valued. If the disturbances $\pi_t$ are independent and identically distributed (i.i.d.), identification is straightforward. To see this, impose an innocuous normalization (e.g., assume a specific marginal distribution for $\pi_t$). Then the relationship $I_{k,t} = q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ can be identified along the lines of the argument of Section 3.2 or 3.3, provided, for instance, that $\pi_t$ is independent from $(\theta_t, \theta_P, y_t)$.

If $\pi_t$ is serially correlated, it is not plausible to assume independence between $\pi_t$ and $\theta_t$, because past values of $\pi_t$ will have an impact on both current $\pi_t$ and on current $\theta_t$ (via the effect of past $\pi_t$ on past $I_{k,t}$). To address this problem, lagged values of income $y_t$ can be used as instruments for $\theta_t$ ($\theta_P$ and $y_t$ could serve as their own instruments). This approach works if $\pi_t$ is independent of $\theta_P$ as well as past and present values of $y_t$. After normalization of the distribution of the disturbance $\pi_t$, the general nonseparable function $q_t$ can be identified using quantile instrumental variable techniques (Chernozhukov,

\(^{26}\)Thus the “multiple measurements” on $y_t$ are all equal to each other in each period $t$.

\(^{27}\)The assumption of a common shock across technologies produces singularity across the investment equations (3.11). This is not a serious problem because, as noted below in Section 4.2.5, we cannot distinguish cognitive investment from noncognitive investment in our data. We assume a single common investment, so $q_{k,t}(\cdot) = q_t(\cdot)$ for $k \in \{C, N\}$.
Imbens, and Newey (2007)) under standard assumptions in that literature, including monotonicity and completeness.\footnote{Complete regularity conditions along with a proof are presented in Appendix A3.3.}

Once the functions $q_{k,t}$ have been identified, one can obtain $q_{k,t}^{-1}(\theta_t, \theta_p, y_t, I_{k,t})$, the inverse of $q_{k,t}(\theta_t, \theta_p, y_t, \pi_t)$ with respect to its last argument, provided $q_{k,t}(\theta_t, \theta_p, y_t, \pi_t)$ is strictly monotone in $\pi_t$ at all values of the arguments. We can then rewrite the technology function (3.11) as

$$
\theta_{k,t+1} = f_{k,s}(\theta_t, I_{k,t}, \theta_p, q_{k,t}^{-1}(\theta_t, \theta_p, y_t, I_{k,t}), \nu_{k,t}) 
$$

Again using standard nonseparable identification techniques and normalizations, one can show that the reduced form $f_{rf}$ is identified. Instruments are unnecessary here, because the disturbance $\nu_{k,t}$ is assumed to be independent of all other variables. However, to identify the technology $f_{k,s}$, we need to disentangle the direct effect of $\theta_t, I_{k,t}$, and $\theta_p$ on $\theta_{t+1}$ from their indirect effect through $\pi_t = q_{k,t}^{-1}(\theta_t, \theta_p, y_t, I_{k,t})$. To accomplish this, we exploit our knowledge of $q_{k,t}^{-1}(\theta_t, \theta_p, y_t, I_{k,t})$ to write

$$
f_{k,s}(\theta_t, I_{k,t}, \theta_p, \pi_t, \nu_{k,t}) = f_{k,s}(\theta_t, I_{k,t}, \theta_p, y_t, \nu_{k,t}) \mid_{\nu_{k,t} = q_{k,t}^{-1}(\theta_t, \theta_p, y_t, I_{k,t}) = \pi_t},
$$

where, on the right-hand side, we set $y_t$ so that the corresponding implied value of $\pi_t$ matches its value on the left-hand side. This does not necessarily require $q_{k,t}^{-1}(\theta_t, \theta_p, y_t, I_{k,t})$ to be invertible with respect to $y_t$, since we only need one suitable value of $y_t$ for each given $(\theta_t, \theta_p, I_{k,t}, \pi_t)$ and do not necessarily require a one-to-one mapping. By construction, the support of the distribution of $y_t$ conditional on $\theta_t, \theta_p$, and $I_{k,t}$, is sufficiently large to guarantee the existence of at least one solution, because, for a fixed $\theta_t, I_{k,t}$, and $\theta_p$, variations in $\pi_t$ are entirely due to $y_t$. We present a more formal discussion of our identification strategy in Appendix A3.3.

In our empirical analysis, we make further parametric assumptions regarding $f_{k,s}$ and $q_{k,t}$, which open the way to a more convenient estimation strategy to account for endogeneity. The idea is to assume that the function $q_{k,t}(\theta_t, \theta_p, y_t, \pi_t)$ is parametrically specified and additively separable in $\pi_t$, so that its identification follows under standard instrumental variables conditions. Next, we replace $I_{k,t}$ by its value given by the policy function in the technology:

$$
\theta_{k,t+1} = f_{k,s}(\theta_t, q_{k,t}(\theta_t, \theta_p, y_t, \pi_t), \theta_p, \pi_t, \nu_{k,t}).
$$

Eliminating $I_{k,t}$ solves the endogeneity problem because the two disturbances $\pi_t$ and $\nu_{k,t}$ are now independent of all explanatory variables, by assumption if the $\pi_t$ are serially independent. Identification is secured by assuming that $f_{k,s}$
is parametric and additively separable in $\nu_{k,t}$ (whose conditional mean is zero) and by assuming a parametric form for $f_{\pi_t}(\pi_t)$, the density of $\pi_t$. We can then write

$$E[\theta_{k,t+1} | \theta_t, \theta_P, y_t] = \int f_{k,s}(\theta_t, q_{k,t}(\theta_t, \theta_P, y_t, \pi_t), \theta_P, \pi_t, 0) f_{\pi_t}(\pi_t) d\pi_t$$

$$\equiv \hat{f}_{k,s}(\theta_t, \theta_P, y_t, \beta).$$

The right-hand side is now known up to a vector of parameters $\beta$ which will be (at least) locally identified if it happens that $\partial \hat{f}_{k,s}(\theta_t, \theta_P, y_t, \beta)/\partial \beta$ evaluated at the true value of $\beta$ is a vector function of $\theta_t, \theta_P, y_t$ that is linearly independent. Section 4.2.5 below describes the specific functional forms used in our empirical analysis, and relaxes the assumption of serial independence of the $\pi_t$.

4. ESTIMATING THE TECHNOLOGY OF SKILL FORMATION

Technology (2.1) and the associated measurement systems are nonparametrically identified. However, we use parametric maximum likelihood to estimate the model and do not estimate it under the most general conditions. We do this for two reasons. First, a fully nonparametric approach is too data hungry to apply to samples of the size that we have at our disposal, because the convergence rates of nonparametric estimators are quite slow. Second, solving a high-dimensional dynamic factor model is a computationally demanding task that can only be made manageable by invoking parametric assumptions. Nonetheless, the analysis of this paper shows that, in principle, the parametric structure used to secure the estimates reported below is not strictly required to identify the technology. The likelihood function for the model is presented in Appendix A5. Appendix A6 describes the nonlinear filtering algorithm we use to estimate the technology. Appendix A7 discusses how we implement anchoring. Appendix A8 reports a limited Monte Carlo study of a version of the general estimation strategy discussed in Section 4.2.5 below.

We estimate the technology on a sample of 2207 firstborn white children from the Children of the NLSY/79 (CNLSY/79) sample. (See Center for Human Resource Research (2004).) Starting in 1986, the children of the NLSY/1979 female respondents, ages 0–14, have been assessed every 2 years. The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-competence of the children as well as their home environments. Data are collected via direct assessment and maternal report during home visits at every biannual wave. Appendix A9 discusses the measurements used to proxy investment and output. Appendix Ta-
bles A9-1–A9-3 present summary statistics on the sample we use. We estimate a model for a single child, and ignore interactions among children and the allocation decisions of multiple child families.

To match the biennial data collection plan, in our empirical analysis, a period is equivalent to 2 years. We have eight periods distributed over two stages of development. We report estimates for a variety of specifications.

Dynamic factor models allow us to exploit the wealth of measures on investment and outcomes available in the CNLSY data. They solve several problems in estimating skill formation technologies. First, there are many proxies for parental investments in children’s cognitive and noncognitive development. Using a dynamic factor model, we let the data pick the best combinations of family input measures that predict levels and growth in test scores. Measured inputs that are not very informative on family investment decisions will have negligible estimated factor loadings. Second, our models help us solve the problem of missing data. Assuming that the data are missing at random, we integrate out the missing items from the sample likelihood.

In practice, we cannot empirically distinguish investments in cognitive skills from investments in noncognitive skills. Accordingly, we assume investment in period $t$ is the same for both skills, although it may have different effects on those skills. Thus we assume $I_{C,t} = I_{N,t}$ and define it as $I_t$.

