Econometrica Supplementary Material

SUPPLEMENT TO "ESTIMATING THE TECHNOLOGY OF COGNITIVE AND NONCOGNITIVE SKILL FORMATION": APPENDIX

(*Econometrica*, Vol. 78, No. 3, May 2010, 883–931)

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

CONTENTS

© 2010 The Econometric Society DOI: 10.3982/ECTA6551

A1. DERIVING THE TECHNOLOGY OF SKILL FORMATION FROM ITS PRIMITIVES

A1.1. *A Model of Skill Formation*

IN THE MODELS PRESENTED in this section of the appendix and in Cunha, Heckman, Lochner, and Masterov (2006) and Cunha and Heckman (2007), parents make decisions about investments in their children. We ignore how the parents get to be who they are and the decisions of the children about their own children.

Suppose that there are two periods in a child's life, 1 and 2, before the child becomes an adult. Adulthood comprises distinct third and fourth periods in the overlapping generations model. The adult works for two periods after the two periods of childhood. Models based on the analysis of Becker and Tomes (1986) assume only one period of childhood. We assume that there are two kinds of skills denoted by C and N. The framework can be modified to accommodate health. (See Heckman (2007).) $\theta_{C,t}$ is cognitive skill at time t. $\theta_{N,t}$ is noncognitive skill at time t. Our treatment of ability is in contrast to the view of the traditional literature on human capital formation that views IQ as innate ability. In our analysis, IQ is just another skill. What differentiates IQ from other cognitive and noncognitive skills is that IQ is subject to accumulation during critical periods. That is, parental and social interventions can increase the IQ of the child, but they can do so successfully only for a limited time in the life of the child.

Let $I_{k,t}$ denote parental investments in child skill k at period t, $k \in \{C, N\}$ and $t \in \{1, 2\}$. Let Y_3 be the level of human capital as the child starts adulthood. It depends on both components of $(\theta_{C,3}, \theta_{N,3})$. The parents fully control investment in the child. A richer model incorporates, among other features, investment decisions of the child as influenced by the parent through preference formation processes (see Carneiro, Cunha, and Heckman (2003)).

We first describe how skills evolve over time. Assume that each agent is born with initial conditions $\theta_1 = (\theta_{C,1}, \theta_{N,1})$. These can be determined by parental environments. At each stage t let $\theta_t = (\theta_{C,t}, \theta_{N,t})$ denote the vector of skill or ability stocks. Thus we ignore parental environmental inputs which are included in the model in the text. The technology of production of skill k at period t is (suppressing the other inputs)

$$\theta_{k,t+1} = f_{k,t}(\theta_t, I_{k,t}) \tag{A1.1}$$

for $k \in \{C, N\}$ and $t \in \{1, 2\}$. We assume that $f_{k,t}$ is twice continuously differentiable, increasing, and concave. In this model, stocks of both skills and abilities produce next period skills and affect the productivity of investments. Cognitive skills can promote the formation of noncognitive skills and vice versa.

Let $(\theta_{C,3}, \theta_{N,3})$ denote the level of skills when adult. We define adult human capital Y_3 of the child as a combination of different adult skills:

$$Y_3 = g(\theta_{C,3}, \theta_{N,3}). \tag{A1.2}$$

The function g is assumed to be continuously differentiable and increasing in $(\theta_{C,3}, \theta_{N,3})$. This model assumes that there is no comparative advantage in the labor market or in other aspects of life.[1]

Equation (2.6) in the text is obtained by assuming $\phi_Y = \phi_{2,C} = \phi_{2,N} = \phi_{1,C} = \phi_{1,N}$, $I_{C,1} = I_{N,1} = I_1$, and $I_{C,2} = I_{N,2} = I_2$ (so investments are general in nature) to derive equation (2.6), and substituting recursively to obtain $\theta_{C,3}$ and $\theta_{N,3}$ using (2.3)–(2.5), $\theta_{C,1}$ and $\theta_{N,1}$, $\theta_{C,P}$ and $\theta_{N,P}$ are initial conditions. An alternative approach is as follows.

To fix ideas, consider the following specialization of our model. Ignore the effect of initial conditions and assume that first period skills are just due to first period investment:

$$\theta_{C,2} = f_{C,1}(\theta_1, I_{C,1}) = I_{C,1}$$

and

$$\theta_{N,2} = f_{N,1}(\theta_1, I_{N,1}) = I_{N,1},$$

where $I_{C,1}$ and $I_{N,1}$ are scalars. For the second period technologies, assume a CES structure

$$\begin{aligned} \theta_{C,3} &= f_{C,2}(\theta_2, I_{C,2}) \\ &= \{\gamma_{2,C,1}(\theta_{C,2})^{\phi_{2,C}} + \gamma_{2,C,2}(\theta_{N,2})^{\phi_{2,C}} + \gamma_{2,C,3}(I_{C,2})^{\phi_{2,C}}\}^{1/\phi_{2,C}}, \\ \text{where}\quad & \gamma_{2,C,1} \geq 0,\ \gamma_{2,C,2} \geq 0,\ \gamma_{2,C,3} \geq 0, \\ & \gamma_{2,C,1} + \gamma_{2,C,2} + \gamma_{2,C,3} = 1 \end{aligned} \tag{A1.3}$$

and

$$\begin{aligned} \theta_{N,3} &= f_{N,2}(\theta_2, I_{N,2}) \\ &= \{\gamma_{2,N,1}(\theta_{C,2})^{\phi_{2,N}} + \gamma_{2,N,2}(\theta_{N,2})^{\phi_{2,N}} + \gamma_{2,N,3}(I_{N,2})^{\phi_{2,N}}\}^{1/\phi_{2,N}}, \\ \text{where}\quad & \gamma_{2,N,1} \geq 0,\ \gamma_{2,N,2} \geq 0,\ \gamma_{2,N,3} \geq 0, \\ & \gamma_{2,N,1} + \gamma_{2,N,2} + \gamma_{2,N,3} = 1, \end{aligned} \tag{A1.4}$$

where $1/(1 - \phi_{2,C})$ is the elasticity of substitution in the inputs producing $\theta_{C,3}$ and $1/(1 - \phi_{2,N})$ is the elasticity of substitution of inputs in producing $\theta_{N,3}$ where $\phi_{2,C} \in (-\infty, 1]$ and $\phi_{2,N} \in (-\infty, 1]$. Notice that $I_{N,2}$ and $I_{C,2}$ are direct complements with $(\theta_{C,2}, \theta_{N,2})$ irrespective of the substitution parameters $\phi_{2,C}$ and $\phi_{2,N}$, except in limiting cases.

[1]Thus we rule out one potentially important avenue of compensation that agents can specialize in tasks that do not require the skills in which they are deficient. Cunha, Heckman, Lochner, and Masterov (2006) considered a more general task function that captures the notion that different tasks require different combinations of skills and abilities. See also the main text of the paper and Cunha and Heckman (2009).

The CES technology is well known and has convenient properties. It imposes direct complementarity even though inputs may be more or less substitutable depending on the share parameters or the elasticity of substitution. We distinguish between direct complementarity (positive cross-partials) and CES substitution/complementarity. Focusing on the technology for producing $\theta_{C,3}$, when $\phi_{2,C} = 1$, the inputs are perfect substitutes in the intuitive use of that term (the elasticity of substitution is infinite). The inputs $\theta_{C,2}$, $\theta_{N,2}$, and $I_{C,2}$ can be ordered by their relative productivity in producing $\theta_{C,3}$. The higher $\gamma_{2,C,1}$ and $\gamma_{2,C,2}$, the higher the productivity of $\theta_{C,2}$ and $\theta_{N,2}$, respectively. When $\phi_{2,C} = -\infty$, the elasticity of substitution is zero. All inputs are required in the same proportion to produce a given level of output so there are no possibilities for technical substitution, and

$$\theta_{C,3} = \min\{\theta_{C,2}, \theta_{N,2}, I_{C,2}\}.$$

In this technology, early investments are a *bottleneck* for later investments. Compensation for adverse early environments through late investments is impossible.

The evidence from numerous studies reviewed in Cunha, Heckman, Lochner, and Masterov (2006), Cunha and Heckman (2007), and Heckman (2007, 2008) shows that IQ is no longer malleable after ages 8–10. Taken at face value, this implies that if θ^C is IQ, for all values of $I_{C,2}$, $\theta_{C,3} = \theta_{C,2}$. Period 1 is a critical period for IQ but not necessarily for other skills and abilities. More generally, period 1 is a critical period if

$$\frac{\partial \theta_{C,t+1}}{\partial I_{C,t}} = 0 \quad \text{for} \quad t > 1.$$

For parameterization (A1.3), this is obtained by imposing $\gamma_{2,C,3} = 0$.

The evidence on adolescent interventions surveyed in Cunha, Heckman, Lochner, and Masterov (2006) shows substantial positive results for such interventions on noncognitive skills ($\theta_{N,3}$) and at most modest gains for cognitive skills. Technologies (A1.3) and (A1.4) rationalize this pattern. Since the populations targeted by adolescent intervention studies tend to come from families with poor backgrounds, we would expect $I_{C,1}$ and $I_{N,1}$ to be below average. Thus, $\theta_{C,2}$ and $\theta_{N,2}$ will be below average. Adolescent interventions make $I_{C,2}$ and $I_{N,2}$ relatively large for the treatment group in comparison to the control group in the intervention experiments. At stage 2, $\theta_{C,3}$ (cognitive ability) is essentially the same in the control and treatment groups, while $\theta_{N,3}$ (noncognitive ability) is higher for the treated group. Large values of $(\gamma_{2,C,1} + \gamma_{2,C,2})$ (associated with a small coefficient on $I_{C,2}$) and small values of $(\gamma_{2,N,1} + \gamma_{2,N,2})$ (so the coefficient on $I_{N,2}$ is large) produce this pattern. Another case that rationalizes the evidence is when $\phi_{2,C} \to -\infty$ and $\phi_{2,N} = 1$. Under these conditions,

$$\theta_{C,3} = \min\{\theta_{C,2}, \theta_{N,2}, I_{C,2}\}, \tag{A1.5}$$

while

$$\text{(A1.6)} \quad \theta_{N,3} = \gamma_{2,N,1}\theta_{N,2} + \gamma_{2,N,2}\theta_{N,2} + \gamma_{2,N,3}I_{N,2}.$$

The attainable period 2 stock of cognitive skill ($\theta_{C,3}$) is limited by the minimum value of $\theta_{C,2}$, $\theta_{N,2}$, and $I_{C,2}$. In this case, any level of investment in period 2 such that $I_{C,2} > \min\{\theta_{C,2}, \theta_{N,2}\}$ is ineffective in incrementing the stock of cognitive skills. Period 1 is a bottleneck period. Unless sufficient skill investments are made in θ_C in period 1, it is not possible to raise skill θ_C in period 2. This phenomenon does not appear in the production of the noncognitive skill, provided that $\gamma_{2,N,3} > 0$. More generally, the larger $\phi_{2,N}$ and the larger $\gamma_{2,N,3}$, the more productive is investment $I_{N,2}$ in producing $\theta_{N,3}$.

To complete the CES example, assume that adult human capital Y_3 is a CES function of the two skills accumulated at stage two:

$$\text{(A1.7)} \quad Y_3 = \{\rho(\theta_{C,3})^{\phi_Y} + (1-\rho)(\theta_{N,3})^{\phi_Y}\}^{1/\phi_Y},$$

where $\rho \in [0, 1]$, and $\phi_Y \in (-\infty, 1]$. In this parameterization, $1/(1-\phi_Y)$ is the elasticity of substitution across different skills in the production of adult human capital. Equation (A1.7) reminds us that the market, or life in general, requires use of multiple skills. Being smart is not the sole determinant of success. In the general case with multiple tasks, different tasks require both skills in different proportions. One way to remedy early skill deficits is to make compensatory investments. Another way is to motivate people from disadvantaged environments to pursue tasks that do not require the skill that deprived early environments do not produce. A richer theory would account for this choice of tasks and its implications for remediation.[2] For the sake of simplifying the argument, we work with equation (A1.7) that captures the notion that skills can trade off against each other in producing effective people. Highly motivated, but not very bright, people may be just as effective as bright but unmotivated people. That is one of the lessons from the GED program. (See Heckman and Rubinstein (2001) and Heckman, Stixrud, and Urzua (2006).)

The analysis is simplified by assuming that investments are general in nature: $I_{C,1} = I_{N,1} = I_1$, $I_{C,2} = I_{N,2} = I_2$.[3] Cunha and Heckman (2007) developed the more general case of skill-specific investments which requires more notational complexity.

With common investment goods, we can solve out for $\theta_{C,2}$ and $\theta_{N,2}$ in terms of I_1 to simplify (A1.3) and (A1.4) to reach

$$\text{(A1.8)} \quad \theta_{C,3} = \{(\gamma_{2,C,1} + \gamma_{2,C,2})(I_1)^{\phi_{2,C}} + \gamma_{2,C,3}(I_2)^{\phi_{2,C}}\}^{1/\phi_{2,C}}$$

[2]See the Appendix in Cunha, Heckman, Lochner, and Masterov (2006).

[3]Thus when a parent buys a book in the first period of childhood, this book may be an investment in all kinds of skills. It is an investment in cognitive skills, as it helps the child get exposure to language and new words. It can also be an investment in noncognitive skills if the book contains a message on the importance of being persistent and patient.

and

$$\theta_{N,3} = \{(\gamma_{2,N,1} + \gamma_{2,N,2})(I_1)^{\phi_{2,N}} + \gamma_{2,N,3}(I_2)^{\phi_{2,N}}\}^{1/\phi_{2,N}}. \tag{A1.9}$$

If we then substitute these expressions into the production function for adult human capital (A1.7), we obtain

$$\begin{aligned} Y_3 = \big\{&\rho[\tilde{\gamma}^C(I_1)^{\phi_{2,C}} + \gamma_{2,C,3}(I_2)^{\phi_{2,C}}]^{\phi_Y/\phi_{2,C}} \\ &+ (1-\rho)[\tilde{\gamma}^N(I_1)^{\phi_{2,N}} + \gamma_{2,N,3}(I_2)^{\phi_{2,N}}]^{\phi_Y/\phi_{2,N}}\big\}^{1/\phi_Y}, \end{aligned} \tag{A1.10}$$

where $\tilde{\gamma}_C = \gamma_{2,C,1} + \gamma_{2,C,2}$ and $\tilde{\gamma}_N = \gamma_{2,N,1} + \gamma_{2,N,2}$. Equation (A1.10) expresses adult human capital as a function of the entire sequence of childhood investments in human capital. Current investments in human capital are combined with the existing stocks of skills to produce the stock of next period skills.

A convenient formulation of the problem arises if it is assumed that $\phi_{2,C} = \phi_{2,N} = \phi_Y = \phi$ so that CES substitution among inputs in producing outputs and CES substitution among skills in producing human capital are the same. This produces the familiar-looking CES expression for adult human capital stocks:

$$Y_3 = \{\tau_1 I_1^{\phi} + \tau_2 I_2^{\phi}\}^{1/\phi}, \tag{A1.11}$$

where $\tau_1 = \rho\tilde{\gamma}_C + (1-\rho)\tilde{\gamma}_N$, $\tau_2 = \rho\gamma_{2,C,3} + (1-\rho)\gamma_{2,N,3}$, and $\phi = \phi_{2,N} = \phi_{2,C} = \phi_Y$. The parameter τ_1 is *a skill multiplier*. It arises in part because I_1 affects the accumulation of $\theta_{C,2}$ and $\theta_{N,2}$. These stocks of skills in turn affect the productivity of I_2 in forming $\theta_{C,3}$ and $\theta_{N,3}$. Thus τ_1 captures the net effect of I_1 on Y_3 through both self-productivity and direct complementarity.[4] $\frac{1}{1-\phi}$ is a measure of how easy it is to substitute between I_1 and I_2, where the substitution arises from both the task performance (human capital) function in equation (A1.7) and the technology of skill formation. Within the CES technology, ϕ is a measure of the ease of substitution of inputs. In this analytically convenient case, the parameter ϕ plays a dual role. First, it informs us how easily one can substitute across different skills to produce one unit of adult human capital Y_3.

