Estimating the Technology of Cognitive and Noncognitive Skill Formation*

Flavio Cunha
The University of Pennsylvania

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

Susanne Schennach
The University of Chicago

First draft, March 2006
This draft, December 29, 2008

*This paper was supported by grants from the National Science Foundation (SES-0241858, SES-0099195, SES-0452089, SES-0752699); the National Institute of Child Health and Human Development (R01HD43411); the J. B. and M. K. Pritzker Foundation; the Susan Buffett Foundation; the American Bar Foundation; the Children's Initiative, a project of the Pritzker Family Foundation at the Harris School of Public Policy Studies at the University of Chicago; and PAES, supported by the Pew Foundation. We thank the editor and three referees for very helpful comments. We have also benefited from comments received from Orazio Attanasio, Gary Becker, Lars Hansen, Kevin Murphy, Petra Todd, and Ken Wolpin, as well as from participants at the Yale Labor/Macro Conference (May 2006), University of Chicago Applications Workshop (June 2006), the New York University Applied Microeconomics Workshop (March 2008), the Indiana University Macroeconomics Workshop (September 2008), the Empirical Microeconomics and Econometrics Seminar at Boston College, the Applied Economics and Econometrics Seminar at the University of Western Ontario, and the IFS Conference on Structural Models of the Labour Market and Policy Analysis. A website containing supplementary material is available at http://jenni.uchicago.edu/elast-sub.

Abstract

This paper formulates and estimates multistage production functions for child cognitive and noncognitive skills. Output is 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 nonlinear factor models. A by-product of our approach is a framework for evaluating childhood interventions that does not rely on arbitrarily scaled test scores as outputs and recognizes the differential effects 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 for the production of cognitive skills. It increases in later stages of the life cycle for the production of noncognitive skills. This finding has important implications for the design of policies that target the disadvantaged. For some configurations of disadvantage and outcomes, it is optimal to invest relatively more in the later stages of childhood.

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

JEL codes C31, J13

Flavio Cunha
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19102
fcunha@sas.upenn.edu
phone: 215-898-5652

James J. Heckman
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
jjh@uchicago.edu
phone: 773-702-0634
fax: 773-702-8490
and Geary Institute
Room B005
University College Dublin
Belfield, Dublin 4, Ireland
phone: +353 1 716 4615
fax: +353 1 716 1108

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

1 Introduction

A large body of research documents the importance of cognitive skills for social and economic success.[1] An emerging body of research establishes the parallel importance of noncognitive skills.[2] Understanding the factors affecting the evolution of cognitive and noncognitive skills is important for understanding how to promote successful lives.[3]

This paper estimates the technology governing the formation of cognitive and noncognitive skills in childhood. We establish identification of general nonlinear factor models which 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.[4] 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.[5] With enough measurements relative to the number of latent skills and investments, it is possible to identify the latent state space dynamics generating the evolution of skills.

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

Drawing on the analyses of Schennach (2004a) and Hu and Schennach (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

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

[2]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 (4), Fall 2008 on noncognitive skills.

[3]See Cunha, Heckman, Lochner, and Masterov (2006) and Cunha and Heckman (2007).

[4]See Knudsen, Heckman, Cameron, and Shonkoff (2006) and Heckman (2008).

[5]See Shumway and Stoffer (1982) and Watson and Engle (1983) for early discussions of such models. Amemiya and Yalcin (2001) survey the literature on nonlinear factor analysis. Our identification analysis is new. For a recent treatment of dynamic factor and related state space models see Durbin, Harvey, Koopman, and Shephard (2004) and the voluminous literature they cite.

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. We allow inputs to interact in producing output. We generalize the factor-analytic index function models used by Carneiro, Hansen, and Heckman (2003) to allow for more general functional forms for measurement equations. We solve the problem of defining a scale for the output of childhood investments by anchoring test scores using the adult outcomes of the child, which have a well-defined cardinal scale. We determine the latent variables that generate test scores by estimating how the latent variables predict adult outcomes.[6] 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.[7]

The plan of this paper is as follows. Section 2 briefly summarizes the previous literature to motivate our generalization of it. Section 3 presents our identification analysis. Section 4 discusses our estimation strategy. Section 5 discusses the data used to estimate the model and the model estimates. Section 6 concludes.

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 J periods of adult working life, $t \in \{T+1, T+2, \ldots, T+J\}$. The T childhood periods are divided in S stages of development, $s \in \{1, \ldots, S\}$, with $S \leq T$. Adult outcomes are produced by cognitive skills, $\theta_{C,T}$, and noncognitive skills, $\theta_{N,T}$.[8] Denote parental investments at age t in child skill k by $I_{k,t}$, $k \in \{C, N\}$.

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 skills θ_t. The technology of production of skill k in period t and developmental stage s

[6] Cawley, Heckman, and Vytlacil (1999) anchor test scores in earnings outcomes.

[7] Cunha 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.

[8] This model generalizes 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.

depends on the stock of skills at date t, investment at t, $I_{k,t}$, parental skills, θ_P, shocks at period t, $\eta_{k,t}$, and the production function at stage s :

$$\theta_{k,t+1} = f_{s,k}\left(\theta_t, I_{k,t}, \theta_P, \eta_{k,t}\right), \tag{2.1}$$

for $k \in \{C, N\}$, $t \in \{1, 2, \ldots, T\}$, and $s \in \{1, \ldots, S\}$. We assume that $f_{s,k}$ is monotone increasing in its arguments, twice continuously differentiable, and concave in $I_{k,t}$. In this model, stocks of skill 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 θ_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

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

Period t stocks of abilities and skills promote 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 CPC programs promoted greater efficiency in learning in high schools and reduce problem behaviors.[9]

Adult outcome j, Q_j, is produced by a combination of different period $T+1$ skills:

$$Q_j = g_j\left(\theta_{C,T+1}, \theta_{N,T+1}\right), \quad j \in \{1, \ldots, J\}.^{10} \tag{2.2}$$

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, wages, achievement in an occupation, hours worked, criminal activity, teenage pregnancy, etc.

In this paper, we focus attention on a *CES* version of technology (2.1) where we assume

[9] See, e.g., Cunha, Heckman, Lochner, and Masterov (2006), Heckman, Moon, Pinto, Savelyev, and Yavitz (2008) and Heckman, Moon, Pinto, and Yavitz (2008).

[10] To focus on the main contribution of this paper, we focus on investment in children. Thus we assume that θ_{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).

that $\theta_{C,t}$, $\theta_{N,t}$, $I_{C,t}$, $I_{N,t}$, $\theta_{C,P}$, $\theta_{N,P}$ are scalars. Output of skills at stage s is governed by

$$\theta_{C,t+1} = \left[\gamma_{s,C,1}\theta_{C,t}^{\phi_{s,C}} + \gamma_{s,C,2}\theta_{N,t}^{\phi_{s,C}} + \gamma_{s,C,3}I_{C,t}^{\phi_{s,C}} + \gamma_{s,C,4}\theta_{C,P}^{\phi_{s,C}} + \gamma_{s,C,5}\theta_{N,P}^{\phi_{s,C}}\right]^{\frac{1}{\phi_{s,C}}}, \tag{2.3}$$

$$\theta_{N,t+1} = \left[\gamma_{s,N,1}\theta_{C,t}^{\phi_{s,N}} + \gamma_{s,N,2}\theta_{N,t}^{\phi_{s,N}} + \gamma_{s,N,3}I_{N,t}^{\phi_{s,N}} + \gamma_{s,N,4}\theta_{C,P}^{\phi_{s,N}} + \gamma_{s,N,5}\theta_{N,P}^{\phi_{s,N}}\right]^{\frac{1}{\phi_{s,N}}}, \tag{2.4}$$

where $\gamma_{s,k,l} \in [0,1]$, $\sum_l \gamma_{s,k,l} = 1$ for $k \in \{C,N\}$, $l \in \{1,\ldots,L\}$, $t \in \{1,\ldots,T\}$ and $s \in \{1,\ldots,S\}$. $\frac{1}{1-\phi_{s,k}}$ is the elasticity of substitution in the inputs producing $\theta_{k,t+1}$, where $\phi_{s,k} \in (-\infty,1]$ for $k \in \{C,N\}$. For the moment, we ignore the shocks η_t in (2.1), although they play an important role in our empirical analysis.

A *CES* specification of adult outcomes in periods after T writes

$$Q_j = \left\{\rho_j\left(\theta_{C,T+1}\right)^{\phi_{Q,j}} + (1-\rho_j)\left(\theta_{N,T+1}\right)^{\phi_{Q,j}}\right\}^{\frac{1}{\phi_{Q,j}}}, \tag{2.5}$$

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

To gain some insight into this model, consider a special case where the elasticities of substitution are the same across technologies (2.3) and (2.4) and in all outcome functions (2.5), so $\phi_{s,C} = \phi_{s,N} = \phi_{Q,j} = \phi$ for all $s \in \{1,\ldots,S\}$ and $j \in \{1,\ldots,J\}$, childhood lasts two periods ($T = 2$), there are no period "0" investments, and investment is scalar. Assume one investment good in each period that increases both cognitive and noncognitive skills, though not necessarily by the same amount. $I_{k,t} \equiv I_t$, $k \in \{C,N\}$. Assume only one adult outcome ("human capital") so $J = 1$. In this case the adult outcome function in terms of investments, initial endowments, and parental characteristics can be written as

$$Q = \left[\tau_1 I_1^{\phi} + \tau_2 I_2^{\phi} + \tau_3\theta_{C,1}^{\phi} + \tau_4\theta_{N,1}^{\phi} + \tau_5\theta_{C,P}^{\phi} + \tau_6\theta_{N,P}^{\phi}\right]^{\frac{1}{\phi}}, \tag{2.6}$$

where τ_i for $i = 1,\ldots,6$ depend on the parameters of equations (2.3)–(2.5).[11] Cunha and Heckman (2007) analyze the optimal timing of investment using a special version of the technology embodied in (2.6).

Suppose that parents maximize the net present value of child wealth, that they can lend and borrow freely at market rate r and that there is no uncertainty. Parents decide how much to invest in period "1", I_1, and period "2", I_2, and how much to transfer in risk-free

[11] See Web Appendix 5 for the derivation of this expression in terms of the parameters of equations (2.3)–(2.5).

assets at a fixed interest rate r, given total parental resources. Assuming an interior solution, and that the price of investment is the same in both periods, the optimal ratio of period 1 investment to period 2 investment is

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

Figure 1 plots the ratio of early to late investment as a function of τ_1/τ_2 for different values of ϕ. *Ceteris paribus*, the higher τ_1 relative to τ_2, the higher first period investment should be relative to second period investment. The parameters τ_1 and τ_2 are affected 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 $k \in \{C,N\}$, and also depend on the relative importance of cognitive skills, ρ, versus noncognitive skills, $1-\rho$, in producing the adult outcome Q. *Ceteris paribus*, if $\frac{\tau_1}{\tau_2} > (1+r)$, the higher the *CES* complementarity, (i.e., the lower ϕ), the greater is the ratio of early to late investment. The greater r, the smaller should be the 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.

To see how these parameters affect the ratio of early to late investment, suppose that early investment only produces cognitive skill, so that $\gamma_{1,N,3} = 0$, and late investment only produces noncognitive skill, so that $\gamma_{2,C,3} = 0$. In this case, the ratio $\left(\frac{\tau_1}{\tau_2}\right)$ can be expressed in terms of the technology and outcome function parameters:

$$\left(\frac{\tau_1}{\tau_2}\right) = \frac{(\rho\gamma_{2,C,1} + (1-\rho)\,\gamma_{2,N,1})}{(1-\rho)}\frac{\gamma_{1,C,3}}{\gamma_{2,N,3}}.$$

For a given value of ρ (the weight placed on cognition in final outcomes), the ratio of early to late investment is higher the greater the ratio $\frac{\gamma_{1,C,3}}{\gamma_{2,N,3}}$. To investigate the role ρ plays in determining the optimal ratio of investments, assume that $\gamma_{2,C,1} \geq \gamma_{2,N,1}$, so that the stock of cognitive skill, $\theta_{C,1}$, is at least as effective in producing next period cognitive skill, $\theta_{C,2}$, as in producing next period noncognitive skill, $\theta_{N,2}$. Under this assumption, the higher ρ, that is, the more important cognitive skills are in producing Q, the higher the equilibrium ratio I_1/I_2. If, on the other hand, Q is more intensive in noncognitive skills, then I_1/I_2 is smaller.

This example builds intuition about the importance of the elasticity of substitution in determining the optimal timing of lifecycle investments. However, it oversimplifies the analysis of skill formation. It is implausible that the elasticity of substitution between skills in adult output $(\frac{1}{1-\phi_Q})$ is the same as the elasticity of substitution for inputs in production, 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 outputs of multiple skills. We allow different elasticities of substitution to govern the technologies of cognitive and noncognitive skills, for these to differ at different stages of the life cycle and for both to be different from the elasticity 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 $s \in \{1, \ldots, S\}$. We do not impose the requirement that either $\phi_{s,C}$ or $\phi_{s,N}$ equals $\phi_{Q,j}$.

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: (1) 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 $\theta_{C,t}$ on $\theta_{C,t+1}$) and cross productivity (the effects of $\theta_{C,t}$ on $\theta_{N,t+1}$ and the effects of $\theta_{N,t}$ on $\theta_{C,t+1}$) at each stage of the life cycle. (2) Develop a non-linear 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 measures. (3) Estimate the elasticities of substitution for the technologies governing the production of cognitive and noncognitive skills. (4) Anchor the scale of test scores using adult outcome measures instead of relying on test scores as measures of output. (5) Account for endogeneity of parental investments. (6) Model parental investment decisions.

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 we 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 $I_{k,t}$:

$$Z_{1,k,t,j} = \mu_{1,k,t,j} + \alpha_{1,k,t,j}\theta_{k,t} + \varepsilon_{1,k,t,j} \tag{3.1}$$

$$Z_{2,k,t,j} = \mu_{2,k,t,j} + \alpha_{2,k,t,j}I_{k,t} + \varepsilon_{2,k,t,j}, \tag{3.2}$$

$$\text{where } E(\varepsilon_{a,k,t,j}) = 0, j \in \{1, \ldots, M_{a,k,t}\}, t \in \{1, \ldots, T\}, k \in \{C, N\}, a \in \{1, 2\}$$

and where $\varepsilon_{a,k,t,j}$ are uncorrelated across the j. Assuming parental endowments are measured only once in period $t = 1$, we write

$$Z_{3,k,1,j} = \mu_{3,k,1,j} + \alpha_{3,k,1,j}\theta_{k,P} + \varepsilon_{3,k,1,j}, \text{[12]} \tag{3.3}$$

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

The αs are factor loadings. The parameters and variables are defined conditional on X which we keep implicit. Following standard conventions in factor analysis, we set the scale of the factors by assuming $\alpha_{a,k,t,1} = 1$ 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.

