

# Estimating the Technology of Cognitive and Noncognitive Skill Formation

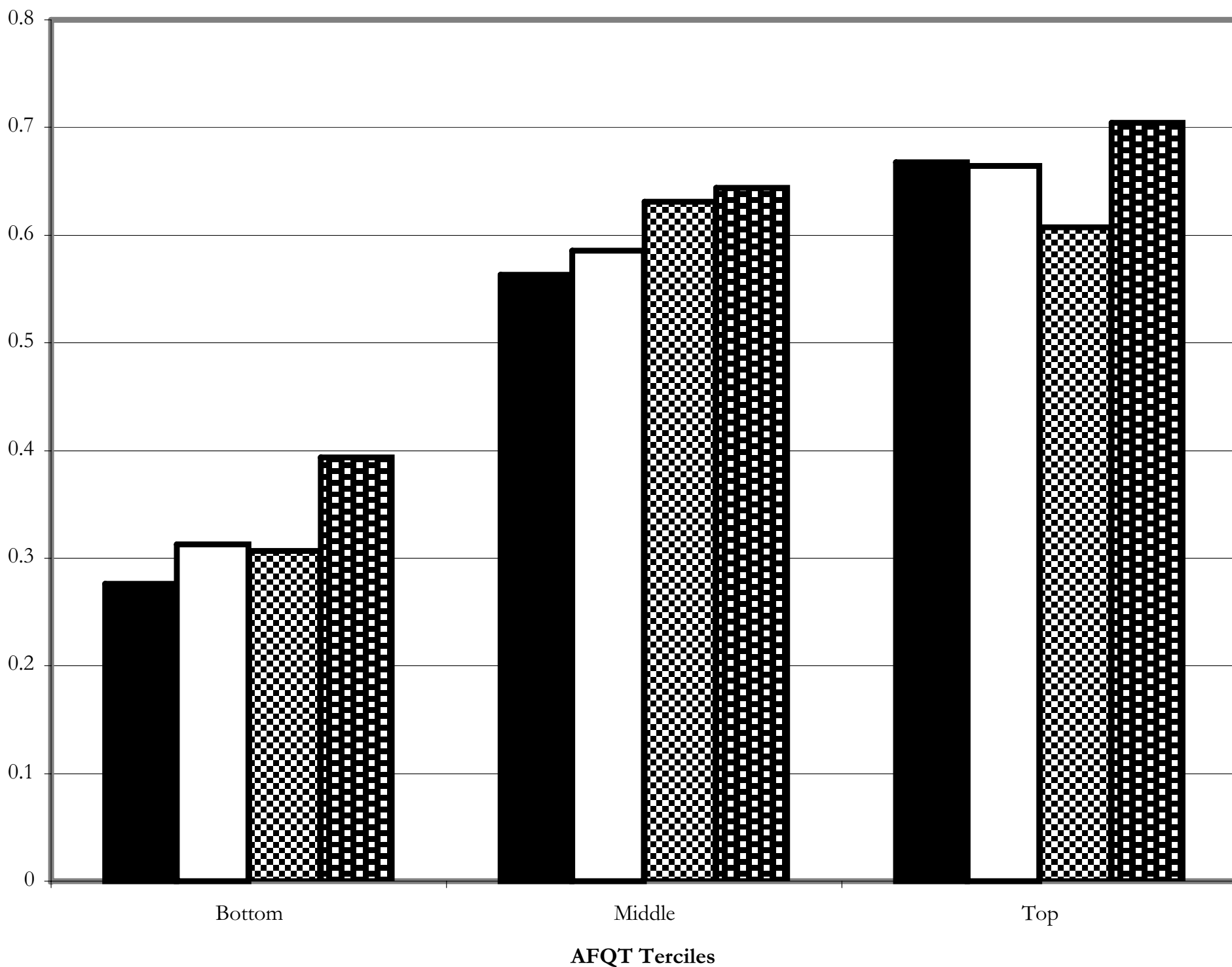
Flávio Cunha, James Heckman, and Susanne Schennach  
University of Pennsylvania and University of Chicago

The Spencer Foundation & Institute of New Economic Thinking Conference  
Series on Individual Differences and Economic Behavior  
University of Chicago

December 10, 2010

# Our Work

- In previous work (Cunha and Heckman, 2007) we proposed simple models that organized the evidence on the development of skills in children.
- We specified technologies for the production of skills that featured:
  1. Self-Productivity: Early investments produce early skills as well as late skills.
  2. Complementarity: Early investments complement late investments.
- This paper is part of a research agenda on estimating and identifying these elements.



