Human Capital Formation in Childhood and Adolescence

Flávio Cunha
Rice University

August 28, 2016
INCOME INEQUALITY IN THE UNITED STATES, 1910-2010

SHARE OF TOP DECILE IN NATIONAL INCOME

25%  30%  35%  40%  45%  50%

Relative Supply and Demand of Skilled Labor

Case 1: Supply and Demand Grow at Same Rate

Relative Skill Premium

Relative stocks of skills

Relative supply of skilled labor

Relative demand for skilled labor
Figure
Relative Supply and Demand of Skilled Labor
Case 1: Supply and Demand Grow at Same Rate
Figure
Relative Supply and Demand of Skilled Labor
Case 1: Supply and Demand Grow at Same Rate
Figure
Relative Supply and Demand of Skilled Labor
Case 1: Supply and Demand Grow at Same Rate
Figure
Relative Supply and Demand of Skilled Labor
Case 2: SBTC - Relative Demand Starts to Grow at a Faster Rate

Relative Demand for Skilled Labor

Relative Supply of Skilled Labor

Relative Skill Premium

Relative stocks of skills
Figure
Relative Supply and Demand of Skilled Labor
Case 2: SBTC - Relative Demand Starts to Grow at a Faster Rate

Relative Skill Premium

Relative supply of skilled labor

Relative demand for skilled labor

Relative stocks of skills
Figure
Relative Supply and Demand of Skilled Labor
Case 2: SBTC - Relative Demand Starts to Grow at a Faster Rate

Relative Skill Premium

Relative supply of skilled labor

Relative demand for skilled labor

Relative stocks of skills
Figure
Relative Supply and Demand of Skilled Labor
Case 2: SBTC - Relative Demand Starts to Grow at a Faster Rate

Relative Skill Premium
Relative supply of skilled labor
Relative demand for skilled labor
Relative stocks of skills
Figure
Relative Demand and Supply of Skilled Labor
Case 3: Relative Supply Starts to Grow at Slower Rate

Relative Skill Premium

Relative supply of skilled labor

Relative demand for skilled labor

Relative stocks of skills
Figure
Relative Demand and Supply of Skilled Labor
Case 3: Relative Supply Starts to Grow at Slower Rate

- Relative Skill Premium
- Relative supply of skilled labor
- Relative demand for skilled labor
- Relative stocks of skills
Figure
Relative Demand and Supply of Skilled Labor
Case 3: Relative Supply Starts to Grow at Slower Rate

Relative Skill Premium

Relative supply of skilled labor

Relative demand for skilled labor

Relative stocks of skills
Relative Demand and Supply of Skilled Labor

Case 3: Relative Supply Starts to Grow at Slower Rate

- Relative stock of skills
- Relative skill premium
- Relative supply of skilled labor
- Relative demand for skilled labor

Figure
Evolution of Inequality in USA

Figure 1
College Graduation Rates (by 35 years) for Men and Women: Cohorts Born from 1876 to 1975

Sources: 1940 to 2000 Census of Population Integrated Public Use Micro-data Samples (IPUMS).
Chart A1.1. Percentage of population that has attained tertiary education, by age group (2009)


Countries are ranked in descending order of the percentage of 25-34 year-olds who have attained tertiary education.


StatLink  
http://dx.doi.org/10.1787/888932459831
Let $L_S$ and $L_U$ denote, respectively, skilled and unskilled labor.

Let $w_S$ and $w_U$ denote, respectively, skilled and unskilled wage rates.

Consider the following problem:

$$\min w_S L_S + w_U L_U$$

subject to the technology of skill formation:

$$Y = \left[ \gamma L_S^\phi + (1 - \gamma) L_U^\phi \right]^{1/\phi}$$

where $\gamma \in [0, 1]$ and $\phi \leq 1$. 
Simple Model

- Taking first-order conditions:

\[
\begin{align*}
    w_S &= \lambda \left[ \gamma L_S^\phi + (1 - \gamma) L_U^\phi \right]^{\frac{1-\phi}{\phi}} \gamma L_S^{\phi-1} \\
    w_U &= \lambda \left[ \gamma L_S^\phi + (1 - \gamma) L_U^\phi \right]^{\frac{1-\phi}{\phi}} (1 - \gamma) L_U^{\phi-1}
\end{align*}
\]

which yields:

\[
\ln \frac{w_S}{w_U} = \ln \frac{\gamma}{1 - \gamma} + (\phi - 1) \ln \frac{L_S}{L_U}
\]
Evolution of Inequality in USA

- Actual values for college wage premium
- Predicted college wage premium
- Predicted wage premium with 1949 dummy
Sources of Economic Inequality

Figure 9.1. Densities of present value of high school earnings under different information sets for the agent. They are calculated for the entire population regardless of schooling choice. Let $Y$ denote the agent’s information set. Let $Y_0$ denote the present value of earnings in the high school sector (discounted at a 3% interest rate). Let $f(y_0 | Y)$ denote the density of the present value of earnings in high school conditioned on the information set $Y$. Then: The solid line plots $f(y_0 | Y)$ under no information, i.e. $Y = \emptyset$. The dashed line plots $f(y_0 | Y)$ when only factor 1 is in the information set, i.e. $Y = \Theta_1$. The dashed-dotted line plots $f(y_0 | Y)$ when factors 1 and 2 are in the information set, i.e. $Y = \Theta_1 \Theta_2$. The crossed line plots $f(y_0 | Y)$ when all factors are in the information set, i.e. $Y = \Theta_1 \Theta_2 \Theta_3$. The $X$ are put at the mean and are assumed to be known. The $\theta_i$ when known, are set at their mean of zero.
Figure 4: The Effect of Cognitive and Socio-emotional endowments, (log) Wages

![Graph showing the effect of cognitive and socio-emotional endowments on log wages.](image)
Figure 1: The Probability of Educational Decisions, by Endowment Levels, Dropping from Secondary School vs. Graduating
Figure 2: The Probability of Educational Decisions, by Endowment Levels, HS Graduate vs. College Enrollment

Figure 3: The Probability of Educational Decisions, by Endowment Levels, Some College vs. 4-year college degree

Figure 5: The Effect of Cognitive and Socio-emotional endowments, Daily Smoking


James Heckman Economics and Econometrics of Human Development
Figure 7: The Effect of Cognitive and Socio-emotional endowments, Participated in 2006 election
Figure 8: The Effect of Cognitive and Socio-emotional endowments on Probability of White-collar occupation (age 30)

James Heckman Economics and Econometrics of Human Development
Ever been in jail by age 30, by ability (males)

Note: This figure plots the probability of a given behavior associated with moving up in one ability distribution for someone after integrating out the other distribution. For example, the lines with markers show the effect of increasing noncognitive ability after integrating the cognitive ability.


James Heckman
Economics and Econometrics of Human Development
Probability of being teenage and single with children (females)

Note: This figure plots the probability of a given behavior associated with moving up in one ability distribution for someone after integrating out the other distribution. For example, the lines with markers show the effect of increasing noncognitive ability after integrating the cognitive ability.


James Heckman
Economics and Econometrics of Human Development
Gaps in Skills in Early Childhood
Hart and Risley (1995)

Figure 2. The widening gap we saw in the vocabulary growth of children from professional, working-class, and welfare families across their first 3 years of life. (See Appendix B for a detailed explanation of this figure.)
Gaps in Skills in Early Childhood
Carneiro and Heckman (2003)

Average percentile rank on anti-social behavior score, by income quartile

Gaps in Investments in Early Childhood
Carneiro and Heckman (2003)

Figure
Unadjusted Mean Home Score
by Quartile of Permanent Income of the Family

Flávio Cunha (Rice University) Human Capital Formation in Childhood and Adolescence August 28, 2016 34 / 235
Gaps in Investments in Early Childhood
Hart and Risley (1995)

The graph shows the number of words addressed to the child by parents in different socioeconomic classes over the age of the child in months. The categories include:
- 13 professional parents
- 23 working-class parents
- 6 welfare parents

The x-axis represents the age of the child in months, ranging from 9 to 36 months. The y-axis represents the number of words addressed to the child, ranging from 0 to 2500 words.
Investments in Human Capital of Children
by Quartiles of Permanent Income

Investments (hours per day)
Gaps in Investments in Early Childhood
Kalil, Ryan, and Corey (2012)

Gaps in Investments in Early Childhood
Kalil, Ryan, and Corey (2012)

Gaps in Investments in Adolescence
Kalil, Ryan, and Corey (2012)

**Figure 15:** Parental Investment over Childhood among Whites by Mother’s Education: Material Resources

Source: Moon (2012).
Figure 16: Parental Investment over Childhood among Whites by Mother’s Education: Cognitive Stimulation

Source: Moon (2012).
Figure 17: Parental Investment over Childhood among Whites by Mother’s Education: Emotional Support

Source: Moon (2012).
Figure 18: Parental Investment over Childhood among Whites by Family Income Quartile: Cognitive Stimulation

Source: Moon (2012).
Figure 19: Parental Investment over Childhood among Whites by Family Type: Cognitive Stimulation

Source: Moon (2012).
How much, and in what ways, do kindergarten teachers matter for learning outcomes?

Two challenges:

- Sorting of students to teachers.
  - Solution: Randomly match students to teachers.
- Data on teachers are weakly correlated with student gain.
  - Improve the quality of data on teachers.
What is the CLASS and why use it?

Classroom observation tool

- Emotional support
  - Climate (positive or negative), teacher sensitivity, and regard for student perspectives

- Classroom organization
  - Behavior management, productivity, and instructional and learning formats

- Instructional support
  - Concept development, quality of feedback, and language modeling
**Example: Teacher Behaviors and CLASS Scores for Behavior Management Dimension**

Behavior Management
Encompasses the teacher's ability to provide clear behavioral expectations and use effective methods to prevent and redirect misbehavior.