[4]Direct complementarity between I_1 and I_2 arises if

$$\frac{\partial^2 Y_3}{\partial I_1\,\partial I_2} > 0.$$

Since $\phi < 1$, I_1 and I_2 are direct complements, because

$$\operatorname{sign}\left(\frac{\partial^2 Y_3}{\partial I_1\,\partial I_2}\right) = \operatorname{sign}(1-\phi).$$

This definition of complementarity is to be distinguished from the notion based on the elasticity of substitution between I_1 and I_2, which is $\frac{1}{1-\phi}$. When $\phi < 0$, I_1 and I_2 are sometimes called complements. When $\phi > 0$, I_1 and I_2 are sometimes called substitutes. I_1 and I_2 are always direct complements, but if $1 > \phi > 0$, they are CES substitutes.

Second, it also represents the degree of complementarity (or substitutability) between early and late investments in producing skills. In this second role, the parameter ϕ dictates how easy it is to compensate for low levels of stage 1 skills in producing late skills.

In principle, compensation can come through two channels: (i) through skill investment or (ii) through choice of market activities, substituting deficits in one skill by the relative abundance in the other through choice of tasks. We do not develop the second channel of compensation in this appendix, deferring it to later work. It is discussed in Carneiro, Cunha, and Heckman (2003).

When ϕ is small, low levels of early investment I_1 are not easily remediated by later investment I_2 in producing human capital. The other face of CES complementarity is that when ϕ is small, high early investments should be followed with high late investments. In the extreme case when $\phi \to -\infty$, (A1.11) converges to $Y_3 = \min\{I_1, I_2\}$. We analyzed this case in Cunha, Heckman, Lochner, and Masterov (2006). The Leontief case contrasts sharply with the case of perfect CES substitutes, which arises when $\phi = 1$: $Y_3 = \tau_1 I_1 + \tau I_2$. If we impose the further restriction that $\tau_1 = \frac{1}{2}$, we generate the model that is implicitly assumed in the existing literature on human capital investments that collapses childhood into a single period. In this special case, only the total amount of human capital investments, regardless of how it is distributed across childhood periods, determines adult human capital. In the case of perfect CES substitutes, it is possible in a physical productivity sense to compensate for early investment deficits by later investments, although it may not be economically efficient to do so.

We can rewrite (A1.11) as

$$Y_3 = I_1\{\tau_1 + \tau_2 \omega^{\phi}\}^{1/\phi},$$

where $\omega = I_2/I_1$. Fixing I_1 (early investment), an increase in ω is the same as an increase in I_2. The marginal productivity of late investment is

$$\frac{\partial Y_3}{\partial \omega} = \tau_2 I_1\{\tau_1 + \tau_2 \omega^{\phi}\}^{(1-\phi)/\phi}\omega^{\phi-1}.$$

For $\omega > 1$ and $\tau_1 < 1$, marginal productivity is increasing in ϕ and τ_2. Thus, provided that late investments are greater than earlier investments, the more substitutable I_2 is with I_1 (the higher ϕ) and the lower is the skill multiplier τ_1, the more productive are late investments.

A1.2. *How the Flourishing of Noncognitive Traits With the Stage of the Life Cycle Can Raise the Elasticity of Substitution for Noncognitive Skills at Later Stages Compared to Earlier Stages*

Suppose that there is one cognitive skill $\theta_{C,t}$ at each stage of the life cycle, but it can change over time.[5] Initially, there is one noncognitive skill $\theta_{N_1,1}$, which

[5] In truth, there are at least two. See Borghans, Duckworth, Heckman, and ter Weel (2008).

can change over time, but a second noncognitive skill trait emerges in period 2: $\theta_{N_2,2}$. The technology changes with the stage of the life cycle and has an expanding set of arguments:

$$\theta_{C,2} = f_{C,1}\big(\theta_{C,1}, \theta_{N_1,1}, I_{C,1}\big),$$
$$\theta_{N_1,2} = f_{N_1,1}\big(\theta_{C,1}, \theta_{N_1,1}, I_{N_1,1}\big).$$

In the second period, a new trait emerges ($\theta_{N_2,2}$):

$$\theta_{C,3} = f_{C,2}\big(\theta_{C,2}, \theta_{N_1,2}, \theta_{N_2,2}, I_{C,2}\big),$$
$$\theta_{N_1,3} = f_{N_1,2}\big(\theta_{C,2}, \theta_{N_1,2}, \theta_{N_2,2}, I_{N_1,2}\big),$$
$$\theta_{N_2,3} = f_{N_2,3}\big(\theta_{C,2}, \theta_{N_1,2}, \theta_{N_2,2}, I_{N_2,2}\big).$$

The emergence of the new trait and new investment options associated with the trait increases the possibility of substitution against the initial trait ($\theta_{N_1,2}$) in period 2 and rationalizes the increase in substitutability of noncognitive skills over the life cycle. This analysis suggests that it might be fruitful to disaggregate noncognitive skills by stage of the life cycle and model evolving technologies, but we leave that task to another occasion.

A2. INVESTMENTS AS A FUNCTION OF INPUTS

Suppose that parents cannot borrow or lend, so at every period t, they use their income y_t to consume or invest in their children. The objective of parents is to maximize lifetime utility, which depends on consumption streams $(c_t)_{t=1}^T$ and the adult skill levels $\theta_{C,T+1}$ and $\theta_{N,T+1}$. In this case, parental investments in skill k and period t, $I_{k,t}$, should depend on parental skills, θ_P, child's skills at the beginning of period t, θ_t, parental income, y_t, and the innovation, π_t.

Our approach exploits the fact that investments in skills are themselves combinations of many intermediate inputs. Suppose that there are $M_{2,k,t}$ inputs in the production of investments in skill k at period t. Following the notation established in Section 3.1 in the text, let $Z_{2,k,t,j}$ denote the intermediate inputs. Let $h_{k,t}$ denote the production function of parental investments in skill k at period t:

$$I_{k,t} = h_{k,t}\big(Z_{2,k,t,1}, Z_{2,k,t,2}, \dots, Z_{2,k,t,M_{2,k,t}}\big),$$

where $h_{k,t}$ is increasing and concave in $(Z_{2,k,t,1}, Z_{2,k,t,2}, \dots, Z_{2,k,t,M_{2,t,k}})$. From the point of view of the parent, she can solve this problem in two stages. In the first stage, she decides how much to consume and how much to invest. Let $I^*_{k,t} = g_{k,t}(\theta_t, \theta_P, \pi_t, y_t)$ denote the investment level chosen by the parent. Then, in the second stage, she decides how much to buy in inputs $(Z_{2,k,t,1}, Z_{2,k,t,2}, \dots, Z_{2,k,t,M_{2,k,t}})$ to produce exactly $I_{k,t}$ units of investment. Let

$p_{k,t,j}$ denote the price of intermediate input $Z_{2,k,t,j}$. In this case, the second stage problem of the parent can be written as

$$\min \sum_{j=1}^{M_{k,t,j}} p_{k,t,j} Z_{2,k,t,j}$$

subject to $h_{k,t}(Z_{2,k,t,1}, Z_{2,k,t,2}, \dots, Z_{2,k,t,M_{2,k,t}}) = I_{k,t}$. The solution to this problem is the set of conditional demand functions

$$Z_{2,k,t,j} = a_{t,k,j}(I_{k,t}, p_{k,t}),$$

where $p_{k,t} = (p_{k,t,1}, \dots, p_{k,t,M_{k,t,j}})$. For example, if $p_{k,t} = 1$, $h_{k,t}(Z_{2,k,t}) = \prod_{j=1}^{M_{2,k,t}} Z_{2,k,t,j}^{\tilde{\alpha}_{2,k,t,j}}$ with $\tilde{\alpha}_{2,k,t,j} \geq 0$, and $\sum_{j=1}^{M_{2,t,k}} \tilde{\alpha}_{2,k,t,j} \leq 1$, then

$$Z_{2,k,t,j} = \alpha_{2,k,t,j} I_{k,t} \quad \text{for all } j, \tag{A2.1}$$

where $\alpha_{2,k,t,j} = \tilde{\alpha}_{2,k,t,j} / \sum_{l=1}^{M_{2,k,t}} \tilde{\alpha}_{2,k,t,l}$. If the intermediate inputs are measured with error, so that the observed input is $Z_{2,k,t,j} = Z^*_{2,k,t,j} + \varepsilon_{2,t,k,j}$, then

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

which is exactly our measurement equation in Section 3.1. This reformulation of the problem shows that the measurement equations for investments can be derived as conditional demand functions for intermediate inputs. Furthermore, when the investment production function $h_{k,t}$ is Cobb–Douglas, the factor loadings in the measurement equations are the share parameters of the intermediary inputs. Cunha and Heckman (2008) develop other cases.

A3. PROOFS OF THEOREMS

A3.1. *Proof of Theorem 1*

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

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

Substituting this expression into the expression for $p_\theta(\theta)$ in Theorem 1, we obtain

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

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

A3.2. *Proof of Theorem 2*

We can use Theorem 1 in Hu and Schennach (2008) (hereafter HS) to prove that the distribution of θ is identified, after setting $x=Z_1$, $y=Z_3$, $z=Z_2$, and $x^*=\theta$ in the notation of that paper. We can show that Assumption 2 in the present paper is equivalent to Assumption 2 in HS. If Z_1, Z_2, and Z_3 are mutually independent conditional on θ, then we have, in particular, the two equalities

$$\begin{aligned}
&p_{Z_3|Z_1,Z_2,\theta}(Z_3\mid Z_1,Z_2,\theta)=p_{Z_3|\theta}(Z_3\mid\theta),\\
&p_{Z_1|Z_2,\theta}(Z_1\mid Z_2,\theta)=p_{Z_1|\theta}(Z_1\mid\theta),
\end{aligned}$$

which constitute the HS Assumption 2. Conversely, by the definition of conditional densities and HS Assumption 2,

$$\begin{aligned}
&p_{Z_3,Z_1,Z_2|\theta}(Z_3,Z_1,Z_2\mid\theta)\\
&\quad=p_{Z_3|Z_1,Z_2,\theta}(Z_3\mid Z_1,Z_2,\theta)p_{Z_1|Z_2,\theta}(Z_1\mid Z_2,\theta)p_{Z_2|\theta}(Z_2\mid\theta)\\
&\quad=p_{Z_3|\theta}(Z_3\mid\theta)p_{Z_1|\theta}(Z_1\mid\theta)p_{Z_2|\theta}(Z_2\mid\theta),
\end{aligned}$$

which states that Z_1, Z_2, and Z_3 are mutually independent conditional on θ. Hence the two assumptions are equivalent. *Q.E.D.*

A3.3. *A More Explicit Discussion of the Analysis of Section 3.6.2*

We assume that suitable repeated measurements of $(\theta_P,\{\theta_t,I_{k,t},y_t\}_{t=1}^T)$, $k\in\{C,N\}$, are available to identify their (joint) distribution. In our application, we assume that y_t is perfectly measured (i.e., its "repeated" measurements can be set equal to each other). We also make the following assumptions:

ASSUMPTION 1: *The random variables* π_t, $\nu_{k,t}$, $I_{k,t}$, $\theta_{k,t}$, *and* y_t, $k \in \{C, N\}$, *are scalar while* θ_P *may be vector-valued*.[6]

ASSUMPTION 2: (i) π_t *is independent from* $(y_{t-2}, y_{t-1}, y_t, \theta_P)$ *and* (ii) $\nu_{k,t}$ *is independent from* $(\theta_t, I_{k,t}, \theta_P, y_t)$.

REMARK: Given our dynamic model structure, component (ii) of Assumption 2 essentially implies that $\nu_{k,t}$ must be independent over time and independent from π_t.[7]

ASSUMPTION 3: $I_{k,t}$ *is continuously distributed conditional on* $(\theta_{k,t}, \theta_P, y_t)$ *with a density* $f(I_{k,t} \mid \theta_t, \theta_P, y_t)$ *that is uniformly Lipschitz in* $I_{k,t}$ *for all* $(\theta_t, \theta_P, y_t)$ *in their joint support*.

ASSUMPTION 4: *For any* $\bar{\pi}$ *in the support of* π_t, $E[D(\theta_t, \theta_P, y_t, \bar{\pi})\Delta(\theta_t, \theta_P, y_t) \mid y_{t-2}, y_{t-1}, \theta_P, y_t] = 0 \Rightarrow \Delta(\theta_t, \theta_P, y_t) = 0$, *where* $D(\theta_t, \theta_P, y_t, \bar{\pi}) = f(q_{k,t}(\theta_t, \theta_P, y_t, \bar{\pi}) \mid \theta_t, \theta_P, y_t)$. (*This is a completeness condition in Chernozhukov, Imbens, and Newey* (2007).)

ASSUMPTION 5: (i) $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ *is strictly increasing in* π_t *and* (ii) $f_{k,s}(\theta_t, I_{k,t}, \theta_P, \pi_t, \nu_{k,t})$ *is strictly increasing in* $\nu_{k,t}$.

ASSUMPTION 6: *The distributions of* π_t *and* $\nu_{k,t}$ *are normalized to be uniform*.[8]

Consider our investment policy equation:

$$I_{k,t} = q_{k,t}(\theta_t, \theta_P, y_t, \pi_t).$$

By Assumption 2, y_{t-2} and y_{t-1} are valid instruments for $\theta_t \equiv (\theta_{C,t}, \theta_{N,t})$ because π_t is independent from y_{t-2} and y_{t-1} and because y_{t-1} has an effect on θ_t through

$$\begin{aligned} \theta_{k,t} &= f_{k,s}(\theta_{t-1}, I_{k,t-1}, \theta_P, \nu_{k,t-1}, \pi_{t-1}) \\ &= f_{k,s}(\theta_{t-1}, q_{k,t-1}(\theta_{t-1}, \theta_P, y_{t-1}, \pi_{t-1}), \theta_P, \pi_{t-1}, \nu_{k,t-1}) \end{aligned}$$

and similarly for y_{t-2}, by induction. The variables θ_P, y_t can instrument for themselves since π_t is independent from θ_P, y_t as well.

[6]The vector-valued $\theta_{k,t}$ and y_t are trivial to allow for, as long as the total dimension of all the lags of y_t used as instruments for $\theta_{k,t}$ is the same as the dimension of $\theta_{k,t}$.

[7]Except for very contrived cases.

[8]One may select any of the alternative normalizations discussed in Matzkin (2003, 2007).

Under Assumptions 3 and 4, Theorem 3.2 in Chernozhukov, Imbens, and Newey (2007) implies that the function $q_{k,t}(\theta_t, \theta_P, y_t, \tau)$ solves the equation

$$P[I_{k,t} \leq q_{k,t}(\theta_t, \theta_P, y_t, \tau) \mid (y_{t-2}, y_{t-1}, \theta_P, y_t)] = \tau \tag{A3.1}$$

for any $\tau \in [0, 1]$ is unique (i.e., identified).

Next, we note that equation (A3.1) is also satisfied after setting $I_{k,t} = q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$, where π_t is uniformly distributed on $[0, 1]$. This follows from

$$\begin{aligned} &P[q_{k,t}(\theta_t, \theta_P, y_t, \pi_t) \leq q_{k,t}(\theta_t, \theta_P, y_t, \tau) \mid (y_{t-2}, y_{t-1}, \theta_P, y_t)] \\ &\quad = P[\pi_t \leq \tau \mid (y_{t-2}, y_{t-1}, \theta_P, y_t)] \end{aligned}$$

since $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ is monotone in π_t by Assumption 5. Then

$$P[\pi_t \leq \tau \mid (y_{t-2}, y_{t-1}, \theta_P, y_t)] = P[\pi_t \leq \tau] = \tau$$

by Assumptions 2 and 6, as desired. It follows that $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ is identified.