Lowest Income Quartile
  Second Income Quartile
  Third Income Quartile
  Highest Income Quartile

**Source: Carneiro and Heckman (2003).**

# This Paper

- We establish conditions to nonparametrically identify the technology even though:
  1. The data set contains lots of measurement error.
  2. Test scores do not have a natural metric.
  3. Investments are chosen by parents who have more information than econometricians.
- We show how estimation can be implemented by nonlinear, non-Gaussian filtering techniques.

## Findings Based on Estimates of the Technologies:

- Early investments should be higher for children from disadvantaged backgrounds.
- The optimal ratio of early to late investments depends on the outcome of interest.
- If early remediation does not occur at early stages of childhood, then remediation at later stages should focus primarily on producing noncognitive skills.

# Outline of the Talk

- The Technology of Skill Formation
- Data (CNLSY/79)
- Intuitive Discussion of Identification
- Empirical Results
- Conclude

# Technology of Skill Formation

- $\theta_{c,t}$  denotes cognitive skills of the child at age  $t$ .
- $\theta_{n,t}$  denotes non-cognitive skills of the child at age  $t$ .
- $x_{k,t}$  is parental investment in skill  $k$  when child is  $t$  years-old.
- $\theta_{c,p}$  represents parental cognitive skills.
- $\theta_{n,p}$  represents parental noncognitive skills.
- $\eta_{k,t}$  are shocks and/or unmeasurable inputs.
- There are  $S$  different developmental stages:  $s = 1, \dots, S$ .
- The technology for skill  $k$ , at period  $t$  and stage  $s$  is:

$$\theta_{k,t+1} = f_{k,s}(\theta_{c,t}, \theta_{n,t}, x_{k,t}, \theta_{c,p}, \theta_{n,p}, \eta_{k,t})$$

# Identification of the Technology of Skill Formation

Consider the estimation of the function:

$$\theta_{k,t+1} = f_{k,s}(\theta_{c,t}, \theta_{n,t}, x_{k,t}, \theta_{c,p}, \theta_{n,p}, \eta_{k,t})$$

We have to deal with three problems:

1. We don't observe  $(\theta_{c,t}, \theta_{n,t}, x_{c,t}, x_{n,t})$  directly.
2. We don't know which scale to use to measure  $\theta_{c,t}, \theta_{n,t}$ .
3. Investments  $x_{c,t}$  and  $x_{n,t}$  are chosen by parents based on information from  $(\eta_{c,t}, \eta_{n,t})$  that is unobserved by the econometrician.



## Data: Sample

- 2207 firstborn white children (birth to age 14) from CNLSY/79.
- Large number of observations: Almost 400,000.
- Large number of parameters: Almost 1250.
- Extensive data collection on parental characteristics and child cognitive and noncognitive development.
- The data collection is every two years.
- There are eight periods: (birth, ages 1-2, ages 3-4, ..., ages 13-14).

# Data: Measurements of Skills and Parental Investments

- Child's Cognitive Skills
  - Bayley, Parts of Body, Memory for Locations, PPVT, PIAT.
  - On average, 3 cognitive measurements per period.
- Child's Noncognitive Skills
  - Temperament and Behavior Problem Index.
  - On average, 5 noncognitive measurements per period.
- Parental Investments
  - Components of the Home Score.
  - On average, 10 investment measures per period.
- Parental Cognitive Skills
  - Components of the ASVAB Tests.
- Parental Noncognitive Skills
  - Mom's Self-Esteem and Mom's Locus of Control

# Measurement Error: Slide 1

- To focus on the important ideas, suppose that we want to estimate a linear production function:

$$\theta_{k,t+1} = \beta x_{k,t} + \eta_{k,t}$$

- If
  1. we know the scale of  $\theta_{k,t+1}$
  2. investments are exogenous,  $E(\eta_{k,t} | x_{k,t}) = 0$
- Then, a consistent estimator for  $\beta$  is:

$$\hat{\beta} = \frac{\text{Cov}(\theta_{k,t+1}, x_{k,t})}{\text{Var}(x_{k,t})}$$

- If we observe  $(\theta_{k,t+1}, x_{k,t})$ , that is it.