[12] This formulation assumes that measurements $a \in \{1, 2, 3\}$ proxy only one factor. Carneiro, Hansen, and Heckman (2003) consider alternative specifications, but in a much less general econometric model. The key idea in all factor approaches is one normalization of the factor loading for each factor in one measurement to set the scale of the factor and *some* measurements for each measurement of type a dedicated to each factor. It is clear that even within the framework of this paper, as long as *some* of each of the measurements of type a satisfy the assumptions in this paper, one can identify the factor loadings of the remaining measurements that do not satisfy the assumptions if, for example, the factors are mutually independent.

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 child skills and investments in each period t, where $T \geq 2$.[13] Without loss of generality, we focus on $\alpha_{1,C,t,j}$ and note that similar expressions can be derived for the loadings for the remaining latent factors.

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

$$Cov\left(Z_{1,C,t,1}, Z_{1,C,t+1,1}\right) = Cov\left(\theta_{C,t}, \theta_{C,t+1}\right). \tag{3.4}$$

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

$$Cov\left(Z_{1,C,t,2}, Z_{1,C,t+1,1}\right) = \alpha_{1,C,t,2} Cov\left(\theta_{C,t}, \theta_{C,t+1}\right). \tag{3.5}$$

If $Cov\left(\theta_{C,t}, \theta_{C,t+1}\right) \neq 0$, one can identify the loading $\alpha_{1,C,t,2}$ from the following ratio of covariances:

$$\frac{Cov\left(Z_{1,C,t,2}, Z_{1,C,t+1,1}\right)}{Cov\left(Z_{1,C,t,1}, Z_{1,C,t+1,1}\right)} = \alpha_{1,C,t,2}.$$

If there are more than two measures of cognitive skill in each period t, one can identify $\alpha_{1,C,t,j}$ for $j \in \{2, 3, \ldots, M_{1,C,t}\}$, $t \in \{1, \ldots, T\}$ up to the normalization $\alpha_{1,C,t,1} = 1$. The assumption that the $\varepsilon_{a,k,t,j}$ 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.[14] 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, one can rewrite (3.1), assuming $\alpha_{1,C,t,j} \neq 0$, as:

$$\frac{Z_{1,C,t,j}}{\alpha_{1,C,t,j}} = \frac{\mu_{1,C,t,j}}{\alpha_{1,C,t,j}} + \theta_{C,t} + \frac{\varepsilon_{1,C,t,j}}{\alpha_{1,C,t,j}}, \; j \in \{1, 2, \ldots, M_{1,C,t}\}. \tag{3.6}$$

[13] In our framework, parental skills are assumed to be constant over time. Consequently, we need only two measures of each parental skill in one period, say the first.

[14] The idea is to write

$$\frac{Cov\left(Z_{1,C,t,2}, Z_{a',k',t',3}\right)}{Cov\left(Z_{1,C,t,1}, Z_{a',k',t',3}\right)} = \frac{\alpha_{1,C,t,2}\alpha_{a',k',t',3} Cov\left(\theta_{C,t}, \theta_{k',t'}\right)}{\alpha_{1,C,t,1}\alpha_{a',k',t',3} Cov\left(\theta_{C,t}, \theta_{k',t'}\right)} = \frac{\alpha_{1,C,t,2}}{\alpha_{1,C,t,1}} = \alpha_{1,C,t,2}$$

This only requires uncorrelatedness across different j but not across t.

In this form, it is clear that the known quantities $\frac{Z_{1,C,t,j}}{\alpha_{1,C,t,j}}$ 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 $\{\theta_{C,t}\}_{t=1}^{T}$. Proceeding in a similar fashion for all types of measurements, $a \in \{1, 2, 3\}$, on abilities $k \in \{C, N\}$, by Schennach (2004a,b), we can identify the joint distribution of all the latent variables. Define the matrix of latent variables by θ, where

$$\theta = \left(\{\theta_{C,t}\}_{t=1}^{T}, \{\theta_{N,t}\}_{t=1}^{T}, \{I_{C,t}\}_{t=1}^{T}, \{I_{N,t}\}_{t=1}^{T}, \theta_{C,P}, \theta_{N,P} \right).$$

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

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 following two vectors

$$W_i = \left(\left\{ \frac{Z_{1,C,t,i}}{\alpha_{1,C,t,i}} \right\}_{t=1}^{T}, \left\{ \frac{Z_{1,N,t,i}}{\alpha_{1,N,t,i}} \right\}_{t=1}^{T}, \left\{ \frac{Z_{2,C,t,i}}{\alpha_{2,C,t,i}} \right\}_{t=1}^{T}, \left\{ \frac{Z_{2,N,t,i}}{\alpha_{2,N,t,i}} \right\}_{t=1}^{T}, \frac{Z_{3,C,1,i}}{\alpha_{3,C,1,i}}, \frac{Z_{3,N,1,i}}{\alpha_{3,N,1,i}} \right)'$$
$$i \in \{1, 2\}.$$

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

$$\omega_i = \left(\left\{ \frac{\varepsilon_{1,C,t,i}}{\alpha_{1,C,t,i}} \right\}_{t=1}^{T}, \left\{ \frac{\varepsilon_{1,N,t,i}}{\alpha_{1,N,t,i}} \right\}_{t=1}^{T}, \left\{ \frac{\varepsilon_{2,C,t,i}}{\alpha_{2,C,t,i}} \right\}_{t=1}^{T}, \left\{ \frac{\varepsilon_{2,N,t,i}}{\alpha_{2,N,t,i}} \right\}_{t=1}^{T}, \frac{\varepsilon_{3,C,1,i}}{\alpha_{3,C,1,i}}, \frac{\varepsilon_{3,N,1,i}}{\alpha_{3,N,1,i}} \right)',$$
$$i \in \{1, 2\}.$$

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

Theorem 1 *Let W_1, W_2, θ, ω_1, ω_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) ω_2 is independent from θ, then the density of θ can be expressed *in terms of observable quantities as:*

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

where $i = \sqrt{-1}$, provided that all the requisite expectations exist and $E\left[e^{i\zeta\cdot W_2}\right]$ is nonvanishing. Note that the innermost integral is the integral of a vector-valued field along a continuous path joining the origin and the point $\chi \in \mathbb{R}^L$, while the outermost integral is over the whole $\mathbb{R}^L$ space. If θ 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.

Proof. See Web Appendix, Part 1.[15]

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 ω_1 and ω_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 ω_1 that may depend on ω_2 and θ. It also allows for potential correlation of components within the vectors ω_1 and ω_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. In order to extract the noise from 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

$$Z_j = a_j(\theta, \varepsilon_j) \text{ for } j \in \{1, \ldots, M\}, \tag{3.7}$$

where $M \geq 3$ and where the indicator Z_j is observed while the latent factor θ and the disturbance ε_j are not. The variables Z_j, θ, and ε_j are assumed to be vectors of the same dimension.

[15] The results of Theorem 1 are sketched informally in Schennach (2004a, footnote 11).

In our application, the vector of observed indicators and corresponding disturbances is

$$\begin{aligned} Z_j &= \left(\{Z_{1,C,t,j}\}_{t=1}^{T}, \{Z_{1,N,t,j}\}_{t=1}^{T}, \{Z_{2,C,t,j}\}_{t=1}^{T}, \{Z_{2,N,t,j}\}_{t=1}^{T}, Z_{3,C,1,j}, Z_{3,N,1,j} \right)' \\ \varepsilon_j &= \left(\{\varepsilon_{1,C,t,j}\}_{t=1}^{T}, \{\varepsilon_{1,N,t,j}\}_{t=1}^{T}, \{\varepsilon_{2,C,t,j}\}_{t=1}^{T}, \{\varepsilon_{2,N,t,j}\}_{t=1}^{T}, \varepsilon_{3,C,1,j}, \varepsilon_{3,C,N,1,j} \right)' \end{aligned}$$

while the vector of unobserved latent factors is:

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

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 θ in Equations (3.7) is identified under the following conditions:*

1. *The joint density*[16] *of θ, Z_1, Z_2, Z_3 is bounded and so are all their marginal and conditional densities.*

2. *Z_1, Z_2, Z_3 are mutually independent conditional on θ.*

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

4. *Whenever $\theta \neq \tilde{\theta}$, $p_{Z_3|\theta}(Z_3 \mid \theta)$ and $p_{Z_3|\theta}\left(Z_3 \mid \tilde{\theta}\right)$ differ over a set of strictly positive probability.*

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

Proof. See Web Appendix, Part 1.[17]

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,

[16]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 μ. Hence θ, Z_1, Z_2 must be continuously distributed while Z_3 may be continuous or discrete.

[17]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.

we work with their densities. This approach offers the advantage that the problem remains linear even when the random vectors are nonlinearly related.

The conditional independence requirement of Assumption 2 is weaker than the full independence assumption traditionally made in standard linear factor models as it allows for heteroskedasticity. Assumption 3 requires θ, Z_1, Z_2 to be vectors of the same dimensions, while Assumption 4 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 3 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 $p_{\theta|Z_1}(\theta|Z_1)$).[18]

Assumption 4 is automatically satisfied, for instance, if θ is univariate and $a_3(\theta, \varepsilon_3)$ is strictly increasing in θ. However, it holds much more generally. Since $a_3(\theta, \varepsilon_3)$ is nonseparable, the distribution of Z_3 conditional on θ can change with θ, thus making it possible for Assumption 4 to be satisfied even if $a_3(\theta, \varepsilon_3)$ is not strictly increasing in θ.

Assumption 5 specifies how the observed Z_1 is used to "anchor" the scale of the unobserved θ. The most common choice of functional Ψ would be 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 ε_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 ε_j or that the functions $a_j(\cdot,\cdot)$ are identified. In fact, it is always possible to alter the distribution of ε_j and the dependence of the function $a_j(\cdot,\cdot)$ on its second argument in ways that cancel each other out, as noted in the literature on nonseparable models.[19] However, lack of identifiability of these features of the model does not prevent identification of the distribution of θ.

[18] In the case of classical measurement error, bounded completeness assumptions can be phrased in terms of primitive conditions requiring 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).

[19] See Matzkin (2003, 2007).

Nevertheless, various normalizations ensuring that the functions $a_j(\theta, \varepsilon_j)$ are fully identified are available. For example, if each element of ε_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 ε_j with zero conditional mean of ε_j given θ.[20]

The conditions justifying 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 θ from general nonseparable measurement equations does not require these strong assumptions.

3.4 Identification of the Technology Function

Once the density of 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 $f_{s,k}$ because, without further normalizations, a change in the density of $\eta_{k,t}$ can be undone by a change in the function $f_{s,k}$.

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 show that $f_{s,k}$ is nonparametrically identified, by noting that, from the knowledge of p_θ, we can calculate, for any $\bar{\theta} \in \mathbb{R}$,

$$\Pr\left[\theta_{k,t+1} \leq \bar{\theta} | \theta_t, I_{k,t}, \theta_P\right] \equiv G\left(\bar{\theta} | \theta_t, I_{k,t}, \theta_P\right).$$

We identify technology (2.1) using the relationship

$$f_{s,k}\left(\theta_t, I_{k,t}, \theta_P\right) = G^{-1}\left(\eta_{k,t} | \theta_t, I_{k,t}, \theta_P\right)$$

where $G^{-1}\left(\eta_{k,t} \mid \theta_t, I_{k,t}, \theta_P\right)$ denotes the inverse of $G\left(\bar{\theta} \mid \theta_t, I_{k,t}, \theta_P\right)$ with respect to its first argument, i.e. the value $\bar{\theta}$ such that $\eta_{k,t} = G\left(\bar{\theta} \mid \theta_t, I_{k,t}, \theta_P\right)$. By construction, this operation

[20] Observe 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.

produces a function $f_{s,k}$ that generates outcomes $\theta_{k,t+1}$ with the appropriate distribution, because a random variable is mapped into a uniformly distributed variable under the mapping defined by its own cdf.

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

$$f_{s,k}(\theta_t, I_{k,t}, \theta_P) \equiv E\left[\theta_{k,t+1} \mid \theta_t, I_{k,t}, \theta_P\right],$$

where the expectation is taken under the density $p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}$, which can be calculated from p_θ. The density of $\eta_{k,t}$ conditional on all variables is identified from

$$p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}(\eta_{k,t} \mid \theta_t, I_{k,t}, \theta_P) = p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}(\eta_{k,t} + E\left[\theta_{k,t+1} \mid \theta_t, I_{k,t}, \theta_P\right] \mid \theta_t, I_{k,t}, \theta_P),$$

since $p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}$ is known once p_θ 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.

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 in period $T+1$:

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

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

$$\begin{aligned} g_{C,j}(\theta_{C,T+1}) &= \mu_{4,j} + \alpha_{4,C,j}\theta_{C,T+1} \quad \text{and} \\ g_{N,j}(\theta_{N,T+1}) &= \mu_{4,j} + \alpha_{4,N,j}\theta_{N,T+1}. \end{aligned} \tag{3.8}$$

We can also anchor using nonlinear functions. One example would be an outcome produced by a latent variable $Z_{4,j}^*$, for $j \in \{1, \ldots, J\}$:

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

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

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

In this notation

$$\begin{aligned} \Pr(Z_{4,j} = 1 | \theta_{C,T+1}, \theta_{N,T+1}) &= \Pr[\varepsilon_{4,j} \leq \tilde{g}_j(\theta_{C,T+1}, \theta_{N,T+1}) | \theta_{C,T+1}, \theta_{N,T+1}] \\ &= F_{\varepsilon_{4,j}}[\tilde{g}_j(\theta_{C,T+1}, \theta_{N,T+1}) | \theta_{C,T+1}, \theta_{N,T+1}] \\ &= g_j(\theta_{C,T+1}, \theta_{N,T+1}). \end{aligned}$$

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

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 θ. 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, e.g.,

$$\begin{aligned} g_{C,j}(\theta_{C,T+1}) &\equiv \int g_j(\theta_{C,T+1}, \theta_{N,T+1}) p_{\theta_{N,T+1}}(\theta_{N,T+1}) d\theta_{N,T+1}, \\ g_{N,j}(\theta_{N,T+1}) &\equiv \int g_j(\theta_{C,T+1}, \theta_{N,T+1}) p_{\theta_{C,T+1}}(\theta_{C,T+1}) d\theta_{C,T+1}, \end{aligned} \tag{3.9}$$

where the marginal densities, $p_{\theta_{j,T+1}}(\theta_{N,T+1})$, $j \in \{C, N\}$ 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

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

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

$$\begin{aligned}&\tilde{f}_{j,s,k}\left(\tilde{\theta}_{j,C,t}, \tilde{\theta}_{j,N,t}, I_{k,t}, \theta_{C,P}, \theta_{N,P}, \eta_{k,t}\right)\\&\qquad\equiv g_{k,j}\left(f_{s,k}\left(g_{C,j}^{-1}\left(\tilde{\theta}_{j,C,t}\right), g_{N,j}^{-1}\left(\tilde{\theta}_{j,N,t}\right), I_{k,t}, \theta_{C,P}, \theta_{N,P}, \eta_{k,t}\right)\right), k \in \{C, N\}\end{aligned}$$

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

$$\begin{aligned}&\tilde{f}_{j,s,k}\left(\tilde{\theta}_{j,C,t}, \tilde{\theta}_{j,N,t}, I_{k,t}, \theta_{C,P}, \theta_{N,P}, \eta_{k,t}\right)\\&\quad= \tilde{f}_{j,s,k}\left(g_{C,j}(\theta_{C,t}), g_{N,j}(\theta_{N,t}), I_{k,t}, \theta_{C,P}, \theta_{N,P}, \eta_{k,t}\right)\\&\quad= g_{k,j}\left(f_{s,k}\left(g_{C,j}^{-1}(g_{C,j}(\theta_{C,t})), g_{N,j}^{-1}(g_{N,j}(\theta_{N,t})), I_{k,t}, \theta_{C,P}, \theta_{N,P}, \eta_{k,t}\right)\right)\\&\quad= g_{k,j}\left(f_{s,k}\left(\theta_{C,t}, \theta_{N,t}, I_{k,t}, \theta_{C,P}, \theta_{N,P}, \eta_{k,t}\right)\right)\\&\quad= g_{k,j}(\theta_{k,t+1}) = \tilde{\theta}_{j,k,t+1},\end{aligned}$$

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

3.6 Allowing for Unobserved Heterogeneity

Thus far, we have maintained the assumption that the error term $\eta_{k,t}$ in the technology (2.1) is independent of all the other inputs $(\theta_t, I_{k,t}, \theta_P)$ as well as $\eta_{l,t}, k \neq l$. 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 strong assumption. The availability of data on adult outcomes can be exploited to relax this assumption and allow for endogeneity of the inputs.