### 4.1. Empirical Specification

We use the separable measurement system (3.1). We estimate versions of the technology (2.3)–(2.4) augmented to include shocks,

$$
\theta_{k,t+1} = \left[ \gamma_{s,k,t} \theta_{C,t}^{\phi_{k}} + \gamma_{s,k,2} \theta_{N,t}^{\phi_{k}} + \gamma_{s,k,3} I_{t}^{\phi_{k}} + \gamma_{s,k,4} \theta_{C,P}^{\phi_{k}} + \gamma_{s,k,5} \theta_{N,P}^{\phi_{k}} \right]^{1/\phi_{k}} e^{\eta_{k,t+1}},
$$

where $\gamma_{s,k,l} \geq 0$ and $\sum_{l=1}^{5} \gamma_{s,k,l} = 1, k \in \{C, N\}, t \in \{1, 2\}, s \in \{1, 2\}$. We assume that the innovations are normally distributed: $\eta_{k,t} \sim N(0, \delta_{\eta,k}^2)$. We further assume that the $\eta_{k,t}$ are serially independent over all $t$ and are independent of...
We assume that measurements $Z_{a,k,t,j}$ proxy the natural logarithms of the factors. In the text, we report only anchored results. For example, for $a = 1$, 

$$Z_{1,k,t,j} = \mu_{1,k,t,j} + \alpha_{1,k,t,j} \ln \theta_{k,t} + \varepsilon_{1,k,t,j},$$

$$j \in \{1, \ldots, M_{a,k,t}\}, t \in \{1, \ldots, T\}, k \in \{C, N\}.$$ 

We use the factors (and not their logarithms) as arguments of the technology. This keeps the latent factors nonnegative, as is required for the definition of technology (4.1). Collect the $\varepsilon$ terms for period $t$ into a vector $\varepsilon_t$. We assume that $\varepsilon_t \sim N(0, \Lambda_t)$, where $\Lambda_t$ is a diagonal matrix. We impose the condition that $\varepsilon_t$ is independent from $\varepsilon_{t'}$ for $t \neq t'$ and all $\eta_k + 1$. Define the $t$th row of $\theta$ as $\theta_r^t$, where $r$ stands for row. Thus

$$\ln \theta_r^t = (\ln \theta_{C,t}, \ln \theta_{N,t}, \ln I_t, \ln \theta_{C,P}, \ln \theta_{N,P}, \ln \pi).$$

Identification of this model follows as a consequence of Theorems 1 and 2 and results in Matzkin (2003, 2007). We estimate the model under different assumptions about the distribution of the factors. Under the first specification, $\ln \theta_r^t$ is normally distributed with mean zero and variance–covariance matrix $\Sigma_t$. Under the second specification, $\ln \theta_r^t$ is distributed as a mixture of $T$ normals. Let $\phi(x; \mu_{t,\tau}, \Sigma_{t,\tau})$ denote the density of a normal random variable with mean $\mu_{t,\tau}$ and variance–covariance matrix $\Sigma_{t,\tau}$. The mixture of normals writes the density of $\ln \theta_r^t$ as

$$p(\ln \theta_r^t) = \sum_{\tau=1}^{T} \omega_{\tau} \phi(\ln \theta_r^t; \mu_{t,\tau}, \Sigma_{t,\tau})$$

subject to $\sum_{\tau=1}^{T} \omega_{\tau} = 1$ and $\sum_{\tau=1}^{T} \omega_{\tau} \mu_{t,\tau} = 0$.

Our anchored results allow us to compare the productivity of investments and stocks of different skills at different stages of the life cycle on the anchored outcome. In this paper, we mainly use completed years of education by age 19, a continuous variable, as an anchor.

### 4.2. Empirical Estimates

This section presents results from an extensive empirical analysis that estimates the multistage technology of skill formation, accounting for measurement error, nonnormality of the factors, endogeneity of inputs, and family

---

31 Appendix A11.1 compares anchored and unanchored results.

32 We use five regressors ($X$) for every measurement equation: a constant, the age of the child at the assessment date, the child’s gender, a dummy variable if the mother was less than 20 years old at the time of the first birth, and a cohort dummy (1 if the child was born after 1987 and 0 otherwise).
investment decisions. The plan of this section is as follows. We first present baseline two-stage models that anchor outcomes in terms of their effects on schooling attainment, that correct for measurement errors, and that assume that the factors are normally distributed. These models do not account for endogeneity of inputs through unobserved heterogeneity components or family investment decisions. The baseline model is far more general than what is presented in previous research on the formation of child skills that uses unanchored test scores as outcome measures and does not account for measurement error.33

We present evidence on the first-order empirical importance of measurement error. When we do not correct for it, the estimated technology suggests that there is no effect of early investment on outcomes. Controlling for endogeneity of family inputs by accounting for unobserved heterogeneity (\(\pi\)) and accounting explicitly for family investment decisions has substantial effects on estimated parameters.

The following empirical regularities emerge across all models that account for measurement error:34 Self-productivity of skills is greater in the second stage than in the first stage. Noncognitive skills are cross-productive for cognitive skills in the first stage of production. The cross-productivity effect is weaker and less precisely determined in the second stage. There is no evidence for a cross-productivity effect of cognitive skills on noncognitive skills at either stage. The estimated elasticity of substitution for inputs in cognitive skill is substantially lower in the second stage of a child’s life cycle than in the first stage. For noncognitive skills, the elasticity in the second period is slightly higher for models that control for unobserved heterogeneity (\(\pi\)). These estimates suggest that it is easier to redress endowment deficits that determine cognition in the first stage of a child’s life cycle than in the second stage. For socioemotional (noncognitive) skills, the opposite is true. For cognitive skills, the productivity parameter associated with parental investment (\(\gamma_{1,c,3}\)) is greater in the first stage than in the second stage (\(\gamma_{2,c,3}\)). For noncognitive skills, the pattern of estimates for the productivity parameter across models is less clear-cut, but there are not dramatic differences across the stages. For both outputs, the parameter associated with the effect of parental noncognitive skills on output is smaller at the second stage than the first stage.

Appendix A11 discusses the sensitivity of estimates of a one-stage two-skill model to alternative anchors and to allowing for nonnormality of the factors. For these and other estimated models which are not reported, allowing for nonnormality has only minor effects on the estimates. However, anchoring af-

33An example is the analysis of Fryer and Levitt (2004).
34Estimated parameters are reported in Appendix A10.
fects the estimates. To facilitate computation, we use years of schooling attained as the anchor in all of the models reported in this section of the paper.

4.2.1. The Baseline Specification

Table I presents evidence on our baseline two-stage model of skill formation. Outcomes are anchored in years of schooling attained. Factors are assumed to be normally distributed and we ignore heterogeneity ($\pi$). The estimates show that for both skills, self-productivity increases in the second stage. Noncognitive skills foster cognitive skills in the first stage, but not in the second stage. Cognitive skills have no cross-productivity effect on noncognitive skills at either stage. The productivity parameter for investment is greater in the first period than in the second period for either skill. The difference across stages in the estimated parameters is dramatic for cognitive skills. The variability in the shocks is greater in the second period than in the first period. The elasticity of substitution for cognitive skills is much greater in the first period than in the second period. However, the estimated elasticity of substitution is about the same in both stages of production.

For cognitive skill production, the parental cognitive skill parameter is about the same in both stages. The opposite is true for parental noncognitive skills. In producing noncognitive skills, parental cognitive skills play no role in the second stage. Parental noncognitive skills play a strong role in stage 1 and a weaker role in stage 2.

4.2.2. The Empirical Importance of Measurement Error

Using our estimated factor model, we can investigate the extent of measurement error on each measure of skill and investment in our data. To simplify the notation, we keep the conditioning on the regressors implicit and, without loss of generality, consider the measurements on cognitive skills in period $t$. For linear measurement systems, the variance can be decomposed as

$$\text{Var}(Z_{1,c,i,j}) = \alpha^2_{1,c,i,j} \text{Var}(\ln \theta_{c,i}) + \text{Var}(\varepsilon_{1,c,i,j}).$$

The fractions of the variance of $Z_{1,c,i,j}$ due to measurement error, $s_{1,c,i,j}^\varepsilon$, and true signal, $s_{1,c,i,j}^\theta$, are, respectively,

$$s_{1,c,i,j}^\varepsilon = \frac{\text{Var}(\varepsilon_{1,c,i,j})}{\alpha^2_{1,c,i,j} \text{Var}(\ln \theta_{c,i}) + \text{Var}(\varepsilon_{1,c,i,j})} \quad \text{(noise)}$$

$^{35}$Cunha and Heckman (2008) show the sensitivity of the estimates to alternative anchors for a linear model specification.

$^{36}$The normalizations for the factors are presented in Appendix A10.

$^{37}$Zero values of coefficients in this and other tables arise from the optimizer attaining a boundary of zero in the parameter space.
### TABLE I

**Using the Factor Model to Correct for Measurement Error: Linear Anchoring on Educational Attainment (Years of Schooling); No Unobserved Heterogeneity (\( \pi \)), Factors Normally Distributed**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>First Stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Technology of Cognitive Skill Formation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Period Cognitive Skills</td>
<td>( \gamma_{1,C,1} )</td>
<td>0.487</td>
</tr>
<tr>
<td>(Self-Productivity)</td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>Current Period Noncognitive Skills</td>
<td>( \gamma_{1,C,2} )</td>
<td>0.083</td>
</tr>
<tr>
<td>(Cross-Productivity)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>( \gamma_{1,C,3} )</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>( \gamma_{1,C,4} )</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>( \gamma_{1,C,5} )</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>( \phi_{1,C} )</td>
<td>0.611</td>
</tr>
<tr>
<td></td>
<td>(0.240)</td>
<td></td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>( 1/(1 - \phi_{1,C}) )</td>
<td>2.569</td>
</tr>
<tr>
<td>Variance of Shocks ( \eta_{C,i} )</td>
<td>( \delta_{1,C}^2 )</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td><strong>The Technology of Noncognitive Skill Formation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Period Cognitive Skills</td>
<td>( \gamma_{1,N,1} )</td>
<td>0.000</td>
</tr>
<tr>
<td>(Cross-Productivity)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>Current Period Noncognitive Skills</td>
<td>( \gamma_{1,N,2} )</td>
<td>0.649</td>
</tr>
<tr>
<td>(Self-Productivity)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>( \gamma_{1,N,3} )</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>( \gamma_{1,N,4} )</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td></td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>( \gamma_{1,N,5} )</td>
<td>0.183</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>( \phi_{1,N} )</td>
<td>-0.674</td>
</tr>
<tr>
<td></td>
<td>(0.324)</td>
<td></td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>( 1/(1 - \phi_{1,N}) )</td>
<td>0.597</td>
</tr>
<tr>
<td>Variance of Shocks ( \eta_{N,i} )</td>
<td>( \delta_{1,N}^2 )</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
</tr>
</tbody>
</table>

*aStandard errors in parentheses.*
and

\[ s_{1,C,t,i,j}^\theta = \frac{\alpha_{1,C,t,i,j}^2 \text{Var}(\ln \theta_{C,t})}{\alpha_{1,C,t,i,j}^2 \text{Var}(\ln \theta_{C,t}) + \text{Var}(\epsilon_{1,C,t,i,j})} \]  

(signal).