## Measurement Error: Slide 2

- We don't observe  $(\theta_{k,t+1}, x_{k,t})$
- We observe proxy variables for  $\theta_{k,t+1}$  and  $x_{k,t}$ :

$$y_{1,k,t+1} = \theta_{k,t+1} + \varepsilon_{1,k,t+1}$$

$$y_{2,k,t+1} = \theta_{k,t+1} + \varepsilon_{2,k,t+1}$$

$$y_{3,k,t} = x_{k,t} + \varepsilon_{3,k,t+1}$$

$$y_{4,k,t} = x_{k,t} + \varepsilon_{4,k,t+1}$$

- The vector  $\varepsilon_{j,k,t+1}$  is measurement error. Assume they are uncorrelated.
- Then, we can estimate  $\beta$  by:

$$\hat{\beta} = \frac{\text{Cov}(\theta_{k,t+1}, x_{k,t})}{\text{Var}(x_{k,t})} = \frac{\text{Cov}(y_{1,k,t+1}, y_{3,k,t})}{\text{Cov}(y_{3,k,t}, y_{4,k,t})}$$

- Key to identification: covariance restrictions (some are testable).

## Measurement Error: Slide 3

- If we have at least two measurements per period, we can identify factor loading  $\alpha_{l,k,t}$  for  $l = 2, \dots, L$ :

$$y_{1,k,t} = \theta_{k,t} + \varepsilon_{1,k,t}$$

$$y_{l,k,t} = \alpha_{l,k,t} \theta_{k,t} + \varepsilon_{2,k,t}$$

- In fact, we estimate factor loadings  $\alpha_{l,k,t}$ .
- If we have at least three measurements per period, we can allow some of the measurement error components to be correlated (see Cunha and Heckman, 2008):

$$\text{Cov}(\varepsilon_{l,k,t}, \varepsilon_{l',k',t'}) \neq 0 \text{ for } l, l' = 2, \dots, L \text{ and } \forall k, k', t, t'$$

- We can also identify nonlinear measurement equations:

$$y_{1,k,t} = \theta_{k,t+1} + \varepsilon_{1,k,t+1}$$

$$y_{l,k,t} = h_{l,k,t}(\theta_{k,t}, \varepsilon_{l,k,t})$$

# Estimation of the Technology of Skill Formation

- When
  1. The transition and measurement equations are linear
  2. The factors and measurement errors are normally distributed
- Then, we use the Kalman Filter to compute the likelihood (Cunha and Heckman, 2008).
- Kalman filter breaks down because of the nonlinearity.
- Two common options are:
  1. Extended Kalman Filter (EKF): Linearize  $f(\theta)$  around  $f(E\theta)$ .
  2. Particle Filter: Sequential Monte Carlo Method.
- We use the Mixture of Normals Unscented Kalman Filter

# Parametric Specification of the Technology

We choose a CES representation for the technology for the formation of skills:

$$\theta_{c,t+1} = \left[ \gamma_{s,c,1} \theta_{c,t}^{\phi_{s,c}} + \gamma_{s,c,2} \theta_{n,t}^{\phi_{s,c}} + \gamma_{s,c,3} x_{c,t}^{\phi_{s,c}} + \dots \right]^{\frac{1}{\phi_{s,c}}} e^{\eta_{c,t+1}}$$

$$\theta_{n,t+1} = \left[ \gamma_{s,n,1} \theta_{c,t}^{\phi_{s,n}} + \gamma_{s,n,2} \theta_{n,t}^{\phi_{s,n}} + \gamma_{s,n,3} x_{c,t}^{\phi_{s,n}} + \dots \right]^{\frac{1}{\phi_{s,n}}} e^{\eta_{n,t+1}}$$