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 heterogeneity component, π:

$$Z_{4,j} = \alpha_{4,C,j}\theta_{C,T+1} + \alpha_{4,N,j}\theta_{N,T+1} + \alpha_{4,\pi,j}\pi + \varepsilon_{4,j}, \text{ for } j \in \{1, 2, \ldots, J\}.$$

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 $(\theta_t, I_t, \theta_P, \pi)$.[21] Write $\eta_{k,t} = (\pi, \nu_{k,t})$. Extending our preceding analysis, we can identify a more general version of the technology:

$$\theta_{k,t} = f_{s,k}\left(\theta_t, I_{k,t}, \theta_P, \pi, \nu_{k,t}\right).$$

π is permitted to be correlated with the inputs $(\theta_t, I_t, \theta_P)$ and $\nu_{k,t}$ is assumed to be independent from the vector $(\theta_t, I_t, \theta_P, \pi)$ as well as $\nu_{l,t}$ for $l \neq k$.

3.7 Adding Parental Investment

Economic theory (see, e.g., Cunha and Heckman, 2007) predicts that parental investments in period t, I_t, should depend on parental skills, $(\theta_{C,P}, \theta_{N,P})$, child's skills at the beginning of period t, $(\theta_{C,t}, \theta_{N,t})$, parental income, y_t, child's unobservable heterogeneity, π, and parental wealth at period t, y_t. We write

$$I_{k,t} = g_{k,t}\left(\theta_{C,t}, \theta_{N,t}, \pi, \theta_{C,P}, \theta_{N,P}, y_t\right) + \zeta_{k,t},\ \ k \in \{C, N\}, t \in \{1, \ldots, T\} \tag{3.10}$$

$\zeta_{k,t} \perp\!\!\!\perp \theta_{t'}$ for all k and t, where the $\zeta_{k,t}$ can be unobserved state variables (such as wealth or unobserved inputs in the technology for the formation of skills) or investment shocks. Our identification analysis covers this case. To see how identification is secured, substitute (3.10) into equation (3.1) to obtain:

$$Z_{2,k,t,j} = \mu_{2,k,t,j} + \alpha_{2,k,t,j} g_{k,t}\left(\theta_{C,t}, \theta_{N,t}, \pi, \theta_{C,P}, \theta_{N,P}, y_t\right) + \alpha_{2,k,t,j}\zeta_{k,t} + \varepsilon_{2,k,t,j} \tag{3.11}$$

for $j \in \{1, \ldots, M_{2,k,t}\}$, $t \in \{1, \ldots, T\}$, and $k \in \{C, N\}$. From measurements on child skills, parental skills, child adult outcomes, and family income, we can obtain the joint distribution of $(\theta_{C,t}, \theta_{N,t}, \pi, \theta_{C,P}, \theta_{N,P}, y_t)$. We can use repeated measurements on investment in the same fashion that we use other measurements to obtain the joint distribution of the $\zeta_{k,t}$, $k \in \{C, N\}$ and $t \in \{1, \ldots, T\}$. The restrictions on the factor loadings required for identification are the same as those required for the case of unobserved heterogeneity that we previously analyzed.

Our analysis of identification of production functions with missing inputs is more general than that of Olley and Pakes (1996), who also consider use of proxies to measure unobserved inputs. They assume that the researcher has access to perfect proxies to measure unobserved inputs, whereas we allow for imperfectly measured proxies, i.e., measurement error.

[21] See Section 2 of the Web Appendix for the formal analysis of identification. We have not systematically investigated identification for a general nonseparable model for $Z_{4,j}$ with π as an argument of the function. A parametric approach appears to require at least one outcome that depends on π and not other factors.

4 Estimation

Technology (3.1) and the associated measurement systems are nonparametrically identified. However, we use parametric maximum likelihood to estimate the model and do not estimate 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.

We now develop the likelihood function for our model. Let $p(\theta)$ denote the density of θ. Although we do not directly observe θ, we observe measurements on it, Z, with realization z. Let $z_{1,k,t,j,h}$ denote measurement j associated with the skill factor $\theta_{k,t}$ for person $h \in \{1, \ldots, H\}$ in period t. Let $z_{2,k,t,j,h}$ represent measurement j associated with the investment factor $I_{k,t}$ for person h in period t. Let $z_{3,k,1,j,h}$ contain the information from measurement j on parental skill $\theta_{k,P}$. $z_{4,T+1,j,h}$ represents the vector of measurements on outcome j (*e.g.* schooling, earnings, and crime). Let $\varepsilon_{l,k,t,j,h}$ denote the measurement error associated with the measurement $z_{l,k,t,j,h}$, $l = 1, 2$. $\varepsilon_{3,k,1,j,h}$ is the measurement error associated with $z_{3,k,1,j,h}$. $\varepsilon_{4,T+1,j,h}$ is the measurement error associated with $z_{4,T+1,j,h}$. Let $p_{\varepsilon_{l,k,t,j,h}}$ denote the density function of $\varepsilon_{l,k,t,j,h}$, $l = 1, 2$. The densities of the other errors are defined in a parallel fashion. In this notation, we can write the likelihood in terms of ingredients that we can measure or identify for the model without the investment equation (but with the investment factor) and without the heterogeneity term π as:

$$p(z) = \prod_{h=1}^{H} \int \dots \int p(\theta) \tag{4.1}$$
$$\times \prod_{k\in\{C,N\}} \prod_{t=1}^{T} \prod_{j=1}^{M_{1,k,t}} p_{\varepsilon_{1,k,t,j,h}} \left(z_{1,k,t,j,h} - \mu_{1,k,t,j} - \alpha_{1,k,t,j}\theta_{k,t} \right) d\theta_{k,t}$$
$$\times \prod_{k\in\{C,N\}} \prod_{t=1}^{T} \prod_{j=1}^{M_{2,k,t}} p_{\varepsilon_{2,k,t,j,h}} \left(z_{2,k,t,j,h} - \mu_{2,k,t,j} - \alpha_{2,k,t,j} I_{k,t} \right) dI_{k,t}$$
$$\times \prod_{k\in\{C,N\}} \prod_{j=1}^{M_{3,k,1}} p_{\varepsilon_{3,k,1,j,h}} \left(z_{3,k,t,j,h} - \mu_{3,k,t,j} - \alpha_{3,k,t,j}\theta_{k,P} \right) d\theta_{k,P}$$
$$\times \prod_{j=1}^{M_{4,T+1}} p\left(z_4, j, h \right) d\theta_{C,T+1} d\theta_{N,T+1},^{22}$$

where

$$p_{4,j,h}\left(z_{4,T+1,j,h}\right) = p_{\varepsilon_{4,T+1,j,h}}\left(z_{4,T+1,j,h} - \mu_{4,T+1,j} - \alpha_{4,C,T+1}\theta_{C,T+1} - \alpha_{4,N,T+1}\theta_{N,T+1}\right)$$
$$\text{for } j = 1, ..., J_1.$$

and

$$\begin{aligned} p_{4,j,h}\left(z_{4,T+1,j,h}\right) &= F_{\varepsilon_{4,j}}\left(\mu_{4,T+1,j} + \alpha_{4,C,T+1}\theta_{C,T+1} + \alpha_{4,N,T+1}\theta_{N,T+1}\right)^{z_{4,T+1,j,h}} \\ &\times \left[1 - F_{\varepsilon_{4,j}}\left(\mu_{4,T+1,j} + \alpha_{4,C,T+1}\theta_{C,T+1} + \alpha_{4,N,T+1}\theta_{N,T+1}\right)\right]^{1-z_{4,T+1,j,h}}. \\ &\quad \text{for } j = J_1 + 1, \dots, M_{4,T+1} \end{aligned}$$

The likelihood is maximized subject to parametric versions of technology constraints (2.1) and the normalizations on the measurements discussed in section 3.1. We assume that the measurement error $\varepsilon_{l,k,t,j,h}$ is classical, and independent of θ. This assumption greatly reduces the number of terms needed to form the likelihood.[23]

In principle, one can estimate the parameters of the model, the parameters of the technology, and the $p(\theta)$ by maximizing (4.1) directly. In order to do that, one can approximate $p(z)$ by computing the integrals numerically in a deterministic fashion. However, if the number

[22]See the Web Appendix for a more detailed derivation of the likelihood function and filtering equations (see Web Appendix Section 3 and Web Appendix Section 6.4). Section 6.4 presents the full model with heterogeneity and investment equations.

[23]Our analysis establishes that we can identify models with correlated measurement errors. However, the computational cost for such a model is substantial.

of integrals is very large, a serious practical problem arises. The number of points required to evaluate the integrals is very large. For example, if there are three latent variables and four time periods, so that $T = 4$, then $\dim(\theta) = 12$ and one has to compute an integral of dimension twelve to obtain the function $p(z)$. This requires computing approximately seventeen million points of evaluation for each individual h if we pick four points of evaluation for each integral. The rate of convergence of the numerical approximation decreases with $\dim(\theta)$. In order to obtain good approximations of $p(z)$ even in the case with three factors and four time periods, we would need more than 4 points of evaluation for each integral.

We avoid this problem by relying on nonlinear filtering methods. They facilitate the approximation of the likelihood by recursive methods, greatly reducing the computational burden. Further details on how we implement nonlinear filtering are presented in Web Appendix, Section 3.

5 Estimating the Technology of Skill Formation

We estimate the technology on a sample of 2207 firstborn white children from the Children of the NLSY/79 (CNLSY/79) sample. Starting in 1986, the children of the NLSY/1979 female respondents, ages 0-14, have been assessed every two years. The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-competence of the children as well as their home environments. Data are collected via direct assessment and maternal report during home visits at every biannual wave. Section 4 of the Web Appendix discusses the measurements used to proxy investment and output. Web Appendix Tables 4-1–4-3 present summary statistics of the sample we use.[24]

To match the biennial data collection plan, in our empirical analysis, a period is equivalent to two years. We have eight periods distributed over two stages of development.[25] We report estimates of 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

[24]While we have rich data on home inputs, the information on schooling inputs is not so rich. Consistent with results reported in Todd and Wolpin (2005), we find that the poorly measured schooling inputs in the CNLSY are estimated to have only weak and statistically insignificant effects on outputs. Even correcting for measurement error, we find no evidence for important effects of schooling inputs on child outcomes. This finding is consistent with the Coleman Report (1966), but we do not push this interpretation. We do not report estimates of the model which include schooling inputs.

[25]The first period is age 0, the second period is ages 1-2, the third period covers ages 3-4, and so on until the eighth period in which children are 13-14 years-old. The first stage of development starts at age 0 and finishes at ages 5-6, while the second stage of development starts at ages 5-6 and finishes at ages 13-14.

cognitive and noncognitive development. Using the dynamic factor model, we let the data pick the best combinations of family input measures that predict the 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.

5.1 Empirical Results

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

$$\theta_{k,t+1} = \left[\gamma_{s,k,1}\theta_{C,t}^{\phi_{s,k}} + \gamma_{s,k,2}\theta_{N,t}^{\phi_{s,k}} + \gamma_{s,k,3}I_t^{\phi_{s,k}} + \gamma_{s,k,4}\theta_{C,P}^{\phi_{s,k}} + \gamma_{s,k,5}\theta_{N,P}^{\phi_{s,k}}\right]^{\frac{1}{\phi_{s,k}}} e^{\eta_{k,t+1}}, \tag{5.1}$$

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: $\eta_{k,t} \sim N\left(0, \delta_{\eta,s}^2\right)$. We further assume that the $\eta_{k,t}$ are serially independent over all t and are independent of $\eta_{\ell,t}$ for $k \neq \ell$. We assume that measurements $Z_{a,k,t,j}$ proxy the *natural logarithms* of the factors. For example, for $a = 1$,

$$\begin{aligned} Z_{1,k,t,j} &= \mu_{1,k,t,j} + \alpha_{1,k,t,j} \ln \theta_{k,t} + \varepsilon_{1,k,t,j} \\ &\quad j \in \{1, \ldots, M_{a,k,t}\}, t \in \{1, \ldots, T\}, k \in \{C, N\}. \end{aligned}$$

We use the factors (and not their logarithms) as arguments of the technology.[26] This keeps the latent factors non-negative, as is required for the definition of technology (5.1). Collect the ε terms for period t into a vector ε_t. We assume that $\varepsilon_t \sim N(0, \Lambda_t)$, where Λ_t is a diagonal matrix. We impose the condition that ε_t is independent from $\varepsilon_{t'}$ for $t \neq t'$ and all $\eta_{k,t+1}$. Define the t^{th} row of θ as θ_t^r where r stands for row. Thus

$$\ln \theta_t^r = \left(\ln \theta_{C,t}, \ln \theta_{N,t}, \ln I_t, \ln \theta_{C,P}, \ln \theta_{N,P}, \ln \pi\right).$$

[26]The modification to likelihood (4.1) from using logs is straightforward and for the sake of brevity we do not show the explicit expression. 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 (one if the child was born after 1987 and zero otherwise).

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, $\ln \theta_t$ is normally distributed with mean zero and variance-covariance matrix Σ_t. Under the second specification, $\ln \theta_t^r$ is distributed as a mixture of $\mathcal{T}$ 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 $\ln \theta_t^r$ as

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

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

We report anchored results in the text. We use the anchoring procedures described in detail in Section 6 of the Web Appendix. The anchored results allow us to compare the productivity of investments and stocks of different skills at different stages of the life cycle on the anchored outcome. In this paper, we mainly use completed years of education by age 19, a continuous variable, as an anchor. We explore the sensitivity of the estimates to alternative anchors for a one stage model in Web Appendix 7.