For each measure of skill and investment used in the estimation, we construct \( s_{1,C,t,i,j}^\varepsilon \) and \( s_{1,C,t,i,j}^\theta \) which are reported in Table IIA. Note that the early proxies tend to have a higher fraction of observed variance due to measurement error. For example, the measure that contains the lowest true signal ratio is the MSD (Motor and Social Developments Score) at year of birth, in which less than 5% of the observed variance is signal. The proxy with the highest signal ratio is the PIAT Reading Recognition Scores at ages 5–6, for which almost 96% of the observed variance is due to the variance of the true signal. Overall, about 54% of the observed variance is associated with the cognitive skill factors \( \theta_{C,t} \).

Table IIA also shows the same ratios for measures of childhood noncognitive skills. The measures of noncognitive skills tend to be lower in informational content than their cognitive counterparts. Overall, less than 40% of the observed variance is due to the variance associated with the factors for noncognitive skills. The poorest measure for noncognitive skills is the “Sociability” measure at ages 3–4, in which less than 1% of the observed variance is signal. The richest is the “Behavior Problem Index (BPI) Headstrong” score, in which almost 62% of the observed variance is due to the variance of the signal.

Table IIA also presents the signal–noise ratio of measures of parental cognitive and noncognitive skills. Overall, measures of maternal cognitive skills tend to have a higher information content than measures of noncognitive skills. While the poorest measurement on cognitive skills has a signal ratio of almost 35%, the richest measurements on noncognitive skills are slightly above 40%.

Analogous estimates of signal and noise for our investment measures are reported in Table IIB. Investment measures are much noisier than either measure of skill. The measures for investments at earlier stages tend to be noisier than the measures at later stages. It is interesting to note that the measure “Number of Books” has a high signal–noise ratio at early years, but not in later years. At earlier years, the measure “How Often Mom Reads to the Child” has about the same informational content as “Number of Books.” In later years, measures such as “How Often Child Goes to the Museum” and “How Often Child Goes to Musical Shows” have higher signal–noise ratios.

These estimates suggest that it is likely to be empirically important to control for measurement error in estimating technologies of skill formation. A general pattern is that at early ages compared to later ages, measures of skill tend to be riddled with measurement error, while the reverse pattern is true for the measurement errors for the proxies for investment.
<table>
<thead>
<tr>
<th>Measurement of Child’s Cognitive Skills</th>
<th>%Signal</th>
<th>%Noise</th>
<th>Measurement of Child’s Noncognitive Skills</th>
<th>%Signal</th>
<th>%Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestation Length</td>
<td>0.501</td>
<td>0.499</td>
<td>Difficulty at Birth</td>
<td>0.151</td>
<td>0.849</td>
</tr>
<tr>
<td>Weight at Birth</td>
<td>0.557</td>
<td>0.443</td>
<td>Friendliness at Birth</td>
<td>0.165</td>
<td>0.835</td>
</tr>
<tr>
<td>Motor-Social Development at Birth</td>
<td>0.045</td>
<td>0.955</td>
<td>Compliance at Ages 1–2</td>
<td>0.232</td>
<td>0.768</td>
</tr>
<tr>
<td>Motor-Social Development at Ages 1–2</td>
<td>0.275</td>
<td>0.725</td>
<td>Insecure at Ages 1–2</td>
<td>0.080</td>
<td>0.920</td>
</tr>
<tr>
<td>Body Parts at Ages 1–2</td>
<td>0.308</td>
<td>0.692</td>
<td>Sociability at Ages 1–2</td>
<td>0.075</td>
<td>0.925</td>
</tr>
<tr>
<td>Memory for Locations at Ages 1–2</td>
<td>0.160</td>
<td>0.840</td>
<td>Difficulty at Ages 1–2</td>
<td>0.382</td>
<td>0.618</td>
</tr>
<tr>
<td>Motor-Social Development at Ages 3–4</td>
<td>0.410</td>
<td>0.590</td>
<td>Friendliness at Ages 1–2</td>
<td>0.189</td>
<td>0.811</td>
</tr>
<tr>
<td>Picture Vocabulary at Ages 3–4</td>
<td>0.431</td>
<td>0.569</td>
<td>Compliance at Ages 3–4</td>
<td>0.133</td>
<td>0.867</td>
</tr>
<tr>
<td>Picture Vocabulary at Ages 5–6</td>
<td>0.225</td>
<td>0.775</td>
<td>Insecure at Ages 3–4</td>
<td>0.122</td>
<td>0.878</td>
</tr>
<tr>
<td>PIAT Mathematics at Ages 5–6</td>
<td>0.314</td>
<td>0.686</td>
<td>Sociability at Ages 3–4</td>
<td>0.008</td>
<td>0.992</td>
</tr>
<tr>
<td>PIAT Reading Recognition at Ages 5–6</td>
<td>0.958</td>
<td>0.042</td>
<td>Behavior Problem Index Antisocial at Ages 3–4</td>
<td>0.405</td>
<td>0.595</td>
</tr>
<tr>
<td>PIAT Reading Comprehension at Ages 5–6</td>
<td>0.938</td>
<td>0.062</td>
<td>Behavior Problem Index Anxiety at Ages 3–4</td>
<td>0.427</td>
<td>0.573</td>
</tr>
<tr>
<td>PIAT Mathematics at Ages 7–8</td>
<td>0.465</td>
<td>0.535</td>
<td>Behavior Problem Index Headstrong at Ages 3–4</td>
<td>0.518</td>
<td>0.482</td>
</tr>
<tr>
<td>PIAT Reading Recognition at Ages 7–8</td>
<td>0.869</td>
<td>0.131</td>
<td>Behavior Problem Index Hyperactive at Ages 3–4</td>
<td>0.358</td>
<td>0.642</td>
</tr>
<tr>
<td>PIAT Reading Comprehension at Ages 7–8</td>
<td>0.797</td>
<td>0.203</td>
<td>Behavior Problem Index Conflict at Ages 3–4</td>
<td>0.336</td>
<td>0.664</td>
</tr>
<tr>
<td>PIAT Mathematics at Ages 9–10</td>
<td>0.492</td>
<td>0.508</td>
<td>Behavior Problem Index Antisocial at Ages 5–6</td>
<td>0.435</td>
<td>0.565</td>
</tr>
<tr>
<td>PIAT Reading Recognition at Ages 9–10</td>
<td>0.817</td>
<td>0.183</td>
<td>Behavior Problem Index Anxiety at Ages 5–6</td>
<td>0.409</td>
<td>0.591</td>
</tr>
<tr>
<td>PIAT Reading Comprehension at Ages 9–10</td>
<td>0.666</td>
<td>0.334</td>
<td>Behavior Problem Index Headstrong at Ages 5–6</td>
<td>0.611</td>
<td>0.389</td>
</tr>
<tr>
<td>PIAT Mathematics at Ages 11–12</td>
<td>0.516</td>
<td>0.484</td>
<td>Behavior Problem Index Hyperactive at Ages 5–6</td>
<td>0.481</td>
<td>0.519</td>
</tr>
<tr>
<td>PIAT Reading Recognition at Ages 11–12</td>
<td>0.781</td>
<td>0.219</td>
<td>Behavior Problem Index Conflict at Ages 5–6</td>
<td>0.290</td>
<td>0.710</td>
</tr>
<tr>
<td>PIAT Reading Comprehension at Ages 11–12</td>
<td>0.614</td>
<td>0.386</td>
<td>Behavior Problem Index Antisocial Ages 7–8</td>
<td>0.446</td>
<td>0.554</td>
</tr>
<tr>
<td>PIAT Mathematics at Ages 13–14</td>
<td>0.537</td>
<td>0.463</td>
<td>Behavior Problem Index Anxiety Ages 7–8</td>
<td>0.475</td>
<td>0.525</td>
</tr>
<tr>
<td>PIAT Reading Recognition at Ages 13–14</td>
<td>0.735</td>
<td>0.265</td>
<td>Behavior Problem Index Headstrong Ages 7–8</td>
<td>0.605</td>
<td>0.395</td>
</tr>
<tr>
<td>PIAT Reading Comprehension at Ages 13–14</td>
<td>0.549</td>
<td>0.451</td>
<td>Behavior Problem Index Hyperactive Ages 7–8</td>
<td>0.497</td>
<td>0.503</td>
</tr>
</tbody>
</table>