<table>
<thead>
<tr>
<th>Clear Behavior Expectations</th>
<th>Low (1,2)</th>
<th>Mid (3,4,5)</th>
<th>High (6,7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Clear expectations are absent, unclear, or inconsistently enforced.</td>
<td>Rules and expectations may be stated clearly, but are inconsistently enforced.</td>
<td>Rules and expectations for behavior are clear and are consistently enforced.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proactive</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Anticipates problem behavior or escalation</td>
<td>Teacher is reactive and monitoring is absent or ineffective.</td>
<td>Teacher uses a mix of proactive and reactive responses; sometimes monitors but at other times misses early indicators of problems.</td>
<td>Teacher is consistently proactive and monitors effectively to prevent problems from developing.</td>
</tr>
<tr>
<td>• Rarely reactive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Monitoring</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Redirection of Misbehavior</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Effectively reduces misbehavior</td>
<td>Attempts to redirect misbehavior are ineffective; teacher rarely focuses on positives or uses subtle cues. As a result, misbehavior continues/escalates and takes time away from learning.</td>
<td>Some attempts to redirect misbehavior are effective; teacher sometimes focuses on positives and uses subtle cues. As a result, there are few times when misbehavior continue/escalate or takes time away from learning.</td>
<td>Teacher effectively redirects misbehavior by focusing on positives and making use of subtle cues. Behavior management does not take time away from learning.</td>
</tr>
<tr>
<td>• Attention to the positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Uses subtle cues to redirect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Efficient</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Behavior</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Frequent compliance</td>
<td>There are frequent instances of misbehavior in the classroom.</td>
<td>There are periodic episodes of misbehavior in the classroom.</td>
<td>There are few, if any, instances of student misbehavior in the classroom.</td>
</tr>
<tr>
<td>• Little aggression &amp; defiance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Source: Pianta, La Paro & Hamre (2008)*
Break analysis in two parts:

- Estimate teacher effects: How much does it matter whether a child was assigned to teacher A or B in a school?
- Estimate the associations between within-school differences in teacher characteristics or behaviors and child learning outcomes
One standard error in teacher quality leads to increases in child learning of

- 11% of standard deviation in math.
- 13% of standard deviation in language.
- 7% of standard deviation in executive function.

Same teachers have their students learn more math and more language year after year.

- Cross-year correlation of teacher effects in math is 0.32
- Cross-year correlation of teacher effects in language is 0.42.
What explains differences in teacher effectiveness?

- One standard deviation in teacher IQ increases child’s performance by 4% of a standard deviation.
- Students randomly assigned to “rookie” teachers learn 16% of standard deviation less.
- No correlation between teacher personality scores (Big Five) and student learning.
- One standard deviation in CLASS explains 59% of a standard deviation in student learning.
- Teachers with better CLASS scores get all their students to learn more: Effects are not concentrated on girls or boys, on children with high or low levels of development when they enter school, or on children of high or low socioeconomic status.
Interestingly, parental reports of teacher quality correlate (very imperfectly) with teacher effectiveness:

- Teachers who produce one standard deviation more learning are given a 0.44 higher score (on a scale from 1 to 5).
- Rookie teachers are given 0.33 lower score by parents.
- Teachers with higher CLASS scores also get higher scores reported by parents.

However, parents do not adjust behaviors in response to differences in teacher quality.

- There is no effect on the quality or quantity of parent-child interaction at home.
- There is no effect on the child’s dropping out or absenteeism.
How teacher ratings relate to a school's poverty level

Teachers who receive the state's top value-added rating — "Most Effective" — are likely to be in schools with fewer poor students, based on value-added ratings for teachers at 1,720 public schools. Of 1,035 teachers at the wealthiest schools, 34 percent got the top rating. In contrast, of 2,411 teachers at the poorest schools, just over 9 percent were rated "Most Effective."

<table>
<thead>
<tr>
<th>Teachers rated Most Effective</th>
<th>Teachers rated Least Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>35%</td>
<td>30%</td>
</tr>
<tr>
<td>30%</td>
<td>25%</td>
</tr>
<tr>
<td>25%</td>
<td>20%</td>
</tr>
<tr>
<td>20%</td>
<td>15%</td>
</tr>
<tr>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>5%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Percent of teachers in rating category

% of students eligible for free or reduced-price lunch

SOURCE: Ohio Department of Education

RICH EXNER, JAMES OWENS | THE PLAIN DEALER
Notes: Effect of schooling on components of the ASVAB. The first four components are averaged to create male’s with average ability. We standardize the test scores to have within-sample mean zero, variance one. The model is estimated using the NLSY79 sample. Solid lines depict average test scores, and dashed lines, confidence intervals.
Source: Heckman, Stixrud and Urzua [2006, Figure 4].
Figure 23: Causal Effect of Schooling on Two Measures of Personality
Evidence is Reinforced from Evidence from RCT

- Heckman has shown that many early childhood programs have positive impacts on long-term outcomes related to inequality.
- This is in spite of the fact that gains in cognitive skills do not tend to persist over time (more about this on “dynamic complementarity”).
- Dobbie and Fryer (2009) show that Harlem Children’s Zone is effective at increasing the achievement of the poorest minority children: It closes the black-white achievement gap in mathematics and reduces it by nearly half in English Language Arts.
- Heller et al (2015) show that Cognitive Behavior Therapy target to disadvantaged adolescents can reduce propensity to participate in crime and even improve educational outcomes.
Increasing Inequality in Skills
Reardon (2013)
Increasing Inequality in Skills

Social Trust
By parents’ education, 12th graders, 1976–2011

“Most people can be trusted” (agree)


Source: Monitoring the Future
Trends in Health: Child obesity

Graph showing trends in child obesity from 1970 to 2010. The graph compares the percentage of children with high school education or less (gray squares) and those with a bachelor's degree or more (black dots). The trend indicates an increase in obesity rates for both groups over time, with a peak around 2000 for the group with less education.
Increasing Inequality in Investments
Altintas (2016)
Increasing Inequality in Investments
Kornrich and Furstenberg (2011)

Increasing Inequality in Investments

Trends in Family Dinners
By parental education, 1978–2005

"Our whole family usually eats dinner together" (agree)

Source: DDB Lifestyle surveys, 1978–2005
Increasing Inequality in Investments

![Graph showing participation in school-based extracurriculars over time. The graph illustrates a decrease in participation for both the highest and lowest SES quartiles, with the highest SES quartile maintaining a higher participation rate throughout the years. The source of the data is the National Longitudinal Study of 1972, High School & Beyond (1980), National Education Longitudinal Study of 1988, Education Longitudinal Study of 2002.]
Full Circle: College Attendance

- Top quartile
- Third quartile
- Second quartile
- Bottom quartile

Graph showing trends from 1965 to 2015.
FIGURE 7.
Share of Population with College Degree, by Income Level and Birth Year

The graduation rate for low-income individuals has not increased very much over the past few decades.

Socioeconomic Distribution at Colleges by Selectivity

A student at one of America’s most-selective universities is fourteen times more likely to be from a high-income family than from a low-income family.

Source: Carnevale and Strohl (2010).
Note: Figure shows college attendance age of 2006. See technical appendix for full description of college selectivity categories.
Application Behavior of High-Achieving Students

Panel A: High-Income Students’ Portfolios of College Applications

- Nonselective
- Less-selective

40%
30%
20%
10%

Relatively few high-income students apply to nonselective schools.
High-income students’ applications are well-distributed among reach, match, and safety schools.

Panel B: Low-Income Students’ Portfolios of College Applications

- Nonselective
- Less-selective

40%
30%
20%
10%

The bulk of low-income students’ applications go to nonselective schools.
Low-income students are less likely than their high-income counterparts to apply to a mix of match and reach schools.

College selectivity, measured as college’s median SAT score—student’s SAT score (in percentiles)

Source: Avery and Hoxby (2012).
Next, I will try to make sense of this data by proposing a very simple model of human capital formation.

At the core of this model, there will be two important parameters:

- Self-productivity of skills: I learn how to read, then I use reading to learn other skills.
- Dynamic complementarity: The returns to the development of advanced skills are higher for the individuals who learned basic skills.
Consider the following cost minimization problem:

$$\min x_E + \frac{1}{1 + r} x_L$$

subject to the technology of skill formation:

$$h = \left[ \gamma x_E^\phi + (1 - \gamma) x_L^\phi \right]^{\frac{1}{\phi}}$$

where $\gamma \in [0, 1]$ and $\phi \leq 1$.

Note that:

- The parameter $\gamma$ captures self-productivity.
- The parameter $\phi$ captures dynamic complementarity.
Boundary Solution when $\phi = 1$

- In this case, $h = \gamma x_E + (1 - \gamma) x_L$.
- Two investment strategies: Invest early and produce $\gamma$ units of human capital per unit of investment.
- Save in physical assets early and invest $1 + r$ late and produce $(1 + r)(1 - \gamma)$ units of human capital.
- Should invest all early if, and only if:

$$\gamma > \frac{1 + r}{2 + r}$$
Boundary Solution when $\phi \to -\infty$

- In this case, $h = \min \{x_E, x_L\}$
- The solution to this problem is $x_E = x_L$ for whatever values of $r$. 
Interior Solution when $-\infty < \phi < 1$

- The solution to this problem is:

$$x_E = \frac{\gamma^{1/\phi}}{\left[\gamma^{1/\phi} + (1-\gamma)^{1/\phi} (1+r)^{\phi/\phi}\right]^{1/\phi}} h$$

$$x_L = \frac{(1-\gamma)^{1/\phi} (1+r)^{1/\phi}}{\left[\gamma^{1/\phi} + (1-\gamma)^{1/\phi} (1+r)^{\phi/\phi}\right]^{1/\phi}} h$$

- Note that we have the following ratio:

$$\ln \frac{x_E}{x_L} = \frac{1}{1-\phi} \ln \left(\frac{\gamma}{1-\gamma}\right) + \frac{1}{1-\phi} \ln \left(\frac{1}{1+r}\right)$$
The Ratio of Early to Late Investment in Human Capital As a Function of the Skill Multiplier for Different Values of Complementarity

This figure shows the optimal ratio of early to late investments, \(L_1/L_2\), as a function of the skill multiplier parameter for different values of the complementarity parameter assuming that the interest rate is zero. The optimal ratio \(L_1/L_2\) is the solution of the parental problem of maximizing the present value of the child's wealth through investments in human capital, \(k\), and transfers of risk-free bonds. In order to do that, parents have to decide how to allocate a total of \(P\) dollars into early and late investments in human capital, \(L_1\) and \(L_2\) respectively, and risk-free bonds. Let \(t\) denote the present value as of period 3 of the future prices of one efficiency unit of human capital:

\[
t = P \frac{W_{w=3}}{w(z + u)^w} = \frac{P}{(1 + u)^w}
\]

The parents solve

\[
\max \mu_1 (1 + u) \left[ tk + e \right]
\]

subject to the budget constraint

\[
L_1 + L_2 (1 + u) + e (1 + u) = P
\]

and the technology of skill formation:

\[
k = h L_1 + (1 - \phi) L_2 i
\]

for \(0 \leq \phi \leq 1\) and \(0 \leq \phi = 0.5\) and for values of the skill multiplier between 0 and 0.9 = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.