5.2 Empirical Estimates

This section presents results from an extensive empirical analysis estimating the multistage technology of skill formation accounting for measurement error, non-normality of the factors, endogeneity of inputs and family investment decisions. The plan of development 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 already 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 (see, e.g., Fryer and Levitt, 2004).

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 child outcomes. Controlling for endogeneity of family inputs by accounting for unobserved heterogeneity (π), 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. 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 ordering is reversed for models that control for unobserved heterogeneity (π). These estimates suggest that it is easier to redress endowment deficits that determine cognition in the first stage of a child's lifecycle 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.

Web Appendix 7 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. Anchoring affects the estimates.[27] Below, we report anchored 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.[28]

5.2.1 The Baseline Specification

Table 1 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 (π). 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.[29] The productivity parameter for investment is greater in the first period than the second period for either skill. The difference in the parameter 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. The opposite is found for cognitive skills.

[27]Cunha and Heckman (2008) show the sensitivity of the estimates to alternative anchors for a linear model specification.

[28]The normalizations for the factors are presented in Web Appendix 8.

[29]Zero values of coefficients in this and other tables arise from the optimizer attaining a boundary of zero in the parameter space.

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

5.2.2 The Magnitude of Measurement Error

Using our factor model, we can investigate the extent of measurement error on each measure of skill and investment in our data. To fix ideas, keep the conditioning on the regressors implicit and, without loss of generality, consider the measurements on cognitive skills in period t. For linear measurement equations

$$Var\left(Z_{1,C,t,j}\right) = \alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right) + Var\left(\varepsilon_{1,C,t,j}\right).$$

The fractions of the variance of $Z_{1,C,t,j}$ due to measurement error, $s^{\varepsilon}_{1,C,t,j}$, and true signal, $s^{\theta}_{1,C,t,j}$ are, respectively,

$$s^{\varepsilon}_{1,C,t,j} = \frac{Var\left(\varepsilon_{1,C,t,j}\right)}{\alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right) + Var\left(\varepsilon_{1,C,t,j}\right)} \text{ (noise)}$$

and

$$s^{\theta}_{1,C,t,j} = \frac{\alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right)}{\alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right) + Var\left(\varepsilon_{1,C,t,j}\right)} \text{ (signal).}$$

For each measure of skill and investment used in the estimation, we construct $s^{\varepsilon}_{1,C,t,j}$ and $s^{\theta}_{1,C,t,j}$ which are reported in Table 2A. Note that 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 2A also shows show the same ratios for measures of child 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 1-2, in which less than 1% of the observed variance is signal. The richest is the "BPI Headstrong" score, in which almost 62% of the observed variance is due to the variance of the signal.

Table 2A also presents the signal-noise ratio of measures of parental cognitive and noncognitive skills. Overall, measures of maternal cognitive skills tend to have 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 2B. 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 "How Often Mom Reads to the Child" has about the same informational content as "Number of Books." In later years, measures such as "Trips to the Museum" and "Attendance of Musical Performances" 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 measures of skill tend to be riddled with measurement error. The general pattern is reversed for measurement error in investments.

5.2.3 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 Table 2.[30]

Not accounting for measurement error has substantial effects on the estimated technology. Comparing the estimates in Table 3 with those in Table 1, 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 1, there are cross productivity effects of noncognitive skills on cognitive skills

[30] At birth we use Cognitive Skill: weight at birth, Noncognitive Skill: Temperament/Difficulty Scale, Parental Investment: Number of books. At ages 1–2 we use Cognitive Skill: Body Parts, Noncognitive Skill: Temperament/Difficulty Scale, Parental Investment: Number of books. At ages 3–4 we use Cognitive Skill: PPVT, Noncognitive Skill: BPI Headstrong, Parental Investment: How often mother reads to the child. At ages 5–6 to ages 13–14 we use Cognitive Skill: Reading Recognition, Noncognitive Skill: BPI Headstrong, Parental Investment: How often child is taken to musical performances. Maternal Skills are time invariant: For Maternal Cognitive Skill: ASVAB Arithmetic Reasoning, For Maternal Noncognitive Skill: Self-Esteem Item: I am a failure.

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 child cognitive skills but not their noncognitive skills. This contrasts with the estimates reported in Table 1 that show strong effects of parental noncognitive skills on child cognitive skills in both stages, and on noncognitive skills in the first stage.

5.2.4 Controlling for Unobserved Heterogeneity in the Estimated Technology

We next consider the effect of controlling for unobserved heterogeneity for the specification with estimates reported in Table 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 π that is correlated with the vector $(\theta_t, I_t, \theta_P)$ and an *i.i.d.* error term $\nu_{i,t}$ that is assumed to be uncorrelated with all other variables.

Table 4 shows that correcting for heterogeneity, the estimated coefficients for parental investments have higher impact on cognitive skills at the first stage. The coefficient on parental investment in the first stage is $\gamma_{1,C,3} \cong 0.17$, while in the second stage $\gamma_{2,C,3} \cong 0.06$. The elasticity of substitution in the first stage is well above one, $\sigma_{1,C} = \frac{1}{1-0.33} \cong 1.5$, and in the second stage it is well below one, $\sigma_{2,C} \cong \frac{1}{1+0.8} \cong 0.55$. These results suggest that early investments are important in producing cognitive skills. Consistent with the estimates reported in Table 1, noncognitive skills increase cognitive skills in the first stage, but not in the second stage. Parental cognitive and noncognitive skills affect the accumulation of child cognitive skills.

Panel B of Table 4 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 from the first stage to the second stage. At the early stage, $\sigma_{1,N} \cong 0.54$ while at the late stage, $\sigma_{2,N} \cong 0.77$. The impact of parental investments is slightly larger at late stages as well ($\gamma_{1,N,3} \cong 0.05$ vs. $\gamma_{2,N,3} \cong 0.07$). While parental noncognitive skills affect the accumulation of a child's noncognitive skills early and late, parental cognitive skills only affect the accumulation of a child's noncognitive skills at early stages. The estimates in Table 1 show no effect of parental cognitive skills on either stage of the production of cognitive skills.

5.2.5 Adjoining an Investment Equation

Table 5 reports estimates of our model when we adjoin investment parameters of the equations (3.10) to the model just discussed and identify g_t along with all of the other parameters estimated in the model reported in Table 4.[31] Estimates of the parameters of g_t are presented in Web Appendix, Part 8. We also report estimates of the anchoring equation and other outcome equations in that appendix.[32] When we introduce an equation for investment, the impact of early investments on the production of cognitive skills is reduced from $\gamma_{1,C,3} \cong 0.17$ (see Table 4, Panel A) to $\gamma_{1,C,3} \cong 0.12$ (see Table 5, Panel A). At the same time, the estimated first stage elasticity of substitution for cognitive skills increases from $\sigma_{1,C} = \frac{1}{1-\phi_{1,C}} \cong 1.5$ to $\sigma_{1,C} = \frac{1}{1-\phi_{1,C}} \cong 2$. Note that the impact of late investments in producing cognitive skills falls in value, reducing slightly from $\gamma_{2,C,3} \cong 0.06$ to $\gamma_{2,C,3} \cong 0.05$ (compare Table 4, Panel A with Table 5, Panel A). The same is true for our estimate of the elasticity of substitution for cognitive skill technology, which falls slightly from $\sigma_{2,C} = \frac{1}{1-\phi_{2,C}} \cong 0.55$ (Table 4, Panel A) to $\sigma_{2,C} = \frac{1}{1-\phi_{2,C}} \cong 0.51$ (see Table 5, Panel A).

Comparable changes in the estimates occur in our estimates of the technology for producing noncognitive skills. The impact of early investments is reduced from $\gamma_{1,N,3} \cong 0.05$ (see Table 4, Panel B) to $\gamma_{1,C,3} \cong 0.02$ (in Table 5, Panel B). The elasticity of substitution in noncognitive skills barely moves, changing from $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.54$ to $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.55$ (in Table 5, Panel B). The estimated impact of late investments in producing noncognitive skills is estimated to be somewhat smaller, falling from $\gamma_{2,C,3} \cong 0.07$ to $\gamma_{2,C,3} \cong 0.05$. Compare Table 4, Panel B with Table 5, Panel B. When we include an equation for investments, the estimated elasticity of substitution increases for noncognitive skills in late stages, from $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.55$ (in Table 4, Panel B) to $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.68$ (in Table 5, Panel B).

5.2.6 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. Table 6 reports estimates of a version of the model in Table 5 where noncognitive skills are excluded.

The estimated self-productivity effect increases from the first stage to the second stage, in accord with the estimates found for all other specifications. However, the estimated first period elasticity of substitution is much smaller than the corresponding parameter in Table

[31] We assume that g_t is linear and separable in its arguments, although this is not a necessary assumption in our identification, but certainly helps to save on computation time. Notice that under our assumption that $I_{C,t} = I_{N,t} = I_t$, $g_{k,t} = g_t$.

[32] We also report the covariance matrix for the initial conditions of the model in that appendix.

5. 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 Table 6, as are the productivity parameters for parental cognitive skills. The simulations discussed in the next subsection suggest dramatically different policies towards disadvantaged families from a model that ignores noncognitive skills compared to a model that does not.

5.3 Interpreting the Estimates

The major findings from our analysis of models with two skills that control for measurement error and endogeneity of inputs are: (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.[33] 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 weaker as children become older. The elasticity of substitution between investment and skills increases between the first stage and the second stage in the production of noncognitive skills. It is easier at *later* stages of childhood to remediate early disadvantage using investments in noncognitive skills.

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. Sixteen percent is due to adolescent cognitive capabilities. Twelve percent is due to adolescent noncognitive capabilities.[34] 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 two social planning problems 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. The first problem determines the cost of investment required to produce high school attainment for children with different initial endowments of their own and parental capabilities. For the same dis-

[33] This is true even in a model that omits noncognitive skills.

[34] The 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 Web Appendix 9.

tribution of endowments, the second problem determines optimal allocations of investments from a fixed budget to maximize aggregate schooling for a cohort of children and to minimize aggregate crime. Our analysis assumes that the state has full control over family investment decisions. For neither problem do we model parental investment responses to the policy or parental investment. 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 parents with cognitive and noncognitive skills denoted by $\theta_{C,P,h}$ and $\theta_{N,P,h}$, respectively. Let π_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 initial distribution $F(\theta_{1,h})$ that we estimate. The price of investment is assumed to be the same in each period.

The criterion adopted for the first problem assumes that the goal of society is to get the schooling of every child to a twelfth grade level. The required investments measure the power of initial endowments in determining inequality and the compensation through investment that is required to eliminate their influence. Let $e(\theta_{1,h})$ be the minimum cost of attaining 12 years of schooling for a child with endowment $\theta_{1,h}$. Assuming no discounting, the problem is formally defined by

$$e(\theta_{1,h}) = \min [I_{1,h} + I_{2,h}]$$

subject to a schooling constraint:

$$S(\theta_{C,3,h}, \theta_{N,3,h}, \pi_h) = 12,$$

where S maps end of childhood capabilities and other relevant factors (π_h) into schooling attainment, subject to the technology of capability formation constraint

$$\theta_{k,t+1,h} = f_{k,t}(\theta_{C,t,h}, \theta_{N,t,h}, \theta_{C,P,h}, \theta_{N,P,h}, I_{t,h}, \pi_h) \text{ for } k \in \{C, N\} \text{ and } t \in \{1, 2\},$$

and the initial endowments of the child and her parents. We have estimated all of the ingredient functions.[35]

Figures 2 (for child endowments) and 3 (for parental endowments) plot the percentage increase in investment over that required for a child with mean parental and personal endowments to attain high school graduation.[36] The shading in the graphs represents different values of investments. The lightly shaded areas of the graph correspond to higher values.

[35]See Web Appendix 8 for the estimates of the schooling equation.

[36]In graphing the investments as a function of the displayed endowments, we set the values of other endowments at mean values.

Eighty percent more investment is required for children with the most disadvantaged personal endowments (Figure 2). The corresponding figure for children with the most disadvantaged parental endowments is 95% (Figure 3). The negative percentages for children with high initial endowments is a measure of their advantage. From the analysis of Moon (2008), investments *received* as a function of a child's endowments are typically in reverse order from what are required. Children born with advantageous endowments typically receive more parental investment than children from less advantaged environments.

A more standard social planner's problem maximizes aggregate human capital subject to a budget constraint $B = 2H$, so that the per capita budget is 2 units of investments. 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

$$\sum_{h=1}^{H} (I_{1,h} + I_{2,h}) = 2H \tag{5.2}$$

subject to the aggregate budget constraint,

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

the technology constraint,

$$\theta_{k,t+1,h} = f_{k,t}(\theta_{C,t,h}, \theta_{N,t,h}, \theta_{C,P,h}, \theta_{N,P,h}, \pi_h) \text{ for } k \in \{C, N\} \text{ and } t \in \{1, 2\},$$

and the initial endowments of the child and her family. Again, 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 4 (for child personal endowments) and 5 (for maternal endowments) show the profiles of early (left hand side graph) and late (right hand side graph) investment as a function of endowments. For the most disadvantaged, the optimal policy is to invest a lot in the early years. The decline in investment by level of advantage is dramatic for early investment. Second period investment profiles are much flatter and slightly favor more advantaged children. A similar profile emerges for investments to reduce aggregate crime, which for the sake of brevity, we do not display.

Figures 6 and 7 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 non-monotonic 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 her endowment is at the mean.

The optimal ratio of early-to-late investment depends on the desired outcome, the endowments of children and budget $B = 2H$. Figure 8 plots the density of the ratio of early-to-late investment for education and crime.[37] Crime is more intensive in noncognitive skill than educational attainment, which depends much more strongly on cognitive skills. Because compensation for adversity in noncognitive skills is less costly in the second period than in the first period, while the opposite is true for cognitive skills, it is optimal to weigh first and second period investments in the directions indicated in the figure.

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.

5.4 Comparison with a Model with Cognitive Skill Formation Only

We now compare the policy implications of the model formulated only for cognitive skills with estimates reported in Table 6. We consider the problem of maximizing aggregate educational attainment using the estimates from a model with only cognitive skills. Figures 9 and 10 compare optimal early investments from the cognitive-skill-only model (left) with investments from the model with both skills (right). As before, less shaded regions of the figures correspond to higher values for investment.

A model of skill formation that focuses solely on cognitive skills suggests that it is optimal to perpetuate inequality. In contrast to the implications from the two skill model, investments are *lower* at the first stage of the life cycle for the most disadvantaged as measured by initial endowments compared to the most advantaged. The cognition-only model ignores the cross productivity of noncognitive skills on cognitive skills and the greater malleability of noncognitive skills in the second stage. By ignoring a central feature of the human skill

[37] The optimal policy is not identical for each h and depends on $\theta_{1,h}$, which varies in the population. The crime outcome is the number of arrests. Estimates of the coefficients of the outcome equations including those for crime are reported in Web Appendix Section 8.

formation process, it produces a misleading guide to public policy.[38]

6 Conclusion

This paper formulates and estimates a multistage model of the evolution of child 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 to 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 enable us to determine the technology of skill formation. A by-product of our approach is a framework for the evaluation of childhood interventions that avoids reliance on arbitrarily scaled test scores. We develop a nonparametric approach to this problem by anchoring test scores in adult outcomes with interpretable scales.