(Continues)
### TABLE IIA—Continued

<table>
<thead>
<tr>
<th>Measurement of Maternal Cognitive Skills</th>
<th>%Signal</th>
<th>%Noise</th>
<th>Measurement of Child’s Noncognitive Skills</th>
<th>%Signal</th>
<th>%Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASVAB Arithmetic Reasoning</td>
<td>0.728</td>
<td>0.272</td>
<td>Behavior Problem Index Conflict Ages 7–8</td>
<td>0.327</td>
<td>0.673</td>
</tr>
<tr>
<td>ASVAB Word Knowledge</td>
<td>0.625</td>
<td>0.375</td>
<td>Behavior Problem Index Antisocial Ages 9–10</td>
<td>0.503</td>
<td>0.497</td>
</tr>
<tr>
<td>ASVAB Paragraph Composition</td>
<td>0.576</td>
<td>0.424</td>
<td>Behavior Problem Index Anxiety Ages 9–10</td>
<td>0.472</td>
<td>0.528</td>
</tr>
<tr>
<td>ASVAB Numerical Operations</td>
<td>0.461</td>
<td>0.539</td>
<td>Behavior Problem Index Headstrong Ages 9–10</td>
<td>0.577</td>
<td>0.423</td>
</tr>
<tr>
<td>ASVAB Coding Speed</td>
<td>0.355</td>
<td>0.647</td>
<td>Behavior Problem Index Hyperactive Ages 9–10</td>
<td>0.463</td>
<td>0.537</td>
</tr>
<tr>
<td>ASVAB Mathematical Knowledge</td>
<td>0.662</td>
<td>0.338</td>
<td>Behavior Problem Index Conflict Ages 9–10</td>
<td>0.369</td>
<td>0.631</td>
</tr>
<tr>
<td>Measurement of Maternal Noncognitive Skills</td>
<td></td>
<td></td>
<td>Behavior Problem Index Antisocial Ages 11–12</td>
<td>0.514</td>
<td>0.486</td>
</tr>
<tr>
<td>Self-Esteem “I am a person of worth”</td>
<td>0.277</td>
<td>0.723</td>
<td>Behavior Problem Index Anxiety Ages 11–12</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>Self-Esteem “I have good qualities”</td>
<td>0.349</td>
<td>0.651</td>
<td>Behavior Problem Index Headstrong Ages 11–12</td>
<td>0.603</td>
<td>0.397</td>
</tr>
<tr>
<td>Self-Esteem “I am a failure”</td>
<td>0.444</td>
<td>0.556</td>
<td>Behavior Problem Index Hyperactive Ages 11–12</td>
<td>0.505</td>
<td>0.495</td>
</tr>
<tr>
<td>Self-Esteem “I have nothing to be proud of”</td>
<td>0.375</td>
<td>0.625</td>
<td>Behavior Problem Index Conflict Ages 11–12</td>
<td>0.370</td>
<td>0.630</td>
</tr>
<tr>
<td>Self-Esteem “I have a positive attitude”</td>
<td>0.406</td>
<td>0.594</td>
<td>Behavior Problem Index Antisocial Ages 13–14</td>
<td>0.494</td>
<td>0.506</td>
</tr>
<tr>
<td>Self-Esteem “I wish I had more self-respect”</td>
<td>0.341</td>
<td>0.659</td>
<td>Behavior Problem Index Anxiety Ages 13–14</td>
<td>0.546</td>
<td>0.454</td>
</tr>
<tr>
<td>Self-Esteem “I feel useless at times”</td>
<td>0.293</td>
<td>0.707</td>
<td>Behavior Problem Index Headstrong Ages 13–14</td>
<td>0.595</td>
<td>0.405</td>
</tr>
<tr>
<td>Self-Esteem “I sometimes think I am no good”</td>
<td>0.375</td>
<td>0.625</td>
<td>Behavior Problem Index Hyperactive Ages 13–14</td>
<td>0.525</td>
<td>0.475</td>
</tr>
<tr>
<td>Locus of Control “I have no control”</td>
<td>0.047</td>
<td>0.953</td>
<td>Behavior Problem Index Conflict Ages 13–14</td>
<td>0.414</td>
<td>0.586</td>
</tr>
<tr>
<td>Locus of Control “I make no plans for the future”</td>
<td>0.064</td>
<td>0.936</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus of Control “Luck is big factor in life”</td>
<td>0.041</td>
<td>0.959</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus of Control “Luck plays big role in my life”</td>
<td>0.020</td>
<td>0.980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurements of Parental Investments</td>
<td>%Signal</td>
<td>%Noise</td>
<td>Measurements of Parental Investments</td>
<td>%Signal</td>
<td>%Noise</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>---------</td>
<td>--------</td>
<td>-------------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>How Often Child Goes on Outings During Year of Birth</td>
<td>0.329</td>
<td>0.671</td>
<td>Child Has Musical Instruments Ages 7–8</td>
<td>0.022</td>
<td>0.978</td>
</tr>
<tr>
<td>Number of Books Child Has During Year of Birth</td>
<td>0.209</td>
<td>0.791</td>
<td>Family Subscribes to Daily Newspapers Ages 7–8</td>
<td>0.023</td>
<td>0.977</td>
</tr>
<tr>
<td>How Often Mom Reads to Child During Year of Birth</td>
<td>0.484</td>
<td>0.516</td>
<td>Child Has Special Lessons Ages 7–8</td>
<td>0.018</td>
<td>0.982</td>
</tr>
<tr>
<td>Number of Soft Toys Child Has During Year of Birth</td>
<td>0.126</td>
<td>0.874</td>
<td>How Often Child Goes to Musical Shows Ages 7–8</td>
<td>0.266</td>
<td>0.734</td>
</tr>
<tr>
<td>Number of Push/Pull Toys Child Has During Year of Birth</td>
<td>0.019</td>
<td>0.981</td>
<td>How Often Child Attends Family Gatherings Ages 7–8</td>
<td>0.125</td>
<td>0.875</td>
</tr>
<tr>
<td>How Often Child Eats With Mom/Dad During Year of Birth</td>
<td>0.511</td>
<td>0.489</td>
<td>How Often Child Is Praised Ages 7–8</td>
<td>0.046</td>
<td>0.954</td>
</tr>
<tr>
<td>How Often Mom Calls From Work During Year of Birth</td>
<td>0.119</td>
<td>0.881</td>
<td>How Often Child Gets Positive Encouragement Ages 7–8</td>
<td>0.053</td>
<td>0.947</td>
</tr>
<tr>
<td>How Often Child Goes on Outings at Ages 1–2</td>
<td>0.148</td>
<td>0.852</td>
<td>Number of Books Child Has Ages 9–10</td>
<td>0.013</td>
<td>0.987</td>
</tr>
<tr>
<td>Number of Books Child Has Ages 1–2</td>
<td>0.055</td>
<td>0.945</td>
<td>Mom Reads to Child Ages 9–10</td>
<td>0.137</td>
<td>0.863</td>
</tr>
<tr>
<td>How Often Mom Reads to Child Ages 1–2</td>
<td>0.186</td>
<td>0.814</td>
<td>Eats With Mom/Dad Ages 9–10</td>
<td>0.162</td>
<td>0.838</td>
</tr>
<tr>
<td>Number of Soft Toys Child Has Ages 1–2</td>
<td>0.240</td>
<td>0.760</td>
<td>How Often Child Goes to Museum Ages 9–10</td>
<td>0.219</td>
<td>0.781</td>
</tr>
<tr>
<td>Number of Push/Pull Toys Child Has Ages 1–2</td>
<td>0.046</td>
<td>0.954</td>
<td>Child Has Musical Instruments Ages 9–10</td>
<td>0.019</td>
<td>0.981</td>
</tr>
<tr>
<td>How Often Child Eats With Mom/Dad Ages 1–2</td>
<td>0.194</td>
<td>0.806</td>
<td>Family Subscribes to Daily Newspapers Ages 9–10</td>
<td>0.019</td>
<td>0.981</td>
</tr>
<tr>
<td>Mom Calls From Work Ages 1–2</td>
<td>0.070</td>
<td>0.930</td>
<td>Child Has Special Lessons Ages 9–10</td>
<td>0.015</td>
<td>0.985</td>
</tr>
<tr>
<td>How Often Child Goes on Outings Ages 3–4</td>
<td>0.123</td>
<td>0.877</td>
<td>How Often Child Goes to Musical Shows Ages 9–10</td>
<td>0.242</td>
<td>0.758</td>
</tr>
<tr>
<td>Number of Books Child Has Ages 3–4</td>
<td>0.012</td>
<td>0.988</td>
<td>How Often Child Attends Family Gatherings Ages 9–10</td>
<td>0.115</td>
<td>0.885</td>
</tr>
<tr>
<td>How Often Mom Reads to Child Ages 3–4</td>
<td>0.088</td>
<td>0.912</td>
<td>How Often Child Is Praised Ages 9–10</td>
<td>0.036</td>
<td>0.964</td>
</tr>
<tr>
<td>How Often Child Eats With Mom/Dad Ages 3–4</td>
<td>0.170</td>
<td>0.830</td>
<td>How Often Child Gets Positive Encouragement Ages 9–10</td>
<td>0.041</td>
<td>0.959</td>
</tr>
<tr>
<td>Number of Magazines at Home Ages 3–4</td>
<td>0.193</td>
<td>0.807</td>
<td>Number of Books Child Has Ages 11–12</td>
<td>0.016</td>
<td>0.984</td>
</tr>
<tr>
<td>Measurements of Parental Investments</td>
<td>%Signal</td>
<td>%Noise</td>
<td>Measurements of Parental Investments</td>
<td>%Signal</td>
<td>%Noise</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------</td>
<td>--------</td>
<td>--------------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>Child Has a CD Player Ages 3–4</td>
<td>0.021</td>
<td>0.979</td>
<td>Eats With Mom/Dad Ages 11–12</td>
<td>0.153</td>
<td>0.847</td>
</tr>
<tr>
<td>How Often Child Goes on Outings Ages 5–6</td>
<td>0.100</td>
<td>0.900</td>
<td>How Often Child Goes to Museum Ages 11–12</td>
<td>0.217</td>
<td>0.783</td>
</tr>
<tr>
<td>Number of Books Child Has Ages 5–6</td>
<td>0.009</td>
<td>0.991</td>
<td>Child Has Musical Instruments Ages 11–12</td>
<td>0.016</td>
<td>0.984</td>
</tr>
<tr>
<td>How Often Mom Reads to Child Ages 5–6</td>
<td>0.086</td>
<td>0.914</td>
<td>Family Subscribes to Daily Newspapers Ages 11–12</td>
<td>0.018</td>
<td>0.982</td>
</tr>
<tr>
<td>How Often Child Eats With Mom/Dad Ages 5–6</td>
<td>0.173</td>
<td>0.827</td>
<td>Child Has Special Lessons Ages 11–12</td>
<td>0.013</td>
<td>0.987</td>
</tr>
<tr>
<td>Number of Magazines at Home Ages 5–6</td>
<td>0.164</td>
<td>0.836</td>
<td>How Often Child Goes to Musical Shows Ages 11–12</td>
<td>0.225</td>
<td>0.775</td>
</tr>
<tr>
<td>Child Has CD Player Ages 5–6</td>
<td>0.015</td>
<td>0.985</td>
<td>How Often Child Attends Family Gatherings Ages 11–12</td>
<td>0.103</td>
<td>0.897</td>
</tr>
<tr>
<td>How Often Child Goes to Museum Ages 5–6</td>
<td>0.296</td>
<td>0.704</td>
<td>How Often Child Is Praised Ages 11–12</td>
<td>0.026</td>
<td>0.974</td>
</tr>
<tr>
<td>Child Has Musical Instruments Ages 5–6</td>
<td>0.026</td>
<td>0.974</td>
<td>How Often Child Gets Positive Encouragement Ages 11–12</td>
<td>0.037</td>
<td>0.963</td>
</tr>
<tr>
<td>Family Subscribes to Daily Newspapers Ages 5–6</td>
<td>0.025</td>
<td>0.975</td>
<td>Number of Books Child Has Ages 13–14</td>
<td>0.023</td>
<td>0.977</td>
</tr>
<tr>
<td>Child Has Special Lessons Ages 5–6</td>
<td>0.020</td>
<td>0.980</td>
<td>Eats With Mom/Dad Ages 13–14</td>
<td>0.152</td>
<td>0.848</td>
</tr>
<tr>
<td>How Often Child Goes to Musical Shows Ages 5–6</td>
<td>0.304</td>
<td>0.696</td>
<td>How Often Child Goes to Museum Ages 13–14</td>
<td>0.201</td>
<td>0.799</td>
</tr>
<tr>
<td>How Often Child Attends Family Gatherings Ages 5–6</td>
<td>0.141</td>
<td>0.859</td>
<td>Child Has Musical Instruments Ages 13–14</td>
<td>0.015</td>
<td>0.985</td>
</tr>
<tr>
<td>How Often Child Is Praised Ages 5–6</td>
<td>0.056</td>
<td>0.944</td>
<td>Family Subscribes to Daily Newspapers Ages 13–14</td>
<td>0.017</td>
<td>0.983</td>
</tr>
<tr>
<td>How Often Child Gets Positive Encouragement Ages 5–6</td>
<td>0.081</td>
<td>0.919</td>
<td>Child Has Special Lessons Ages 13–14</td>
<td>0.012</td>
<td>0.988</td>
</tr>
<tr>
<td>Number of Books Child Has Ages 7–8</td>
<td>0.007</td>
<td>0.993</td>
<td>How Often Child Goes to Musical Shows Ages 13–14</td>
<td>0.224</td>
<td>0.776</td>
</tr>
<tr>
<td>How Often Mom Reads to Child Ages 7–8</td>
<td>0.113</td>
<td>0.887</td>
<td>How Often Child Attends Family Gatherings Ages 13–14</td>
<td>0.099</td>
<td>0.901</td>
</tr>
<tr>
<td>How Often Child Eats With Mom/Dad Ages 7–8</td>
<td>0.166</td>
<td>0.834</td>
<td>How Often Child Is Praised Ages 13–14</td>
<td>0.031</td>
<td>0.969</td>
</tr>
<tr>
<td>How Often Child Goes to Museum Ages 7–8</td>
<td>0.240</td>
<td>0.760</td>
<td>How Often Child Gets Positive Encouragement Ages 13–14</td>
<td>0.032</td>
<td>0.968</td>
</tr>
</tbody>
</table>
4.2.3. *The Effect of Ignoring Measurement Error on the Estimated Technology*