Once the function $q_{k,t}$ has been identified, one can obtain $q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$, the inverse of $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ with respect to its last argument, thanks to Assumption 5. After substituting $\pi_t = q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$, we can rewrite the technology function as

$$\begin{aligned} \theta_{k,t+1} &= f_{k,s}(\theta_t, I_{k,t}, \theta_P, q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t}), \nu_{k,t}) \\ &\equiv f_{k,t}^{\mathrm{rf}}(\theta_t, I_{k,t}, \theta_P, y_t, \nu_{k,t}). \end{aligned}$$

By Assumption 2, $\nu_{k,t}$ is independent from $(\theta_t, I_{k,t}, \theta_P, y_t)$ and by Assumption 5, $f_{k,t}^{\mathrm{rf}}(\theta_t, I_{k,t}, \theta_P, y_t, \nu_{k,t})$ is monotone in $\nu_{k,t}$, which has a uniform distribution by Assumption 6. By standard arguments (see Matzkin (2003, 2007)), it follows that element k $(= C, N)$ of the vector-valued function $f^{\mathrm{rf}}(\theta_t, I_{k,t}, \theta_P, y_t, \tau)$ is identified from the τ-quantile of $\theta_{k,t+1}$ conditional on $\theta_t, I_{k,t}, \theta_P, y_t$.

To identify the technology $f_{k,s}$, we need to disentangle the direct effect of θ_t, I_t, θ_P on θ_{t+1} from their indirect effect through $\pi_t = q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$. To accomplish this, we exploit our knowledge of $q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$ to identify the technology through the equality

$$f_{k,s}(\theta_t, I_{k,t}, \theta_P, \pi_t, \nu_{k,t}) = f_{k,t}^{\mathrm{rf}}(\theta_t, I_{k,t}, \theta_P, y_t, \nu_{k,t})\big|_{y_t : q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t}) = \pi_t},$$

where, on the right-hand side, we set y_t to a specific value such that the corresponding implied value of π_t matches its value on the left-hand side. This does not necessarily require $q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$ to be invertible with respect to y_t, since we merely need at least one suitable value of y_t for each given $(\theta_t, \theta_P, I_{k,t}, \pi_t)$ and not necessarily a one-to-one mapping. Such a value y_t necessarily exists for any given $\theta_t, I_{k,t}, \theta_P, \pi_t$ because, for a fixed $\theta_t, I_{k,t}, \theta_P$, the variations in π_t are entirely due to y_t, since $\pi_t = q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$. Hence, by construction, any value of π_t in its support conditional on $\theta_t, I_{k,t}, \theta_P$ is reachable via a suitable choice of y_t.

A4. IDENTIFICATION OF FACTOR LOADINGS FOR THE CASE WITH UNOBSERVED HETEROGENEITY

Suppose that the relationship between adult outcomes $Z_{4,j}$ with skills $(\theta_{C,T+1}, \theta_{N,T+1})$ and unobserved heterogeneity π can be written as

$$\begin{aligned} Z_{4,j} &= \mu_{4,j} + \alpha_{4,C,j}\theta_{C,T+1} + \alpha_{4,N,j}\theta_{N,T+1} + \alpha_{4,\pi,j}\pi + \varepsilon_{4,j} \\ &\quad \text{for} \quad j = 1, 2, \ldots, J. \end{aligned}$$

Normalize $\alpha_{4,\pi,1} = 1$. If $J \geq 3$, we can identify the entire model as long as one of the following two conditions holds:

(i) The loading $\alpha_{4,\pi,j} = 1$ for $j \in \{2, 3\}$.

(ii) For some period t, $\mathrm{Cov}(\theta_{k,t}, \pi) = 0$ for $k \in \{C, N\}$.

To see why we need one of these two conditions, suppose that we have $J = 3$ and we normalize $\alpha_{4,\pi,1} = 1$. Consider the covariances

$$\begin{aligned} \mathrm{Cov}(Z_{4,j}, Z_{1,C,t,1}) &= \alpha_{4,C,j}\,\mathrm{Cov}(\theta_{C,t}, \theta_{C,T+1}) + \alpha_{4,N,j}\,\mathrm{Cov}(\theta_{C,t}, \theta_{N,T+1}) \\ &\quad + \alpha_{4,\pi,j}\,\mathrm{Cov}(\theta_{C,t}, \pi), \\ \mathrm{Cov}(Z_{4,j}, Z_{1,N,t,1}) &= \alpha_{4,C,j}\,\mathrm{Cov}(\theta_{N,t}, \theta_{C,T+1}) + \alpha_{4,N,j}\,\mathrm{Cov}(\theta_{N,t}, \theta_{N,T+1}) \\ &\quad + \alpha_{4,\pi,j}\,\mathrm{Cov}(\theta_{N,t}, \pi), \\ \mathrm{Cov}(Z_{4,j}, Z_{2,C,t,1}) &= \alpha_{4,C,j}\,\mathrm{Cov}(I_t, \theta_{C,T+1}) + \alpha_{4,N,j}\,\mathrm{Cov}(I_t, \theta_{N,T+1}) \\ &\quad + \alpha_{4,\pi,j}\,\mathrm{Cov}(I_t, \pi). \end{aligned}$$

We can identify $\mathrm{Cov}(\theta_{C,t}, \theta_{C,T+1})$ from the $\mathrm{Cov}(Z_{1,C,t,1}, Z_{1,C,T+1,1})$. The same line of argument applies to $\mathrm{Cov}(\theta_{C,t}, \theta_{N,T+1})$, $\mathrm{Cov}(\theta_{N,t}, \theta_{C,T+1})$, $\mathrm{Cov}(\theta_{N,t}, \theta_{N,T+1})$, $\mathrm{Cov}(I_t, \theta_{C,T+1})$, and $\mathrm{Cov}(I_t, \theta_{N,T+1})$. The factor loadings $\{\alpha_{4,C,j}, \alpha_{4,N,j}\}_{j=1}^{3}$ and $\{\alpha_{4,\pi,j}\}_{j=2}^{3}$ as well as the covariances $\mathrm{Cov}(\theta_{C,t}, \pi)$, $\mathrm{Cov}(\theta_{N,t}, \pi)$, and $\mathrm{Cov}(I_t, \pi)$ need to be identified. We then have 11 unknowns and 9 equations. Adding terms such as $\mathrm{Cov}(Z_{4,j}, Z_{1,C,t+1,1})$ is not helpful because they would add other unknown quantities such as $\mathrm{Cov}(\theta_{C,t+1}, \pi)$. We need two extra normalizations. This can be accomplished, for example, by imposing the restriction $\alpha_{4,\pi,j} = 1$ for $j \in \{2, 3\}$ or $\mathrm{Cov}(\theta_{k,0}, \pi) = 0$ for $k \in \{C, N\}$, provided that rank conditions hold.

A5. THE LIKELIHOOD IN THE NOTATION OF THE TEXT

We now develop the likelihood function for our model using the formal notation used in the text. We present the specification for the model that generates the estimates in Table V, used in the simulations. The likelihood for the other models can be constructed along similar lines. Let $p_\theta(\theta)$ denote the density of θ. For simplicity and notational consistency with the text, we use θ. In constructing the estimates we use $\ln\theta$ (i.e., the log of each component of θ) in the measurement and anchoring equations and use θ in the technology. 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 $h \in \{1, \dots, H\}$ 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 $\theta_{k,P}$. Let $z_{4,T+1,j,h}$ represent 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 $z_{l,k,t,j,h}$, $l = 1, 2$, let $\varepsilon_{3,k,1,j,h}$ denote the measurement error associated with $z_{3,k,1,j,h}$, and let $\varepsilon_{4,T+1,j,h}$ denote 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:

$$
\begin{aligned}
\text{(A5.1)} \quad p(z) = \prod_{h=1}^{H} \int \cdots \int p_\theta(\theta) \\
&\times \prod_{k\in\{C,N\}} \Bigg\{ \prod_{t=1}^{T} \Bigg[\prod_{j=1}^{M_{1,k,t}} p_{\varepsilon_{1,k,t,j,h}}(z_{1,k,t,h} - \mu_{1,k,t,j} - \alpha_{1,k,t,j}\theta_{k,t}) \\
&\times \prod_{j=1}^{M_{2,k,t}} p_{\varepsilon_{2,k,t,j,h}}\big(z_{2,k,t,j,h} - \mu_{2,k,t,j} \\
&- \alpha_{2,k,t,j} g_{k,t}(\theta_{C,t}, \theta_{N,t}, \pi, \theta_{C,P}, \theta_{N,P}) - \alpha_{2,k,t,j}\zeta_{k,t}\big) \\
&\times p_{\zeta_{k,t}}(\zeta_{k,t})\, d\theta_{k,t}\, d\zeta_{k,t} \Bigg] \\
&\times \prod_{j=1}^{M_{3,k,1}} p_{\varepsilon_{3,k,1,j,h}}(z_{3,k,1,j,h} - \mu_{3,k,1,j} - \alpha_{3,k,1,j}\theta_{k,P})\, d\theta_{k,P} \Bigg\} \\
&\times \prod_{j=1}^{M_{4,T+1}} p(z_{4,j,h})\, d\theta_{C,T+1}\, d\theta_{N,T+1}\, d\pi,
\end{aligned}
$$

where

$$
p(z_{4,j,h}) = p_{\varepsilon_{4,j,h}}(z_{4,j,h} - \mu_{4,j} - \alpha_{4,C,j}\theta_{C,T+1} - \alpha_{4,N,j}\theta_{N,T+1} - \alpha_{4,\pi,j}\pi) \quad \text{for} \quad j = 1, \dots, J_1
$$

and

$$p(z_{4,j,h}) = \left[1 - F_{\varepsilon_{4,j}}(\mu_{4,j} + \alpha_{4,C,j}\theta_{C,T+1} + \alpha_{4,N,j}\theta_{N,T+1} + \alpha_{4,\pi,j}\pi)\right]^{z_{4,j,h}} \times \left[F_{\varepsilon_{4,j}}(\mu_{4,j} + \alpha_{4,C,j}\theta_{C,T+1} + \alpha_{4,N,j}\theta_{N,T+1} + \alpha_{4,\pi,j}\pi)\right]^{1-z_{4,j,h}}$$
$$\text{for} \quad j = J_1 + 1, \dots, J.$$

In practice, we assume that $g_{k,t}$ is linear in the variables. 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.[9]

In principle, one can estimate the parameters of the model, the parameters of the technology, and the $p_\theta(\theta)$ by maximizing (A5.1) directly. To do that, one can approximate $p(z)$ by computing the integrals numerically in a deterministic fashion. However, if the number of integrals is very large, a serious 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 12 to obtain the function $p(z)$. This requires computing approximately 17 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 $\dim(\theta)$. To obtain good approximations of $p(z)$ even in the case with three factors and four time periods, we would need more than four 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 the next section.

A6. NONLINEAR FILTERING[10]

Let the vector θ_t denote the unobserved state vector which evolves according to the transition equation

$$\theta_{t+1} = f(\theta_t) + \eta_{t+1}. \tag{A6.1}$$

[9]Our analysis establishes that we can identify models with correlated measurement errors. However, the computational cost for such a model is substantial and we do not do that in this paper. Furthermore, Cunha and Heckman (2008) test and do not reject the null hypothesis that the measurement errors are uncorrelated.

[10]The notation in this section is self-contained and not closely related to the notation in the text or the notation in other parts of the Appendix. It is generic notation used in the filtering literature, and to connect with that literature, we use it here.

We denote by N_θ the dimension of θ_t. The vector η_t of dimension N_θ is the process noise that drives the dynamic system. It is assumed to be independently distributed over time with $E(\eta_t) = 0$ and $\mathrm{Var}(\eta_t) = \mathcal{H}_t$. Let z_t denote the vector of dimension N_z of observable variables which are related to the state θ_t via the measurement equations

$$z_t = h(\theta_t) + \varepsilon_t, \tag{A6.2}$$

where ε_t is the observation noise, which is a vector of dimension N_z. The noise ε_t is assumed to be independent random variables with $E(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = \mathcal{E}_t$. The functions $f : \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_\theta}$ and $h : \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_z}$ are possibly nonlinear.[11] The goal of nonlinear filtering is to estimate the density of θ_t conditional on the history $z^t = (z_1, \dots, z_t)$, which we denote by $p(\theta_t \mid z^t)$. As discussed in Arulampalam, Ristic, and Gordon (2004), the conceptual solution of nonlinear filtering is simple. Break the problem into a prediction step and an update step, and then proceed recursively. The prediction step generates $p(\theta_t \mid z^{t-1})$ given knowledge of $p(\theta_{t-1} \mid z^{t-1})$. This is accomplished by applying the Chapman–Kolmogorov equation:

$$p(\theta_t \mid z^{t-1}) = \int p(\theta_t \mid \theta_{t-1}) p(\theta_{t-1} \mid z^{t-1})\, d\theta_{t-1},$$

where $p(\theta_t \mid \theta_{t-1})$ is the density of θ_t conditional on θ_{t-1}. The update step computes $p(\theta_t \mid z^t)$ given $p(\theta_t \mid z^{t-1})$ via the Bayes rule:

$$p(\theta_t \mid z^t) = \frac{p(z_t \mid \theta_t) p(\theta_t \mid z^{t-1})}{p(z_t \mid z^{t-1})}.$$

A simple solution to the filtering problem exists when the functions f and h are linear and separable in each of their arguments, the unobserved state θ_t is Gaussian, and the noise terms ε_t, η_t are Gaussian, independent random variables. In this case, one can use the Kalman filter to derive the equations used in the prediction and update steps analytically. However, simple departures of this framework (e.g., f is nonlinear) make the Kalman filter unsuitable. It is possible to adapt this approach by considering the first-order Taylor series approximation of the function f and then applying the standard Kalman filter prediction and update rules. This is known in the filtering literature as the extended Kalman filter (EKF). The problem with this approach is that the EKF generally generates biased expressions for means and variances.

[11] A more general formulation is one in which these functions are nonseparable and indexed by time, and there are observable covariates x_t and w_t which affect the measurement or the transition equation, so that $z_t = h_t(x_t, \theta_t, \varepsilon_t)$ and $\theta_{t+1} = g_t(w_t, \theta_t, \eta_t)$. These are important extensions, but in this appendix we discuss a simpler formulation because it covers the model estimated in this paper. See Cunha (2008) for a discussion on the more general case.

More recently, researchers have used general particle filtering techniques.[12] However, in the context of panel data with a large cross-section dimension, the particle filter can be computationally costly. Furthermore, the particle filter may not be a good tool if the goal of the researcher is to estimate the (parameters of the) functions f or h, especially when these functions are time invariant (see discussion in Maskell (2004)).

Another approach is to consider the unscented Kalman filter (UKF) as proposed by Julier and Uhlmann (1997). The crucial assumption in this algorithm is that both $p(\theta_t \mid z^t)$ and $p(\theta_{t+1} \mid z^t)$ can be accurately approximated by the density of a normal random variable with mean

$$a_{t+k,t} = E(\theta_{t+k} \mid z^t)$$

and variance

$$\Sigma_{t+k,t} = \mathrm{Var}(\theta_{t+k} \mid z^t)$$

for $k \in \{0, 1\}$. Because of this assumption, the only objects that have to undergo the prediction and update steps are the means and variances of the approximating normal distribution, just as in the standard Kalman filter algorithm.

Obviously, in some situations the normal approximation may not be a good one. It is possible that nonlinear functions of normally distributed random variables generate random variables that have densities that are not symmetric around their means or have many modes, which would be inconsistent with a normal approximation. We introduce a more flexible approach which considers approximations that use mixture of normals:

$$p(\theta_{t+k} \mid z^t) \simeq \sum_{l=1}^{L} \tau_{l,t} \phi(\theta_t; a_{l,t+k,t}, \Sigma_{l,t+k,t}),$$

where $\phi(\theta_t; a_{l,t+k,t}, \Sigma_{l,t+k,t})$ is the probability density function of a normal random variable with mean $a_{l,t+k,t}$ and variance $\Sigma_{l,t+k,t}$, for $k \in \{0, 1\}$. The weights $\tau_{l,t}$ are such that $\tau_{l,t} \in [0, 1]$ and $\sum_l^L \tau_{l,t} = 1$. Under this formulation, within each stage, we break the filtering problem into parallel problems and obtain the final result at the end.