## On Measurement Error

- Is measurement error a problem here?
- Suppose that factors and measurement errors are independent. Then:

$$\text{Var}(y_{1,k,t+1}) = \text{Var}(\theta_{k,t}) + \text{Var}(\varepsilon_{1,k,t+1})$$

- Let's compare:

$$\begin{aligned} \text{signal} &= \frac{\text{Var}(\theta_{k,t})}{\text{Var}(\theta_{k,t}) + \text{Var}(\varepsilon_{1,k,t+1})} \\ \text{noise} &= \frac{\text{Var}(\varepsilon_{1,k,t+1})}{\text{Var}(\theta_{k,t}) + \text{Var}(\varepsilon_{1,k,t+1})} \end{aligned}$$



**Figure 3**  
**Share of Residual Variance in Measurements of Cognitive Skills**  
**Due to the Variance of Cognitive Factor (Signal)**  
**and Due to the Variance of Measurement Error (Noise)**

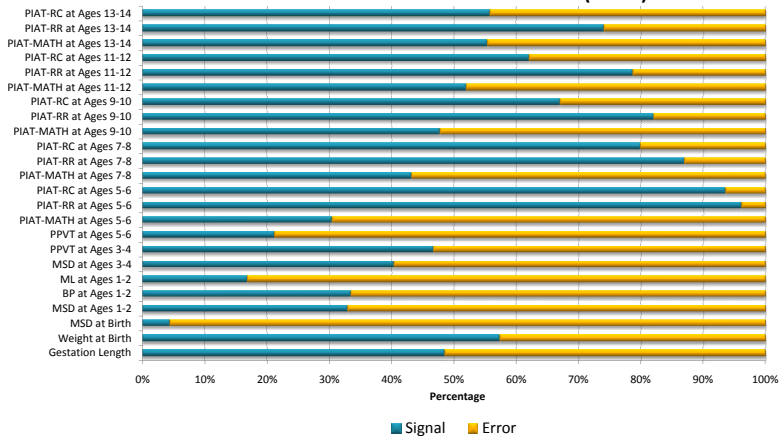
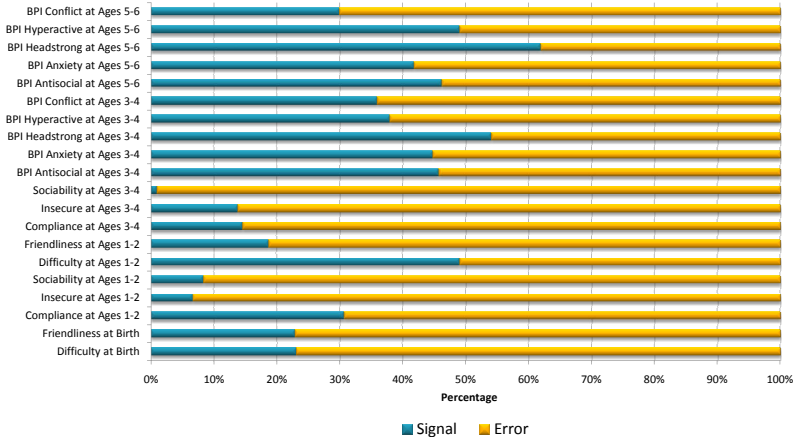


Figure 4A

**Share of Residual Variance in Measurements of Noncognitive Skills  
Due to the Variance of Noncognitive Factor (Signal)  
and Due to the Variance of Measurement Error (Noise)**



**Figure 4B**  
**Share of Residual Variance in Measurements of Noncognitive Skills**  
**Due to the Variance of Noncognitive Factor (Signal)**  
**and Due to the Variance of Measurement Error (Noise)**