We now demonstrate the impact of neglecting measurement error on estimates of the technology. To make the most convincing case for the importance of measurement error, we use the least error-prone proxies as determined in our estimates of Tables IIA and IIB.\(^{38}\) We continue to assume no heterogeneity.

Not accounting for measurement error has substantial effects on the estimated technology. Comparing the estimates in Table III with those in Table I, the estimated first stage investment effects are much less precisely estimated in a model that ignores measurement errors than in a model that corrects for them. In the second stage, the estimated investment effects are generally stronger. Unlike all of the specifications that control for measurement error, we estimate strong cross-productivity effects of cognitive skills on noncognitive skill production. As in Table I, there are cross-productivity effects of noncognitive skills on cognitive skills at both stages, although the estimated productivity parameters are somewhat smaller. The estimated elasticities of substitution for cognitive skills at both stages are comparable across the two specifications. The elasticities of substitution for noncognitive skills are substantially lower at both stages in the specification that does not control for measurement error. The error variances of the shocks are substantially larger. Parental cognitive skills are estimated to have substantial effects on childhood cognitive skills, but not on their noncognitive skills. This contrasts with the estimates reported in Table I that show strong effects of parental noncognitive skills on childhood cognitive skills in both stages, and on noncognitive skills in the first stage.

4.2.4. *Controlling for Time-Invariant Unobserved Heterogeneity in the Estimated Technology*

We next consider the effect of controlling for unobserved heterogeneity in the model, with estimates reported in Table I. We follow the method discussed in Section 3.6.1. Doing so allows for endogeneity of the inputs. We break the error term for the technology into two parts: a time-invariant unobserved heterogeneity factor \(\pi\) that is correlated with the vector \((\theta_t, I_t, \theta_P)\) and an i.i.d. error term \(\nu_{k,t}\) that is assumed to be uncorrelated with all other variables.

TABLE III
THE TECHNOLOGY FOR COGNITIVE AND NONCOGNITIVE SKILL FORMATION: NOT CORRECTING FOR MEASUREMENT ERROR; LINEAR ANCHORING ON EDUCATIONAL ATTAINMENT (YEARS OF SCHOOLING); NO UNOBSERVED HETEROGENEITY ($\pi$); FACTORS NORMALLY DISTRIBUTED

<table>
<thead>
<tr>
<th>Parameters</th>
<th>First Stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Period Cognitive Skills</td>
<td>$\gamma_{1,C,1}$ 0.403</td>
<td>$\gamma_{2,C,1}$ 0.657</td>
</tr>
<tr>
<td>(Self-Productivity)</td>
<td>(0.058)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Current Period Noncognitive Skills</td>
<td>$\gamma_{1,C,2}$ 0.218</td>
<td>$\gamma_{2,C,2}$ 0.009</td>
</tr>
<tr>
<td>(Cross-Productivity)</td>
<td>(0.105)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>$\gamma_{1,C,3}$ 0.067</td>
<td>$\gamma_{2,C,3}$ 0.167</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>$\gamma_{1,C,4}$ 0.268</td>
<td>$\gamma_{2,C,4}$ 0.047</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>$\gamma_{1,C,5}$ 0.044</td>
<td>$\gamma_{2,C,5}$ 0.119</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.150)</td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>$\phi_{1,C}$ 0.375</td>
<td>$\phi_{2,C}$ −0.827</td>
</tr>
<tr>
<td></td>
<td>(0.294)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>$1/(1 - \phi_{1,C})$ 1.601</td>
<td>$1/(1 - \phi_{2,C})$ 0.547</td>
</tr>
<tr>
<td>Variance of Shocks $\eta_{C,t}$</td>
<td>$\delta_{1,C}^2$ 0.941</td>
<td>$\delta_{2,C}^2$ 0.358</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>First Stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Period Cognitive Skills</td>
<td>$\gamma_{1,N,1}$ 0.193</td>
<td>$\gamma_{2,N,1}$ 0.058</td>
</tr>
<tr>
<td>(Cross-Productivity)</td>
<td>(0.095)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Current Period Noncognitive Skills</td>
<td>$\gamma_{1,N,2}$ 0.594</td>
<td>$\gamma_{2,N,2}$ 0.638</td>
</tr>
<tr>
<td>(Self-Productivity)</td>
<td>(0.090)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>$\gamma_{1,N,3}$ 0.099</td>
<td>$\gamma_{2,N,3}$ 0.239</td>
</tr>
<tr>
<td></td>
<td>(0.296)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>$\gamma_{1,N,4}$ 0.114</td>
<td>$\gamma_{2,N,4}$ 0.065</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>$\gamma_{1,N,5}$ 0.000</td>
<td>$\gamma_{2,N,5}$ 0.000</td>
</tr>
<tr>
<td></td>
<td>(0.821)</td>
<td>(0.203)</td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>$\phi_{1,N}$ −0.723</td>
<td>$\phi_{2,N}$ −0.716</td>
</tr>
<tr>
<td></td>
<td>(0.441)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>$1/(1 - \phi_{1,N})$ 0.580</td>
<td>$1/(1 - \phi_{2,N})$ 0.583</td>
</tr>
<tr>
<td>Variance of Shocks $\eta_{N,t}$</td>
<td>$\delta_{1,N}^2$ 0.767</td>
<td>$\delta_{2,N}^2$ 0.597</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

*aStandard errors in parentheses.*
Table IV shows that correcting for heterogeneity, the estimated coefficients for parental investments have a greater impact on cognitive skills at the first stage. The coefficient on parental investment in the first stage is \( \gamma_{1,C,3} \approx 0.16 \), while in the second stage \( \gamma_{2,C,3} \approx 0.04 \). The elasticity of substitution in the first stage is well above 1, \( \sigma_{1,C} = \frac{1}{1-0.31} \approx 1.45 \), and in the second stage it is well below 1, \( \sigma_{2,C} \approx \frac{1}{1+1.24} \approx 0.44 \). These estimates are statistically significantly different from each other and from the estimates of the elasticities of substitution \( \sigma_{1,N} \) and \( \sigma_{2,N} \). These results suggest that early investments are important in producing cognitive skills. Consistent with the estimates reported in Table I, noncognitive skills increase cognitive skills in the first stage, but not in the second stage. Parental cognitive and noncognitive skills affect the accumulation of childhood cognitive skills.