A6.1. *The Update Step*

First, we compute the update density for each element of the mixture. Namely, let $\hat{z}_{l,t}$ denote the predicted measurement by the lth element of the mixture:

$$\hat{z}_{l,t} = E_l(z_t \mid z^{t-1}) = E_l[h(\theta_t) \mid z^{t-1}] + E_l[\varepsilon_t \mid z^{t-1}] = E_l[h(\theta_t) \mid z^{t-1}]. \tag{A6.3}$$

[12]See, for example, Doucet, de Freitas, and Gordon (2001) and Fernandez-Villaverde and Rubio-Ramirez (2006).

Below, we show how to compute the moment displayed above. For now, consider the updating equations

$$a_{l,t,t} = a_{l,t,t-1} + K_{l,t}(z_t - \hat{z}_{l,t}), \tag{A6.4a}$$

$$\Sigma_{l,t,t} = \Sigma_{l,t,t-1} - K_{l,t}F_{l,t}K'_{l,t}, \tag{A6.4b}$$

where

$$K_{l,t} = \mathrm{Cov}[\theta_t, z_t \mid z^{t-1}]F_{l,t}^{-1} \tag{A6.4c}$$

and

$$F_{l,t} = \mathrm{Var}[h(\theta_t, \varepsilon_t) \mid z^{t-1}]. \tag{A6.4d}$$

We can then approximate the posterior density $p(\theta_t \mid z^t)$ with a linear combination of densities $\phi(\theta_t; a_{l,t,t}, \Sigma_{l,t,t})$ with weights given by

$$\tau_{r,t} = \frac{\tau_{r,t-1}\phi(z_t; \hat{z}_{r,t}, F_{r,t})}{\sum_{l=1}^{L} \tau_{l,t-1}\phi(z_t; \hat{z}_{l,t}, F_{l,t})}, \quad r \in \{1, \dots, L\}. \tag{A6.4e}$$

A6.2. *The Prediction Step*

With knowledge of a good approximation for the density $p(\theta_t \mid z^t)$ expressed as the mixture of normals and knowledge of the transition equation (A6.1) one can approximately compute the one-step-ahead prediction density $p(\theta_{t+1} \mid z^t)$ also expressed as a mixture of normals. More precisely, let

$$a_{l,t+1,t} = E_l(\theta_{t+1} \mid z^t) = E_l(f(\theta_t) + \eta_{t+1} \mid z^t) = E_l(f(\theta_t) \mid z^t), \tag{A6.5}$$

$$\begin{aligned} \Sigma_{l,t+1,t} &= \mathrm{Var}_l[\theta_{t+1} \mid z^t] = \mathrm{Var}(f(\theta_t) + \eta_{t+1} \mid z^t) \\ &= \mathrm{Var}(f(\theta_t) \mid z^t) + \mathcal{H}_{t+1}. \end{aligned} \tag{A6.6}$$

Then an approximation to $p(\theta_{t+1} \mid z^t)$ is given by

$$p(\theta_{t+1} \mid z^t) \approx \sum_{l=1}^{L} \tau_{l,t}\phi(z_t; a_{l,t+1,t}, \Sigma_{l,t+1,t}).$$

A6.3. *Unscented Transform*

A difficulty arises in the implementation of the filtering because in the prediction and update stages one has to compute integrals that involve nonlinear transformations of random variables whose distributions are approximated by mixtures of normals. The unscented transform (UT) is a convenient tool to compute the mean and variance of a random variable that undergoes a nonlinear transformation. For example, consider computing the expressions (A6.5) and (A6.6). Then, by definition,

$$a_{l,t+1,t} = \int f(\theta_t)\phi(\theta_t; a_{l,t,t}, \Sigma_{l,t,t})\, d\theta_t, \tag{A6.7}$$

$$\begin{aligned}\Sigma_{l,t+1,t} = \int &(f(\theta_t) - a_{l,t+1,t})(f(\theta_t) - a_{l,t+1,t})' \\ &\times \phi(\theta_t; a_{l,t,t}, \Sigma_{l,t,t})\, d\theta_{t-1} + \mathcal{H}_{t+1}.\end{aligned} \tag{A6.8}$$

The expressions (A6.7) and (A6.8) involve the computation of N_θ integrals. One way to proceed is to consider the product of integrals estimated by quadrature. The difficulty with this approach is that as N_θ becomes larger, the number of evaluations increases exponentially.

Another approach, discussed in Judd (1998), is to consider monomial rules. The unscented transform proposed by Julier and Uhlmann (1997) is a monomial rule that approximates the expressions (A6.7) and (A6.8). To do so, one picks deterministically $2N_\theta + 1$ points $x_{n,l,t,t}$ and corresponding weights $w_{n,l,t}$, $n = 0, 1, \dots, 2N_\theta$. Let $\Sigma_{l,t,t}^{1/2}$ denote the square root of the positive definite $(N_\theta \times N_\theta)$ matrix $\Sigma_{l,t,t}$. Let $\Sigma_{l,t,t}^{1/2}(n, :)$ denote the nth row of $\Sigma_{l,t,t}^{1/2}$. Let $\kappa \in \mathbb{R}$ such that $\kappa + N_\theta \neq 0$. The UT proposes the points

$$x_{n,l,t,t} = \begin{cases} a_{l,t,t}, & \text{for } n = 0, \\ a_{l,t,t} + (N_\theta + \kappa)^{1/2}\Sigma_{l,t,t}^{1/2}(n, :), & \text{for } n = 1, \dots, N_\theta, \\ a_{l,t,t} - (N_\theta + \kappa)^{1/2}\Sigma_{l,t,t}^{1/2}(n, :), & \text{for } n = N_\theta + 1, \dots, 2N_\theta, \end{cases} \tag{A6.9}$$

and the weights

$$w_{n,l,t} = \begin{cases} \dfrac{\kappa}{N_\theta + \kappa}, & \text{for } n = 0, \\ \dfrac{1}{2(N_\theta + \kappa)}, & \text{for } n = 1, \dots, N_\theta, \\ \dfrac{1}{2(N_\theta + \kappa)}, & \text{for } n = N_\theta + 1, \dots, 2N_\theta. \end{cases}$$

We approximate $E_l[f(\theta_t) \mid z^t]$ and $\mathrm{Var}_l[f(\theta_t) \mid z^t]$ by computing

$$a_{l,t+1,t} = E_l[f(\theta_t) \mid z^t] \approx \sum_{n=0}^{2N_\theta} w_{n,l,t} f(x_{n,l,t,t})$$

and

$$\Sigma_{l,t,t} = \mathrm{Var}[f(\theta_t) \mid z^t]$$
$$\approx \sum_{n=0}^{2N_\theta} w_{n,l,t}[f(x_{n,l,t,t}) - a_{l,t+1,t}][f(x_{n,l,t,t}) - a_{l,t+1,t}]' + \mathcal{H}_{t+1}.$$

A6.4. *Implementation of Nonlinear Filtering*

Let $p(z)$ denote the likelihood (4.1). The following representation is always valid:

$$p(z) = p(z_1)\prod_{t=1}^{T} p(z_{t+1} \mid z^t). \tag{A6.10}$$

The idea is to use the nonlinear filtering to obtain a recursive algorithm which we can use to calculate $p(z_{t+1} \mid z^t)$. To see how, note that we assume that

$$p(\theta_1) \approx \sum_{l=1}^{L} \tau_{l,0}\phi(\theta_1; a_{l,1,0}, \Sigma_{l,1,0}).$$

It follows that

$$p(z_1) \approx \sum_{l=1}^{L} \tau_{l,0}\phi(z_1; \hat{z}_1, F_{l,1}),$$

where $\hat{z}_1$ and $F_{l,1}$ are defined in (A6.3) and (A6.4d). Now, applying (A6.4a), (A6.4b), (A6.4c), and (A6.4e) allows us to obtain $a_{l,1,1}$, $\Sigma_{l,1,1}$, and $\tau_{l,1}$ which are really helpful because now we can characterize the posterior density as

$$p(\theta_1 \mid z_1) \approx \sum_{l=1}^{L} \tau_{l,1}\phi(\theta_1; a_{l,1,1}, \Sigma_{l,1,1}).$$

We then apply the prediction steps to obtain $a_{l,2,1}$ and $\Sigma_{l,2,1}$. With knowledge of these quantities, we can approximate the predicted density as

$$p(\theta_2 \mid z_1) \approx \sum_{l=1}^{L} \tau_{l,1}\phi(\theta_2; a_{l,2,1}, \Sigma_{l,2,1}).$$

Now we complete the cycle, because by using (A6.3)–(A6.4d), we can compute $\hat{z}_2$ and $F_{l,2}$, with which we can compute

$$p(z_2 \mid z_1) \approx \sum_{l=1}^{L} \tau_{l,1}\phi(z_2; \hat{z}_2, F_{l,2}).$$

Furthermore, we use (A6.4e) to update the weights $\tau_{l,2}$. By proceeding in a recursive manner, we can construct the right-hand side of (A6.10), which is just equal to the likelihood.

A7. IMPLEMENTATION OF ANCHORING

A7.1. *Linear Anchoring*

To fix ideas, let $Z_{4,1}$ denote a continuous outcome that can be written as a linear, separable function of period $T+1$ cognitive and noncognitive skills. Let $\theta_{N,T+1}$ and $\theta_{C,T+1}$ denote the stocks of noncognitive and cognitive skills of the agent upon completion of childhood. π denotes the heterogeneity component. Then

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

We define linear anchoring functions

$$\begin{aligned} g_{N,1}(\ln \theta_{N,T+1}) &= \mu_{4,1} + \alpha_{4,N,1} \ln \theta_{N,T+1}, \\ g_{C,1}(\ln \theta_{C,T+1}) &= \mu_{4,1} + \alpha_{4,C,1} \ln \theta_{C,T+1}. \end{aligned}$$

We rewrite the technology as

$$\begin{aligned} \ln \theta_{k,t+1} = {} & -\frac{\mu_{4,1}}{\alpha_{4,k,1}} + \frac{1}{\alpha_{4,k,1}} \\ & \times f_{k,s}[e^{\mu_{4,1}+\alpha_{4,C,1} \ln \theta_{C,t}}, e^{\mu_{4,1}+\alpha_{4,N,1} \ln \theta_{N,t}}, e^{\ln I_{k,t}}, e^{\ln \theta_P}, \pi, \eta_{k,t}] \\ & \text{for} \quad k \in \{C, N\}, \end{aligned}$$

where $\ln \theta_P = (\ln \theta_{C,P}, \ln \theta_{N,P})$, and $\alpha_{4,k,1} \neq 0$.

A7.2. *Nonlinear Anchoring*

Let $Z^*_{4,1}$ denote the net benefit of graduating from high school. Let $\theta_{N,T+1}$ and $\theta_{C,T+1}$ denote the stocks of noncognitive and cognitive skills of the agent at the end of childhood. We assume that

$$Z^*_{4,1} = \mu_{4,1} + \alpha_{4,C,1} \ln \theta_{C,T+1} + \alpha_{4,N,1} \ln \theta_{N,T+1} + \alpha_{4,\pi,1}\pi + \varepsilon_{4,1}.$$

We do not observe $Z^*_{4,1}$, but the variable $Z_{4,1}$ which is defined as

$$Z_{4,1} = \begin{cases} 0, & \text{if } \mu_{4,1} + \alpha_{4,C,1} \ln \theta_{C,T+1} + \alpha_{4,N,1} \ln \theta_{N,T+1} \\ & \qquad + \alpha_{4,\pi,1}\pi + \varepsilon_{4,1} \leq 0, \\ 1, & \text{otherwise.} \end{cases}$$

We can define the anchoring functions in terms of probabilities. Assuming that $\varepsilon_{4,1} \sim N(0,1)$, it follows that

$$\begin{aligned}&\Pr(Z_{4,1}=1 \mid \mu_{4,1}, \ln\theta_{N,T+1}, \ln\theta_{C,T+1}, \pi)\\&\quad=\Phi(\mu_{4,1}+\alpha_{4,C,1}\ln\theta_{C,T+1}+\alpha_{4,N,1}\ln\theta_{N,T+1}+\alpha_{4,\pi,1}\pi).\end{aligned}$$

For any $\theta_{k,T}$ we can define the anchoring functions $g_{k,1}$ for $k=C,N$ as

$$\begin{aligned}&g_{C,1}(\ln\theta_{C,T+1})\\&\quad=\int\Phi(\mu_{4,1}+\alpha_{4,C,1}\ln\theta_{C,T+1}+\alpha_{4,N,1}\ln\theta_{N,T+1}+\alpha_{4,\pi,1}\pi)\\&\qquad\times p(\ln\theta_{N,T+1},\pi)\,d\ln\theta_{N,T+1}\,d\pi\end{aligned}$$

and

$$\begin{aligned}&g_{N,1}(\ln\theta_{N,T+1})\\&\quad=\int\Phi(\mu_{4,1}+\alpha_{4,C,1}\ln\theta_{C,T+1}+\alpha_{4,N,1}\ln\theta_{N,T+1}+\alpha_{4,\pi,1}\pi)\\&\qquad\times p(\ln\theta_{C,T+1},\pi)\,d\ln\theta_{C,T+1}\,d\pi,\end{aligned}$$

where $\mu_{4,1}$ can depend on covariates on which we condition. If $\ln\theta_{k,T+1} \sim N(0,\sigma^2_{\theta_{k,T+1}})$ for $k\in\{C,N\}$ and $\pi\sim N(0,\sigma^2_\pi)$, we obtain

$$\begin{aligned}&g_{C,1}(\ln\theta_{C,T+1})\\&\quad=\Phi\left(\frac{\mu_{4,1}+\alpha_{4,C,1}\ln\theta_{C,T+1}}{\sqrt{1+\alpha^2_{4,N,1}\sigma^2_{\theta_{N,T+1}}+\alpha^2_{4,\pi,1}\sigma^2_\pi+2\alpha_{4,N,1}\alpha_{4,\pi,1}\operatorname{Cov}(\pi,\ln\theta_{N,T+1})}}\right),\\&g_{N,1}(\ln\theta_{N,T+1})\\&\quad=\Phi\left(\frac{\mu_{4,1}+\alpha_{4,N,1}\ln\theta_{N,T+1}}{\sqrt{1+\alpha^2_{4,C,1}\sigma^2_{\theta_{C,T+1}}+\alpha^2_{4,\pi,1}\sigma^2_\pi+2\alpha_{4,C,1}\alpha_{4,\pi,1}\operatorname{Cov}(\pi,\ln\theta_{C,T+1})}}\right).\end{aligned}$$

We seek to estimate the technologies $\tilde{f}_{k,s}$ in terms of the anchored factors:

$$\begin{aligned}\text{(A7.1)}\quad&\tilde{f}_{k,s}(g_{C,1}(\ln\theta_{C,t}), g_{N,1}(\ln\theta_{N,t}), I_{k,t}, \theta_P, \pi, \eta_{k,t})\\&\quad=g_{k,1}\big[f_{k,s}\big(g^{-1}_{C,1}(g_{C,1}(\ln\theta_{C,t})), g^{-1}_{N,1}(g_{N,1}(\ln\theta_{N,t})), I_{k,t}, \theta_P, \pi, \eta_{k,t}\big)\big],\\&\quad k\in\{C,N\}.\end{aligned}$$

The implementation of (A7.1) is not without problems. Note that $g^{-1}(x)$ is the inverse of a c.d.f. function. It will be defined as long as $x\in[0,1]$. However, there is no guarantee that $f_{k,s}(g_{C,1}(\theta_{C,t}), g_{N,1}(\theta_{N,t}), I_{k,t}, \theta_P, \pi, \eta_{k,t})\in[0,1]$.