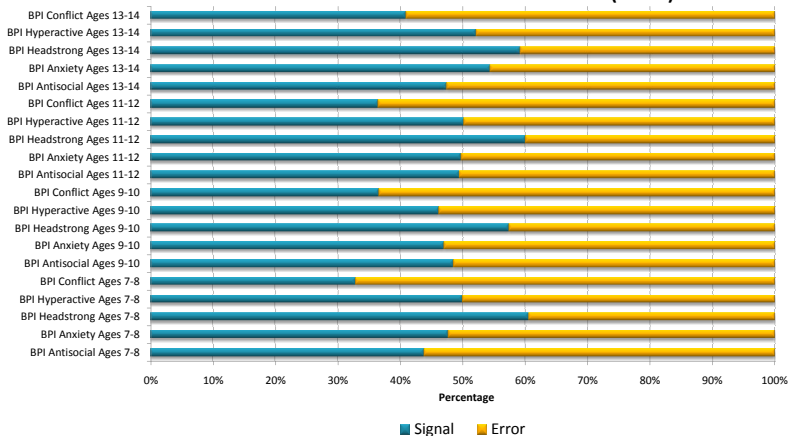
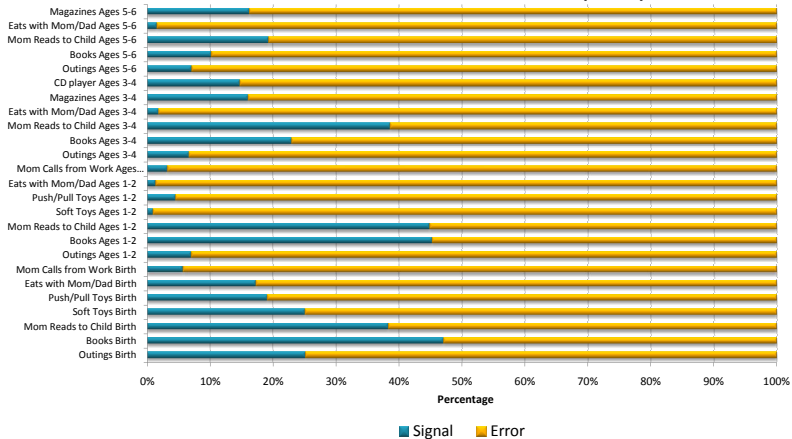
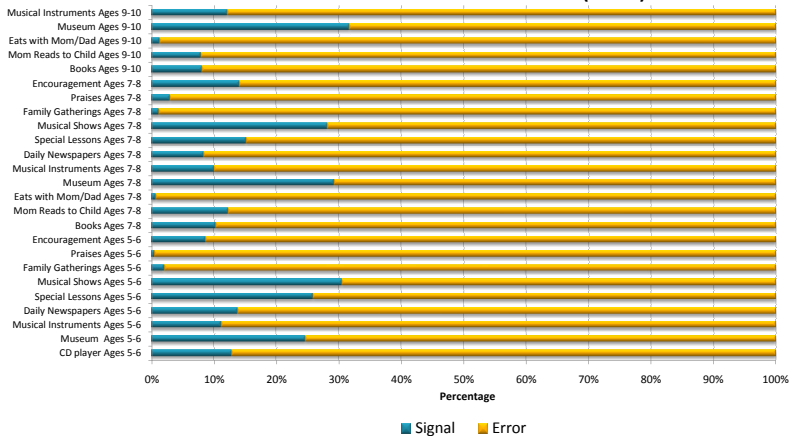


Figure 5A

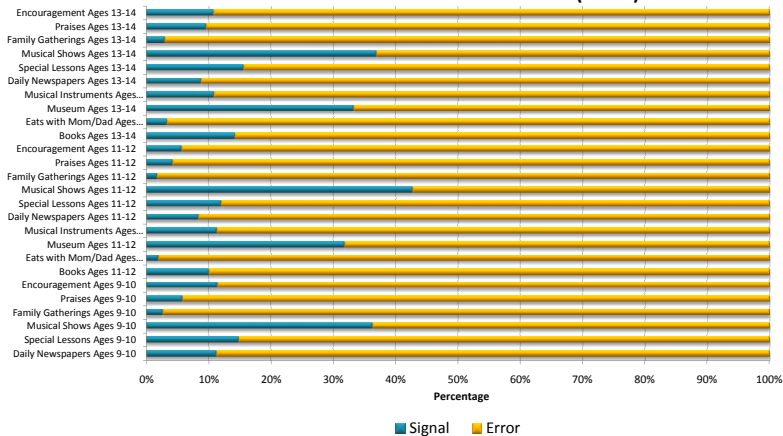
**Share of Residual Variance in Measurements of Investments  
Due to the Variance of Investment Factor (Signal)  
and Due to the Variance of Measurement Error (Noise)**