(Assumes \(r = 0\))
Dual Side of Dynamic Complementarity

- Returns to late investments are higher for the individuals that have high early investments.
- BUT
- Returns to early investments are higher for the individuals who will also have high late investments.
- In other words, if the child will not receive high late investments, then the impacts of early investments will be diminished.
Estimating Parameters of the Technology of Skill Formation: Parameterization

- $\theta_{c,t}$ 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: $s = 1, \ldots, S$. The technology for skill $k$, at period $t$ and stage $s$ is:

$$\theta_{k,t+1} = e^{\eta_{c,t+1}} \times f_{s,k}$$

where

$$f_{s,k} = \left[ \gamma_{s,k,1}\phi_{c,t} + \gamma_{s,k,2}\phi_{n,t} + \gamma_{s,k,3}\phi_{k,t} + \gamma_{s,k,4}\phi_{c,p} + \gamma_{s,k,5}\phi_{n,p} \right] \frac{1}{\phi_{s,c}}$$
Estimating Parameters of the Technology of Skill Formation: Data

- 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).
Estimating Parameters of the Technology of Skill Formation: Measures

- Child’s Cognitive Skills:
  - MSD
  - Parts of Body
  - Memory for Locations
  - PPVT
  - PIAT

- Child’s Noncognitive Skills:
  - Temperament and Behavior Problem Index.

- Parental Investments
  - Components of the Home Score

- Parental Cognitive Skills
  - Components of the ASVAB Tests.

- Parental Noncognitive Skills
  - Mother’s Self-Esteem and Mom’s Locus of Control
We need to deal with three problems:

- We don’t observe $\theta_{c,t}$, $\theta_{n,t}$ and $x_{k,t}$ directly.
- We don’t know which scale to use to measure $\theta_{c,t}$ and $\theta_{n,t}$.
- Investments $x_{k,t}$ are chosen by parents based on information from $\eta_{k,t}$ that is unobserved by the econometrician.
Addressing Measurement Error

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

\[ \beta = \frac{Cov(\theta_{k,t+1}, x_{k,t})}{Var(x_{k,t})} \]

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

- When we don’t observe $\theta_{k,t+1}$ and $x_{k,t}$, we can use observed proxy variables:

  $$y_{1,k,t+1} = \theta_{k,t+1} + \varepsilon_{1,k,t+1}$$
  $$y_{2,k,t+1} = \theta_{k,t+1} + \varepsilon_{2,k,t+1}$$
  $$y_{3,k,t} = x_{k,t} + \varepsilon_{3,k,t}$$
  $$y_{4,k,t} = x_{k,t} + \varepsilon_{4,k,t}$$

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

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

- Key to identification: covariance restrictions in measurement error (some are testable).
Addressing Measurement Error

- It is possible to relax the model so we can give each measure its own “information parameter” (also called factor loading) $\alpha_{l,k,t}$ for $l = 2, ..., L$:

\[
y_{1,k,t} = \theta_{k,t} + \varepsilon_{1,k,t} y_{l,k,t} = \alpha_{l,k,t} \theta_{k,t} + \varepsilon_{2,k,t}
\]

- 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):

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

- We can also identify nonlinear measurement equations:

\[
y_{1,k,t} = \theta_{k,t+1} + \varepsilon_{1,k,t}
\]
\[
y_{l,k,t} = h_{l,k,t}(\theta_{k,t}, \varepsilon_{l,k,t})
\]
Estimation algorithm

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

\begin{align*}
y_1 & = \mu_1 + x + \varepsilon_1 \\
y_2 & = \mu_2 + x + \varepsilon_2 \\
y_3 & = \mu_3 + \theta + \varepsilon_3 \\
y_4 & = \mu_4 + \theta + \varepsilon_4
\end{align*}

As well as the following technology of skill formation (nonlinear):

\[ \theta = \alpha + \beta x^2 + \eta \]

The goal is to construct the contribution to the likelihood for each individual: \( L_i = L(y_1, y_2, y_3, y_4) \)
The filtering algorithm provides a set of steps that allows us to compute:

\[ L(y_1, y_2, y_3, y_4) = L(y_4 | y_1, y_2, y_3) \times L(y_3 | y_2, y_1) \times L(y_2 | y_1) \times L(y_1) \]

While computation of \( L(y_1, y_2, y_3, y_4) \) is time-consuming, the computation of each one of the terms \( L(y_4 | y_3, y_2, y_1) \), \( L(y_3 | y_2, y_1) \), \( L(y_2 | y_1) \), and \( L(y_1) \) is fast.
Let’s consider the situation in which both $x$ and $\epsilon_l$, for $l = 1, 2$, are normally distributed. This means that $y_1$ and $y_2$ are jointly normally distributed.

More concretely, let $x \sim N(\mu_x, \sigma_x^2)$ and $\epsilon_l \sim N(0, \sigma_l^2)$ for $l = 1, 2$.

Note that: $y_1 \sim N(\mu_1 + \mu_x, \sigma_x^2 + \sigma_1^2)$.

Therefore, the first term in the contribution to the likelihood is:

$$L(y_1) = \sqrt{\frac{1}{2\pi (\sigma_x^2 + \sigma_1^2)}} \exp \left\{ -\frac{1}{2} \frac{(y_1 - \mu_1 - \mu_x)^2}{(\sigma_x^2 + \sigma_1^2)} \right\}$$
To derive the second term in the contribution to the likelihood, it will be very helpful to compute $E(x|y_1)$ and $Var(x|y_1)$.

Because of the normality assumption, it follows that:

$$E(x|y_1) = E(x) + \frac{Cov(x, y_1)}{Var(y_1)} [y_1 - E(y_1)]$$

$$E(x|y_1) = \mu_x + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_1^2} [y_1 - \mu_1 - \mu_x]$$

And:

$$Var(x|y_1) = Var(x) - \frac{Cov(x, y_1)^2}{Var(y_1)}$$

$$Var(x|y_1) = \sigma_x^2 - \frac{\sigma_x^4}{\sigma_x^2 + \sigma_1^2}$$
Illustration of Estimation Algorithm

- Note that

\[ E (y_2 \mid y_1) = \mu_2 + E (x \mid y_1) + E (\varepsilon_2 \mid y_1) \]

\[ E (y_2 \mid y_1) = \mu_2 + \mu_x + \frac{\sigma^2_x}{\sigma^2_x + \sigma^2_1} (y_1 - \mu_1 - \mu_x) \]

- And:

\[ \text{Var} (y_2 \mid y_1) = \text{Var} (x \mid y_1) + \text{Var} (\varepsilon_2 \mid y_1) \]

\[ \text{Var} (y_2 \mid y_1) = \sigma^2_2 + \sigma^2_x - \frac{\sigma^4_x}{\sigma^2_x + \sigma^2_1} \]

- Therefore:

\[ L (y_2 \mid y_1) = \sqrt{\frac{1}{2\pi \text{Var} (y_2 \mid y_1)}} \exp \left\{ -\frac{1}{2} \frac{[y_2 - E (y_2 \mid y_1)]^2}{\text{Var} (y_2 \mid y_1)} \right\} \]
Now, we need to compute the moments $E(x|y_1, y_2)$ and $\text{Var}(x|y_1, y_2)$

$$E(x|y_1, y_2) = E(x|y_1) + \frac{\text{Cov}(x, y_2|y_1)}{\text{Var}(y_2|y_1)} [y_2 - E(y_2|y_1)]$$

And:

$$\text{Var}(x|y_1, y_2) = \text{Var}(x|y_1) - \frac{\text{Cov}(x, y_2|y_1)}{\text{Var}(y_2|y_1)}$$

Define $\mu_{x|y_1,y_2} = E(x|y_1, y_2)$ and $\sigma^2_{x|y_1,y_2} = \text{Var}(x|y_1, y_2)$. Great exercise to make sure you understand steps is to derive the expressions for
Note that $y_3$ is a function of $\theta$, not $x$.

For this reason, we need to compute $E(\theta|y_1,y_2)$ and $\text{Var}(\theta|y_1,y_2)$.

One way to compute these moments is by exploring the information from the technology of skill formation which defines the relationship between $\theta$ and $x$. 
If I use the technology of skill formation, note that:

$$E(\theta|y_1, y_2) = \alpha + \beta E(x^2|y_1, y_2) + E(\eta|y_1, y_2)$$

Endogeneity means $E(\eta|y_1, y_2) \neq 0$. For now, let’s assume away endogeneity so that $E(\eta|y_1, y_2) = 0$.

If the technology were linear in $x$, I could use the steps described above to derive a closed-form expression.

This is where the Unscented Transform is helpful.
If I use the technology of skill formation, note that:

\[
E(\theta|y_1,y_2) = \alpha + \beta E(x^2|y_1,y_2) + E(\eta|y_1,y_2)
\]

and

\[
Var(\theta|y_1,y_2) = \beta^2 Var(x^2|y_1,y_2) + Var(\eta|y_1,y_2)
\]

Endogeneity means \(E(\eta|y_1,y_2) \neq 0\). For now, let’s assume away endogeneity so that \(E(\eta|y_1,y_2) = 0\).

Heteroskedasticity means that \(Var(\eta|y_1,y_2)\) is a function of \(y_1,y_2\). For now, let’s assume homoskedasticity so that \(Var(\eta|y_1,y_2) = \sigma^2_\eta\)
Illustration of Estimation Algorithm

- Under these assumptions:

\[ E(\theta | y_1, y_2) = \alpha + \beta E(x^2 | y_1, y_2) \]

and

\[ \text{Var}(\theta | y_1, y_2) = \beta^2 \text{Var}(x^2 | y_1, y_2) + \sigma^2_\eta \]

- If the technology were linear in \( x \), I could use the steps described above to derive a closed-form expression.

- This is where the Unscented Transform is helpful.
The Unscented Transform approximates the mean and variance of a random variable that undergoes a nonlinear transformation.