We confront this problem by proposing slightly modified nonlinear anchoring functions:

$$
g_{C,1}(\ln\theta_{C,T+1}) = \frac{\Phi\left(\dfrac{\mu_{4,1}+\alpha_{4,C,1}\ln\theta_{C,T+1}}{\sqrt{1+\alpha_{4,N,1}^2\sigma_{\theta_{N,T+1}}^2+\alpha_{4,\pi,1}^2\sigma_\pi^2+2\alpha_{4,N,1}\alpha_{4,\pi,1}\operatorname{Cov}(\pi,\ln\theta_{N,T+1})}}\right)}{1-\Phi\left(\dfrac{\mu_{4,1}+\alpha_{4,C,1}\ln\theta_{C,T+1}}{\sqrt{1+\alpha_{4,N,1}^2\sigma_{\theta_{N,T+1}}^2+\alpha_{4,\pi,1}^2\sigma_\pi^2+2\alpha_{4,N,1}\alpha_{4,\pi,1}\operatorname{Cov}(\pi,\ln\theta_{N,T+1})}}\right)}
$$

and

$$
g_{N,1}(\theta_{N,T+1}) = \frac{\Phi\left(\dfrac{\mu_{4,1}+\alpha_{4,N,1}\ln\theta_{N,T+1}}{\sqrt{1+\alpha_{4,C,1}^2\sigma_{\theta_{C,T+1}}^2+\alpha_{4,\pi,1}^2\sigma_\pi^2+2\alpha_{4,C,1}\alpha_{4,\pi,1}\operatorname{Cov}(\pi,\ln\theta_{C,T+1})}}\right)}{1-\Phi\left(\dfrac{\mu_{4,1}+\alpha_{4,N,1}\ln\theta_{N,T+1}}{\sqrt{1+\alpha_{4,C,1}^2\sigma_{\theta_{C,T+1}}^2+\alpha_{4,\pi,1}^2\sigma_\pi^2+2\alpha_{4,C,1}\alpha_{4,\pi,1}\operatorname{Cov}(\pi,\ln\theta_{C,T+1})}}\right)}.
$$

If we define $\ln\tilde{\theta}_{k,1} = g_{k,1}(\ln\theta_k)$, then

$$
\ln\theta_k = g_{k,1}^{-1}(\ln\tilde{\theta}_{k,1}) = \Phi^{-1}\left(\frac{\ln\tilde{\theta}_{k,1}}{1+\ln\tilde{\theta}_{k,1}}\right)
$$

and because by definition $\ln\tilde{\theta}_{k,1}\ge 0$, it follows that $\ln\tilde{\theta}_{k,1}/(1+\ln\tilde{\theta}_{k,1})\in[0,1]$.

A7.3. *Adjusting the Likelihood*

Finally, we need to include the anchoring equations in the likelihood. To that end, let $z_{4,j,h}$ denote the jth anchoring outcome for person h. Note that, for $j\in\{1,\dots,J_1\}$, we have linear equations. In this case, we can write the contribution of measurement $z_{4,j,h}$, $j=1,\dots,J_1$, as

$$
\begin{aligned}
p(z_{4,j,h}) = p_{\varepsilon_{4,j,h}}(&z_{4,j,h}-\mu_{4,j}-\alpha_{4,C,j}\ln\theta_{C,T+1}\\
&-\alpha_{4,N,j}\ln\theta_{N,T+1}-\alpha_{4,\pi,j}\pi).
\end{aligned}
$$

For $j\in\{J_1+1,\dots,J\}$ we have discrete dependent variables. In our case, they are binary dependent variables. We write the contribution of measurement $Z_{4,j,h}$ for $j=J_1+1,\dots,J$ as

$$
\begin{aligned}
p(z_{4,j,h}) = \big[1-F_{\varepsilon_{4,j}}(&\mu_{4,j}+\alpha_{4,C,j}\ln\theta_{C,T+1}\\
&+\alpha_{4,N,j}\ln\theta_{N,T+1}+\alpha_{4,\pi,j}\pi)\big]^{z_{4,j,h}}
\end{aligned}
$$

$$\times \left[F_{\varepsilon_{4,j}}(\mu_{4,j} + \alpha_{4,C,j} \ln \theta_{C,T+1} + \alpha_{4,N,j} \ln \theta_{N,T+1} + \alpha_{4,\pi,j}\pi)\right]^{1-z_{4,j,h}}.$$

A8. A LIMITED MONTE CARLO STUDY OF THE EMPIRICAL MODEL OF SECTION 4.2.5

We assume $T = 8$ periods and $N = 2200$, which is our sample size. We use two measures for each factor and they are written as in (3.1)–(3.3). This is very conservative, because we have many more measures for parental investments than only two. We performed 10 simulations, because this is a very time-consuming algorithm. Results are for a model with one latent factor. We do not report the estimates of the intercept parameter.

	Parameter	True Value	Mean Estimate	Standard Error (10 Simulations)
Eq. (4.2)	κ_θ	0.4	0.4191	0.0942
	κ_y	0.4	0.3658	0.0851
Technology	γ_1 (θ_t)	0.7	0.7178	0.1903
	γ_3 (investment)	0.4	0.4150	0.1437
	ϕ	−0.2	−0.2282	0.0472
Eq. (4.3)	ρ_y	0.6	0.6236	0.0914
Eq. (4.4)	ρ_π	0.4	0.3638	0.0686

A9. DATA APPENDIX

A9.1. *Survey Measures on Parental Investments*

The measures regarding parental investments are those that describe the quality of a child's home environment that are included in the CNLSY/79 Home Observation Measurement of the Environment—Short Form (HOME-SF). They are a subset of the measures used to construct the HOME scale designed by Bradley and Caldwell (1980, 1984) to assess the emotional support and cognitive stimulation children receive through their home environment, planned events, and family surroundings. These measurements have been used extensively as inputs to explain child characteristics and behaviors (see, e.g., Todd and Wolpin (2005)). As discussed in Linver, Brooks-Gunn, and Cabrera (2004), some of these items are not useful because they do not vary much among families (i.e., more than 90–95% of all families have the same response). Our empirical study uses measurements on the following parental investments: How often child gets out of house, the number of books the child has, how often the mother reads to the child, the number of soft/role play toys, the number of push/pull toys, how often child eats with mom/dad, how often mom talks to child from work, number of magazines, whether the child has tape recorder/CD player, how often child is taken to museum, whether the child has musical instrument, whether the family receives daily newspaper, whether the child receives special lessons/activities, whether the child is taken to musical performances, how often the child sees family friends, the number of times praised child last week, the number of times said positive things last week.

A9.2. *Survey Measures on Child's Cognitive Skills*

We use several measures on the child's cognitive skills. These measures vary with the age of the child. Before age 4, we use the Motor and Social Development Scale, the Parts of the Body Score, the Memory for Location Score, and the Peabody Picture Vocabulary Test. We briefly describe them next.

The Motor and Social Development scale (MSD) was developed by the National Center for Health Statistics to measure dimensions of the motor, social, and cognitive development of young children from birth through the age of 47 months. The items were derived from standard measures of child development (Bayley Scales of Infant Development, the Gesell Scale, Denver Developmental Screening Test), which have high reliability and validity as shown by Poe (1986). The scale has been used in the National Health Interview Survey (a large national health survey that included 2714 children up to age 4) and in the third National Health and Nutrition Examination Survey (NHANES, 1988–1994). Analyses by Child Trends, a nonprofit, nonpartisan research organization, of the scale in the 1981 Child Health Supplement to the National Health Interview Survey established the age ranges at which each item's developmental milestone is generally reached by U.S. children, as shown in Peterson and Moore (1987). Based on the child's age, NLSY79 mothers answer 15 age-appropriate items out of 48 motor and social development items. These items have been used with a full spectrum of minority children with no apparent difficulty.

The NLSY79 Motor and Social Development assessment has eight components (Parts A–H) that a mother completes contingent on the child's ages. Part A is appropriate for infants during the first 4 months of life (i.e., 0–3 months) and the most advanced section, Part H, is addressed to children between the ages of 22 and 47 months. All of the items are dichotomous (scored either 0 or 1) and the total raw score for children of a particular age is obtained by a simple summation (with a range 0–15) of the affirmative responses in the age-appropriate section.

The Parts of the Body assessment was completed by age-eligible NLSY79 children in 1986 and 1988 only. Developed by Kagan (1981), Parts of the Body attempts to measure a 1- or 2-year-old child's receptive vocabulary knowledge of orally presented words as a means of estimating verbal intellectual development. The interviewer names each of 10 body parts and asks the child to point to that part of his or her body. The child's score is computed by summing the items that a child correctly identifies. Thus, a minimum score is 0 and a maximum score is 10. No proration was attempted since the later items in the sequence are more difficult than the earlier items.

The Memory for Locations assessment was completed by age-eligible NLSY79 children in 1986 and 1988 only. It was developed as a measure of a child's short-term memory and has been extensively used by Kagan (1981). The child, aged 8 months through 3 years, watches as a figure is placed under one of two to six cups. The cups are screened from a child's view for 1–15 seconds; the child is then asked to find the location of the figure. Items increase in difficulty as the number of cups and/or the length of time during which the cups are hidden from view increases. A child's score is based on his or her ability to select the cup hiding the figure.

The number of individual items that a child can potentially answer in this assessment is contingent on the age of the child. Children between the ages of 8 and 23 months start with item 1, the easiest question; children who are at least 2 years of age begin with item 4, and children age 3 start with item 7. A child's score is based on the highest (most difficult) question answered. A child who cannot answer the entry item receives a raw score of zero regardless of where he or she enters.

The Body Parts and Memory for Locations assessments were no longer used in the NLSY79 Child surveys following the 1988 Child data collection effort, partly because of funding constraints and partly because of the greater difficulty in administering them to children in a home setting. Interviewers found it difficult to make an unambiguous determination as to whether a child was unable to respond or whether he/she was just shy. It was sometimes difficult to be definitive regarding the direction in which a child was pointing, either toward a cup or toward a body part. More recent research by Mott, Baker, Ball, Keck, and Lenhart (1998) suggests that these two assessments may be useful independent predictors of cognitive development since Body Parts and Memory for Location scores in 1986 are highly significant predictors of Peabody assessments in 1992. It appears that, in standard multivariate analyses, these early child cognitive measures may indeed be useful predictors of aptitude and achievement measures 6 years later.

Finally, starting at age 5, we use as measurements of cognitive skills the Peabody Individual Achievement Test (PIAT), which is a wide-ranging measure of academic achievement of children aged 5 and over. It is widely used in developmental research. Todd and Wolpin (2005) used the raw PIAT test score as their measure of cognitive outcomes. The CNLSY/79 includes two subtests from the full PIAT battery: PIAT Mathematics and PIAT Reading Recognition.[13] The PIAT Mathematics measures a child's attainment in mathematics as taught in mainstream education. It consists of 84 multiple-choice items of

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

increasing difficulty. It begins with basic skills such as recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The PIAT Reading Recognition subtest measures word recognition and pronunciation ability. Children read a word silently, then say it aloud. The test contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include the ability to match letters, name names, and read single words aloud.

A9.3. *Survey Measures on Child's Noncognitive Skills*

We use two systems of measurements on the child's noncognitive skills. Up to age 4, we use the Temperament Scale. Starting at age 4, we use the Behavior Problem Index. We give a brief description of these measures next.

The Temperament Scale is based on Rothbart's Infant Behavior Questionnaire, Campos, Barrett, Lamb, Goldsmith, and Stenberg (1983) and Kagan (1981) compliance scale. Because the child's temperament is partially a parental perception, the behavioral style of children in the NLSY79 was measured by a set of maternal-report items. The maternal scale "How My Infant Usually Acts" addresses the activity, predictability, fearfulness, positive affect, and friendliness of infants below age 1. "How My Toddler Usually Acts" addresses the fearfulness, positive affect, and friendliness of 1-year-olds. "How My Child Usually Acts" measures the compliance and attachment of 2- and 3-year-olds and, additionally, the friendliness of children aged 4–6.

As measurements of noncognitive skills we use components of the Behavior Problem Index (BPI) created by Peterson and Zill (1986) and designed to measure the frequency, range, and type of childhood behavior problems for children age 4 and over. The Behavior Problem score is based on responses from the mothers to 28 questions about specific behaviors that children age 4 and over may have exhibited in the previous 3 months. Three response categories are used in the questionnaire: often true, sometimes true, and not true. In our empirical analysis we use the following subscores of the behavioral problems index: (i) antisocial, (ii) anxious/depressed, (iii) headstrong, (iv) hyperactive, (v) peer problems. Among other characteristics, a child who scores low on the antisocial subscore is a child who often cheats or tell lies, is cruel or mean to others, and does not feel sorry for misbehaving. A child who displays a low score on the anxious/depressed measurement is a child who experiences sudden changes in mood, feels no one loves him/her, is fearful, or feels worthless or inferior. A child with low scores on the headstrong measurement is tense, nervous, argues too much, and is disobedient at home, for example. Children will score low on the hyperactivity subscale if they have difficulty concentrating, act without thinking, and are restless or overly active. Finally, a child will be assigned a low score on the peer problem subscore if they have problems getting along with others, are not liked by other children, and are not involved with others. Tables A9-1–A9-3 present basic statistics for the data used in our analysis.

TABLE A9-1

MEASUREMENTS ON CHILD COGNITIVE AND NONCOGNITIVE SKILLS

	Period 1: Year of Birth of Child			Period 2: Ages 1–2			Period 3: Ages 3–4			Period 4: Ages 5–6		
	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error
Measurements on Child Cognitive Skills												
Gestation Length (10 Weeks)	2118	3.878	0.234									
Weight at Birth	2159	3.345	0.561									
Motor-Social Development Score	209	0.123	0.998	1043	0.094	0.968	915	0.185	0.923			
Body Parts				317	0.267	1.002						
Memory for Locations				373	0.221	0.939						
Peabody Picture Vocabulary Test							738	0.543	0.954	809	0.475	0.907
PIAT Math										1101	0.271	1.040
PIAT Reading Recognition										1074	0.246	1.016
PIAT Reading Comprehension										1025	0.241	0.980
Measurements on Child Noncognitive Skills												
Temperament/Compliance Raw Score				274	0.213	0.926	1253	0.119	0.941			
Temperament/ Insecure Attachment Raw Score				299	0.033	0.888	1282	−0.015	0.869			
Temperament/Sociability Raw Score				422	0.281	0.911	893	0.100	1.002			
Temperament/Difficulty Raw Score	207	−0.046	1.035	788	0.111	1.019						
Temperament/Friendliness Raw Score	224	0.183	0.920	794	0.245	0.847						
Behavior Problem Index/ Antisocial Raw Score							353	0.073	0.983	1453	0.093	0.937
Behavior Problem Index/ Anxiety Raw Score							353	−0.124	1.092	1461	−0.066	1.033
Behavior Problem Index/ Headstrong Raw Score							351	−0.161	0.990	1462	−0.099	0.997
Behavior Problem Index/ Hyperactive Raw Score							352	−0.084	0.991	1461	0.010	0.973
Behavior Problem Index/ Conflict Raw Score							354	−0.004	0.975	1463	0.064	0.906

(*Continues*)

TABLE A9-1—*Continued*

	Period 5: Ages 7–8			Period 6: Ages 9–10			Period 7: Ages 11–12			Period 8: Ages 13–14		
	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error
Measurements on Child Cognitive Skills												
Gestation Length (10 Weeks)												
Weight at Birth												
Motor-Social Development Score												
Body Parts												
Memory for Locations												
Peabody Picture Vocabulary Test												
PIAT Math	1433	0.285	0.976	1379	0.321	0.911	1238	0.372	0.920	1063	0.425	0.922
PIAT Reading Recognition	1433	0.222	1.054	1380	0.299	0.945	1236	0.342	0.915	1064	0.336	0.876
PIAT Reading Comprehension	1383	0.246	1.058	1361	0.337	0.936	1221	0.392	0.932	1056	0.427	0.937
Measurements on Child Noncognitive Skills												
Temperament/Compliance Raw Score												
Temperament/ Insecure Attachment Raw Score												
Temperament/Sociability Raw Score												
Temperament/Difficulty Raw Score												
Temperament/Friendliness Raw Score												
Behavior Problem Index/ Antisocial Raw Score	1489	0.083	0.951	1422	0.112	0.929	1293	0.137	0.932	1125	0.117	0.971
Behavior Problem Index/ Anxiety Raw Score	1517	−0.135	1.063	1438	−0.098	1.032	1321	−0.076	1.045	1138	−0.088	1.053
Behavior Problem Index/ Headstrong Raw Score	1512	−0.124	0.995	1444	−0.110	0.995	1319	−0.067	1.009	1143	−0.070	0.998
Behavior Problem Index/ Hyperactive Raw Score	1513	0.011	0.975	1443	0.042	0.941	1321	0.069	0.960	1138	0.044	0.974
Behavior Problem Index/ Conflict Raw Score	1517	0.011	0.988	1446	0.035	0.962	1319	−0.016	1.028	1142	−0.024	1.033