Using measures of parental investment and child outcomes from the Children of the National Longitudinal Survey of Youth, we estimate the parameters governing the substitutability between early and late investments in cognitive and noncognitive skills. In our preferred specification, we find greater malleability and substitutability for noncognitive skills in later stages of a child's life cycle than for cognitive skills, consistent with evidence reported in Cunha, Heckman, Lochner, and Masterov (2006). These estimates imply that successful adolescent remediation strategies for disadvantaged children should focus on noncognitive skills. Investments in the early years are important for the formation of adult cognitive skills. Policy simulations from the model suggest that there is no tradeoff between equity and efficiency. The optimal investment strategy to maximize aggregate schooling attainment is to target the most disadvantaged at younger ages. Accounting for noncognitive skills is important. A model that ignores the impact of noncognitive skills on productivity and outcomes suggests an equity-efficiency tradeoff and that to maximize aggregate productivity those born with the most advantage should receive relatively more investment in the early years.

[38]Web Appendix 10 shows that this contrast is stronger if we assume a one stage-one cognitive skill model.

References

Amemiya, Y. and I. Yalcin (2001). Nonlinear factor analysis as a statistical method. *Statistical Science 16*(3), 275–294.

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

Ben-Porath, Y. (1967, August). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4, Part 1), 352–365.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

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

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

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

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

Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

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

d'Haultfoeuille, X. (2006). On the completeness condition in nonparametric instrumental problems. Working Paper, ENSAE, CREST-INSEE and Université de Paris I.

Durbin, J., A. C. Harvey, S. J. Koopman, and N. Shephard (2004). *State Space and Unobserved Component Models: Theory and Applications: Proceedings of a Conference in Honour of James Durbin*. New York NY: Cambridge University Press.

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

Heckman, J. J. (2008, July). Schools, skills and synapses. *Economic Inquiry 46*(3), 289–324.

Heckman, J. J., L. J. Lochner, and C. Taber (1998, January). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics 1*(1), 1–58.

Heckman, J. J., S. H. Moon, R. R. Pinto, and A. Q. Yavitz (2008). The rate of return to the Perry Preschool program. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., S. H. Moon, R. R. A. Pinto, P. A. Savelyev, and A. Q. Yavitz (2008). Cost-benefit analysis of the Perry preschool program. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

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

Hu, Y. and S. M. Schennach (2008, January). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Kniesner, T. J. and B. ter Weel (2008, Fall). Special issue on noncognitive skills and their development. *Journal of Human Resources 43*(4), 729–1059.

Knudsen, E. I., J. J. Heckman, J. Cameron, and J. P. Shonkoff (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences 103*(27), 10155–10162.

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

Mattner, L. (1993, December). Some incomplete but boundedly complete location families. *Annals of Statistics 21*(4), 2158–2162.

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

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

Moon, S. H. (2008). *Skill Formation Technology and Multi-Dimensional Parental Investment.* Ph. D. thesis, University of Chicago, Department of Economics.

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

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

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

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

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

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

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

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

Todd, P. E. and K. I. Wolpin (2005). The production of cognitive achievement in children: Home, school and racial test score gaps. Working paper, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Todd, P. E. and K. I. Wolpin (2007, Winter). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital 1*(1), 91–136.

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

Table 1
Estimates of the Technology
Using the Factor Model to Correct for Measurement Error
Linear Anchoring on Educational Attainment (Years of Schooling)
No Unobserved Heterogeneity (π), Factors Normally Distributed

The Technology of Cognitive Skill Formation				
		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.424	$\gamma_{2,C,1}$	0.852
		(0.028)		(0.011)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.146	$\gamma_{2,C,2}$	0.007
		(0.036)		(0.010)
Current Period Investments	$\gamma_{1,C,3}$	0.266	$\gamma_{2,C,3}$	0.032
		(0.026)		(0.006)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.048	$\gamma_{2,C,4}$	0.102
		(0.018)		(0.016)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.117	$\gamma_{2,C,5}$	0.007
		(0.038)		(0.025)
Complementarity Parameter	$\phi_{1,C}$	0.298	$\phi_{2,C}$	-1.033
		(0.117)		(0.107)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	1.425	$1/(1-\phi_{2,C})$	0.492
Variance of Shocks $\eta_{c,t}$	$\delta^2_{1,C}$	0.180	$\delta^2_{2,C}$	0.087
		(0.007)		(0.003)

The Technology of Noncognitive Skill Formation				
		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.000
		(0.017)		(0.007)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.731	$\gamma_{2,N,2}$	0.867
		(0.029)		(0.010)
Current Period Investments	$\gamma_{1,N,3}$	0.078	$\gamma_{2,N,3}$	0.067
		(0.017)		(0.006)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.000	$\gamma_{2,N,4}$	0.000
		(0.011)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.190	$\gamma_{2,N,5}$	0.066
		(0.026)		(0.014)
Complementarity Parameter	$\phi_{1,N}$	-0.083	$\phi_{2,N}$	-0.150
		(0.187)		(0.064)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.923	$1/(1-\phi_{2,N})$	0.869
Variance of Shocks $\eta_{n,t}$	$\delta^2_{1,N}$	0.179	$\delta^2_{2,N}$	0.104
		(0.011)		(0.004)

Note: Standard errors in parenthesis

Table 2A

Percentage of Total Variance in Measurements due to Signal and Noise

Measurement of Child's Cognitive Skills	**%Signal**	**%Noise**	**Measurement of Child's Noncognitive Skills**	**%Signal**	**%Noise**
Gestation Length	0.485	0.515	Difficulty at Birth	0.231	0.769
Weight at Birth	0.574	0.426	Friendliness at Birth	0.229	0.771
Motor-Social Development at Birth	0.045	0.955	Compliance at Ages 1-2	0.307	0.693
Motor-Social Development at Ages 1-2	0.330	0.670	Insecure at Ages 1-2	0.067	0.933
Body Parts at Ages 1-2	0.335	0.665	Sociability at Ages 1-2	0.084	0.916
Memory for Locations at Ages 1-2	0.169	0.831	Difficulty at Ages 1-2	0.490	0.510
Motor-Social Development at Ages 3-4	0.404	0.596	Friendliness at Ages 1-2	0.187	0.813
Picture Vocabulary at Ages 3-4	0.467	0.533	Compliance at Ages 3-4	0.146	0.854
Picture Vocabulary at Ages 5-6	0.213	0.787	Insecure at Ages 3-4	0.138	0.862
PIAT-Mathematics at Ages 5-6	0.305	0.695	Sociability at Ages 3-4	0.010	0.990
PIAT-Reading Recognition at Ages 5-6	0.961	0.039	Behavior Problem Index Antisocial at Ages 3-4	0.457	0.543
PIAT-Reading Comprehension at Ages 5-6	0.936	0.064	Behavior Problem Index Anxiety at Ages 3-4	0.448	0.552
PIAT-Mathematics at Ages 7-8	0.432	0.568	Behavior Problem Index Headstrong at Ages 3-4	0.541	0.459
PIAT-Reading Recognition at Ages 7-8	0.870	0.130	Behavior Problem Index Hyperactive at Ages 3-4	0.380	0.620
PIAT-Reading Comprehension at Ages 7-8	0.799	0.201	Behavior Problem Index Conflict at Ages 3-4	0.359	0.641
PIAT-Mathematics at Ages 9-10	0.478	0.522	Behavior Problem Index Antisocial at Ages 5-6	0.462	0.538
PIAT-Reading Recognition at Ages 9-10	0.820	0.180	Behavior Problem Index Anxiety at Ages 5-6	0.418	0.582
PIAT-Reading Comprehension at Ages 9-10	0.670	0.330	Behavior Problem Index Headstrong at Ages 5-6	0.619	0.381
PIAT-Mathematics at Ages 11-12	0.520	0.480	Behavior Problem Index Hyperactive at Ages 5-6	0.490	0.510
PIAT-Reading Recognition at Ages 11-12	0.787	0.213	Behavior Problem Index Conflict at Ages 5-6	0.299	0.701
PIAT-Reading Comprehension at Ages 11-12	0.621	0.379	Behavior Problem Index Antisocial Ages 7-8	0.439	0.561
PIAT-Mathematics at Ages 13-14	0.554	0.446	Behavior Problem Index Anxiety Ages 7-8	0.477	0.523
PIAT-Reading Recognition at Ages 13-14	0.740	0.260	Behavior Problem Index Headstrong Ages 7-8	0.606	0.394
PIAT-Reading Comprehension at Ages 13-14	0.558	0.442	Behavior Problem Index Hyperactive Ages 7-8	0.500	0.500
Measurement of Maternal Cognitive Skills			Behavior Problem Index Conflict Ages 7-8	0.329	0.671
ASVAB Arithmetic Reasoning	0.730	0.270	Behavior Problem Index Antisocial Ages 9-10	0.485	0.515
ASVAB Word Knowledge	0.631	0.369	Behavior Problem Index Anxiety Ages 9-10	0.471	0.529
ASVAB Paragraph Composition	0.577	0.423	Behavior Problem Index Headstrong Ages 9-10	0.574	0.426
ASVAB Numerical Operations	0.457	0.543	Behavior Problem Index Hyperactive Ages 9-10	0.462	0.538
ASVAB Coding Speed	0.350	0.650	Behavior Problem Index Conflict Ages 9-10	0.366	0.634
ASVAB Mathematical Knowledge	0.657	0.343	Behavior Problem Index Antisocial Ages 11-12	0.495	0.505
Measurement of Maternal Noncognitive Skills			Behavior Problem Index Anxiety Ages 11-12	0.498	0.502
Self-Esteem "I am a person of worth"	0.271	0.729	Behavior Problem Index Headstrong Ages 11-12	0.601	0.399
Self-Esteem " I have good qualities"	0.349	0.651	Behavior Problem Index Hyperactive Ages 11-12	0.502	0.498
Self-Esteem "I am a failure"	0.445	0.555	Behavior Problem Index Conflict Ages 11-12	0.365	0.635
Self-Esteem "I have nothing to be proud of"	0.373	0.627	Behavior Problem Index Antisocial Ages 13-14	0.475	0.525
Self-Esteem "I have a positive attitude"	0.406	0.594	Behavior Problem Index Anxiety Ages 13-14	0.544	0.456
Self-Esteem "I wish I had more self-respect"	0.339	0.661	Behavior Problem Index Headstrong Ages 13-14	0.592	0.408
Self-Esteem "I feel useless at times"	0.290	0.710	Behavior Problem Index Hyperactive Ages 13-14	0.522	0.478
Self-Esteem "I sometimes think I am no good"	0.374	0.626	Behavior Problem Index Conflict Ages 13-14	0.409	0.591
Locus of Control "I have no control"	0.047	0.953			
Locus of Control "I make no plans for the future"	0.064	0.936			
Locus of Control "Luck is big factor in life"	0.041	0.959			
Locus of Control "Luck plays big role in my life"	0.020	0.980			

Table 2B

Percentage of Total Variance in Measurements due to Signal and Noise

Measurements of Parental Investments	%Signal	%Noise	**Measurements of Parental Investments**	%Signal	%Noise
How Often Child Goes on Outings during Year of Birth	0.251	0.749	Child Has Musical Instruments Ages 7-8	0.101	0.899
Number of Books Child Has during Year of Birth	0.471	0.529	Family Subscribes to Daily Newspapers Ages 7-8	0.084	0.916
How Often Mom Reads to Child during Year of Birth	0.383	0.617	Child Has Special Lessons Ages 7-8	0.152	0.848
Number of Soft Toys Child Has during Year of Birth	0.251	0.749	How Often Child Goes to Musical Shows Ages 7-8	0.283	0.717
Number of Push/Pull Toys Child Has during Year of Birth	0.191	0.809	How Often Child Attends Family Gatherings Ages 7-8	0.013	0.987
How Often Child Eats with Mom/Dad during Year of Birth	0.173	0.827	How Often Child is Praised Ages 7-8	0.030	0.970
How Often Mom Calls from Work during Year of Birth	0.057	0.943	How Often Child Gets Positive Encouragement Ages 7-8	0.141	0.859
How Often Child Goes on Outings at Ages 1-2	0.070	0.930	Number of Books Child Has Ages 9-10	0.081	0.919
Number of Books Child Has Ages 1-2	0.453	0.547	Mom Reads to Child Ages 9-10	0.080	0.920
How Often Mom Reads to Child Ages 1-2	0.449	0.551	Eats with Mom/Dad Ages 9-10	0.014	0.986
Number of Soft Toys Child Has Ages 1-2	0.010	0.990	How Often Child Goes to Museum Ages 9-10	0.318	0.682
Number of Push/Pull Toys Child Has Ages 1-2	0.045	0.955	Child Has Musical Instruments Ages 9-10	0.122	0.878
How Often Child Eats with Mom/Dad Ages 1-2	0.014	0.986	Family Subscribes to Daily Newspapers Ages 9-10	0.113	0.887
Mom Calls from Work Ages 1-2	0.033	0.967	Child Has Special Lessons Ages 9-10	0.149	0.851
How Often Child Goes on Outings Ages 3-4	0.067	0.933	How Often Child Goes to Musical Shows Ages 9-10	0.363	0.637
Number of Books Child Has Ages 3-4	0.230	0.770	How Often Child Attends Family Gatherings Ages 9-10	0.027	0.973
How Often Mom Reads to Child Ages 3-4	0.386	0.614	How Often Child is Praised Ages 9-10	0.059	0.941
How Often Child Eats with Mom/Dad Ages 3-4	0.018	0.982	How Often Child Gets Positive Encouragement Ages 9-10	0.115	0.885
Number of Magazines at Home Ages 3-4	0.161	0.839	Number of Books Child Has Ages 11-12	0.101	0.899
Child Has a CD player Ages 3-4	0.148	0.852	Eats with Mom/Dad Ages 11-12	0.020	0.980
How Often Child Goes on Outings Ages 5-6	0.071	0.929	How Often Child Goes to Museum Ages 11-12	0.318	0.682
Number of Books Child Has Ages 5-6	0.102	0.898	Child Has Musical Instruments Ages 11-12	0.113	0.887
How Often Mom Reads to Child Ages 5-6	0.193	0.807	Family Subscribes to Daily Newspapers Ages 11-12	0.084	0.916
How Often Child Eats with Mom/Dad Ages 5-6	0.016	0.984	Child Has Special Lessons Ages 11-12	0.121	0.879
Number of Magazines at Home Ages 5-6	0.163	0.837	How Often Child Goes to Musical Shows Ages 11-12	0.428	0.572
Child Has CD player Ages 5-6	0.129	0.871	How Often Child Attends Family Gatherings Ages 11-12	0.018	0.982
How Often Child Goes to Museum Ages 5-6	0.247	0.753	How Often Child is Praised Ages 11-12	0.043	0.957
Child Has Musical Instruments Ages 5-6	0.112	0.888	How Often Child Gets Positive Encouragement Ages 11-12	0.057	0.943
Family Subscribes to Daily Newspapers Ages 5-6	0.139	0.861	Number of Books Child Has Ages 13-14	0.143	0.857
Child Has Special Lessons Ages 5-6	0.259	0.741	Eats with Mom/Dad Ages 13-14	0.034	0.966
How Often Child Goes to Musical Shows Ages 5-6	0.305	0.695	How Often Child Goes to Museum Ages 13-14	0.333	0.667
How Often Child Attends Family Gatherings Ages 5-6	0.021	0.979	Child Has Musical Instruments Ages 13-14	0.109	0.891
How Often Child is Praised Ages 5-6	0.006	0.994	Family Subscribes to Daily Newspapers Ages 13-14	0.088	0.912
How Often Child Gets Positive Encouragement Ages 5-6	0.087	0.913	Child Has Special Lessons Ages 13-14	0.157	0.843
Number of Books Child Has Ages 7-8	0.103	0.897	How Often Child Goes to Musical Shows Ages 13-14	0.370	0.630
How Often Mom Reads to Child Ages 7-8	0.123	0.877	How Often Child Attends Family Gatherings Ages 13-14	0.030	0.970
How Often Child Eats with Mom/Dad Ages 7-8	0.008	0.992	How Often Child is Praised Ages 13-14	0.097	0.903
How Often Child Goes to Museum Ages 7-8	0.293	0.707	How Often Child Gets Positive Encouragement Ages 13-14	0.108	0.892