For a one-dimensional problem, it is possible to approximate both $E(x^2 | y_1, y_2)$ and $Var(x^2 | y_1, y_2)$ with the following three points:

<table>
<thead>
<tr>
<th>Points ($\chi_j$)</th>
<th>Weights ($\omega_j$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_x - \sigma_x \sqrt{(1 + \kappa)}$</td>
<td>$\frac{1}{2(1+\kappa)}$</td>
</tr>
<tr>
<td>$\mu_x$</td>
<td>$\frac{\kappa}{(1+\kappa)}$</td>
</tr>
<tr>
<td>$\mu_x + \sigma_x \sqrt{(1 + \kappa)}$</td>
<td>$\frac{1}{2(1+\kappa)}$</td>
</tr>
</tbody>
</table>

where $\kappa$ is a parameter whose value is to be determined below.
Illustration of Unscented Transform

- We use these points and weights to compute the following moments:

\[
\tilde{E}(\theta|y_1, y_2) = \sum_{j=1}^{3} \omega_j \left(\alpha + \beta \chi_j^2\right)
\]

\[
\tilde{\text{Var}}(\theta|y_1, y_2) = \sum_{j=1}^{3} \omega_j \left(\alpha + \beta \chi_j^2\right) - \tilde{E}(\theta|y_1, y_2) \right)^2 + \sigma_\eta^2
\]
Illustration of Unscented Transform

- It is easy to show that these expressions are equal to:

\[ \widetilde{E}(\theta | y_1, y_2) = \alpha + \beta \left( \mu_{x|y_1,y_2}^2 + \sigma_{x|y_1,y_2}^2 \right) \]

\[ \widetilde{\text{Var}}(\theta | y_1, y_2) = \beta^2 \left( \kappa \sigma_{x|y_1,y_2}^4 + 4 \mu_{x|y_1,y_2}^2 \sigma_{x|y_1,y_2}^2 \right) + \sigma_{\eta}^2 \]

- How do we pick a value for \( \kappa \)?
Illustration of Unscented Transform

- To understand how to pick the value for \( \kappa \), note that the assumption of normality of \( x \), together with the assumption that the technology function is quadratic in \( x \), implies that:

\[
E(\theta | y_1, y_2) = \alpha + \beta \left( \mu^2_{x|y_1,y_2} + \sigma^2_{x|y_1,y_2} \right)
\]

\[
\text{Var}(\theta | y_1, y_2) = \beta^2 \left( 2\sigma^4_{x|y_1,y_2} + 4\mu^2_{x|y_1,y_2} \sigma^2_{x|y_1,y_2} \right) + \sigma^2_\eta
\]

- In contrast, the approximation via the Unscented Transform yields:

\[
\tilde{E}(\theta | y_1, y_2) = \alpha + \beta \left( \mu^2_{x|y_1,y_2} + \sigma^2_{x|y_1,y_2} \right)
\]

\[
\tilde{\text{Var}}(\theta | y_1, y_2) = \beta^2 \left( \kappa\sigma^4_{x|y_1,y_2} + 4\mu^2_{x|y_1,y_2} \sigma^2_{x|y_1,y_2} \right) + \sigma^2_\eta
\]

- The approximation is exact when \( \kappa = 2 \). Kmenta (1969) showed one can approximate CES function with quadratic function (in logs).
This means that we can approximate the first and second conditional moments of $y_3$

$$
\tilde{E} ( y_3 | y_1, y_2 ) = \mu_3 + \alpha + \beta \left( \mu^2_{x | y_1, y_2} + \sigma^2_{x | y_1, y_2} \right)
$$

$$
\tilde{\text{Var}} ( y_3 | y_1, y_2 ) = \beta^2 \left( \kappa \sigma^4_{x | y_1, y_2} + 4 \mu^2_{x | y_1, y_2} \sigma^2_{x | y_1, y_2} \right) + \sigma^2_{\eta} + \sigma^2_{3}
$$

We can use these moments to approximate the contribution to the likelihood term $L ( y_3 | y_2, y_1 )$.

In the last step, we can follow the rules above to derive $\tilde{E} ( \theta | y_1, y_2, y_3 )$ and $\tilde{\text{Var}} ( \theta | y_1, y_2, y_3 )$.

We then use these moments to obtain $\tilde{E} ( y_4 | y_1, y_2, y_3 )$, $\tilde{\text{Var}} ( y_4 | y_1, y_2, y_3 )$, and, finally, $L ( y_4 | y_3, y_2, y_1 )$. 
The second problem in the estimation of the technology of skill formation is the lack of metric in test scores.

Our approach: anchor skills on adult outcomes that have clear metric. Consider linear anchor, $z$, say log earnings (measured in dollars):

$$z = \mu + \alpha_c \theta_{c,T} + \alpha_n \theta_{n,T} + \nu$$

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

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

Anchoring functions can be linear or nonlinear.
Estimation of Technology of Skill Formation: Endogeneity

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

Write:

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

Suppose the policy function for investment 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 the distribution of $\zeta_t$.

Intuition: Nonlinear version of 2SLS.
### Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)

| Current Period Cognitive Skills (Self-Productivity) | $\gamma_{1,C,1}$ | 0.426 | $\gamma_{2,C,1}$ | 0.901 |
| Current Period Noncognitive Skills (Cross-Productivity) | $\gamma_{1,C,2}$ | 0.127 | $\gamma_{2,C,2}$ | 0.014 |
| Current Period Investments | $\gamma_{1,C,3}$ | 0.322 | $\gamma_{2,C,3}$ | 0.024 |
| Parental Cognitive Skills | $\gamma_{1,C,4}$ | 0.059 | $\gamma_{2,C,4}$ | 0.062 |
| Parental Noncognitive Skills | $\gamma_{1,C,5}$ | 0.066 | $\gamma_{2,C,5}$ | 0.000 |
| Complementarity Parameter | $\phi_{1,C}$ | 0.748 | $\phi_{2,C}$ | -1.207 |
| Implied Elasticity Parameter | $1/(1-\phi_{1,C})$ | 3.968 | $1/(1-\phi_{2,C})$ | 0.453 |
| Variance of Shocks $\eta_{C,t}$ | $\delta^2_{1,C}$ | 0.159 | $\delta^2_{2,C}$ | 0.092 |

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

| Current Period Noncognitive Skills (Self-Productivity) | $\gamma_{1,N,1}$ | 0.000 | $\gamma_{2,N,1}$ | 0.000 |
| Current Period Investments | $\gamma_{1,N,3}$ | 0.195 | $\gamma_{2,N,3}$ | 0.121 |
| Parental Noncognitive Skills | $\gamma_{1,N,5}$ | 0.093 | $\gamma_{2,N,5}$ | 0.011 |
| Complementarity Parameter | $\phi_{1,N}$ | 0.017 | $\phi_{2,N}$ | -0.323 |
| Elasticity Parameter | $1/(1-\phi_{1,N})$ | 1.017 | $1/(1-\phi_{2,N})$ | 0.756 |
| Variance of Shocks $\eta_{N,t}$ | $\delta^2_{1,N}$ | 0.170 | $\delta^2_{2,N}$ | 0.104 |

Note: Standard errors in parenthesis.
### Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>First Stage Parameters</th>
<th>Second Stage Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current Period Cognitive Skills (Cross-Productivity)</strong></td>
<td>( \gamma_{1,N,1} )</td>
<td>( \gamma_{2,N,1} )</td>
</tr>
<tr>
<td>( \gamma_{1,N,1} )</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>**Current Period Noncognitive Skills (Self-Productivity)</td>
<td>( \gamma_{1,N,2} )</td>
<td>( \gamma_{2,N,2} )</td>
</tr>
<tr>
<td>( \gamma_{1,N,2} )</td>
<td>0.712</td>
<td>0.868</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><strong>Current Period Investments</strong></td>
<td>( \gamma_{1,N,3} )</td>
<td>( \gamma_{2,N,3} )</td>
</tr>
<tr>
<td>( \gamma_{1,N,3} )</td>
<td>0.195</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td><strong>Parental Cognitive Skills</strong></td>
<td>( \gamma_{1,N,4} )</td>
<td>( \gamma_{2,N,4} )</td>
</tr>
<tr>
<td>( \gamma_{1,N,4} )</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><strong>Parental Noncognitive Skills</strong></td>
<td>( \gamma_{1,N,5} )</td>
<td>( \gamma_{2,N,5} )</td>
</tr>
<tr>
<td>( \gamma_{1,N,5} )</td>
<td>0.093</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><strong>Complementarity Parameter</strong></td>
<td>( \phi_{1,N} )</td>
<td>( \phi_{2,N} )</td>
</tr>
<tr>
<td>( \phi_{1,N} )</td>
<td>0.017</td>
<td>-0.323</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.21)</td>
</tr>
<tr>
<td><strong>Elasticity Parameter</strong></td>
<td>1/(1-( \phi_{1,N} ))</td>
<td>1/(1-( \phi_{2,N} ))</td>
</tr>
<tr>
<td></td>
<td>1.017</td>
<td>0.756</td>
</tr>
<tr>
<td><strong>Variance of Shocks</strong></td>
<td>( \delta^2_{1,N} )</td>
<td>( \delta^2_{2,N} )</td>
</tr>
<tr>
<td>( \delta^2_{1,N} )</td>
<td>0.170</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

*Note: Standard errors in parenthesis.*
Interpretation of Findings: Maximizing Average Education

- Suppose that $H$ children are born, $h = 1, \ldots, H$.
- 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:

$$S^* = \max \frac{1}{H} \left[ \sum_{h=1}^{H} S(\theta_{c,3}, \theta_{n,3}, \pi_h) \right]$$

subject to the technologies for the formation of cognitive and noncognitive skills as well as:

$$\sum_{h=1}^{H} (x_{1,h} + x_{2,h}) = B$$
Another possibility is to minimize aggregate crime (average crime per individual).

This will lead to different optimal ratios as crime is more sensitive to changes in noncognitive skills.

Relative to cognitive skills, noncognitive skills are more malleable at later ages.
FIGURE 5.—Ratio of early to late investments by maternal cognitive and noncognitive skills maximizing aggregate education (left) and minimizing aggregate crime (right) (other endowments held at mean levels).

FIGURE 6.—Densities of ratio of early to late investments maximizing aggregate education versus minimizing aggregate crime.

FIGURE 6.—Densities of ratio of early to late investments maximizing aggregate education versus minimizing aggregate crime.
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)
Figure 6
Share of Variance in Measurements of Maternal Cognitive and Noncognitive Skills due to the Variances of Cognitive and Noncognitive Factors (Signal) versus Variance of Measurement Error (Noise)
Home Observation for the Measurement of the Environment

- Created by Bettye Caldwell and Robert Bradley in late 1960s, early 1970s (first published in 1980s)
- Evaluates a child’s home environment as well as parent-child interaction.
- Administered by trained professional at the child’s home with both child and primary caregiver present.
- Semi-structured interview and observation period: 45-60 minutes.
HOME: Strengths and Weaknesses

**Strengths**

- Easy to administer and score.
- Reliability and validity.
- Easy to adapt for specific purposes.
- Provides objective information on home, child, and parent-child interaction.