TABLE A9-2

MEASUREMENTS ON PARENTAL INVESTMENTS ON CHILD SKILLS

	Period 1: Year of Birth of Child			Period 2: Ages 1–2			Period 3: Ages 3–4			Period 4: Ages 5–6		
	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error
How Often Child Gets Out of House	228	3.596	1.546	1125	4.418	1.000	1023	3.720	1.001	717	3.615	0.997
Number of Books	229	2.590	1.198	1125	3.600	0.787	1027	3.933	0.325	1117	3.944	0.301
How Often Mom Reads to Child	223	3.085	2.094	1120	4.990	1.412	1026	5.198	1.041	1116	5.044	1.032
Number of Soft/Role Play Toys	229	1.167	0.825	1114	1.843	1.302						
Number of Push/Pull Toys	228	0.151	0.294	1115	0.670	0.536						
How Often Child Eats With Mom/Dad	212	3.623	2.122	1086	4.892	1.117	986	4.792	1.182	1054	4.677	1.333
How Often Mom Talks to Child From Work	221	4.434	0.727	1123	4.484	0.632						
Number of Magazines							1023	3.246	1.378	718	3.164	1.417
Child Has Tape Recorder/CD Player							1022	0.771	0.420	716	0.809	0.394
How Often Child Is Taken to Museum										1110	2.247	0.977
Child Has Musical Instrument										392	0.434	0.496
Family Receives Daily Newspaper										393	0.506	0.501
Child Receives Special Lessons/Activities										392	0.579	0.494
Child Is Taken to Musical Performances										393	1.936	0.835
How Often Child Sees Family Friends										395	3.901	1.183
Number of Times Praised Child Last Week										273	2.524	2.353
Number of Times Said Positive Things Last Week										233	3.931	1.933

(*Continues*)

TABLE A9-2—*Continued*

	Period 5: Ages 7–8			Period 6: Ages 9–10			Period 7: Ages 11–12			Period 8: Ages 13–14		
	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error	Number of Observations	Mean	Standard Error
How Often Child Gets Out of House												
Number of Books	1525	3.949	0.278	1176	3.889	0.411	1322	3.815	0.489	1142	3.701	0.619
How Often Mom Reads to Child	1525	4.589	1.255	725	3.917	1.405						
Number of Soft/Role Play Toys												
Number of Push/Pull Toys												
How Often Child Eats With Mom/Dad	1476	4.552	1.409	1143	4.491	1.486	1268	4.350	1.510	1086	4.233	1.575
How Often Mom Talks to Child From Work												
Number of Magazines												
Child Has Tape Recorder/CD Player												
How Often Child Is Taken to Museum	1521	2.373	0.917	1174	2.350	0.870	1322	2.299	0.857	1137	2.182	0.839
Child Has Musical Instrument	1522	0.442	0.497	1174	0.520	0.500	1323	0.621	0.485	1140	0.635	0.482
Family Receives Daily Newspaper	1522	0.528	0.499	1176	0.505	0.500	1320	0.523	0.500	1139	0.507	0.500
Child Receives Special Lessons/Activities	1519	0.630	0.483	1172	0.708	0.455	1321	0.751	0.433	1142	0.736	0.441
Child is Taken to Musical Performances	1516	1.922	0.832	1175	1.926	0.849	1323	1.929	0.802	1141	1.918	0.856
How Often Child Sees Family Friends	1517	3.792	1.190	1176	3.725	1.228	1321	3.656	1.207	1142	3.509	1.235
Number of Times Praised Child Last Week	955	2.929	2.249	809	2.907	2.266	967	3.127	2.150	856	3.197	2.145
Number of Times Said Positive Things Last Week	910	3.849	1.886	779	3.705	1.898	964	3.567	1.908	897	3.479	1.946

TABLE A9-3

MEASUREMENTS ON MATERNAL COGNITIVE AND NONCOGNITIVE SKILLS

	Number of Observations	Mean	Standard Error
Measurements on Maternal Cognitive Skills			
Mom's Arithmetic Reasoning Test Score	2207	0.172	0.933
Mom's Word Knowledge Test Score	2207	0.302	0.822
Mom's Paragraph Composition Test Score	2207	0.377	0.827
Mom's Numerical Operations Test Score	2207	0.343	0.875
Mom's Coding Speed Test Score	2207	0.468	0.879
Mom's Mathematical Knowledge Test Score	2207	0.185	0.972
Measurements on Maternal Noncognitive Skills			
Mom's Self-Esteem: "I am a person of worth"	2207	3.534	0.516
Mom's Self-Esteem: "I have good qualities"	2207	3.382	0.530
Mom's Self-Esteem: "I am a failure"	2207	3.477	0.580
Mom's Self-Esteem: "I have nothing to be proud of"	2207	3.480	0.625
Mom's Self-Esteem: "I have a positive attitude"	2207	3.200	0.576
Mom's Self-Esteem: "I wish I had more self-respect"	2207	2.876	0.787
Mom's Self-Esteem: "I feel useless at times"	2207	2.650	0.774
Mom's Self-Esteem: "I sometimes think I am no good"	2207	3.005	0.808
Mom's Rotter Score: "I have no control"	2207	2.897	1.156
Mom's Rotter Score: "I make no plans for the future"	2207	2.543	1.159
Mom's Rotter Score: "Luck is big factor in life"	2207	3.154	0.974
Mom's Rotter Score: "Luck plays big role in my life"	2207	2.426	1.144

A10. ESTIMATED PARAMETERS FOR OUTCOME AND INVESTMENT EQUATIONS AND INITIAL CONDITIONS; NORMALIZATIONS FOR THE MEASUREMENT SYSTEM

TABLE A10-1

ESTIMATES OF THE ADULT OUTCOME EQUATIONS[a]

	Linear Anchoring Model	
	Cognitive Skill	Noncognitive Skill
Educational Attainment (Years of Schooling)[b]	1.007 (0.060)	0.993 (0.067)
Crime Participation[c]	−0.123 (0.101)	−0.463 (0.108)
Drug Consumption[d]	−0.055 (0.061)	−0.341 (0.069)
Teenage Pregnancy[e]	−0.087 (0.091)	−0.241 (0.113)

[a] Standard errors in parentheses.

[b] Completed years of schooling by 19 years of age.

[c] Ever on probation as reported in 2004.

[d] Consumption of marijuana and amphetamines.

[e] Ever pregnant by 19 years of age.

TABLE A10-2

ESTIMATES OF THE PARENTAL INVESTMENT EQUATIONS (FACTOR LOADINGS AND REGRESSION COEFFICIENTS)[a]

	Period 1 Birth	Period 2 Ages 1–2	Period 3 Ages 3–4	Period 4 Ages 5–6	Period 5 Ages 7–8	Period 6 Ages 9–10	Period 7 Ages 11–12	Period 8 Ages 13–14
Current Period Cognitive Skill	0.255	0.392	0.150	−0.007	0.030	0.070	0.036	0.062
	(0.230)	(0.091)	(0.058)	(0.043)	(0.031)	(0.030)	(0.028)	(0.026)
Current Period Noncognitive Skill	0.233	0.359	0.089	0.136	0.118	0.107	0.172	0.171
	(0.107)	(0.085)	(0.056)	(0.028)	(0.023)	(0.023)	(0.024)	(0.025)
Maternal Cognitive Skill	−0.147	0.279	0.267	0.128	0.062	0.051	0.069	0.056
	(0.190)	(0.055)	(0.043)	(0.034)	(0.026)	(0.028)	(0.029)	(0.030)
Maternal Noncognitive Skill	−0.778	0.892	0.429	−0.262	−0.838	−0.835	−0.828	−0.921
	(1.313)	(0.279)	(0.256)	(0.256)	(0.177)	(0.160)	(0.155)	(0.168)
Log Family Income	0.002	0.189	0.186	0.152	0.131	0.137	0.096	0.126
	(0.205)	(0.056)	(0.042)	(0.032)	(0.022)	(0.022)	(0.022)	(0.022)
Unobserved Heterogeneity	−2.719	0.890	1.890	−2.447	−2.522	−2.390	−2.579	−2.497
	(0.642)	(0.121)	(1.093)	(1.174)	(0.807)	(0.735)	(0.718)	(0.770)

[a]Standard errors in parentheses.

TABLE A10-3

COVARIANCE AND CORRELATION MATRICES OF INITIAL CONDITIONS ($\theta_{1,h}$)

	Child Cognitive Skill at Birth	Child Noncognitive Skill at Birth	Maternal Cognitive Skill	Maternal Noncognitive Skill	Unobserved Heterogeneity
			Covariance Matrix		
Child Cognitive Skill at Birth	0.1777				
Child Noncognitive Skill at Birth	−0.0204	0.2002			
Maternal Cognitive Skill	0.0182	0.0592	0.5781		
Maternal Noncognitive Skill	0.0050	0.0261	0.0862	0.0667	
Unobserved Heterogeneity (π_h)	0.0000	0.0000	−0.0340	−0.0211	0.0087
			Correlation Matrix		
Child Cognitive Skill at Birth	1.0000				
Child Noncognitive Skill at Birth	−0.1081	1.0000			
Maternal Cognitive Skill	0.0569	0.1741	1.0000		
Maternal Noncognitive Skill	0.0463	0.2260	0.4390	1.0000	
Unobserved Heterogeneity (π_h)	0.0000	0.0000	−0.4774	−0.8739	1.0000

TABLE A10-4

NORMALIZATIONS OF THE FACTOR LOADINGS; COEFFICIENTS ARE SET TO 1 TO SET THE SCALE OF THE FACTORS

	Child Cognitive Skills	Child Noncognitive Skills	Parental Investments	Parental Cognitive Skills	Parental Noncognitive Skills
Period 1: Year of Birth	Weight at Birth	Temperament Difficulty	Frequency Mother Reads to Child	ASVAB Arithmetic Reasoning	Self-Esteem: "I am a person of worth"
Period 2: Ages 1–2	Motor-Social Development	Temperament Difficulty	Frequency Mother Reads to Child	–	–
Period 3: Ages 3–4	Peabody Picture Vocabulary	BPI Antisocial Behavior	Frequency Mother Reads to Child	–	–
Period 4: Ages 5–6	PIAT Mathematics	BPI Antisocial Behavior	Frequency Child Goes to Musical Shows	–	–
Period 5: Ages 7–8	PIAT Mathematics	BPI Antisocial Behavior	Frequency Child Goes to Musical Shows	–	–
Period 6: Ages 9–10	PIAT Mathematics	BPI Antisocial Behavior	Frequency Child Goes to Musical Shows	–	–
Period 7: Ages 11–12	PIAT Mathematics	BPI Antisocial Behavior	Frequency Child Goes to Musical Shows	–	–
Period 8: Ages 13–14	PIAT Mathematics	BPI Antisocial Behavior	Frequency Child Goes to Musical Shows	–	–

TABLE A10-5

POINT ESTIMATES (MODEL REPORTED IN TABLE IV)

$\phi_{1,C}$	$\phi_{2,C}$	$\phi_{1,N}$	$\phi_{2,N}$
0.3135	−1.2433	−0.6100	−0.5507
Variance–Covariance Matrix of Elasticity Parameters			
$\phi_{1,C}$	$\phi_{2,C}$	$\phi_{1,N}$	$\phi_{2,N}$
0.0180	−0.0004	−0.0009	0.0010
−0.0004	0.0463	0.0001	−0.0021
−0.0009	0.0001	0.0155	0.0007
0.0010	−0.0021	0.0007	0.0288
Null Hypothesis	Test Statistic	Null Hypothesis	Test Statistic
$\phi_{1,C} = \phi_{1,N}$	4.9056	$\phi_{1,N} = \phi_{2,C}$	2.5495
$\phi_{1,C} = \phi_{2,C}$	6.0959	$\phi_{1,N} = \phi_{2,N}$	−0.2863
$\phi_{1,C} = \phi_{2,N}$	4.0806	$\phi_{2,C} = \phi_{2,N}$	−2.4607

TABLE A10-6

ESTIMATES OF THE PARENTAL INVESTMENT EQUATION[a]

Current Period Cognitive Skill	0.045 (0.028)
Current Period Noncognitive Skill	0.130 (0.027)
Maternal Cognitive Skill	0.124 (0.030)
Maternal Noncognitive Skill	0.164 (0.074)
Log Family Income	0.143 (0.026)

[a] Standard errors in parentheses.

TABLE A10-7

POINT ESTIMATES (MODEL REPORTED IN TABLE V)

$\phi_{1,C}$	$\phi_{2,C}$	$\phi_{1,N}$	$\phi_{2,N}$
0.5317	−1.3000	−0.9596	−0.4713
Variance–Covariance Matrix of Elasticity Parameters			
$\phi_{1,C}$	$\phi_{2,C}$	$\phi_{1,N}$	$\phi_{2,N}$
0.0763	−0.0093	−0.0019	0.0022
−0.0093	0.1383	−0.0030	−0.0076
−0.0019	−0.0030	0.0270	0.0014
0.0022	−0.0076	0.0014	0.0487
Null Hypothesis	Test Statistic	Null Hypothesis	Test Statistic
$\phi_{1,C} = \phi_{1,N}$	4.5561	$\phi_{1,N} = \phi_{2,C}$	0.8225
$\phi_{1,C} = \phi_{2,C}$	3.7927	$\phi_{1,N} = \phi_{2,N}$	−1.8080
$\phi_{1,C} = \phi_{2,N}$	2.8891	$\phi_{2,C} = \phi_{2,N}$	−1.8426

A11. SENSITIVITY ANALYSES FOR A ONE-STAGE MODEL

A11.1. *Sensitivity to Alternative Anchors*

Table A11-1 compares the estimated parameter values and their standard errors in an unanchored system against two anchored systems. The results from the unanchored model are in column 1 of Table A11-1. From these columns, we see that (i) both cognitive and noncognitive skills show strong persistence over time; (ii) noncognitive skills affect the accumulation of the next period's cognitive skills, but noncognitive skills do not determine cognitive skills; (iii) the estimated parental investment factor affects cognitive skills somewhat more strongly than noncognitive skills; (iv) the mother's cognitive ability affects the child's cognitive ability; (v) the mother's noncognitive ability affects both cognitive and noncognitive skills.

The estimated coefficients for the system of adult outcome equations associated with each system are reported in Table A11-2. An anchor sets the scale of the factor, as discussed in the text. The elasticities of substitution between investments and stocks of skills are both below 1, with noncognitive investments more substitutable across stages of the life cycle than cognitive investments. This finding is consistent with the evidence on plasticity of noncognitive skills and the lesser plasticity of cognitive skills discussed in Cunha, Heckman, Lochner, and Masterov (2006), Cunha and Heckman (2007), and Heckman (2008).

To circumvent the problem that scales of test score are intrinsically arbitrary, we use two different anchors. The first anchor is years of education completed by age 19 which is assumed to be a linear function of the factors (results reported in column 2 of Table A11-1). The second anchor is a probit model for high school graduation by age 19 (column 3). Compared to the unanchored case, the estimated elasticity of substitution for cognitive skills slightly increases. The elasticity of substitution for noncognitive skills considerably decreases. However, both estimates are still below 1 ($\phi_C \cong -1.03$, $\phi_N \cong -0.97$). When we use nonlinear anchoring, the estimated elasticities of substitution are different ($\phi_C \cong -1.21$, $\phi_N \cong -0.79$), but the standard errors are also larger. The qualitative conclusions of Table A11-1 survive.