Table 3

Estimates of the Technology that Ignore Measurement Error

(Model of Table 1 with No Measurement Error Corrections)

Linear Anchoring on Educational Attainment (Years of Schooling)

No Unobserved Heterogeneity (π), Factors Normally Distributed

Panel A: Technology of Cognitive Skill Formation

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

Panel B: Technology of Noncognitive Skill Formation

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

Note: Standard errors in parenthesis

Table 4
Estimated Technology Allowing for Heterogeneity
Linear Anchoring on Educational Attainment (Years of Schooling)
Allowing for Unobserved Heterogeneity (π), Factors Normally Distributed

Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.359	$\gamma_{2,C,1}$	0.656
		(0.022)		(0.014)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.045	$\gamma_{2,C,2}$	0.007
		(0.020)		(0.005)
Current Period Investments	$\gamma_{1,C,3}$	0.174	$\gamma_{2,C,3}$	0.059
		(0.014)		(0.004)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.041	$\gamma_{2,C,4}$	0.100
		(0.013)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.381	$\gamma_{2,C,5}$	0.178
		(0.024)		(0.018)
Complementarity Parameter	$\phi_{1,C}$	0.331	$\phi_{2,C}$	-0.797
		(0.107)		(0.077)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	1.496	$1/(1-\phi_{2,C})$	0.557
Variance of Shocks $\eta_{C,t}$	$\delta^2_{1,C}$	0.165	$\delta^2_{2,C}$	0.084
		(0.007)		(0.003)

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

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.000
		(0.019)		(0.008)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.451	$\gamma_{2,N,2}$	0.662
		(0.030)		(0.018)
Current Period Investments	$\gamma_{1,N,3}$	0.046	$\gamma_{2,N,3}$	0.066
		(0.016)		(0.005)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.037	$\gamma_{2,N,4}$	0.000
		(0.012)		(0.009)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.466	$\gamma_{2,N,5}$	0.272
		(0.036)		(0.019)
Complementarity Parameter	$\phi_{1,N}$	-0.852	$\phi_{2,N}$	-0.308
		(0.199)		(0.077)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.540	$1/(1-\phi_{2,N})$	0.765
Variance of Shocks $\eta_{N,t}$	$\delta^2_{1,N}$	0.220	$\delta^2_{2,N}$	0.098
		(0.013)		(0.003)

Note: Standard errors in parenthesis

Table 5

Estimates of the Technology for Cognitive and Noncognitive Skill Formation

Estimated Along with Investment Equation with

Linear Anchoring on Educational Attainment (Years of Schooling)

Allowing for Unobserved Heterogeneity (π), Factors Normally Distributed

Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.384	$\gamma_{2,C,1}$	0.770
		(0.022)		(0.018)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.071	$\gamma_{2,C,2}$	0.009
		(0.023)		(0.005)
Current Period Investments	$\gamma_{1,C,3}$	0.124	$\gamma_{2,C,3}$	0.049
		(0.015)		(0.011)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.054	$\gamma_{2,C,4}$	0.072
		(0.012)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.368	$\gamma_{2,C,5}$	0.099
		(0.026)		(0.016)
Complementarity Parameter	$\phi_{1,C}$	0.480	$\phi_{2,C}$	-0.961
		(0.109)		(0.115)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	1.925	$1/(1-\phi_{2,C})$	0.510
Variance of Shocks $\eta_{C,t}$	$\delta^2_{1,C}$	0.151	$\delta^2_{2,C}$	0.090
		(0.006)		(0.003)

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

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.000
		(0.020)		(0.009)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.526	$\gamma_{2,N,2}$	0.748
		(0.032)		(0.019)
Current Period Investments	$\gamma_{1,N,3}$	0.021	$\gamma_{2,N,3}$	0.055
		(0.010)		(0.015)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.057	$\gamma_{2,N,4}$	0.000
		(0.012)		(0.009)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.396	$\gamma_{2,N,5}$	0.196
		(0.031)		(0.021)
Complementarity Parameter	$\phi_{1,N}$	-0.818	$\phi_{2,N}$	-0.483
		(0.253)		(0.157)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.550	$1/(1-\phi_{2,N})$	0.674
Variance of Shocks $\eta_{N,t}$	$\delta^2_{1,N}$	0.210	$\delta^2_{2,N}$	0.101
		(0.012)		(0.003)

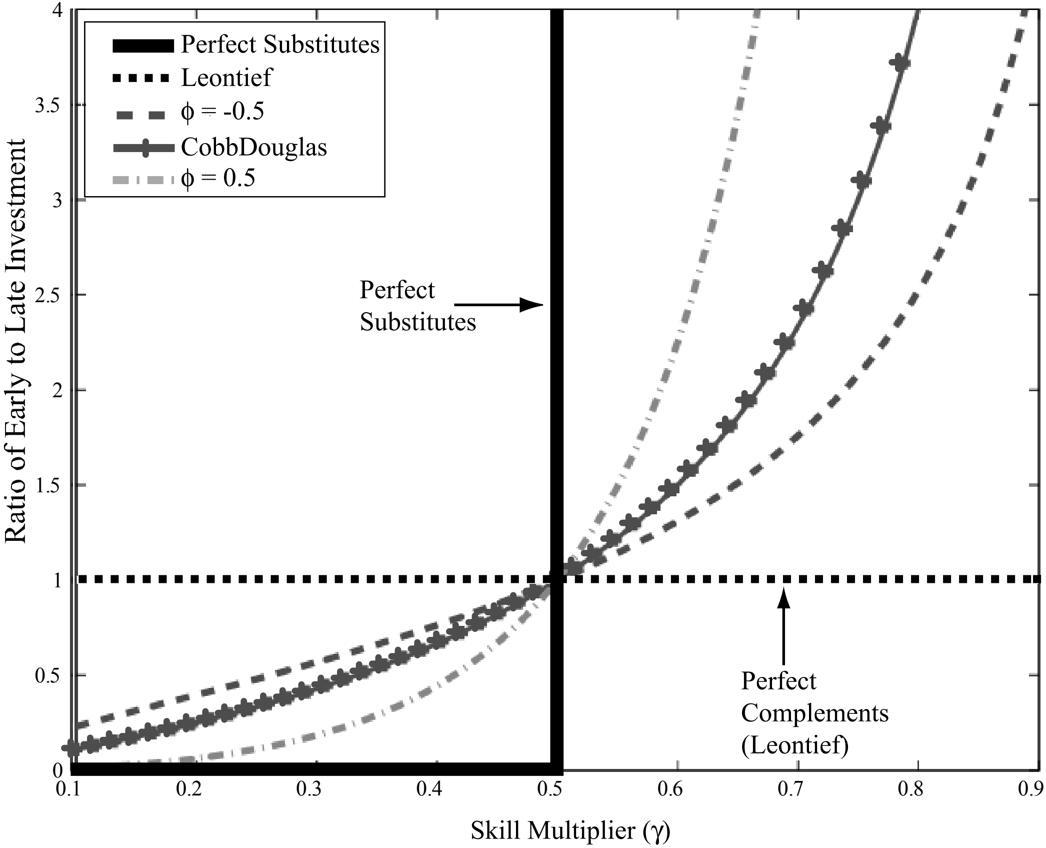
Note: Standard errors in parenthesis

Table 6
Technology of Cognitive Skill Formation
Model with Cognitive Skills Only
Estimated Along with Investment Equation with
Linear Anchoring on Educational Attainment (Years of Schooling)
Allowing for Unobserved Heterogeneity (π), Factors Normally Distributed

Current Period Cognitive Skills	$\gamma_{1,C,1}$	0.303	$\gamma_{2,C,1}$	0.448
		0.026		0.015
Current Period Investments	$\gamma_{1,C,3}$	0.319	$\gamma_{2,C,3}$	0.098
		0.025		0.015
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.378	$\gamma_{2,C,4}$	0.454
		0.022		0.017
Complementarity Parameter	$\phi_{1,C}$	-0.180	$\phi_{2,C}$	-0.781
		0.130		0.096
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	0.847	$1/(1-\phi_{2,C})$	0.562
Variance of Shocks $\eta_{C,t}$	δ^2_η	0.193	δ^2_η	0.050
		0.006		0.002

Note: Standard errors in parenthesis

Figure 1: Ratio of early to late investment in human capital
as a function of the ratio of first period to second period investment productivity
for different values of the complementarity parameter



Note: Assumes $r = 0$.
Source: Cunha and Heckman (2007).

Figure 2

Percentage Increase in Total Investments as a Function of Child Initial Conditions of Cognitive and Noncognitive Skills

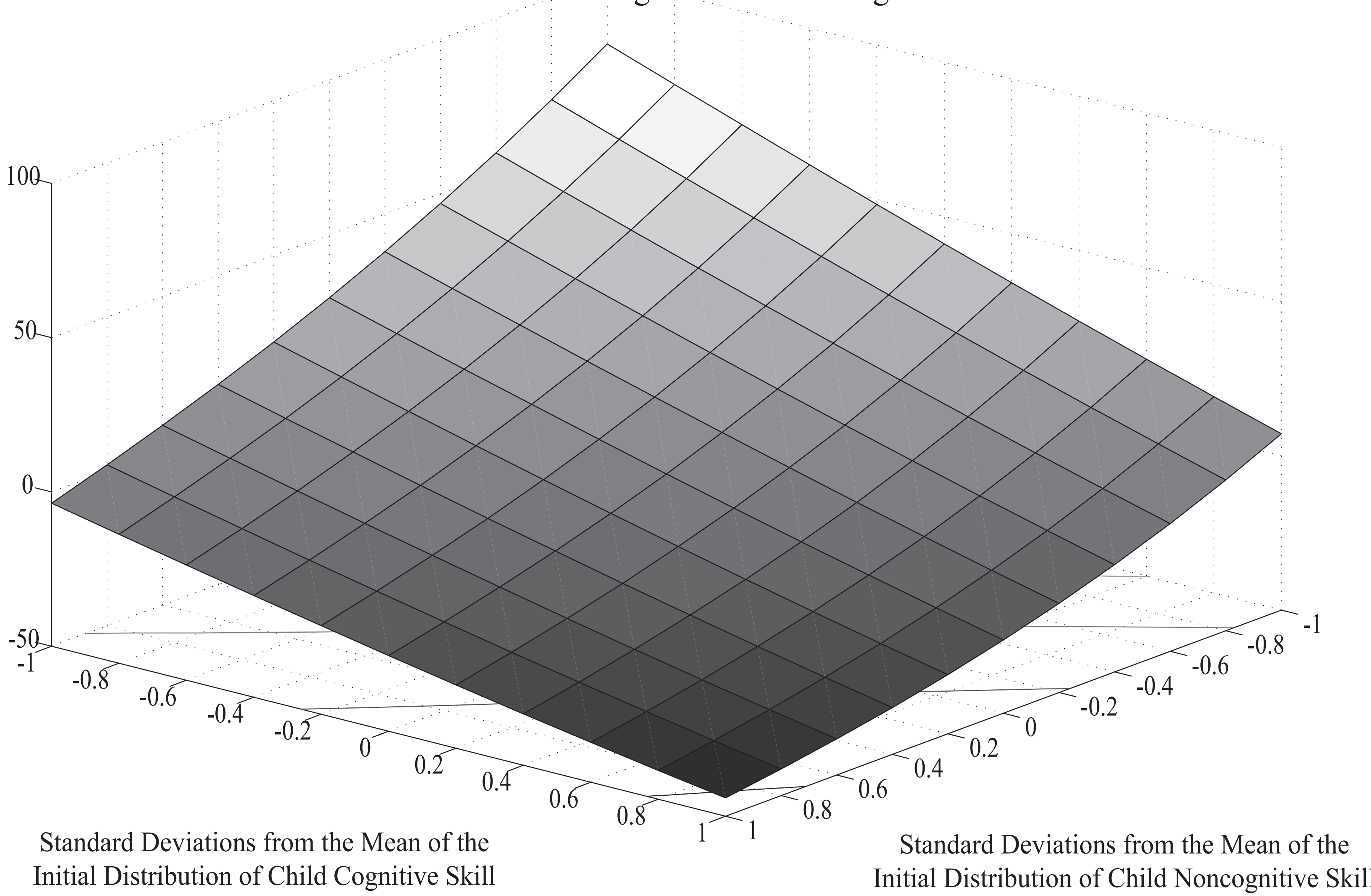


Figure 3
Percentage Increase in Total Investments
as a Function of Maternal Cognitive and Noncognitive Skills