**Weaknesses:**

- Training of administrators to follow standardized measurement.
- Only Yes/No questions.
- Score: Simple summation gives “too much” weight to items that do not vary a lot across households.
Let $\theta_i$ denote the latent quality of the environment experienced by child $i$.

Let $d^*_{i,j} = a_j (\theta_i - b_j) + \epsilon_{i,j}$ and define $d_{i,j} = 0$ if $d^*_{i,j} \leq 0$ and $d_{i,j} = 1$, otherwise.

Assume $\epsilon_{i,j}$ has logistic distribution and let $\theta_i$ be normally distributed with mean zero and variance $\sigma^2$.

Parameter $a_j$ is item discrimination while $b_j$ is item difficulty.
Interpretation of IRT Parameters

Section 5: The two-parameter logistic (2PL) model

... tradition; the same is true of the symbol $a$ since slopes are usually denoted with $b$ in statistics.

We saw on Figure 6 that the IRF in a 1PL model run parallel to each other and never cross; different difficulty parameters solely shift the curve to the left or to the right while its shape remains unchanged.

$P_i(\theta, b_i, a_i)$

Figure 13: The item response functions of three 2PL items

Flávio Cunha (Rice University) Human Capital Formation in Childhood and Adolescence August 28, 2016 118 / 235
Properties of an Informative IRT Scale

Informative Scale

Probability

Theta

0

-4

-2

0

2

4
Properties of an Informative IRT Scale: IIF
Properties of an Informative IRT Scale: TCC

Informative Scale

- Expected Score vs. Theta

theta

- Model for expected score

2.26

- Scale parameters

45

42.9

37.3

29.8

22.6

15.4

8.05

0
Properties of an Informative IRT Scale: IIF

Informative Scale

- Test information
- Standard error

Theta

Information

Standard Error

-4
-2
0
2
4
0
5
10
15

.4
.5
.6
.7

Flávio Cunha (Rice University) Human Capital Formation in Childhood and Adolescence August 28, 2016 122 / 235
In few words, an informative scale (as presented in the last four graphs) would have items that have good discriminatory power as well as variability in difficulty.

This combination allows us to identify, with a lot of precision, households that have low, medium, and high quality environments.

Unfortunately, the HOME Scale does not have this property.

As I will show below, that are “too many” easy items and “too few” medium and difficult items.

For this reason, the HOME Scale will be able to separate very low quality home environments from okay ones, but it will not have power to separate okay from great home environments.
IRT Properties of Full Scale HOME

Item Characteristic Curve - All Items

PHD Study

Probability

Theta
IRT Properties of Full Scale HOME

Item Information Function - All Items

PHD Study

Information

Theta
IRT Properties of Full Scale HOME

Test Characteristic Curve of the HOME Scale - All Items

PHD Study
Information Function of the HOME Scale - All Items

PHD Study

Test information

Standard error
Why does the IRT Properties of the HOME Matter?

- It probably affects the estimation of the technology of skill formation.
- Why? Medium and high quality environments are difficult to separate.
- It is possible that differences between medium and high quality environments are more (or less) important for child development than differences between medium and low quality environments.
- Either case may lead to biases in the estimation of the technology of skill formation.
Monte Carlo Exercise

- Let $h_1$ denote human capital, $\theta$ denote investments, and $\zeta$ denote uncorrelated shocks. Consider the simple technology of skill formation:

$$h_1 = 1.0 + 0.5\theta - 0.25\theta^2 + \zeta$$

(1)

- To obtain an idea about potential problems of using the HOME as a measure of investment to be used in the estimation of (1):
  - Generate a HOME Scale with desirable IRT properties as the “desired” HOME Scale;
  - Generate a HOME Scale that has “flawed” IRT properties as the “actual” HOME Scale;
  - Estimate $\theta_{desired}$ from “desired” HOME Scale and $\theta_{flawed}$ from “actual” HOME Scale;
  - Regress $h_1$ on quadratic function of $\theta_{desired}$ and compare estimated with true coefficients;
  - Regress $h_1$ on quadratic function of $\theta_{flawed}$ and compare estimated with true coefficients.
Monte Carlo Exercise

**Coeff. on theta, Desired HOME**

- Density
- $b_{1\_inf}$: $0.48, 0.49, 0.5, 0.51, 0.52$
- Kernel: epanechnikov, Bandwidth: 0.0014

**Coeff. on theta squared, Desired HOME**

- Density
- $b_{2\_inf}$: $-0.27, -0.26, -0.25, -0.24, -0.23$
- Kernel: epanechnikov, Bandwidth: 0.0016

**Coeff. on theta, Actual HOME**

- Density
- $b_{1\_uninf}$: $0.43, 0.44, 0.45, 0.46, 0.47, 0.48$
- Kernel: epanechnikov, Bandwidth: 0.0025

**Coeff. on theta squared, Actual HOME**

- Density
- $b_{2\_uninf}$: $-0.36, -0.34, -0.32, -0.3, -0.28$
- Kernel: epanechnikov, Bandwidth: 0.0031
Measuring Quality and Quantity of Time: LENA Pro

![Graph showing words addressed to children by different types of parents over age in months.]

- **13 professional parents**
- **23 working-class parents**
- **6 welfare parents**
Measuring Quality and Quantity of Time: LENA Pro

1. Turn on the DLP and place it in the pocket of the child's LENA clothing.

2. After completing recording, plug the DLP into a PC running LENA Pro. The sophisticated language environment analysis software automatically uploads and processes the audio file.

3. The software generates the LENA reports and other analyses.

4. Export data from LENA Pro to mine your LENA data and perform custom in-depth analyses.
Step 1: Segmenting Audio File

Where it all starts: An audio file

Last night?
You had blueberries?
You didn’t have pancakes last night!

I-I-I did.

Uh huh.
Child: Uh huh.
Noise
In my pancakes.

Noise +
Child: Uh huh! +
Dad (laughing)
You did not!

Silence
Steps 2 and 3: Assign and Confirm

Silence

- Adult
- Male
- Overlap
- Noise
- TV
- Other Child
- Key Child
- Female
- Adult Child

Flávio Cunha (Rice University)
Figure 1. Human and LENA-based AWC estimates for 70 test files.
Measuring Quality and Quantity of Time: LENA Pro

Audio Environment per Hour

Adult Word Counts per Hour

Conversation Turn Counts per Hour

Child Vocalization Counts per Hour
Measuring Quantity of Time: Meaningful Time
Philadelphia Human Development Study

Philadelphia Human Development Study
Audio Environment Data

perc_meaningful  perc_distant  perc_tv  other
Measuring Quality of Time: Conversation Turn Counts
Philadelphia Human Development Study
Measuring Quality and Quantity of Time: LENA Pro
Philadelphia Human Development Study

![Graph showing variance and mean data points.](chart.png)
This dependence between mean and variance (in hours when the mean is high, the variance is also high) is typical in count data.

One may think of taking the natural log of conversation turn counts and proceed with OLS-type analysis.

Not a good idea with count data:

- There are many zeros; taking the logs will eliminate the zeros from the analysis, which means it reduces cases of poor language environment.
- We want to identify households in terms of expected number of counts, not the expected log of number of counts (nonlinear transformation).
Analysis of LENA Conversation Turn Counts Data

- Let $Y_{i,j}$ denote the $j$th observation on conversation turn counts between an adult and child $i$.

- Because these are counts, we model each observation as a Poisson random variable with parameter $\epsilon_i \lambda_{i,j}$ where $\epsilon_i$ is a random effect term and $\lambda_{i,j}$ is such that:

$$\ln \lambda_{i,j} = X_{i,j} \delta_j + \ln s_{i,j}$$ (2)

- Vector $X_{i,j}$ contains variables that describe the context of measurement and $s_{i,j}$ is “exposure” (i.e., number of seconds that the LENA device was on during the $j$th measurement).
LENA Measurement in Practice

8:00  9:00  10:00  11:00  12:00  13:00  14:00  15:00  16:00  17:00  19:00  20:00  21:00  22:00

Child #1  Child #2
Conditional on $\epsilon_i$, the probability of observing a count equal to:

$$
\Pr (y_{i,j} | \epsilon_i) = \frac{(\epsilon_i \lambda_{i,j})^{y_{i,j}}}{y_{i,j}!} e^{-\epsilon_i \lambda_{i,j}}
$$

where $\Pr (y_{i,j} | \epsilon_i) = \Pr (Y_{i,j} = y_{i,j} | \epsilon_i)$ is the probability that the count of variable $Y_{i,j}$ is equal to $y_{i,j}$ conditional on $\epsilon_i$.

Assume that, conditional on $\epsilon_i$, the events are independent. Thus:

$$
\Pr (y_{i,1}, \ldots, y_{i,J} | \epsilon_i) = \left\{ \frac{\prod_{j=1}^{J} (\lambda_{i,j})^{y_{i,j}}}{y_{i,j}!} \right\}^{\epsilon_i} \sum_{\sum_{j=1}^{J} y_{i,j}} e^{-\epsilon_i \sum_{j=1}^{J} \lambda_{i,j}}
$$

Because we don’t observe the random effect $\epsilon_i$, we need to integrate it out.

We assume that $\epsilon_i$ has gamma distribution with mean one and variance $\frac{1}{\alpha}$.
Density of Quality of Interaction

The graph shows the density of interactions measured on the Y-axis against some variable on the X-axis. The density peaks around the 200 mark and gradually decreases as the variable increases, reaching a minimum near 1000.
Let $M_{i,j}$ denote the share of meaningful time of adult-child interaction in $j$th observation.

Because these are proportion data, we model each observation as the following logistic regression:

$$\ln \left\{ \frac{M_{i,j}}{1 - M_{i,j}} \right\} = X_{i,j} \rho_j + \mu_i + \nu_{i,j}$$

where $\mu_i$ is a random effect with mean zero and variance $\sigma_{\mu}^2$.

We are interested in estimating the unobserved heterogeneity captured by $\mu_i$ across families.
Quality and Quantity of Interaction by HOME

PHD Study

Quality of Interaction

Quantity of Interaction

Low HOME Score
Medium HOME Score
High HOME Score
In the standard measurement paradigm:

- The goal is to measure a trait or skill $\theta_i$ (could be scalar or vector).
- Each task is a self-contained situation.
- The task evokes a response meant to provide evidence about the construct.
- Each response is evaluated to provide an item score.
- A test score accumulates evidence over the many different tasks.
Say: Point to mowing.
1) At Joe's Restaurant, one-fourth of the patrons are male and one-fifth of the patrons are from out of town. What proportion would you expect to be male and out of town?