**Figure 5B**  
**Share of Residual Variance in Measurements of Investments**  
**Due to the Variance of Investment Factor (Signal)**  
**and Due to the Variance of Measurement Error (Noise)**



**Figure 5C**  
**Share of Residual Variance in Measurements of Investments**  
**Due to the Variance of Investment Factor (Signal)**  
**and Due to the Variance of Measurement Error (Noise)**



## Problem #2: Lack of a Natural Metric in Test Scores

- Test scores (cognitive and noncognitive) have no metric.
- Different test scores may measure different types of the same skill.
- Our approach: anchor skills on adult outcomes that have clear metric.
- Consider linear anchor,  $z_1$ , say log earnings (measured in dollars):

$$z_1 = \mu + \alpha_c \theta_{c,T} + \alpha_n \theta_{n,T} + \nu_4$$

- Note that  $\alpha_c \theta_{c,T}$  and  $\alpha_n \theta_{n,T}$  are in dollar units. Consequently:

$$\alpha_c \theta_{c,t+1} = f(\alpha_c \theta_{c,t}, \alpha_n \theta_{n,t}, x_{c,t}, \theta_{c,p}, \eta_{c,t})$$

- Anchoring functions can be linear or nonlinear (as we estimate in the paper).

## Problem #3: Endogeneity of Investments

- We have already proposed ways to deal with:
  1. Measurement Error
  2. Lack of Metric
- We have to deal with the fact that  $\eta_{k,t}$  is probably correlated with  $x_{k,t}$ .
- We go through different approaches.



## Approach 1: Unobservable Heterogeneity

- Suppose that we observe multiple adult outcomes:

$$z_1 = \mu_1 + \alpha_c \theta_{c,T} + \alpha_n \theta_{n,T} + \varepsilon_{z,1}$$

$$z_2 = \mu_2 + \delta_c \theta_{c,T} + \delta_n \theta_{n,T} + \varepsilon_{z,2}$$

- We can break the error terms  $\varepsilon_z$  into two parts:

$$\varepsilon_{zj} = \pi + \zeta_{zj} \text{ for } j = 1, 2$$

- We break the error term  $\eta_{k,t}$  in the technology as well:

$$\alpha_c \theta_{c,t+1} = f(\alpha_c \theta_{c,t}, \alpha_n \theta_{n,T}, x_{c,t}, \theta_{c,p}, \theta_{n,p}, \pi, \nu_{c,t})$$

- Our identification approach allows for  $(\alpha_c \theta_{c,T}, \alpha_n \theta_{n,T}, x_{c,t}, \theta_{c,p})$  to be correlated with  $\pi$ .

## Approach 2: Estimate Investment Equation

- Suppose that  $\Omega_t$  are the state variables at period  $t$ .
- $\{\theta_{c,T}, \theta_{n,T}, \theta_{c,p}, \theta_{c,p}, \theta_{n,p}, \pi\} \subset \Omega_t$ , but we need the reverse not to be true.
- Write:

$$\alpha_c \theta_{c,T} = f(\alpha_c \theta_{c,t}, \alpha_n \theta_{n,T}, x_{c,t}, \theta_{c,p}, \theta_{n,p}, \pi, v_{c,t})$$

- Suppose the function is:

$$x_{k,t} = g(\Omega_t) + \zeta_t$$

- We need exclusion restrictions, say  $z_t \in \Omega_t$ , but  $z_t \notin \{\theta_{c,T}, \theta_{n,T}, \theta_{c,p}, \theta_{c,p}, \theta_{n,p}, \pi\}$ .
- Repeated measurements on  $x_{k,t}$  allows us to identify  $\zeta_t$ .