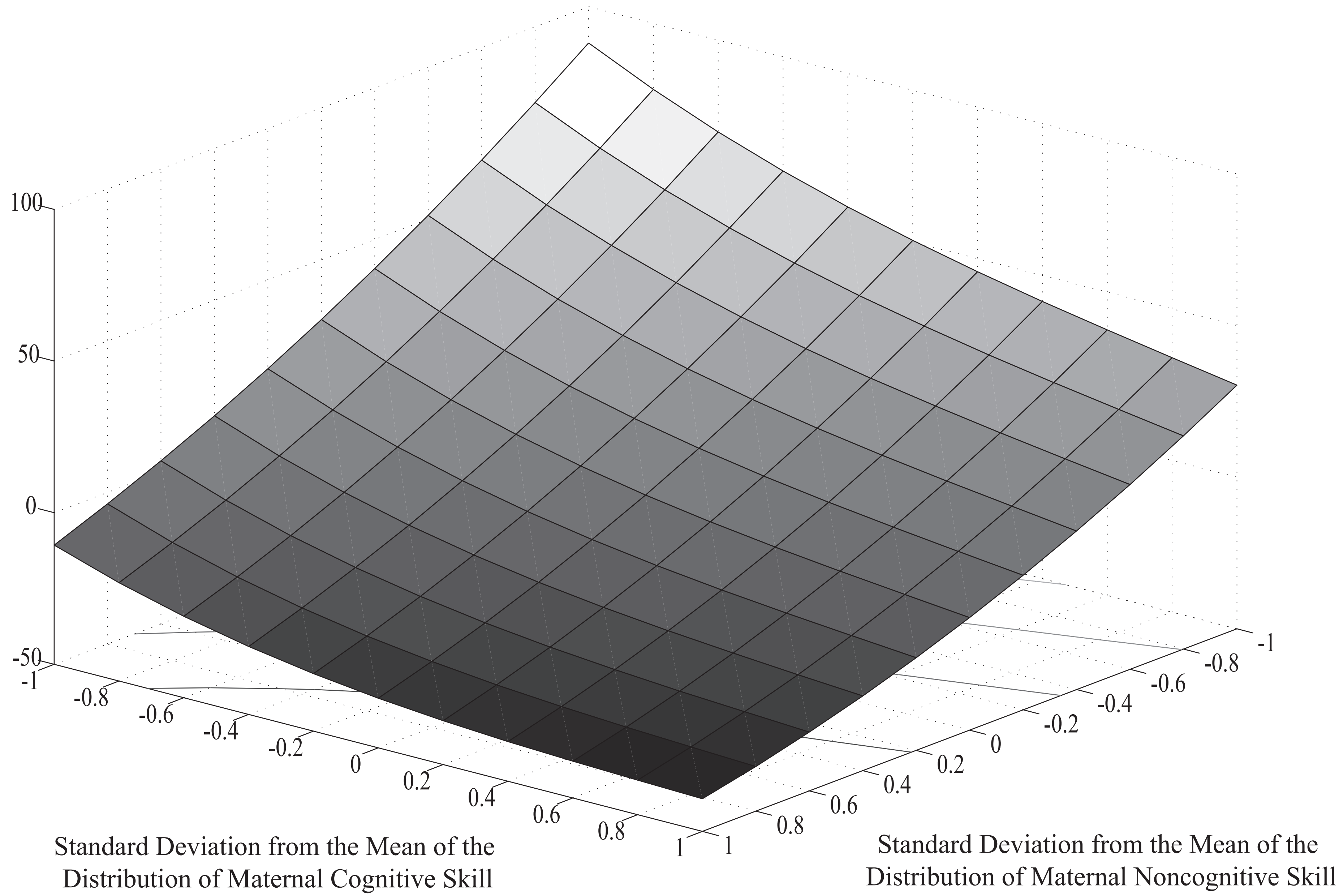
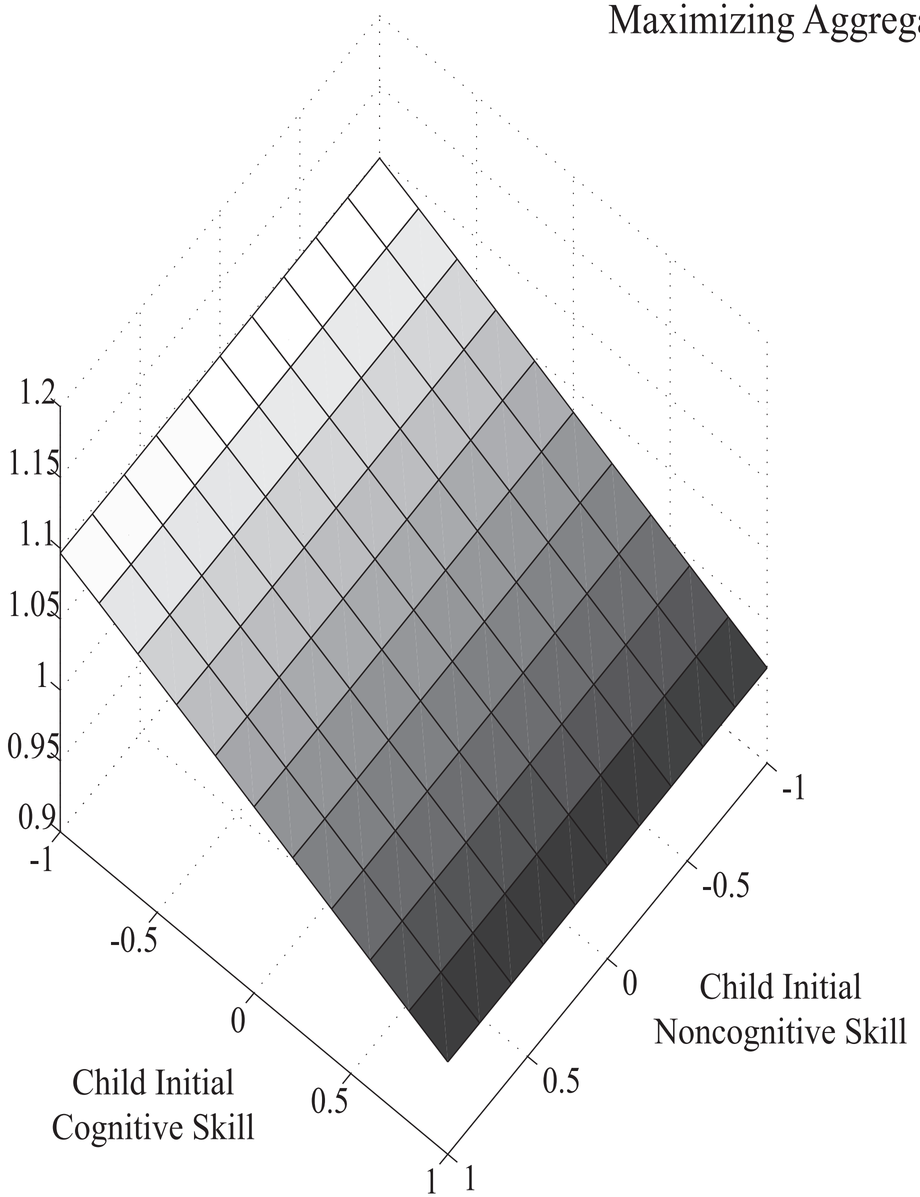


Figure 4
Optimal Early (Left) and Late (Right) Investments by
Child Initial Conditions of Cognitive and Noncognitive Skills
Maximizing Aggregate Education



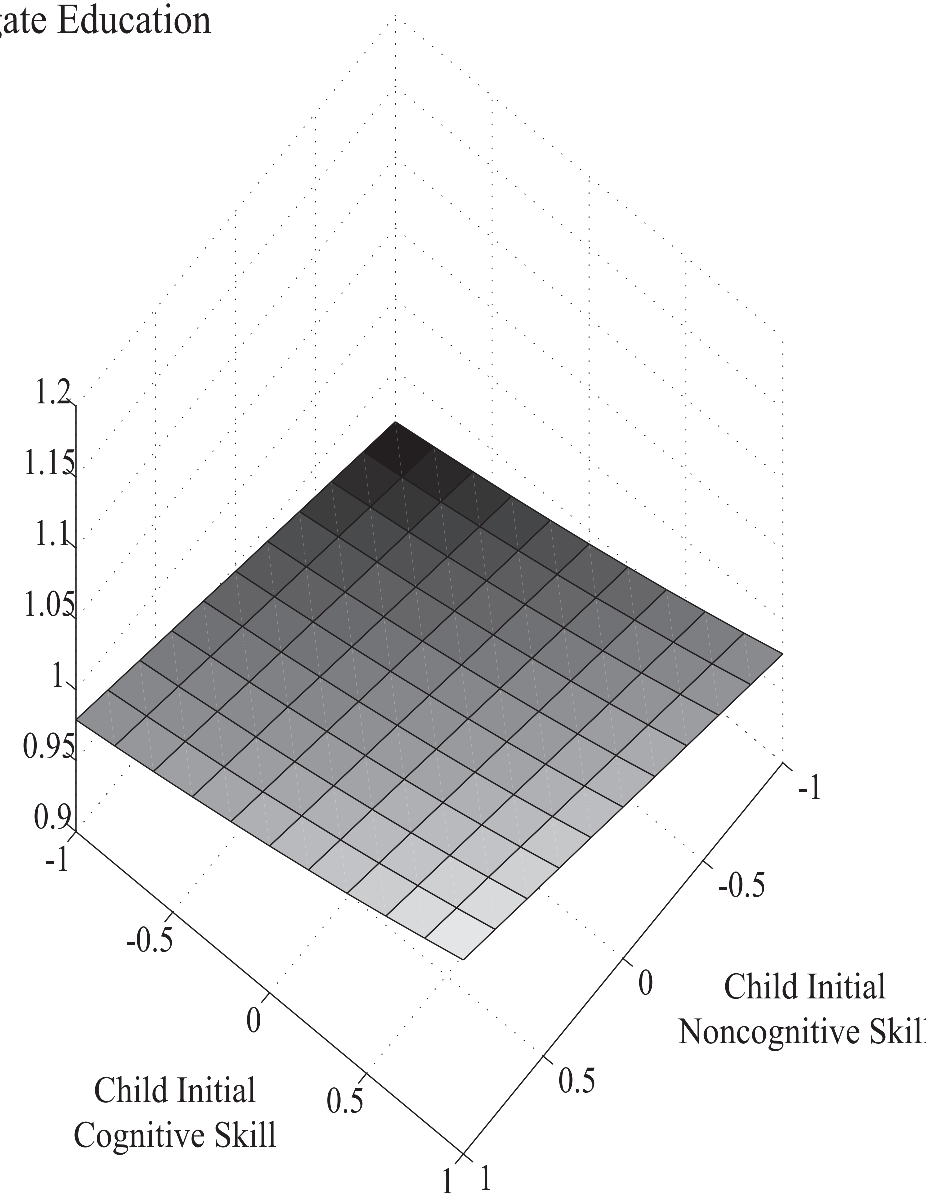


Figure 5
Optimal Early (Left) and Late (Right) Investments by
Maternal Cognitive and Noncognitive Skills
Maximizing Aggregate Education

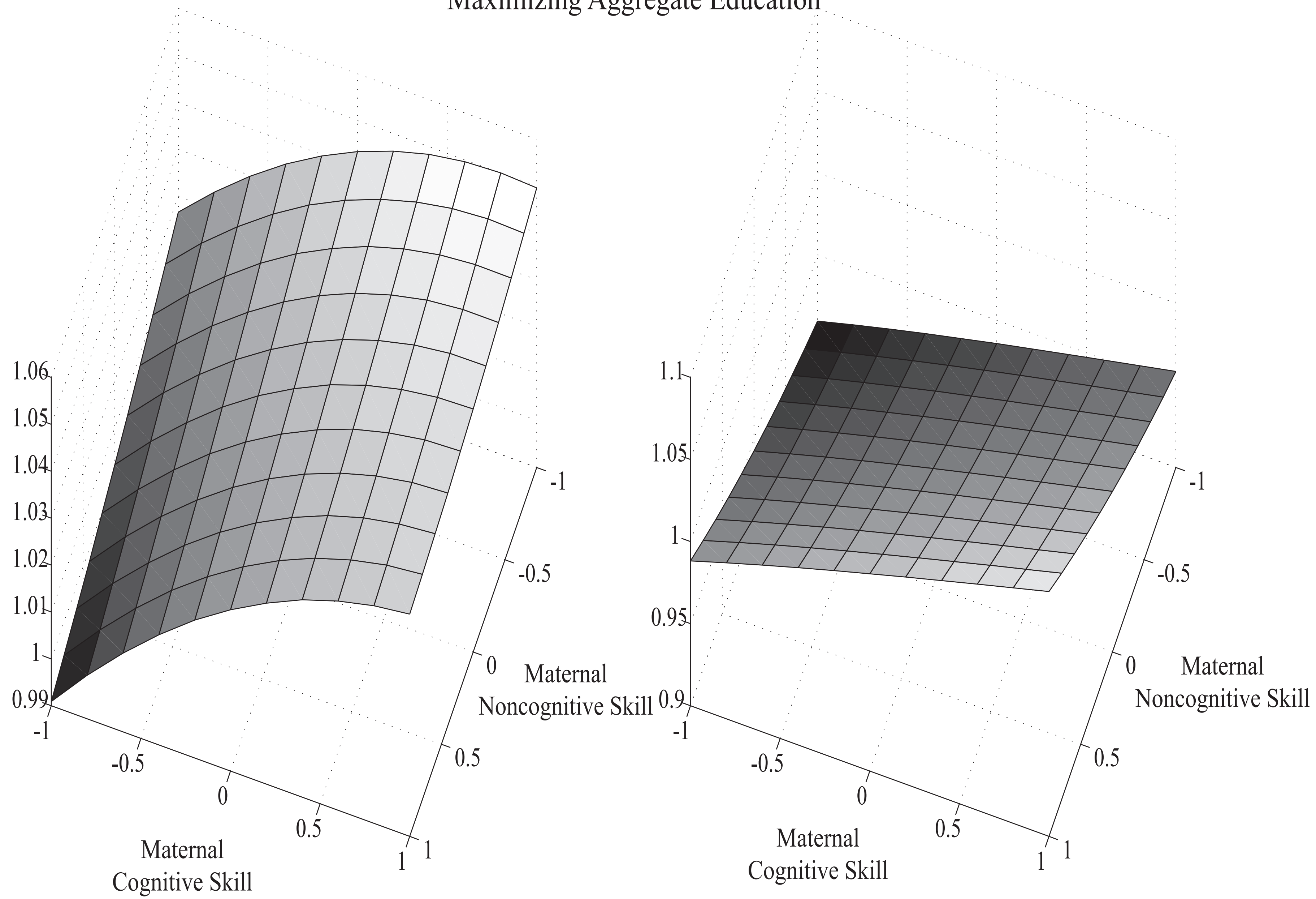


Figure 6
Ratio of Early to Late Investments by
Child Initial Conditions of Cognitive and Noncognitive Skills
Maximizing Aggregate Education (Left) and Minimizing Aggregate Crime (Right)