○ 1/5
○ 1/10
○ 1/20
○ 1/25
○ 1/50
The Big Five Inventory (BFI)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

<table>
<thead>
<tr>
<th>Disagree strongly</th>
<th>Disagree a little</th>
<th>Neither agree nor disagree</th>
<th>Agree a little</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

I see Myself as Someone Who...

- _____1. Is talkative
- _____2. Tends to find fault with others
- _____3. Does a thorough job
- _____4. Is depressed, blue
- _____23. Tends to be lazy
- _____24. Is emotionally stable, not easily upset
- _____25. Is inventive
- _____26. Has an assertive personality
1. When it is mealtime, how often does your child eat what you want him/her to eat?

1  2  3  4  5
Almost  Less than  1/2 the time  More than  Almost
Never   1/2 the time  time         1/2 the time  Always

2. When your child doesn't eat what you want him/her to eat and you tell him/her to do so, how often does he/she obey and eat?

1  2  3  4  5
Almost  Less than  1/2 the time  More than  Almost
Never   1/2 the time  time         1/2 the time  Always
Within a substance, atoms that collide frequently and move independently of one another are *most* likely in a

A  liquid.
B  solid.
C  gas.
D  crystal.
Concerns about Standardized Tests
Concerns about Standardized Tests

- To be able to do science, we need to have:
  - Clear understanding of goals;
  - Motivation;
  - Collaboration skills;
  - Understanding of steps and beliefs;
  - Science process skills;
  - Science content knowledge.

- Traditional tests measure only the science knowledge.
- Simulation tests aim to measure all of the factors of interest.
Concerns about Self Reports from Surveys

- Motivated responding (e.g., wanting to look good, or bad)
- Reference group bias (to whom you compare yourself)
- Response style bias (e.g., extreme responses, modesty)
- Lack of differentiation in others’ ratings (halo, horn)
- Cross-cultural comparability (to compare countries x and y)
Concerns about Self Reports from Surveys

Within-country

Pooled sample (3 countries)

Country means (63 countries)
Peg Solitaire Game

- Very interesting work by Rafferty, LaMar, and Griffiths (2015)
- The goal is to leave as few pegs on the board as possible.
- Jump pegs to remove them.
- Way to evaluate an individual’s decision-making process, strategic thinking, motivation.

![Peg Solitaire Game Diagram]
State Space

22 Reachable States
States

Each state presents a choice:

Available Actions $A_s$

- $(3,3) \rightarrow (1,3)$
- $(3,3) \rightarrow (5,3)$
- $(3,2) \rightarrow (3,4)$
- $(4,3) \rightarrow (2,3)$

Score
Reset
Markov Decision Process

State Space: $S = \{s_1, s_2, ..., s_S\}$

Action Set: $A = \{a_1, a_2, ..., a_A\}$

Transition Function: $T(s, a, s') = p(s'|s, a)$

Reward Structure: $R(s, a, s')$

Policy: $p(a|s, \xi)$
Let $V(s|a)$ denote the value of choosing action $a$ when state is $s$:

$$V(s|a) = R(a|s) + p(s'|s,a) V(s'|a)$$

where $R(a|s)$ is the reward of choosing action $a$ and $p(s'|s,a)$ is the state transition function.
Let $\theta_i \in [0, \infty)$ denote the individual “capability”. Then, define

$$\Pr \left( \text{chooses } a | s \right) = \frac{e^{\theta_i V(s|a)}}{\sum_b e^{\theta_i V(s|b)}}$$

If $\theta_i = 0$,

$$\Pr \left( \text{chooses } a | s \right) = \frac{1}{\sum_b}$$

As $\theta_i$ approaches infinity:

$$\Pr \left( \text{chooses } a | s \right) = \begin{cases} 
0, & \text{if } a \text{ is not optimal} \\
1, & \text{if } a \text{ is optimal}
\end{cases}$$
Flávio Cunha (Rice University)  Human Capital Formation in Childhood and Adolescence  August 28, 2016  167 / 235

Statistics: MDP / Economics: DDCM
Rafferty, LaMar, and Griffiths (2015)
Challenge: Clouds over Jackson City

Welcome to Jackson City!

We're so glad you're here. For the last ten years our city has been growing— a boom town, you might say! But lately some of our citizens have started to complain about the air pollution. We need to REDUCE air pollution (icamente) without LOSING jobs (economically). What do you say, are you up for it?

http://vimeo.com/69007945
DDCM Applied to SimCity
Rafferty, LaMar, and Griffiths (2015)

• Designed to assess systems thinking.
• There are 17 actions and 25,420 reachable states.
• Students must optimize two variables simultaneously.
A New class Pet

Have you ever considered getting a new class pet? Well, if you happen to be searching for one now I’ve got a few suggestions for you. If you’re looking for a pet that’s really easy to take care of try a chameleon. Just get a jar or tank and put in a few twigs, a bit of grass, some leaves, and plenty of live prey. And if that’s not what you had in mind and want a more active animal try a rabbit. They are cute, cuddly, and active. You can buy rabbit mix at Walmart or at a pet shop.

But if you want to traumatize the children (which I hope you don’t) get an untamed king cobra. If you wanted something more like a dog you could get a golden retriever in my opinion the are the best kind of dog. They are gentle and calm and love children.

But I will always say that a rat is the best pet. Not a greasy, slimy, bloodsucking...
<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency Spectrum $V(m,N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>$V(m,N)$</td>
</tr>
<tr>
<td>1</td>
<td>94</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
</tr>
</tbody>
</table>
Estimating Vocabulary Size

- Think of writing text as selecting words from an urn.
- Let $\pi_i$ denote the probability of selecting word $\omega_i$ from the urn.
- The frequency of word $\omega_i$ in a sample of $N$ tokens is binomially distributed $(N, \pi_i)$ distributed.
Expected Frequency Spectrum (expected number of words that will be used $m$ times in a text of $N$ words):

$$E[V(m, N)] = \sum_{i=1}^{S} \binom{N}{m} \pi_i^m (1 - \pi_i)^{N-m}$$

Expected Vocabulary Size (given a text of $N$ words):

$$E[V(N)] = \sum_{m=1}^{N} \sum_{i=1}^{S} \binom{N}{m} \pi_i^m (1 - \pi_i)^{N-m}$$
Estimating Vocabulary Size

- Formulas are not too helpful because we can’t use observed data to estimate \( \pi_i \) (too many words in the urn, regardless of size \( N \), the sample will always be too small).
- Make distributional assumptions (requires lots and lots of text data - thousands of words, which is unlike to get from children in Grade 1).
Estimating Vocabulary Size

- Use outside data to estimate $\pi_i$.
  - Web 1T 5-gram Version 1: 1 trillion tokens from the web.

- Text “New Class Pet” is from 3rd grade student, but vocabulary is more like someone in 5th grade.

- Possible to use tools to analyze grammar.
Estimating Vocabulary Size

- Use outside data to estimate \( \pi_i \).
  - The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century.
  - Use Adam Kilgarriff estimates of \( \pi_i \) (http://www.kilgarriff.co.uk/bnc-readme.html).
- Text “New Class Pet” is from 3rd grade student, but vocabulary is more like someone in 5th grade.
- Possible to use tools to analyze grammar.
We can compare this to the analysis carried out by the Boston Globe (Boston Globe) using the Flesch-Kincaid measure on the candidates’ 2015 speeches as shown in Figure 2. They performed their analysis only on each candidate’s campaign announcements. It would appear that an analysis more geared toward spoken language gives both Mr. Trump and Mrs. Clinton higher scores for their choice of words.
Presidential Campaign: Grammar

Figure 3. REAP lexical measure standard deviation per candidate

Figure 3 shows the standard deviation of the scores in Figure 1. This reveals the degree to which the candidate changes their choice of words from one speech to another. This could reflect an effort to take into account the different audiences or circumstances (winning or concession speech in a state, for example). We can see that Hillary Clinton has the highest standard deviation and so the biggest change of choice of words from one speech to another, while Ted Cruz varies the least in his choices.

We also compared the grammar levels for all of the candidates and past presidents as shown in Figure 4.

Figure 4. REAP grammar measure

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Lincoln</th>
<th>Reagan</th>
<th>Clinton</th>
<th>GWBush</th>
<th>Obama</th>
<th>Cruz</th>
<th>H Clinton</th>
<th>Rubio</th>
<th>Sanders</th>
<th>Trump</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>2.0000</td>
<td>4.0000</td>
<td>6.0000</td>
<td>8.0000</td>
<td>10.0000</td>
<td>12.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Benefits of Text Data

- Better measure of communication.
- Usually not used to evaluate teachers and schools, so not so sensitive to “teaching to test.”
02 Thinking about the mathematics teacher who taught your last mathematics class. To what extent do you agree with the following statements?
(Please check only one box on each row.)

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) My teacher lets students know they need to work hard.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Teacher Support Scale
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.
*(Please check only one box on each row.)*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. <strong>Ms. Anderson is concerned about her students’ learning.</strong></td>
<td>☐️</td>
<td>☐️</td>
<td>☐️</td>
<td>☐️</td>
</tr>
<tr>
<td>b) Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. <strong>Mr. Crawford is concerned about his students’ learning.</strong></td>
<td>☐️</td>
<td>☐️</td>
<td>☐️</td>
<td>☐️</td>
</tr>
<tr>
<td>c) Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. <strong>Ms. Dalton is concerned about her students’ learning.</strong></td>
<td>☐️</td>
<td>☐️</td>
<td>☐️</td>
<td>☐️</td>
</tr>
</tbody>
</table>
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

*(Please check only one box on each row.)*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Homework Frequency</th>
<th>Answers Back</th>
<th>Concerned About Students’ Learning</th>
<th>Student “A’s” responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Anderson</td>
<td>Assigns homework every other day. She always gets the answers back to students before examinations.</td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Disagree</td>
</tr>
<tr>
<td>Mr. Crawford</td>
<td>Assigns mathematics homework once a week. He always gets the answers back to students before examinations.</td>
<td>Disagree</td>
<td>Strongly disagree</td>
<td>Agree</td>
</tr>
<tr>
<td>Ms. Dalton</td>
<td>Assigns mathematics homework once a week. She never gets the answers back to students before examinations.</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly disagree</td>
</tr>
</tbody>
</table>
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. <strong>Ms. Anderson is concerned about her students’ learning.</strong></td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. <strong>Mr. Crawford is concerned about his students’ learning.</strong></td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. <strong>Ms. Dalton is concerned about her students’ learning.</strong></td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

02 My teacher lets students know they need to work hard.

For Student “A” this can be interpreted as “at the same level as the best hypothetical teacher”
01 Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. <strong>Ms. Anderson is concerned about her students’ learning.</strong></td>
<td>☒,</td>
<td>☐,</td>
<td>☐,</td>
<td>☐,</td>
</tr>
<tr>
<td>b) Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. <strong>Mr. Crawford is concerned about his students’ learning.</strong></td>
<td>☐,</td>
<td>☒,</td>
<td>☐,</td>
<td>☐,</td>
</tr>
<tr>
<td>c) Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. <strong>Ms. Dalton is concerned about her students’ learning.</strong></td>
<td>☐,</td>
<td>☐,</td>
<td>☐,</td>
<td>☒,</td>
</tr>
</tbody>
</table>

02 My teacher lets students know they need to work hard.

**For Student “A” this can be interpreted as “like the middle hypothetical teacher”**
Other Measurements Using Technology

- Altruism
- Reciprocity
- Risk Preference
- Patience
- Trust
- Decision Making Capacity (Rationality)
Introducing Heterogeneity in Beliefs

Figure Choice Set and Preferences

Red: Mother underestimates returns
Blue: Mother has unbiased returns

Flávio Cunha (Rice University)

Human Capital Formation in Childhood and Adolescence
August 28, 2016 186 / 235
Why Heterogeneity in Beliefs?