A11.2. *Relaxing the Normality Assumption*

Thus far, we have assumed normality for the unobservables. We investigate how our findings change when we allow for a more flexible representation of the joint distribution of factors using a mixtures of normal model.[14] However,

[14]We constrain variances to be the same across different elements of the mixture. The data are not rich enough to afford identification of different variance–covariance matrices by mixture components. When we allow for different variance–covariance matrices, the model does not converge. When we penalize the likelihood to eliminate small values of the probability weights on the mixture components, the estimated variance–covariance matrices turns out to be very similar.

TABLE A11-1

COMPARISON OF DIFFERENT ANCHORING ALTERNATIVES; ONE DEVELOPMENT STAGE MODEL; NO UNOBSERVED HETEROGENEITY (π); FACTORS NORMALLY DISTRIBUTED[a]

		(1) Unanchored	(2) Linear Anchor (Years of Schooling)	(3) Nonlinear Anchor (Probability of High School Graduation)
The Technology of Cognitive Skill Formation				
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.693 (0.014)	0.727 (0.012)	0.778 (0.016)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.029 (0.004)	0.018 (0.004)	0.021 (0.007)
Current Period Investments	$\gamma_{1,C,3}$	0.238 (0.024)	0.222 (0.022)	0.179 (0.022)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.023 (0.006)	0.028 (0.005)	0.019 (0.006)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.018 (0.007)	0.005 (0.007)	0.003 (0.009)
Complementarity Parameter	$\phi_{1,C}$	−1.173 (0.004)	−1.032 (0.002)	−1.214 (0.048)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	0.460	0.492	0.452
Variance of Shocks $\eta_{C,t}$	δ_η^2	0.103 (0.002)	0.105 (0.002)	0.067 (0.001)
The Technology of Noncognitive Skill Formation				
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000 (0.008)	0.000 (0.009)	0.000 (0.014)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.802 (0.019)	0.804 (0.020)	0.841 (0.053)
Current Period Investments	$\gamma_{1,N,3}$	0.144 (0.019)	0.156 (0.021)	0.113 (0.021)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.000 (0.004)	0.000 (0.004)	0.000 (0.004)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.054 (0.012)	0.040 (0.013)	0.046 (0.019)
Complementarity Parameter	$\phi_{1,N}$	−0.462 (0.101)	−0.972 (0.101)	−0.788 (0.154)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.684	0.507	0.559
Variance of Shocks $\eta_{N,t}$	δ_η^2	0.114 (0.002)	0.122 (0.003)	0.058 (0.001)

[a]Standard errors in parentheses.

TABLE A11-2
ESTIMATES OF THE ADULT OUTCOME EQUATIONS[a]

	Unanchored Model		Linear Andchoring Model (Years of Schooling)		Nonlinear Anchoring (Probability of High School Graduation)	
	Cognitive Skill	Noncognitive Skill	Cognitive Skill	Noncognitive Skill	Cognitive Skill	Noncognitive Skill
Educational Attainment (Years of Schooling)[b]	1.239	0.927	1.008	1.02		
	(0.068)	(0.064)	(0.021)	(0.023)		
High School Graduation[c]					1.014	1.007
					(0.132)	(0.141)
Crime Participation[d]	−0.142	−0.443	−0.126	−0.458	−0.153	−0.476
	(0.102)	(0.106)	(0.115)	(0.113)	(0.114)	(0.134)
Drug Consumption[e]	−0.054	−0.321	−0.053	−0.337	−0.064	−0.371
	(0.063)	(0.069)	(0.071)	(0.076)	(0.071)	(0.088)
Teenage Pregnancy[f]	−0.091	−0.237	−0.086	−0.234	−0.108	−0.251
	(0.099)	(0.107)	(0.116)	(0.110)	(0.110)	(0.122)

[a] Standard errors in parentheses.
[b] Completed years of schooling by 19 years of age.
[c] High school graduation status by 2004.
[d] Ever on probation as reported in 2004.
[e] Consumption of marijuana and amphetamines.
[f] Ever pregnant by 19 years of age.

the added flexibility does not affect the estimates obtained from a model estimated under a more restrictive normality assumption. See the estimates reported in Table A11-3.

A12. VARIANCE DECOMPOSITION

A12.1. *Cognitive and Noncognitive Skills*

We model educational attainment (years of schooling), $Z_{4,1}$ in the notation of Section 3.5, as a linear, separable function of observable characteristics X, cognitive skills, $\ln\theta_{C,T}$, noncognitive skills, $\ln\theta_{N,T}$, unobserved heterogeneity, $\ln\pi$, and an error term $\varepsilon_{4,1}$ as

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

where we normalize the loading on $\ln\pi$, $\alpha_{4,\pi,1} = 1$ for identification purposes. Note that the total residual variance is given by the term

$$\begin{aligned} R_{4,1} &= \alpha^2_{4,C,1}\operatorname{Var}(\ln\theta_{C,T+1}) + 2\alpha_{4,C,1}\alpha_{4,N,1}\operatorname{Cov}(\ln\theta_{C,T+1}, \ln\theta_{N,T+1}) \\ &\quad + 2\alpha_{4,C,1}\operatorname{Cov}(\ln\theta_{C,T+1}, \ln\pi) + \alpha^2_{4,N,1}\operatorname{Var}(\ln\theta_{N,T+1}) \\ &\quad + 2\alpha_{4,N,1}\operatorname{Cov}(\ln\theta_{N,T+1}, \ln\pi) + \operatorname{Var}(\ln\pi) + \operatorname{Var}(\varepsilon_{4,1}). \end{aligned}$$

We compute the share of $R_{4,1}$ that is due to cognitive and noncognitive skills, $p_{4,1}$, as

$$\begin{aligned} p_{4,1} &= [\alpha^2_{4,C,1}\operatorname{Var}(\ln\theta_{C,T+1}) + 2\alpha_{4,C,1}\alpha_{4,N,1}\operatorname{Cov}(\ln\theta_{C,T+1}, \ln\theta_{N,T+1}) \\ &\quad + \alpha^2_{4,N,1}\operatorname{Var}(\ln\theta_{N,T+1})]/R_{4,1} \\ &\approx 34\%. \end{aligned}$$

We compute the share of $R_{4,1}$ that is due exclusively to skill k, $s_{4,k,1}$ as

$$s_{4,k,1} = \frac{\alpha^2_{4,k,1}\operatorname{Var}(\ln\theta_{C,T+1})}{R_{4,1}}, \quad k \in \{C, N\}.$$

We find that $s_{4,C,1} \approx 16\%$ and $s_{4,N,1} \approx 12\%$.

A12.2. *Parental Investment*

To compute how much of total residual variance is due to parental investment, we proceed in two ways. First, we generate filtered estimates of parental investments for each child i, $i = 1, \dots, N$. Let $\ln\hat{I}_{i,t}$ denote the filtered investment for child i at period t. We then proceed by applying the Frisch–Waugh–Lovell theorem. First, we regress

$$Z_{4,1} = X\beta_{4,1} + \nu_{Z,4,1}$$

TABLE A11-3

ALLOWING FOR UNOBSERVED HETEROGENEITY AND FACTORS JOINT DISTRIBUTION IS MIXTURE OF NORMALS; ONE DEVELOPMENT STAGE MODEL; LINEAR ANCHORING ON EDUCATIONAL ATTAINMENT (YEARS OF SCHOOLING)[a]

		(1) No Unobserved Heterogeneity and Normal Distribution	(2) Unobserved Heterogeneity and Normal Distribution	(3) Unobserved Heterogeneity and Mixture of Normals Distribution
The Technology of Cognitive Skill Formation				
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.727 (0.012)	0.689 (0.014)	0.693 (0.016)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.018 (0.004)	0.003 (0.005)	0.004 (0.007)
Current Period Investments	$\gamma_{1,C,3}$	0.222 (0.022)	0.275 (0.022)	0.267 (0.022)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.028 (0.005)	0.030 (0.005)	0.032 (0.007)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.005 (0.007)	0.003 (0.007)	0.004 (0.008)
Complementarity Parameter	$\phi_{1,C}$	−1.032 (0.002)	−1.012 (0.002)	−0.987 (0.004)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	0.492	0.497	0.503
Variance of Shocks $\eta_{C,t}$	δ_η^2	0.105 (0.002)	0.097 (0.002)	0.093 (0.002)
The Technology of Noncognitive Skill Formation				
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000 (0.009)	0.000 (0.010)	0.001 (0.012)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.804 (0.020)	0.787 (0.022)	0.794 (0.025)
Current Period Investments	$\gamma_{1,N,3}$	0.156 (0.021)	0.171 (0.024)	0.163 (0.024)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.000 (0.004)	0.000 (0.004)	0.001 (0.006)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.040 (0.013)	0.042 (0.014)	0.041 (0.014)
Complementarity Parameter	$\phi_{1,N}$	−0.972 (0.101)	−0.751 (0.102)	−0.768 (0.123)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.507	0.571	0.566
Variance of Shocks $\eta_{N,t}$	δ_η^2	0.122 (0.003)	0.116 (0.003)	0.131 (0.004)

[a]Standard errors in parentheses.

and obtain the residual

$$\hat{\nu}_{Z,4,1} = Z_{4,1} - X\hat{\beta}_{4,1}.$$

Then, for each period t we regress

$$\ln \hat{I}_t = X\kappa_t + \nu_{I,t},$$

where κ_t is the coefficient on the regressors, and obtain the residual

$$\hat{\nu}_{I,t} = \ln \hat{I}_t - X\hat{\zeta}_t.$$

Finally, we regress

$$\hat{\nu}_{Z,4,1} = m(\hat{\nu}_{I,1}, \hat{\nu}_{I,2,}, \dots, \hat{\nu}_{I,T}) + \varepsilon_{4,1},$$

where the function m is a complete polynomial of degree 2. We find that the R^2 of such regression is around 15%.

A13. FURTHER SIMULATIONS OF THE MODEL

Suppose that the goal of society is to get the schooling of every child to the same 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

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

and the initial endowments of the child and her parents. We have estimated all of the ingredient functions.[15] We use the estimates reported in Table V.

[15]See Appendix A10 for the estimates of the schooling equation.

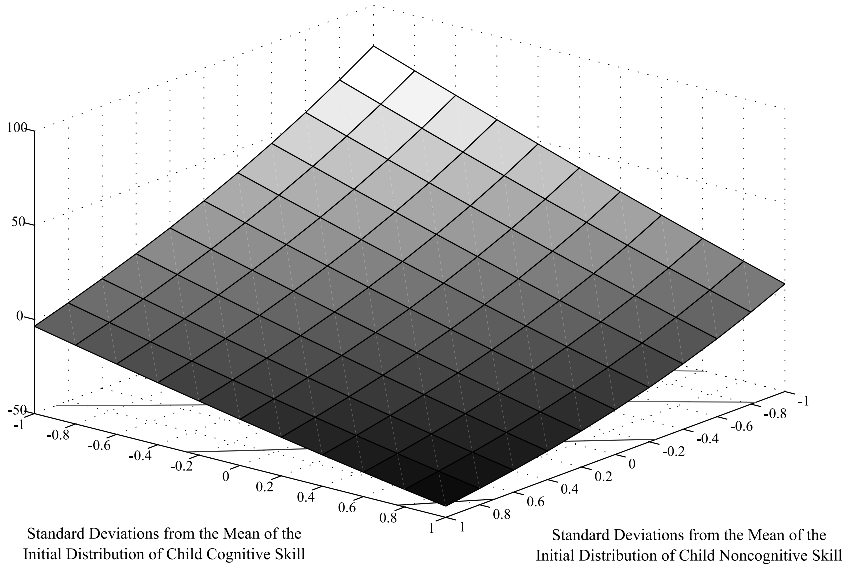

FIGURE A13-1.—Percentage increase in total investments as a function of child initial conditions of cognitive and noncognitive skills.

Figures A13-1 (for child endowments) and A13-2 (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.[16] The shading in the graphs represents different values of investments. The lightly shaded areas of the graph correspond to higher values. Eighty percent more investment is required for children with the most disadvantaged personal endowments (Figure A13-1). The corresponding figure for children with the most disadvantaged parental endowments is 95% (Figure A13-2). The negative percentages for children with high initial endowments is a measure of their advantage. From the analysis of Moon (2009), 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.

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

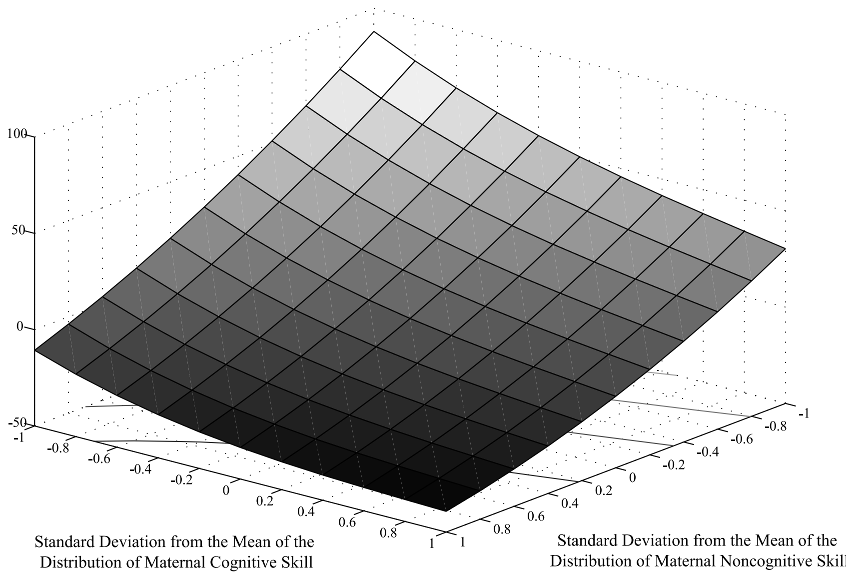

FIGURE A13-2.—Percentage increase in total investments as a function of maternal cognitive and noncognitive skills.

A14. ANALYSIS OF ONE-SKILL (COGNITIVE) MODELS

A14.1. *Implications of Estimates From a Two-Stage–One-Skill Model of Investment*

We examine the policy implications of a traditional model formulated only in terms of cognitive skills. Our estimates of this model are reported in Table A14-1. We consider the problem of maximizing aggregate educational attainment using the estimates from a model with only cognitive skills. Figures A14-1 and A14-2 compare optimal early investments from the cognitive-skill-only model (left) with investments from the model with both skills (right). As in the figures reported in Section 4.3, less shaded regions of the figures correspond to higher values for investment.

An empirical model of skill formation that focuses solely on cognitive skills suggests that it is optimal to perpetuate and reinforce initial inequality. In Section 4.3, we established that even in a one-skill model it is *theoretically* possible to obtain the result that is optimal to invest more in the early years of the initially disadvantaged. However, the empirical estimates of the one-skill model suggest the opposite. In contrast to the empirical implications of the two-skill model reported in the text, investments are *lower* at the first stage of the life cycle for those initially most disadvantaged (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 mal-

TABLE A14-1

TECHNOLOGY OF COGNITIVE SKILL FORMATION; MODEL WITH COGNITIVE SKILLS ONLY AND TWO STAGES OF CHILDHOOD; ESTIMATED ALONG WITH INVESTMENT EQUATION WITH LINEAR ANCHORING ON EDUCATIONAL ATTAINMENT (YEARS OF SCHOOLING); ALLOWING FOR UNOBSERVED HETEROGENEITY (π); FACTORS NORMALLY DISTRIBUTED[a]

Current Period Cognitive Skills	$\gamma_{1,C,1}$	0.303 (0.026)	$\gamma_{2,C,1}$	0.448 (0.015)
Current Period Investments	$\gamma_{1,C,3}$	0.319 (0.025)	$\gamma_{2,C,3}$	0.098 (0.015)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.378 (0.022)	$\gamma_{2,C,4}$	0.454 (0.017)
Complementarity Parameter	$\phi_{1,C}$	−0.180 (0.130)	$\phi_{2,C}$	−0.781 (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_η^2	0.193 (0.006)	δ_η^2	0.050 (0.002)

[a]Standard errors in parentheses.