Panel B of Table IV presents estimates of the technology of noncognitive skills. Note that, contrary to the estimates reported for the technology for cognitive skills, the elasticity of substitution increases slightly from the first stage to the second stage. For the early stage, \( \sigma_{1,N} \approx 0.62 \), while for the late stage, \( \sigma_{2,N} \approx 0.65 \). The elasticity is about 50% higher for investments in noncognitive skills for the late stage in comparison to the elasticity for investments in cognitive skills. The estimates of \( \sigma_{1,N} \) and \( \sigma_{2,N} \) are not statistically significantly different from each other, however. The impact of parental investments is about the same at early and late stages (\( \gamma_{1,N,3} \approx 0.06 \) vs. \( \gamma_{2,N,3} \approx 0.05 \)). Parental noncognitive skills affect the accumulation of a child’s noncognitive skills both in early and late periods, but parental cognitive skills have no effect on noncognitive skills at either stage. The estimates in Table IV show a strong effect of parental cognitive skills at both stages of the production of noncognitive skills.

4.2.5. A More General Approach to Solving the Problem of the Endogeneity of Inputs

This section relaxes the invariant heterogeneity assumption and reports empirical results from a more general model of time-varying heterogeneity. Our approach to estimation is motivated by the general analysis of Section 3.6.2, but, in the interest of computational tractability, we make parametric and distributional assumptions.

We augment the measurement system (3.1)–(3.3) by investment equation (3.11), which is motivated by economic theory. Our investment equation is

\[
(4.2) \quad I_t = k_C \theta_{C,i} + k_N \theta_{N,i} + k_{C,P} \theta_{C,P} + k_{N,P} \theta_{N,P} + k_y y_i + \pi_t. \]

39See Table A10-5.
40See Table A10-5.
TABLE IV
THE TECHNOLOGY FOR COGNITIVE AND NONCOGNITIVE SKILL FORMATION: LINEAR ANCHORING ON EDUCATIONAL ATTAINMENT (YEARS OF SCHOOLING); ALLOWING FOR UNOBSERVED HETEROGENEITY ($\pi$); FACTORS NORMALLY DISTRIBUTED$^a$

<table>
<thead>
<tr>
<th>Parameters</th>
<th>First Stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Period Cognitive Skills</td>
<td>$\gamma_{1,C,1}$</td>
<td>0.479</td>
</tr>
<tr>
<td>(Self-Productivity)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Current Period Noncognitive Skills (Cross-Productivity)</td>
<td>$\gamma_{1,C,2}$</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>$\gamma_{1,C,3}$</td>
<td>0.161</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>$\gamma_{1,C,4}$</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>$\gamma_{1,C,5}$</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td></td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>$\phi_{1,C}$</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td></td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>$1/(1 - \phi_{1,C})$</td>
<td>1.457</td>
</tr>
<tr>
<td>Variance of Shocks $\eta_{C,t}$</td>
<td>$\delta_{1,C}^2$</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Period Cognitive Skills (Cross-Productivity)</td>
<td>$\gamma_{1,N,1}$</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Current Period Noncognitive Skills (Self-Productivity)</td>
<td>$\gamma_{1,N,2}$</td>
<td>0.585</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td></td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>$\gamma_{1,N,3}$</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>$\gamma_{1,N,4}$</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>$\gamma_{1,N,5}$</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>$\phi_{1,N}$</td>
<td>-0.610</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td></td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>$1/(1 - \phi_{1,N})$</td>
<td>0.621</td>
</tr>
<tr>
<td>Variance of Shocks $\eta_{N,t}$</td>
<td>$\delta_{1,N}^2$</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Standard errors in parentheses.
We substitute (4.2) into equations (3.2) and (3.10). We specify the income process as

\[(4.3) \quad \ln y_t = \rho_y \ln y_{t-1} + \nu_{y,t}\]

and the equation of motion for \(\pi_t\) as

\[(4.4) \quad \pi_t = \rho_{\pi} \pi_{t-1} + \nu_{\pi,t}\]

We assume that \(\nu_{y,t} \perp (\theta'_t, \nu_{y,t}, \theta'_t, \nu_{y,t})\) for all \(t' \neq t\) and \(\nu_{y,t} \perp (y'_t, \nu_{k,t}, \theta_t, \nu_{k,t})\), \(t > t', k \in \{C, N\}\), where \(\perp\) means independence. We further assume that \(\nu_{\pi,t} \perp (\theta'_t, \theta_t, \nu_{k,t}, \theta'_t, \nu_{k,t})\) and that \((\theta_1, y_1) \perp \pi\). In addition, \(\nu_{y,t} \sim N(0, \sigma^2_y)\) and \(\nu_{\pi,t} \sim N(0, \sigma^2_{\pi})\). In Appendix A8, we report favorable results from a Monte Carlo study of the estimator based on these assumptions.

Table V reports estimates of this model. Allowing for time-varying heterogeneity does not greatly affect the estimates from the model that assumes fixed heterogeneity reported in Table IV. In the results that we describe below, we allow the innovation \(\pi_t\) to follow an AR(1) process and we estimate the investment equation \(q_{k,t}\) along with all of the other parameters estimated in the model reported in Table IV. Estimates of the parameters of equation (4.2) are presented in Appendix A10. We also report estimates of the anchoring equation and other outcome equations in that appendix.

When we introduce an equation for investment, the impact of early investments on the production of cognitive skill increases from \(\gamma_{1,C,3} \approx 0.17\) (see Table IV, panel A) to \(\gamma_{1,C,3} \approx 0.26\) (see Table V, panel A). At the same time, the estimated first stage elasticity of substitution for cognitive skills increases from \(\sigma_{1,C} = 1/(1 - \phi_{1,C}) \approx 1.5\) to \(\sigma_{1,C} = 1/(1 - \phi_{1,C}) \approx 2.4\). Note that for this specification the impact of late investments in producing cognitive skills remains largely unchanged at \(\gamma_{2,C,3} \approx 0.045\) (compare Table IV, panel A) to \(\gamma_{2,C,3} \approx 0.209\) (in Table V, panel A). The estimate of the elasticity of substitution for cognitive skill technology is about the same as \(\sigma_{2,C} = 1/(1 - \phi_{2,C}) \approx 0.44\) (Table IV, panel A) and \(\sigma_{2,C} = 1/(1 - \phi_{2,C}) \approx 0.45\) (see Table V, panel A).

We obtain comparable changes in our estimates of the technology for producing noncognitive skills. The estimated impact of early investments increases from \(\gamma_{1,N,3} \approx 0.065\) (see Table IV, panel B) to \(\gamma_{1,N,3} \approx 0.209\) (in Table V, panel B).

---

41The intercept of the equation is absorbed into the intercept of the measurement equation.
42This assumption enables us to identify the parameters of equation (4.2).
43Table A10-6 reports estimates of the parameters of the investment equation (4.2).
44We model \(q\) as time invariant, linear, and separable in its arguments, although this is not a necessary assumption in our identification, but certainly helps to save on computation time and to obtain tighter standard errors for the policy function and the production function parameters. Notice that under our assumption \(I_{C,t} = I_{N,t} = I\), and time invariance of the investment function, it follows that \(q_{C,t} = q = q\) for all \(t\).
45We also report the covariance matrix for the initial conditions of the model in the Appendix.
<table>
<thead>
<tr>
<th>Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)</th>
<th>First Stage Parameters</th>
<th>Second Stage Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Period Cognitive Skills</td>
<td>$\gamma_{1,C,1}$</td>
<td>0.485 (0.031)</td>
</tr>
<tr>
<td>(Self-Productivity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Period Noncognitive Skills (Cross-Productivity)</td>
<td>$\gamma_{1,C,2}$</td>
<td>0.062 (0.026)</td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>$\gamma_{1,C,3}$</td>
<td>0.261 (0.026)</td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>$\gamma_{1,C,4}$</td>
<td>0.035 (0.015)</td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>$\gamma_{1,C,5}$</td>
<td>0.157 (0.033)</td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>$\phi_{1,C}$</td>
<td>0.585 (0.225)</td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>$1/(1 - \phi_{1,C})$</td>
<td>2.410 (0.007)</td>
</tr>
<tr>
<td>Variance of Shocks $\eta_{C,t}$</td>
<td>$\delta_{1,C}^2$</td>
<td>0.165 (0.007)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)</th>
<th>First Stage Parameters</th>
<th>Second Stage Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Period Cognitive Skills</td>
<td>$\gamma_{1,N,1}$</td>
<td>0.000 (0.028)</td>
</tr>
<tr>
<td>(Cross-Productivity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Period Noncognitive Skills (Self-Productivity)</td>
<td>$\gamma_{1,N,2}$</td>
<td>0.602 (0.034)</td>
</tr>
<tr>
<td>Current Period Investments</td>
<td>$\gamma_{1,N,3}$</td>
<td>0.209 (0.031)</td>
</tr>
<tr>
<td>Parental Cognitive Skills</td>
<td>$\gamma_{1,N,4}$</td>
<td>0.014 (0.013)</td>
</tr>
<tr>
<td>Parental Noncognitive Skills</td>
<td>$\gamma_{1,N,5}$</td>
<td>0.175 (0.033)</td>
</tr>
<tr>
<td>Complementarity Parameter</td>
<td>$\phi_{1,N}$</td>
<td>-0.464 (0.263)</td>
</tr>
<tr>
<td>Implied Elasticity of Substitution</td>
<td>$1/(1 - \phi_{1,N})$</td>
<td>0.683 (0.012)</td>
</tr>
<tr>
<td>Variance of Shocks $\eta_{N,t}$</td>
<td>$\delta_{1,N}^2$</td>
<td>0.203 (0.012)</td>
</tr>
</tbody>
</table>