- Time spent in activities that are appropriate for the child’s age (Kalil et al, 2012).
- Home visitation programs on parenting:
  - Nurse-Family Partnership (Olds et al, 2012).
  - HIPPY Program (Baker et al, 2002).
  - Parent as Teachers (PAT, Wagner et al, 1998)
  - Thirty Million Words Program (Suskind and Lefler, 2013).
  - Many others (Healthy Families, Healthy Start, CHIP of Virginia, MOM of Philadelphia, etc.)
Current Research

- My current research aims to answer the following questions:
  - Can we measure parental beliefs about the technology of skill formation?
  - If so:
    - How do parental beliefs compare with objective estimates of the technology of skill formation?
    - Is there heterogeneity in parental beliefs?
  - If so, does the heterogeneity in beliefs predict heterogeneity in investments?
  - If so, can we change parental investments by affecting parental beliefs?
The technology of skill formation is:

$$\ln h_{i,1} = \psi_0 + \psi_1 \ln h_{0,i} + \psi_2 \ln x_i + \psi_3 \ln h_{0,i} \ln X_i + \nu_i$$
Let $\Psi_i$ denote the mother’s information set.

Let $E(\psi_j|h_{0,i}, x_i, \Psi_i) = \mu_{i,j}$ and assume that $E(\nu_i|\Psi_i) = 0$.

From the point of view of the mother:

$$E(\ln h_{i,1}|h_{0,i}, x_i, \Psi_i) = \mu_{i,0} + \mu_{i,1} \ln h_{0,i} + \mu_{i,2} \ln x_i + \mu_{i,3} \ln h_{0,i} \ln x_i$$
Consider a simple static model. Parent’s utility is:

\[ u(c_i, h_i; \alpha_{i,1}, \alpha_{i,2}) = \ln c_i + \alpha_{i,1} \ln h_i + \alpha_{i,2} \ln x_i \]

Budget constraint is:

\[ c_i + px_i = y_i. \]
The problem of the mother is to maximize expected utility subject to the mother’s information set, the budget constraint, and the technology of skill formation.

The solution is

\[ x_i = \left[ \frac{\alpha_{i,1} (\mu_{i,2} + \mu_{i,3} \ln h_{0,i}) + \alpha_{i,2}}{1 + \alpha_{i,1} (\mu_{i,2} + \mu_{i,3} \ln h_{0,i}) + \alpha_{i,2}} \right] \frac{y_i}{p} \]

Clearly, we cannot separately identify \( \alpha_i \) from \( \mu_{i,\gamma} \) if we only observe \( x_i, y_i, \) and \( p \).
Identification

- Elicit maternal beliefs.
- Elicit maternal preferences.
- Estimate the technology of skill formation.
Eliciting beliefs: Steps

- Measure actual child development: MSD and Item Response Theory (IRT).
- Develop the survey instrument to elicit beliefs $E \left[ \ln h_{i,1} \mid h_0, x, \psi_i \right]$:  
  - Reword MSD items.
  - Create hypothetical scenarios of $h_0$ and $x$.
- Estimate beliefs from answers allowing for error in responses.
SECTION 3: MOTOR AND SOCIAL DEVELOPMENT

PART H: (22 MONTHS - 3 YEARS, 11 MONTHS)

MOTHER/GUARDIAN:

If ___________________________ is at least 22 months old, but not yet 4 years old, please answer these 15 questions.

Child's Name

1. Has your child ever let someone know, without crying, that wearing wet (soiled) pants or diapers bothered him/her?  
   YES.... 1  
   NO..... 0  
   72/

2. Has your child ever spoken a partial sentence of 3 words or more?  
   YES.... 1  
   NO..... 0  
   73/

3. Has your child ever walked upstairs by himself/herself without holding on to a rail?  
   YES.... 1  
   NO..... 0  
   74/

4. Has your child ever washed and dried his/her hands without any help except for turning the water on and off?  
   YES.... 1  
   NO..... 0  
   75/

5. Has your child ever counted 3 objects correctly?  
   YES.... 1  
   NO..... 0  
   76/
Eliciting beliefs: Item response theory

- Let $d_{i,j}^* = b_{0,j} + b_{1,j} \left( \ln a_i + \frac{b_{2,j}}{b_{1,j}} \theta_i \right) + \eta_{i,j}$
- We observe $d_{i,j} = 1$ if $d_{i,j}^* \geq 0$ and $d_{i,j} = 0$, otherwise.
- Measure of (log of) human capital: $\ln h_i = \ln a_i + \frac{b_{2,j}}{b_{1,j}} \theta_i$.
- In this sense, $\theta_i$ is deviation from typical development for age.
Figure 4
Probability as a Function of Child's Age

- **Speak partial sentence, data**
- **Speak partial sentence, predicted**
Eliciting beliefs: Changing wording of the MSD Instrument

- In order to measure $E \left[ \ln h_{i,1} \mid h_0, x, \psi_i \right]$, we take the tasks from the MSD Scale, but instead of asking: “Has your child ever spoken a partial sentence with three words or more?”, we ask:
  - Method 1: How likely is it that a baby will speak a partial sentence with three words or more by age 24 months?
  - Method 2: What is the youngest and oldest age a baby learns to speak a partial sentence with three words or more?
Eliciting beliefs: Scenarios of human capital and investments

We consider four scenarios:

- Scenario 1: Child is healthy at birth (e.g., normal gestation, birth weight, and birth length) and investment is high (e.g., six hours per day).
- Scenario 2: Child is healthy at birth and investment is low (e.g., two hours per day).
- Scenario 3: Child is not healthy at birth (e.g., premature, low birth weight, and small at birth) and investment is high.
- Scenario 4: Child is not healthy at birth and investment is low.

Scenarios are described to survey respondents through a video.
Method 1: How likely is it that a baby will speak a partial sentence with three words or more by age 24 months?

Let’s say that when investment is high – that is, when $x = \bar{x}$ – the mother states that there is a 75% chance that the child will learn how to speak a partial sentence with three words or more.

And when investment is low – that is, when $x = \underline{x}$ – the mother states that there is a 25% chance that the child will learn how to speak a partial sentence with three words or more.

We convert this probability statement into an age-equivalent statement using the NHANES data.
Figure 4

Probability as a Function of Child's Age

- **Probability**
  - 0
  - 0.25
  - 0.5
  - 0.75
  - 1

- **Child Age (Months)**
  - 0
  - 4
  - 8
  - 12
  - 16
  - 20
  - 24
  - 28
  - 32
  - 36
  - 40
  - 44
  - 48

- **Legend**:
  - ○ Speak partial sentence, data
  - – Speak partial sentence, predicted
Method 2: Transforming age ranges into probabilities

- **Method 2:** What is the youngest and oldest age a baby learns to speak a partial sentence with three words or more?

- Let’s say that when investment is high, so that $x = \bar{x}$, the mother states that the youngest and oldest ages a baby will learn how to speak a sentence with three words or more are, respectively, 18 and 28 months.

- And when investment is low, so that $x = \underline{x}$, the mother states that the ages are 20 and 30 months.

- We need to transform the age ranges into probabilities. We use the age ranges to estimate a mother-specific IRT model.
Figure 3
Transforming age range into probability

![Graph showing probability vs age in months with a point marked as High investment.](image-url)
Figure 3
Transforming age range into probability

- Logistic prediction, high
- High investment
Figure 3
Transforming age range into probability

Logistic prediction, high
High investment
Figure 3
Transforming age range into probability

Probability
0 .25 .5 .75 1
Age (in months)
Logistic prediction, high High investment
Logistic prediction, low Low investment
Transforming age range into probability

Logistic prediction, high  High investment
Logistic prediction, low  Low investment
Transforming age range into probability

Figure 3
Method 2: Transforming probabilities into mean beliefs

- Method 2: Given scenario for $h_0$ and $x$, how likely is it that a baby will speak a partial sentence with three words or more by age 24 months?

  Given maternal supplied age range and the logistic assumption, we conclude that when $x = \bar{x}$, the mother believes that there is a 75% chance that the child will learn how to speak a partial sentence with three words or more.

  Analogously, when $x = \bar{x}$, the mother believes that there is a 25% chance that the child will learn how to speak a partial sentence with three words or more.