Table 5A

## The Technology for Cognitive and Noncognitive Skill Formation

Estimated Along with Investment Equation with

Linear Anchoring on Educational Attainment (Years of Schooling)

Allowing for Unobserved Heterogeneity ( $\pi$ ), Factors Normally Distributed

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

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

Note: Standard errors in parenthesis

Table 5B

The Technology for Cognitive and Noncognitive Skill Formation

Estimated Along with Investment Equation with

Linear Anchoring on Educational Attainment (Years of Schooling)

Allowing for Unobserved Heterogeneity ( $\pi$ ), Factors Normally Distributed

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

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

Note: Standard errors in parenthesis

## Maximizing Aggregate Education

- Suppose that  $H$  children are born,  $h = 1, \dots, H$ .
- These children represent draws from the distribution of initial conditions  $F(\theta_{c,1,h}, \theta_{n,1,h}, \theta_{c,p}, \theta_{n,p}, \pi)$ .
- We want to allocate finite resources  $B$  across these children for early and late investments.
- Formally:

$$\text{Max } \bar{S} = \frac{1}{H} \sum_{h=1}^H S(\theta_{c,3}, \theta_{n,3}, \pi_h)$$

$$\text{subject to : } \sum_{h=1}^H (I_{1,h} + I_{2,h}) = B$$

$$\theta_{k,t+1,h} = f_t(\theta_{c,t,h}, \theta_{n,t,h}, I_{t,h}, \theta_{c,p}, \theta_{n,p}, \pi_h)$$

$$\text{for } k = c, n \text{ and } t = 1, 2.$$

Figure 4

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

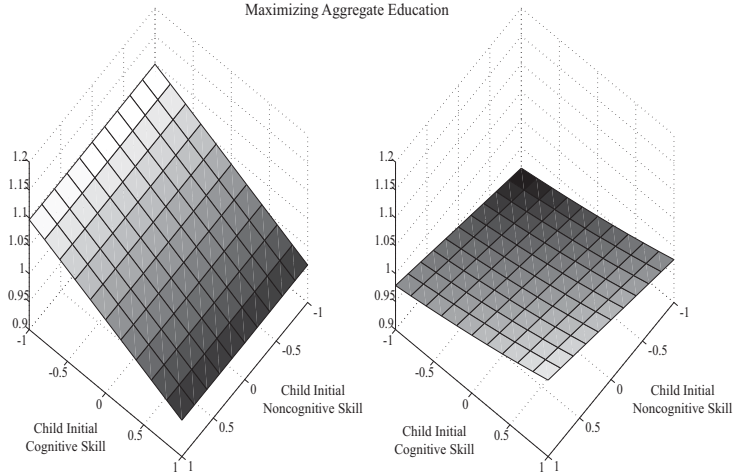
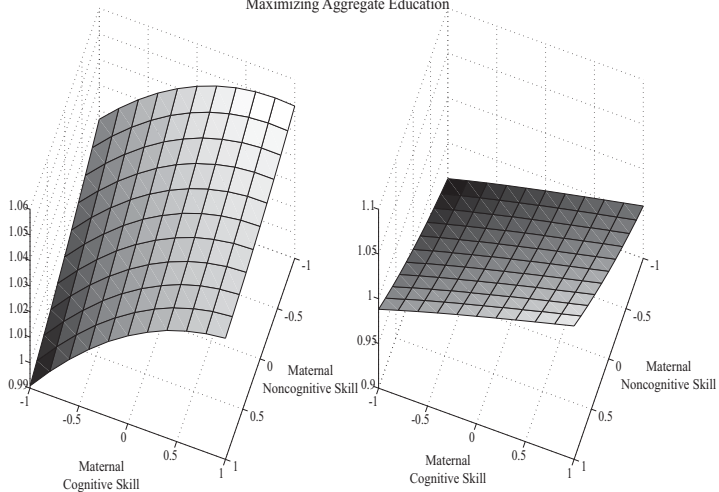


Figure 5  
Optimal Early (Left) and Late (Right) Investments by  
Maternal Cognitive and Noncognitive Skills  
Maximizing Aggregate Education



# This Paper

- We show that it is possible to nonparametrically identify the technology even though:
  1. Measurement error.
  2. Lack of Metric.
  3. Endogeneity of Investments.
- We find that:
  1. Early investments should be higher for children from disadvantaged backgrounds.
  2. The optimal ratio of early to late investments depends on the outcome of interest.
  3. If early remediation does not occur at early stages of childhood, then remediation at later stages should focus primarily on producing noncognitive skills.