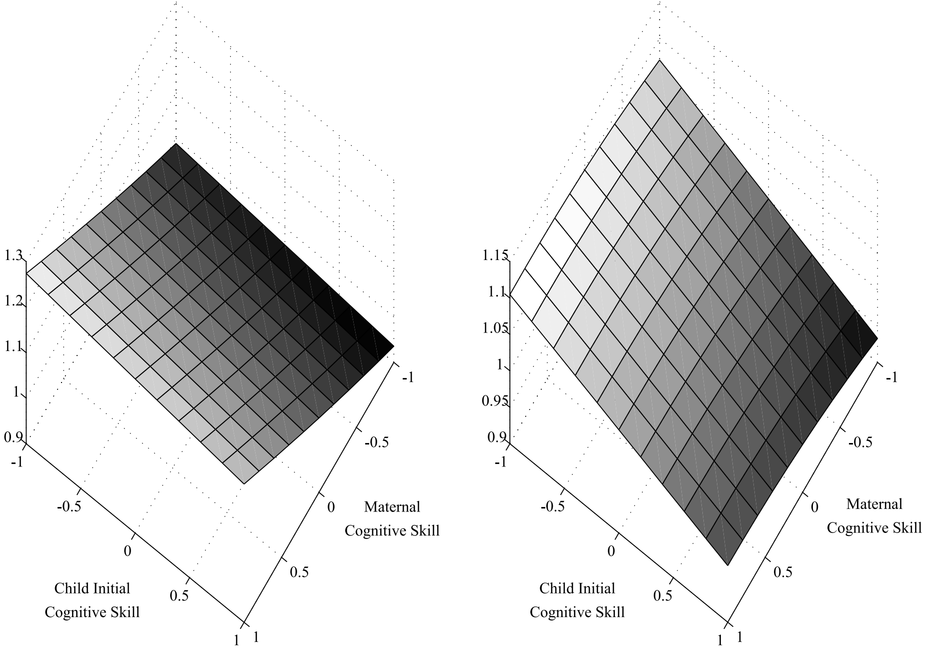

FIGURE A14-1.—Optimal early investments by child initial cognitive skills and maternal cognitive skills model with cognitive skill only (left) and the model with cognitive and noncognitive skills (right).

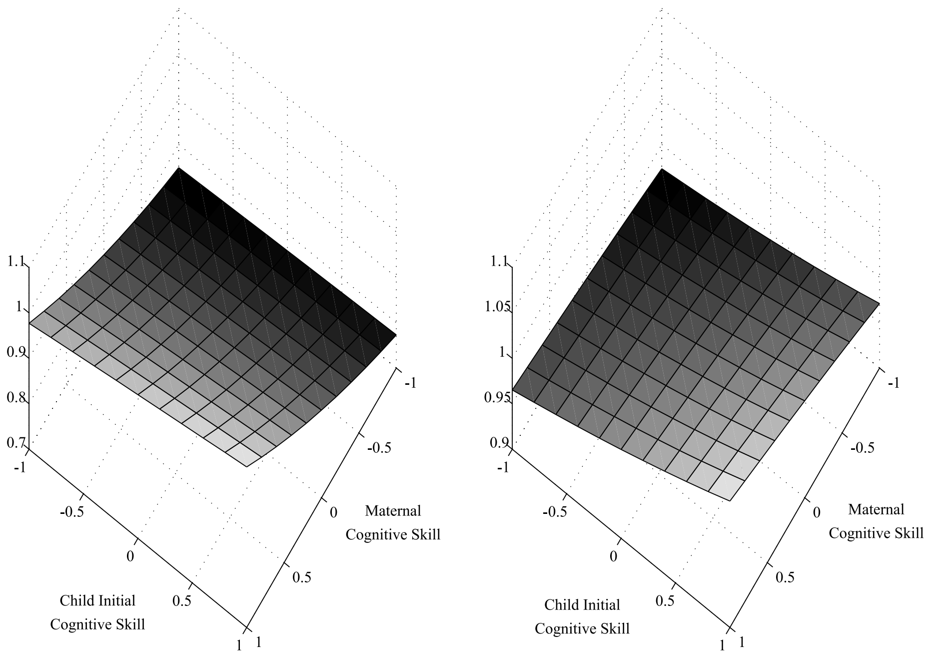

FIGURE A14-2.—Optimal late investments by child initial cognitive skills and maternal cognitive skills model with cognitive skill only (left) and the model with cognitive and noncognitive skills (right).

leability of noncognitive skills in the second stage of childhood. By ignoring a central feature of the human skill formation process, a one-skill model produces misleading public policy conclusions.

A14.2. *The Implications of Estimates From a Single-Stage–One-Skill (Cognitive) Model of Investment*

This appendix discusses implications of the estimates of the technology reported in Table A14-2 that assumes a single cognitive skill and a single stage. The model is comparable to the one whose estimates are reported in Table A14-1, except that model has two stages while the model used to generate Table A14-2 has only a single stage of the life cycle. The estimated elasticity of substitution is lower than the first stage estimate reported in Table A14-1. Estimated investment effects are comparable for the first stage of Table A14-1 and the corresponding parameter in Table A14-2. Estimated parental skill estimates are weaker.

This specification shows an even more dramatic efficiency–equity trade-off for maximizing aggregate schooling than the specification discussed in Section A14.1. See Figures A14-3 and A14-4. The left-hand side figures are based

TABLE A14-2

TECHNOLOGY OF COGNITIVE SKILL FORMATION; MODEL WITH COGNITIVE SKILLS ONLY AND ONE STAGE OF CHILDHOOD; LINEAR ANCHORING ON EDUCATIONAL ATTAINMENT (YEARS OF SCHOOLING); ALLOWING FOR UNOBSERVED HETEROGENEITY (π) AND INVESTMENT; FACTORS NORMALLY DISTRIBUTED[a]

Current Period Cognitive Skills	$\gamma_{1,C,1}$	0.626 (0.013)
Current Period Investments	$\gamma_{1,C,3}$	0.293 (0.022)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.081 (0.009)
Complementarity Parameter	$\phi_{1,C}$	−0.701 (0.070)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	0.588
Variance of Shocks $\eta_{C,t}$	δ_η^2	0.081 (0.002)

[a]Standard errors in parentheses.

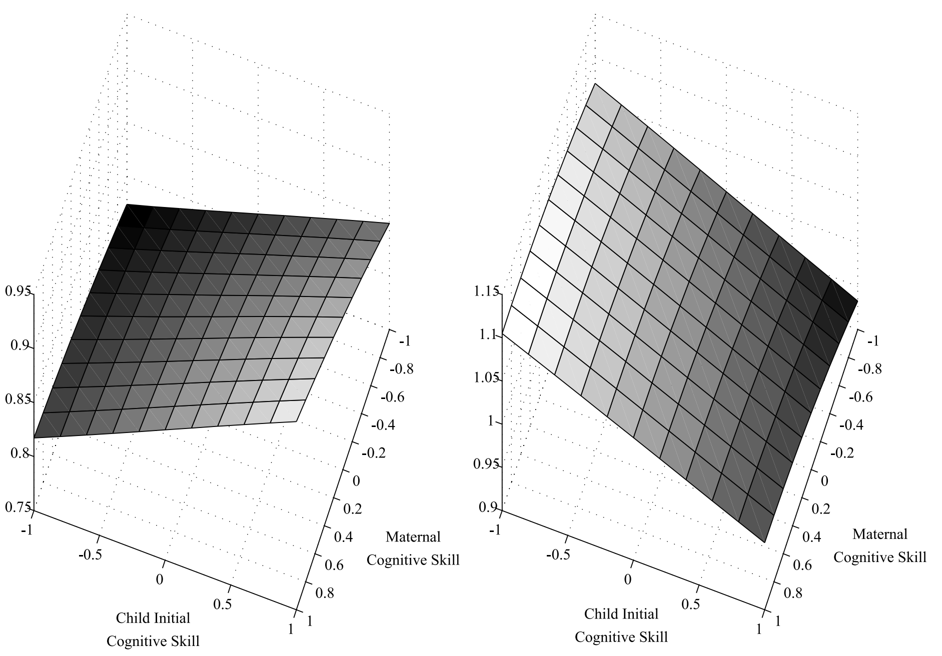

FIGURE A14-3.—Optimal early investments by child initial cognitive skills and maternal cognitive skills model with cognitive skill only (right) and the model with cognitive and noncognitive skills (right).

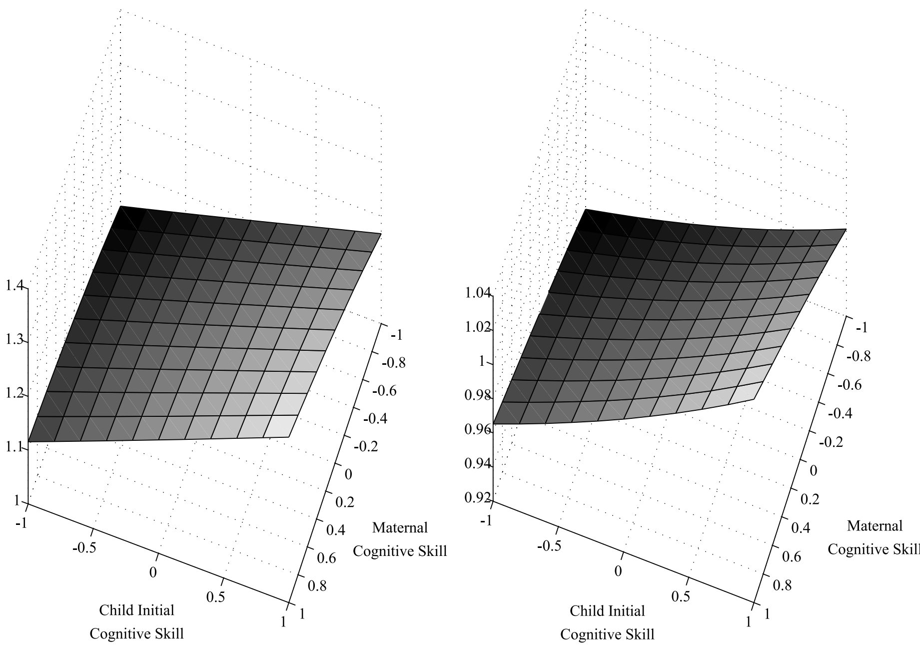

FIGURE A14-4.—Optimal early investments by child initial cognitive skills and maternal cognitive skills model with cognitive skill only (right) and the model with cognitive and noncognitive skills (right).

on the model of Table A14-2. The right-hand side figures are based on the model with estimates reported in Table V. Ignoring noncognitive skills and multiple stages, it is socially optimal to *perpetuate* rather than remediate initial disadvantage, especially in the early stage of the child life cycle.

REFERENCES

ARULAMPALAM, S., B. RISTIC, AND N. GORDON (2004): *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. New York: Artech House Radar Library. [16]

BECKER, G. S., AND N. TOMES (1986): "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4, S1–S39. [2]

BORGHANS, L., A. L. DUCKWORTH, J. J. HECKMAN, AND B. TER WEEL (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43, 972–1059. [7]

BRADLEY, R. H., AND B. M. CALDWELL (1980): "The Relation of Home Environment, Cognitive Competence, and IQ Among Males and Females," *Child Development*, 51, 1140–1148. [24]

——— (1984): "The Relation of Infants' Home Environments to Achievement Test Performance in First Grade: A Follow-Up Study," *Child Development*, 55, 803–809. [24]

CAMPOS, J. J., K. BARRETT, M. E. LAMB, H. H. GOLDSMITH, AND C. STENBERG (1983): "Socioemotional Development," in *Handbook of Child Psychology*, Vol. 2, ed. by M. M. Haith and J. J. Campos. New York: Wiley, 783–915. [27]

CARNEIRO, P., F. CUNHA, AND J. J. HECKMAN (2003): "Interpreting the Evidence of Family Influence on Child Development," in *Proceedings of Conference of the Economics of Early Childhood Development: Lessons for Economic Policy*, Federal Reserve Bank of Minneapolis, Minneapolis, MN. [2,7]

CUNHA, F. (2008): "Two Notes on the Unscented Kalman Filter," Working Paper, University of Pennsylvania. [16]

CUNHA, F., AND J. J. HECKMAN (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47. [2,4,5,37]

——— (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782. [9,15]

——— (2009): "The Economics and Psychology of Inequality and Human Development," *Journal of the European Economics Association*, 7, 320–364. [3]

CUNHA, F., J. J. HECKMAN, L. J. LOCHNER, AND D. V. MASTEROV (2006): "Interpreting the Evidence on Life Cycle Skill Formation," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek and F. Welch. Amsterdam: North-Holland, Chapter 12, 697–812. [2-5,7,37]

DOUCET, A., N. DE FREITAS, AND N. GORDON (2001): *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag. [17]

FERNANDEZ-VILLAVERDE, J., AND J. F. RUBIO-RAMIREZ (2006): "Estimating Macroeconomic Models: A Likelihood Approach," Technical Working Paper 321, NBER. [17]

HECKMAN, J. J. (2007): "The Economics, Technology and Neuroscience of Human Capability Formation," *Proceedings of the National Academy of Sciences*, 104, 13250–13255. [2,4]

——— (2008): "Schools, Skills and Synapses," *Economic Inquiry*, 46, 289–324. [4,37]

HECKMAN, J. J., AND Y. RUBINSTEIN (2001): "The Importance of Noncognitive Skills: Lessons From the GED Testing Program," *American Economic Review*, 91, 145–149. [5]

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482. [5]

HU, Y., AND S. M. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216. [10]

JUDD, K. L. (1998): *Numerical Methods in Economics*. Cambridge, MA: MIT Press. [19]

JULIER, S. J., AND J. K. UHLMANN (1997): "A New Extension of the Kalman Filter to Nonlinear Systems," in *Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls*. [17,19]

KAGAN, J. (1981): *The Second Year: The Emergence of Self Awareness*. Cambridge, MA: Harvard University Press. [25-27]

LINVER, M. R., J. BROOKS-GUNN, AND N. CABRERA (2004): "The Home Observation for Measurement of the Environment (HOME) Inventory: The Derivation of Conceptually Designed Subscales," *Parenting: Science & Practice*, 4, 99–114. [24]

MASKELL, S. (2004): "An Introduction to Particle Filters," in *State Space and Unobserved Component Models: Theory and Applications*, ed. by A. Harvey, S. J. Koopman, and N. Shephard. Cambridge, U.K.: Cambridge University Press. [17]

MOON, S. H. (2009): "Multi-Dimensional Human Skill Formation With Multi-Dimensional Parental Investment," Unpublished Manuscript, University of Chicago, Department of Economics. [43]

MOTT, F. L., P. C. BAKER, D. E. BALL, C. K. KECK, AND S. M. LENHART (1998): *The NLSY Children 1992: Description and Evaluation*. Columbus, OH: The Ohio State University, Center for Human Resource Research. [26]

PETERSON, J. L., AND K. A. MOORE (1987): "Motor and Social Development in Infancy: Some Results From a National Survey," Report 87-03, Child Trends, Inc., Washington, DC. [25]

PETERSON, J. L., AND N. ZILL (1986): "Marital Disruption, Parent–Child Relationships, and Behavior Problems in Children," *Journal of Marriage and the Family*, 48, 295–307. [27]

POE, G. S. (1986): "Design and Procedures for the 1981 Child Health Supplement to the National Health Interview Survey," Working Paper, National Center for Health Statistics, Hyattsville, MD. [25]

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. [24,26]

Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19102, U.S.A.; fcunha@sas.upenn.edu,

Dept. of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, U.S.A. and Geary Institute, University College Dublin, Belfield, Dublin 4, Ireland and American Bar Foundation, Chicago, IL 60611, U.S.A.; jjh@uchicago.edu,

and

Dept. of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, U.S.A.; smschenn@uchicago.edu.

Manuscript received June, 2006; final revision received November, 2009.