*Standard errors in parentheses.
panel B). The elasticity of substitution for noncognitive skills in the early period rises, changing from $\sigma_{1,N} = 1/(1 - \phi_{1,N}) \approx 0.62$ to $\sigma_{1,N} = 1/(1 - \phi_{1,N}) \approx 0.68$ (in Table V, panel B). The estimated share parameter for late investments in producing noncognitive skills increases from $\gamma_{2,N,3} \approx 0.05$ to $\gamma_{2,N,3} \approx 0.10$. Compare Table IV, panel B with Table V, panel B. When we include an equation for investments, the estimated elasticity of substitution for noncognitive skills slightly increases at the later stage, from $\sigma_{2,N} = 1/(1 - \phi_{2,N}) \approx 0.645$ (in Table IV, panel B) to $\sigma_{2,N} = 1/(1 - \phi_{2,N}) \approx 0.66$ (in Table V, panel B), but this difference is not statistically significant. Thus, the estimated elasticities of substitution from the more general procedure show roughly the same pattern as those obtained from the procedure that assumes time-invariant heterogeneity.

The general pattern of decreasing substitution possibilities across stages for cognitive skills and roughly constant or slightly increasing substitution possibilities for noncognitive skills is consistent with the literature on the evolution of cognitive and personality traits (see Borghans, Duckworth, Heckman, and ter Weel (2008), Shiner (1998), Shiner and Caspi (2003)). Cognitive skills stabilize early in the life cycle and are difficult to change later on. Noncognitive traits flourish, that is, more traits are exhibited at later ages of childhood and there are more possibilities (more margins to invest in) for compensation of disadvantage. For a more extensive discussion, see Appendix A1.2.

4.2.6. A Model Based Only on Cognitive Skills

Most of the empirical literature on skill production focuses on cognitive skills as the output of family investment (see, e.g., Todd and Wolpin (2005, 2007) and the references they cite). It is of interest to estimate a more traditional model that ignores noncognitive skills and the synergism between cognitive and noncognitive skills and between investment and noncognitive skills in production. Appendix Table A14-1 reports estimates of a version of the model in Table IV, based on a model with time-invariant heterogeneity, where noncognitive skills are excluded from the analysis.

The estimated self-productivity effect increases from the first stage to the second stage, as occurs with the estimates found for all other specifications estimated in this paper. However, the estimated first period elasticity of substitution is much smaller than the corresponding parameter in Table IV. The estimated second period elasticity is slightly higher. The estimated productivity parameters for investment are substantially higher in both stages of the model reported in Appendix Table A14-1, as are the productivity parameters for parental cognitive skills. We note in the next section that the policy implications from a cognitive-skill-only model are very different from the policy implications for a model with cognitive and noncognitive skills.

We cannot reject the null hypothesis that $\sigma_{1,N} = \sigma_{2,N}$, but we reject the null hypothesis that $\sigma_{1,C} = \sigma_{2,C}$ and that the elasticities of different skills are equal. See Table A10-7.
4.3. Interpreting the Estimates

The major findings from our analysis of models with two skills that control for measurement error and endogeneity of inputs are as follows: (a) Self-productivity becomes stronger as children become older, for both cognitive and noncognitive skill formation. (b) Complementarity between cognitive skills and investment becomes stronger as children become older. The elasticity of substitution for cognition is smaller in second stage production. It is more difficult to compensate for the effects of adverse environments on cognitive endowments at later ages than it is at earlier ages. This pattern of the estimates helps to explain the evidence on ineffective cognitive remediation strategies for disadvantaged adolescents reported in Cunha, Heckman, Lochner, and Masterov (2006). (c) Complementarity between noncognitive skills and investments becomes slightly weaker as children become older, but the estimated effects are not that different across stages of the life cycle. The elasticity of substitution between investment and current endowments increases slightly between the first stage and the second stage in the production of noncognitive skills. It is somewhat easier at later stages of childhood to remediate early disadvantage using investments in noncognitive skills.

Using the estimates present in Table IV, we find that 34% of the variation in educational attainment in the sample is explained by the measures of cognitive and noncognitive capabilities that we use: 16% is due to adolescent cognitive capabilities; 12% is due to adolescent noncognitive capabilities. Measured parental investments account for 15% of the variation in educational attainment. These estimates suggest that the measures of cognitive and noncognitive capabilities that we use are powerful, but not exclusive, determinants of educational attainment and that other factors, besides the measures of family investment that we use, are at work in explaining variation in educational attainment.

To examine the implications of these estimates, we analyze a standard social planning problem that can be solved solely from knowledge of the technology of skill formation and without knowledge of parental preferences and parental access to lending markets. We determine optimal allocations of investments from a fixed budget to maximize aggregate schooling for a cohort of children. We also consider a second social planning problem that minimizes aggregate crime. Our analysis assumes that the state has full control over family investment decisions. We do not model parental investment responses to the policy. These simulations produce a measure of the investment that is needed from whatever source to achieve the specified target.

Suppose that there are $H$ children indexed by $h \in \{1, \ldots, H\}$. Let $(\theta_{C,1,h}, \theta_{N,1,h})$ denote the initial cognitive and noncognitive skills of child $h$. She has

47This is true even in a model that omits noncognitive skills.
48The skills are correlated so the marginal contributions of each skill do not add up to 34%. The decomposition used to produce these estimates is discussed in Appendix A12.
parents with cognitive and noncognitive skills denoted by $\theta_{C,P,h}$ and $\theta_{N,P,h}$, respectively. Let $\pi_h$ denote additional unobserved determinants of outcomes. Denote $\theta_{1,h} = (\theta_{C,1,h}, \theta_{N,1,h}, \theta_{C,P,h}, \theta_{N,P,h}, \pi_h)$ and let $F(\theta_{1,h})$ denote its distribution. We draw $H$ people from the estimated initial distribution $F(\theta_{1,h})$. We use the estimates reported in Table IV in this simulation. The key substitution parameters are basically the same in this model and the more general model with estimates reported in Table V.\textsuperscript{49} The price of investment is assumed to be the same in each period.

The social planner maximizes aggregate human capital subject to a budget constraint $B = 2H$, so that the per capita budget is 2 units of investment. We draw $H$ children from the initial distribution $F(\theta_{1,h})$, and solve the problem of how to allocate finite resources $2H$ to maximize the average education of the cohort. Formally, the social planner maximizes aggregate schooling

$$\max \bar{S} = \frac{1}{H} \sum_{h=1}^{H} S(\theta_{C,3,h}, \theta_{N,3,h}, \pi_h),$$

subject to the aggregate budget constraint

$$\sum_{h=1}^{H} (I_{1,h} + I_{2,h}) = 2H,$$

the technology constraint

$$\theta_{k,t+1,h} = f_{k,t}(\theta_{C,t,h}, \theta_{N,t,h}, \theta_{C,P,h}, \theta_{N,P,h}, \pi_h)$$

for $k \in \{C, N\}$ and $t \in \{1, 2\}$,

and the initial endowments of the child and her family. We assume no discounting. Solving this problem, we obtain optimal early and late investments, $I_{1,h}$ and $I_{2,h}$, respectively, for each child $h$. An analogous social planning problem is used to minimize crime.

Figures 2 (for the child’s personal endowments) and 3 (for maternal endowments) show the profiles of early (left-hand side graph) and late (right-hand side graph) investment as a function of child and maternal endowments. (Lighter shading corresponds to higher values of investment.) Endowments are measured in units of standard deviations from the means. In each figure, the endowments not plotted are fixed at sample mean values. The optimal policy is to invest relatively more in the disadvantaged compared to the advantaged in the early years. Moon (2009) shows that, in actuality, society and family together invest much more in the early years of the advantaged compared to the disadvantaged. The decline in investment by level of advantage is dramatic

\textsuperscript{49}Simulation from the model of Section 3.6.2 (with estimates reported in Section 4.2.5) that has time-varying child quality is considerably more complicated because of the high dimensionality of the state space. We leave this for another occasion.
FIGURE 2.—Optimal early (left) and late (right) investments by child initial conditions of cognitive and noncognitive skills maximizing aggregate education (other endowments held at mean levels).