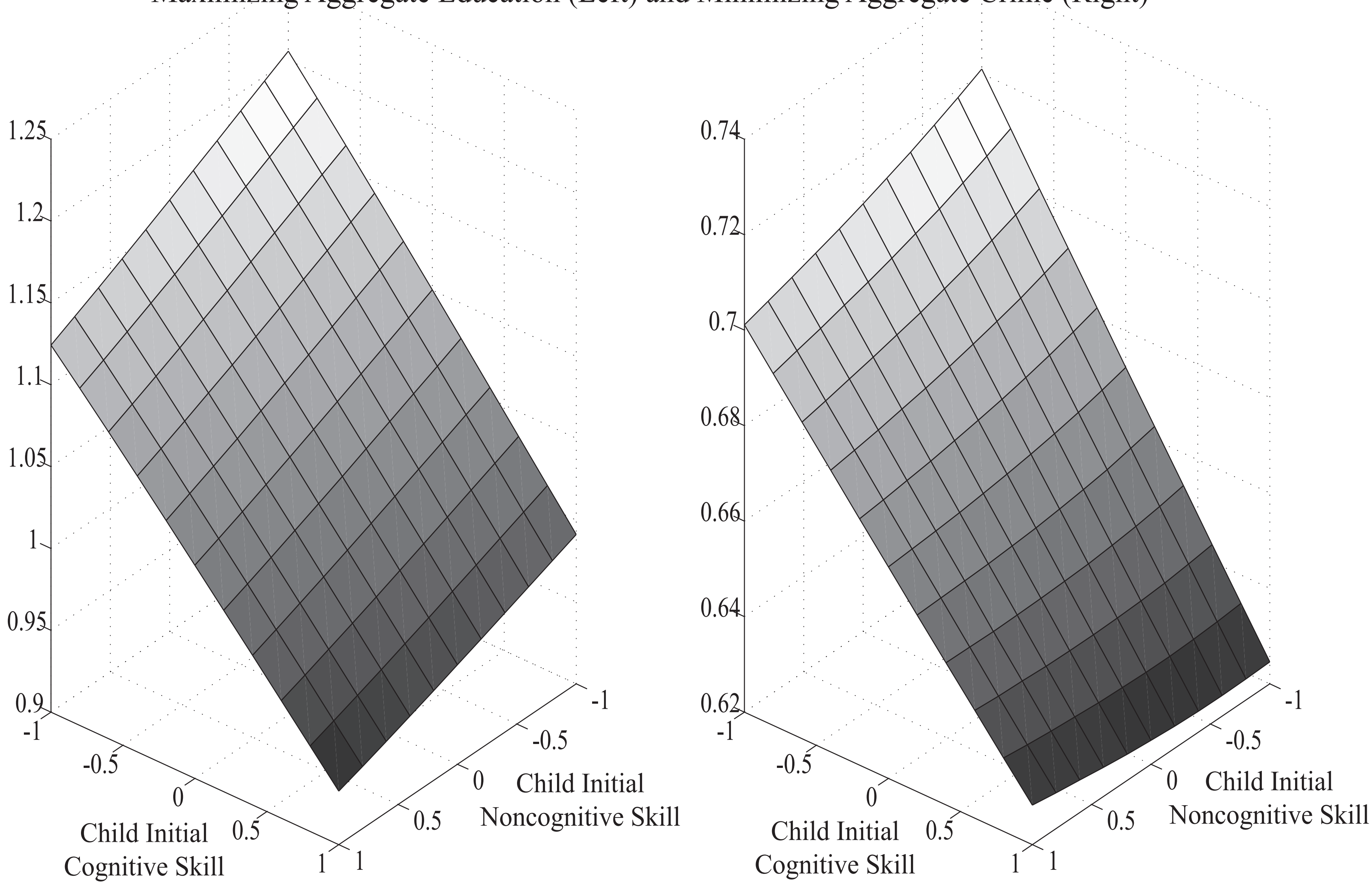


Figure 7
Ratio of Early to Late Investments by
Maternal Cognitive and Noncognitive Skills
Maximizing Aggregate Education (Left) and Minimizing Aggregate Crime (Right)

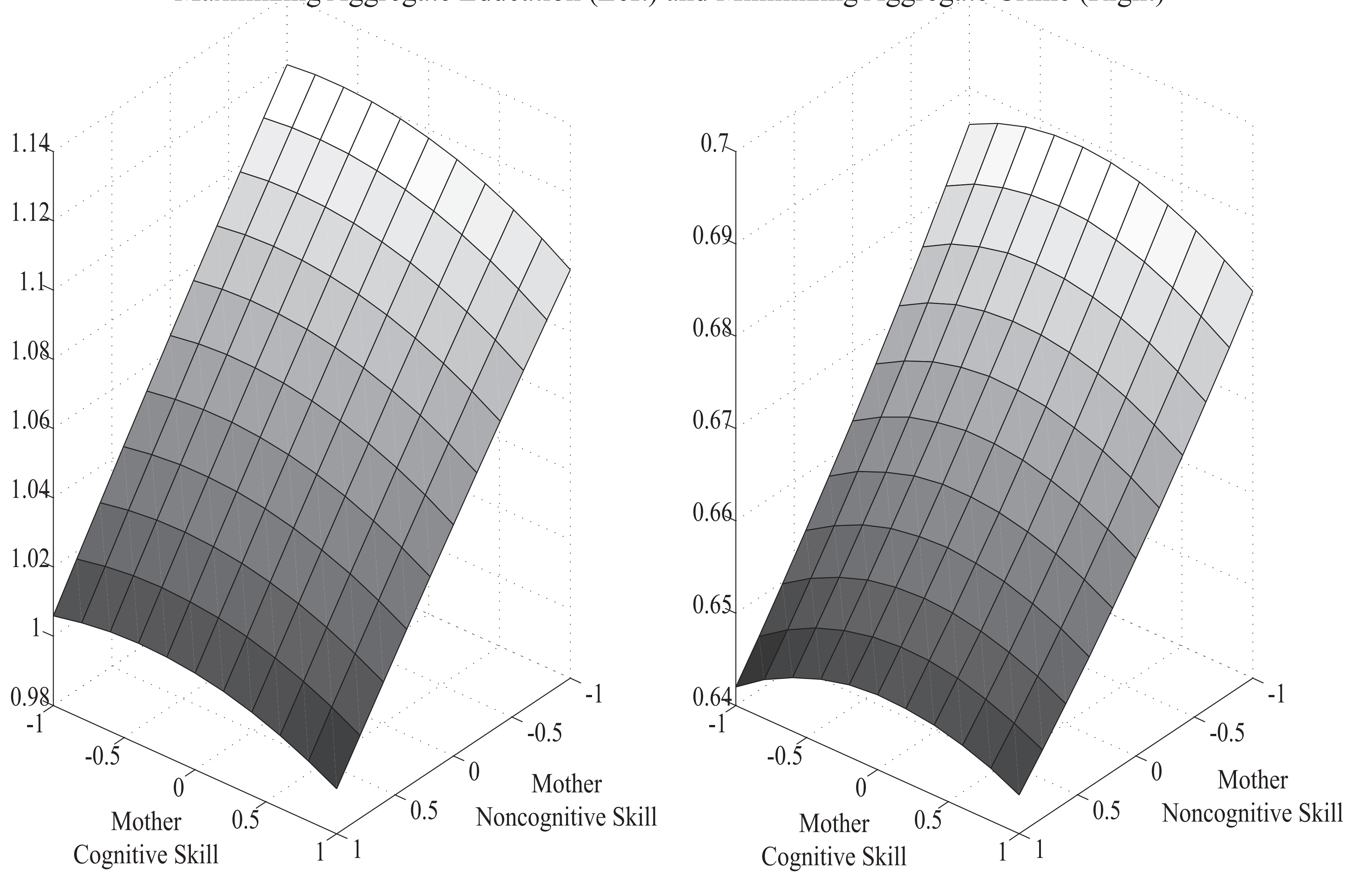


Figure 8
Densities of Ratio of Early to Late Investments
Maximizing Aggregate Education versus Minimizing Aggregate Crime

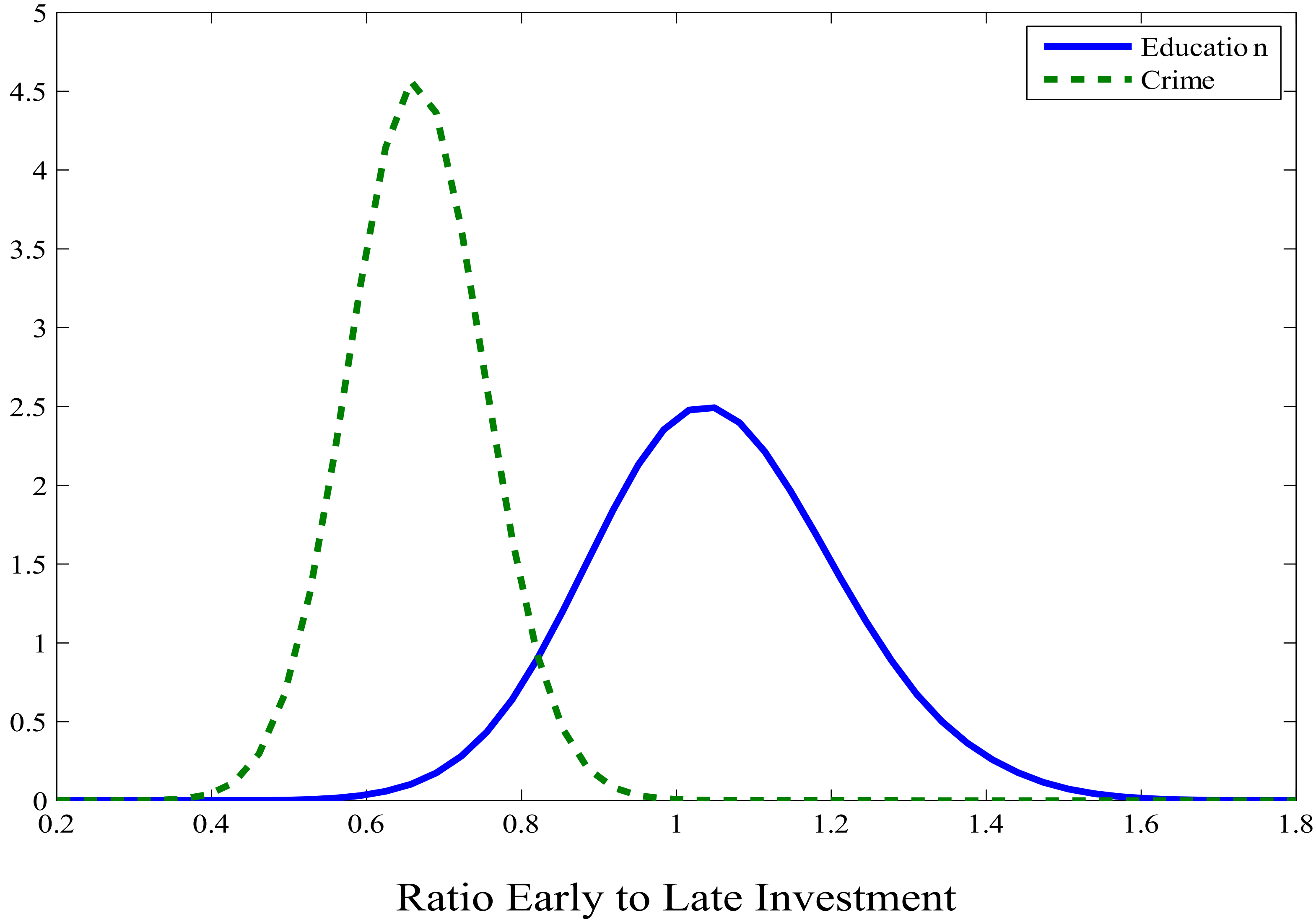


Figure 9

Optimal Early Investments by Child Initial Cognitive Skills and Maternal Cognitive Skills

Model with Cognitive Skill Only (Left) and the Model with Cognitive and Noncognitive Skills (Right)

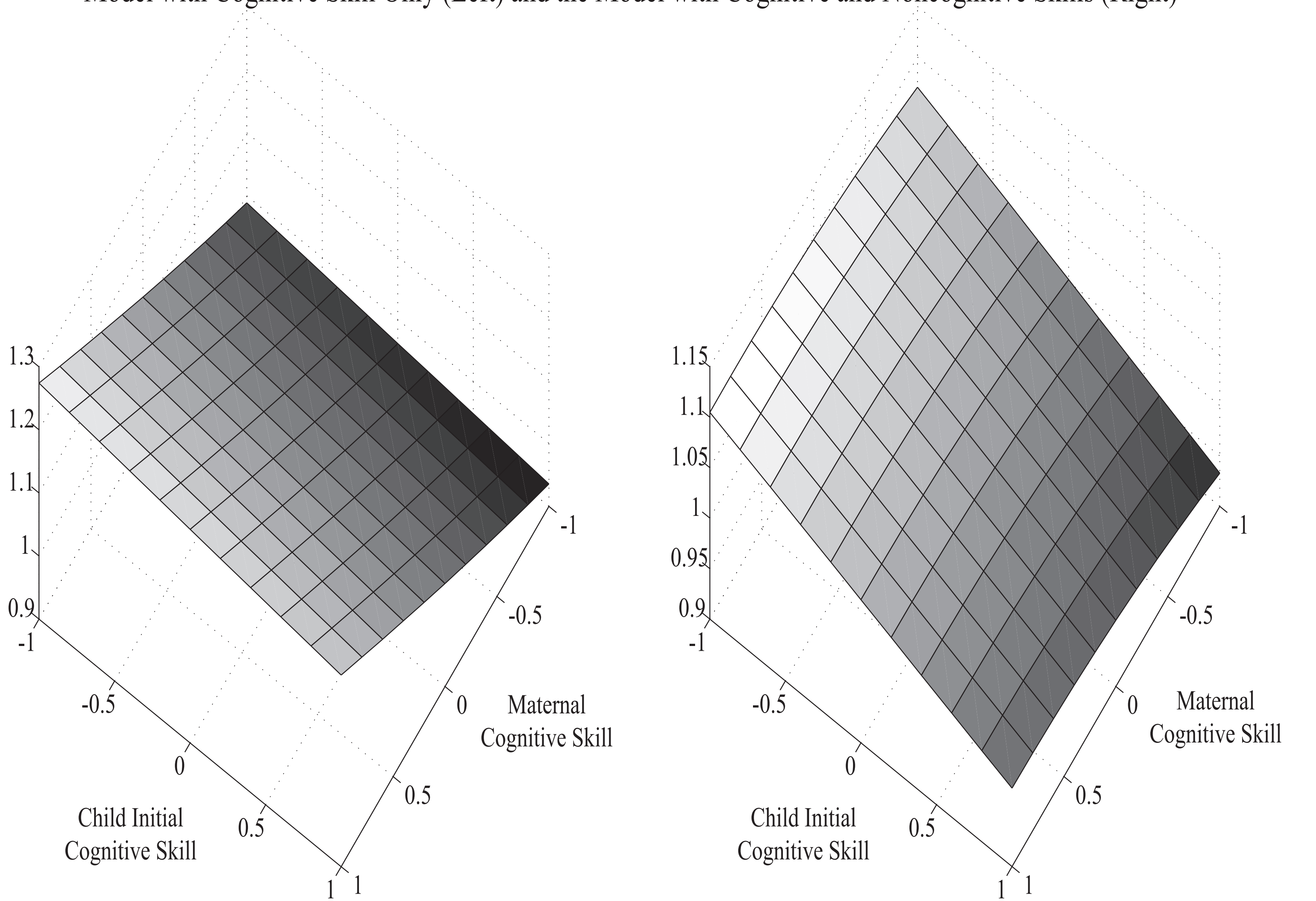


Figure 10

Optimal Late Investments by Child Initial Cognitive Skills and Maternal Cognitive Skills

Model with Cognitive Skill Only (Left) and the Model with Cognitive and Noncognitive Skills (Right)