FIGURE 3.—Optimal early (left) and late (right) investments by maternal cognitive and noncognitive skills maximizing aggregate education (other endowments held at mean levels).
for early investment. Second period investment profiles are much flatter and slightly favor relatively more investment in more advantaged children. A similar profile emerges for investments to reduce aggregate crime, which for the sake of brevity, we do not display.

Figures 4 and 5 reveal that the ratio of optimal early to late investment as a function of the child’s personal endowments declines with advantage whether the social planner seeks to maximize educational attainment (left-hand side) or to minimize aggregate crime (right-hand side). A somewhat similar pattern emerges for the optimal ratio of early to late investment as a function of maternal endowments with one interesting twist. The optimal investment ratio is nonmonotonic in the mother’s cognitive skill for each level of her noncognitive skills. At very low or very high levels of maternal cognitive skills, it is better to invest relatively more in the second period than if the mother’s cognitive endowment is at the mean.

The optimal ratio of early to late investment depends on the desired outcome, the endowments of children, and the budget. Figure 6 plots the density of the optimal ratio of early to late investment for education and crime.\(^{50}\) For

\(^{50}\)The optimal policy is not identical for each \(\eta\) and depends on \(\theta_{1,\eta}\), which varies in the population. The education outcome is the number of years of schooling attainment. The crime outcome is whether or not the individual has been on probation. Estimates of the coefficients of the outcome equations including those for crime are reported in Appendix A10.
FIGURE 5.—Ratio of early to late investments by maternal cognitive and noncognitive skills maximizing aggregate education (left) and minimizing aggregate crime (right) (other endowments held at mean levels).

FIGURE 6.—Densities of ratio of early to late investments maximizing aggregate education versus minimizing aggregate crime.
both outcomes and for most initial endowments, it is optimal to invest relatively more in the first stage. Crime is more intensive in noncognitive skills than educational attainment, which depends much more strongly on cognitive skills. Because compensation for adversity in cognitive skills is much more costly in the second stage than in the first stage, it is efficient to invest relatively more in cognitive traits in the first stage relative to the second stage to promote education. Because crime is more intensive in noncognitive skills and for such skills the increase in second stage compensation costs is less steep, the optimal policy for preventing crime is relatively less intensive in first stage investment.

These simulations suggest that the timing and level of optimal interventions for disadvantaged children depend on the conditions of disadvantage and the nature of desired outcomes. Targeted strategies are likely to be effective especially for different targets that weight cognitive and noncognitive traits differently.51

4.3.1. Some Economic Intuition That Explains the Simulation Results

This subsection provides an intuition for the simulation results just discussed. Given the (weak) complementarity implicit in technology (2.3) and (2.4), how is it possible to obtain our result that it is optimal to invest relatively more in the early years of the most disadvantaged? The answer hinges on the interaction between different measures of disadvantage.

Consider the following example, where individuals have a single capability, \( \theta \). Suppose that there are two children, \( A \) and \( B \), born with initial skills \( \theta_A^1 \) and \( \theta_B^1 \), respectively. Let \( \theta_A^P \) and \( \theta_B^P \) denote the skills of the parents \( A \) and \( B \), respectively. Suppose that there are two periods for investment, which we denote by periods 1 (early) and 2 (late). For each period, there is a different technology that produces skills. Assume that the technology for period 1 is

\[
\theta_2 = \gamma_1 \theta_1 + \gamma_2 I_1 + (1 - \gamma_1 - \gamma_2) \theta_P;
\]

for period 2 it is

\[
\theta_3 = \min\{\theta_2, I_2, \theta_P\}.
\]

These patterns of complementarity are polar cases that represent, in extreme form, the empirical pattern found for cognitive skill accumulation: that substitution possibilities are greater early in life compared to later in life.

The problem of society is to choose how much to invest in child \( A \) and child \( B \) in periods 1 and 2 to maximize total aggregate skills, \( \theta_A^3 + \theta_B^3 \), subject to the

51 Appendix A13 presents additional simulations of the model for an extreme egalitarian criterion that equalizes educational attainment across all children. We reach the same qualitative conclusions about the optimality of differentially greater investment in the early years for disadvantaged children.
resource constraint \( I^A_1 + I^A_2 + I^B_1 + I^B_2 \leq M \), where \( M \) is total resources available for investment. Formally, the problem is

\[
\begin{align*}
\text{max} & \quad \left[ \min \{ \gamma_1 \theta^A_1 + \gamma_2 I^A_1 + (1 - \gamma_1 - \gamma_2) \theta^A_p, I^A_2, \theta^A_p \} + \min \{ \gamma_1 \theta^B_1 + \gamma_2 I^B_1 + (1 - \gamma_1 - \gamma_2) \theta^B_p, I^B_2, \theta^B_p \} \right] \\
\text{subject to} & \quad I^A_1 + I^A_2 + I^B_1 + I^B_2 \leq M.
\end{align*}
\]

(4.6)

When the resource constraint in (4.6) does not bind, which it does not if \( M \) is above a certain threshold (determined by \( \theta^P \)), optimal investments are

\[
\begin{align*}
I^A_1 &= \frac{(\gamma_1 + \gamma_2) \theta^A_p - \gamma_1 \theta^A_1}{\gamma_2}, & I^B_1 &= \frac{(\gamma_1 + \gamma_2) \theta^B_p - \gamma_1 \theta^B_1}{\gamma_2}, \\
I^A_2 &= \theta^A_p, & I^B_2 &= \theta^B_p.
\end{align*}
\]

Notice that if child \( A \) is disadvantaged compared to \( B \) on both measures of disadvantage (\( \theta^A_1 < \theta^B_1 \) and \( \theta^A_p < \theta^B_p \)), it can happen that

\[
I^A_1 > I^B_1, \quad \text{but} \quad I^A_2 < I^B_2
\]

if

\[
\theta^A_p - \theta^B_p > \frac{\gamma_1}{\gamma_1 + \gamma_2} (\theta^A_1 - \theta^B_1).
\]

Thus, if parental endowment differences are less negative than child endowment differences (scaled by \( \gamma_1/(\gamma_1 + \gamma_2) \)), it is optimal to invest more in the early years for the disadvantaged and less in the later years. Notice that since \( (1 - \gamma_1 - \gamma_2) = \gamma_P \) is the productivity parameter on \( \theta_P \) in the first period technology, we can rewrite this condition as \( (\theta^A_p - \theta^B_p) > \gamma_1/(1 - \gamma_P)(\theta^A_1 - \theta^B_1) \). The higher the self-productivity (\( \gamma_1 \)) and the higher the parental environment productivity, \( \gamma_P \), the more likely will this inequality be satisfied for any fixed level of disparity.

### 4.4. Implications of a One Cognitive Skill Model

Appendix A14.1 considers the policy implications of the social planner’s problem from our estimates of a model formulated solely in terms of cognitive skills. This is the traditional focus in the analysis of educational production functions. (See, e.g., Todd and Wolpin (2003, 2007) and Hanushek and Woessmann (2008).) The optimal policy is to invest relatively more in the early years of the initially advantaged. Our estimates of two-stage and one-stage models based solely on cognitive skills would indicate that it is optimal to perpetuate initial inequality and not to invest relatively more in disadvantaged young children.
5. CONCLUSION

This paper formulates and estimates a multistage model of the evolution of children’s cognitive and noncognitive skills as determined by parental investments at different stages of the life cycle of children. We estimate the elasticity of substitution between contemporaneous investment and stocks of skills inherited from previous periods and determine the substitutability between early and late investments. We also determine the quantitative importance of early endowments and later investments in determining schooling attainment. We account for the proxy nature of the measures of parental inputs and of outputs, and find evidence for substantial measurement error which, if not accounted for, leads to badly distorted characterizations of the technology of skill formation. We establish nonparametric identification of a wide class of nonlinear factor models which enables us to determine the technology of skill formation. We present an analysis of the identification of production technologies with endogenous missing inputs that is more general than the replacement function analysis of Olley and Pakes (1996) and allows for measurement error in the proxy variables.\footnote{See Heckman and Robb (1985), Heckman and Vytlacil (2007), and Matzkin (2007) for a discussion of replacement functions.} A by-product of our approach is a framework for the evaluation of childhood interventions that avoids reliance on arbitrarily scaled test scores. We develop a nonparametric approach to this problem by anchoring test scores in adult outcomes with interpretable scales.

Using measures of parental investment and children’s outcomes from the Children of the National Longitudinal Survey of Youth, we estimate the parameters that govern the substitutability between early and late investments in cognitive and noncognitive skills. In our preferred empirical specification, we find much less evidence of malleability and substitutability for cognitive skills in later stages of a child’s life cycle, while malleability for noncognitive skills is about the same at both stages. These estimates are consistent with the evidence reported in Cunha, Heckman, Lochner, and Masterov (2006).

These estimates imply that successful adolescent remediation strategies for disadvantaged children should focus on fostering noncognitive skills. Investments in the early years are important for the formation of adult cognitive skills. Furthermore, policy simulations from the model suggest that there is no trade-off between equity and efficiency. The optimal investment strategy to maximize aggregate schooling attainment or to minimize aggregate crime is to target the most disadvantaged at younger ages.

Accounting for both cognitive and noncognitive skills makes a difference. An empirical model that ignores the impact of noncognitive skills on productivity and outcomes yields the opposite conclusion that an economically efficient policy that maximizes aggregate schooling would perpetuate initial advantages.
REFERENCES


CENTER FOR HUMAN RESOURCE RESEARCH (2004): NLSY79 Child and Young Adult Data User’s Guide. Columbus, OH: Ohio State University. [903]