  We convert this probability statement into an age-equivalent statement using the NHANES data.
Figure 3

Expected development for two levels of investments (x)

Age range to probability
Speak partial sentence - MKIDS

Probability to expected development
Speak partial sentence - NHANES

Expected development for two levels of investments (x)
Recovering mean beliefs: Measurement error model

- Let $\ln q^L_{i,j,k}$ denote an error-ridden measure of $E[\ln h_{i,1} | h_{0,k}, x_k, \psi_i]$ generated by “how likely” questions:

$$\ln q^L_{i,j,k} = E[\ln h_{i,1} | h_{0,k}, x_k, \psi_i] + \epsilon^L_{i,j,k}.$$ 

- Let $\ln q^A_{i,j,k}$ denote an error-ridden measure of $E[\ln h_{i,1} | h_{0,k}, x_k, \psi_i]$ generated by “age range” questions:

$$\ln q^A_{i,j,k} = E[\ln h_{i,1} | h_{0,k}, x_k, \psi_i] + \epsilon^A_{i,j,k}.$$ 

- For each scenario, we have multiple measures of the same underlying latent variable.
Recovering mean beliefs:

- Use technology of skill formation, and the mother’s information set, to obtain:

\[
\ln q_{i,j,k}^L = \mu_{i,0} + \mu_{i,1} \ln h_{0,k} + \mu_{i,2} \ln x_k + \mu_{i,3} \ln h_{0,k} \ln x_k + \epsilon_{i,j,k}^L.
\]

\[
\ln q_{i,j,k}^A = \mu_{i,0} + \mu_{i,1} \ln h_{0,k} + \mu_{i,2} \ln x_k + \mu_{i,3} \ln h_{0,k} \ln x_k + \epsilon_{i,j,k}^A.
\]

- We have a factor model where:
  - \( \mu_i = (\mu_{i,0}, \mu_{i,1}, \mu_{i,2}, \mu_{i,3}) \) are the latent factors;
  - \( \lambda_k = (1, h_{0,k}, \ln x_k, \ln h_{0,k} \ln x_k) \) are the factor loadings;
  - \( \epsilon_{i,j,k} = (\epsilon_{i,j,k}^L, \epsilon_{i,j,k}^A) \) are the uniquenesses.
Eliciting beliefs: Intuitive explanation

- Let $E[\ln h_{i,1} | h_0, h, \Psi_i]$ denote maternal expectation of child development at age 24 months conditional on the child’s initial level of human capital, investments, and the mother’s information set.

- Assume, for now, technology is Cobb-Douglas.

- Suppose we measure $E[\ln h_{i,1} | h_0, x, \Psi_i]$ at two different levels of investments:

  $$E[\ln h_{i,1} | h_0, x, \Psi_i] = \mu_{i,0} + \mu_{i,1} \ln h_0 + \mu_{i,2} \ln x$$

  $$E[\ln h_{i,1} | h_0, \bar{x}, \Psi_i] = \mu_{i,0} + \mu_{i,1} \ln h_0 + \mu_{i,2} \ln \bar{x}$$

- Subtracting and re-organizing terms:

  $$\mu_{i,2} = \frac{E[\ln h_{i,1} | h_0, \bar{x}, \Psi_i] - E[\ln h_{i,1} | h_0, x, \Psi_i]}{\ln \bar{x} - \ln x}$$
Important issue

- We could use only one MSD item to elicit beliefs.
- But, if we use more items, we can relax assumptions about measurement error.
- And, we can check for consistency in answers.
Figure 5
Comparing answers across scenarios

Age range into probability

Probability into expected development

Speaks partial sentence
Knows own age and sex

Comparing answers across scenarios

Figure 5
The investment policy function is:

\[ x_i = \left[ \frac{\alpha_{i,1} (\mu_{i,2} + \mu_{i,3} \ln h_{0,i}) + \alpha_{i,2}}{1 + \alpha_{i,1} (\mu_{i,2} + \mu_{i,3} \ln h_{0,i}) + \alpha_{i,2}} \right] \frac{y_i}{p} \]

where \( \alpha_{i,1} \) and \( \alpha_{i,2} \) captures heterogeneity in preferences.

The usual procedure is to work with observed investment data.

We are in the field collecting these investment data.
Estimation of Preferences

- Today, we elicit the preference parameters by stated-choice data (as it is commonly applied in Marketing).
- We tell the respondent to assume that the child’s initial level of human capital is high.
- Then, we create nine hypothetical scenarios of monthly income and prices:

<table>
<thead>
<tr>
<th>Income</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1500</td>
<td>$30</td>
</tr>
<tr>
<td>$2000</td>
<td>$45</td>
</tr>
<tr>
<td>$2500</td>
<td>$60</td>
</tr>
<tr>
<td>Scenario 1</td>
<td>Scenario 2</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>Scenario 5</td>
</tr>
<tr>
<td>Scenario 7</td>
<td>Scenario 8</td>
</tr>
</tbody>
</table>
Estimation of Preferences

In order to link investment to time, we prepared a three-minute video in which we explain to the respondent that the more time that the mother interacts with the child, the more money she has to spend every month buying educational goods such as child books and educational toys.

Our goal is to pass on to the respondent the idea that investment is costly.

Respondents are not familiar with the concept of “opportunity cost.”
Estimation of Preferences

- For each combination of prices and income, we ask the respondents the following question: *Suppose that your household income is $y per month and that for each hour per day that the mother spends interacting with the child she has to spend $p per month on educational goods. Consider the following four options:*

- The four options correspond to two, three, four, and five hours of investments per day.

- Thus, if the respondent reports $x_{i,m,n}$ hours of investment per day when price is $p_m$ and income is $y_n$, then share of income allocated to investments, $s_{m,n}$ is:

$$s_{i,m,n} = \frac{p_m x_{i,m,n}}{y_n}$$
Note that the ratio, \( r_{i,m,n} \) is:

\[
\frac{s_{i,m,n}}{1-s_{i,m,n}} = \alpha_{i,1} \left( \mu_{i,2} + \mu_{i,3} \ln h_{0,i} \right) + \alpha_{i,2} + \xi_{i,m,n}
\]

The parameters \( \alpha_{i,1} \) and \( \alpha_{i,2} \) can be estimated as a simple random-effects model.
Descriptive Information about Participants: MKIDS and PHD

Pilot Study: Maternal Knowledge of Infant Development Study (MKIDS)

- 777 participants, all African-American.
- MKIDS: 60% are primiparous; PHD: 100% are primiparous.
- 80% are single (not cohabiting or married).
- 80% are at most 25 years-old.
- Median household income is below the second decile of U.S. distribution.
- Low education sample: only 12% of respondents have a two-year college degree or more.
<table>
<thead>
<tr>
<th>Number of observations</th>
<th>MKIDS</th>
<th>PHD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td><strong>Type of Elicitation Method</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only probability</td>
<td>20</td>
<td>6.2</td>
<td>0</td>
</tr>
<tr>
<td>Only age ranges</td>
<td>233</td>
<td>72.1</td>
<td>0</td>
</tr>
<tr>
<td>Both methods</td>
<td>70</td>
<td>21.7</td>
<td>454</td>
</tr>
<tr>
<td><strong>MSD Items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wearing wet pants bothers child</td>
<td>323</td>
<td>100.0</td>
<td>0</td>
</tr>
<tr>
<td>Speak partial sentence</td>
<td>323</td>
<td>100.0</td>
<td>454</td>
</tr>
<tr>
<td>Say first and last name</td>
<td>323</td>
<td>100.0</td>
<td>454</td>
</tr>
<tr>
<td>Count 3 objects correctly</td>
<td>323</td>
<td>100.0</td>
<td>454</td>
</tr>
<tr>
<td>Know own age and sex</td>
<td>323</td>
<td>100.0</td>
<td>454</td>
</tr>
<tr>
<td>Says the names of 4 colors</td>
<td>323</td>
<td>100.0</td>
<td>0</td>
</tr>
<tr>
<td>Count out loud up to 10</td>
<td>323</td>
<td>100.0</td>
<td>0</td>
</tr>
<tr>
<td>Draw picture of man/woman</td>
<td>323</td>
<td>100.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Hypothetical scenarios</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>158</td>
<td>48.9</td>
<td>454</td>
</tr>
<tr>
<td>Alternative scenario #1</td>
<td>42</td>
<td>13.0</td>
<td>0</td>
</tr>
<tr>
<td>Alternative scenario #2</td>
<td>91</td>
<td>28.2</td>
<td>0</td>
</tr>
<tr>
<td>Alternative scenario #3</td>
<td>32</td>
<td>9.9</td>
<td>0</td>
</tr>
<tr>
<td><strong>Stated choice data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothetical scenarios for prices of investment and income</td>
<td>158</td>
<td>48.9</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 2
Basic Features of Raw Data

<table>
<thead>
<tr>
<th>Rank</th>
<th>Item Description</th>
<th>Probability Scenarios</th>
<th>Age ranges Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NHANES Obs.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1  2  3  4</td>
<td>1  2  3  4</td>
</tr>
<tr>
<td>1</td>
<td>Chils lets someone know that wearing wet pants bothers him/her?</td>
<td>0.78 0.55 0.70 0.51</td>
<td>0.64 0.55 0.50 0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.24) (0.27) (0.27) (0.26)</td>
<td>(0.33) (0.36) (0.36) (0.37)</td>
</tr>
<tr>
<td>2</td>
<td>Child speaks a partial sentence of 3 words or more</td>
<td>0.81 0.63 0.61 0.45</td>
<td>0.60 0.44 0.42 0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.18) (0.22) (0.20) (0.20)</td>
<td>(0.36) (0.38) (0.38) (0.36)</td>
</tr>
<tr>
<td>3</td>
<td>Child counts 3 objects correctly?</td>
<td>0.84 0.67 0.62 0.47</td>
<td>0.41 0.32 0.26 0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.18) (0.22) (0.20) (0.20)</td>
<td>(0.38) (0.36) (0.33) (0.30)</td>
</tr>
<tr>
<td>4</td>
<td>Child knows own age and sex</td>
<td>0.83 0.66 0.62 0.47</td>
<td>0.33 0.26 0.23 0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.19) (0.23) (0.21) (0.21)</td>
<td>(0.36) (0.33) (0.31) (0.29)</td>
</tr>
<tr>
<td>5</td>
<td>Child says first and last name together without someone's help</td>
<td>0.80 0.64 0.60 0.46</td>
<td>0.31 0.24 0.22 0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.20) (0.22) (0.21) (0.21)</td>
<td>(0.36) (0.33) (0.31) (0.29)</td>
</tr>
<tr>
<td>6</td>
<td>Child says the names of at least 4 colors</td>
<td>0.81 0.59 0.74 0.56</td>
<td>0.26 0.22 0.19 0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.23) (0.28) (0.22) (0.27)</td>
<td>(0.31) (0.29) (0.28) (0.26)</td>
</tr>
<tr>
<td>7</td>
<td>Child counts out loud up to 10?</td>
<td>0.80 0.58 0.75 0.53</td>
<td>0.24 0.20 0.19 0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.20) (0.27) (0.19) (0.27)</td>
<td>(0.30) (0.28) (0.28) (0.27)</td>
</tr>
<tr>
<td>8</td>
<td>Child draws a picture of a man/woman, 2 parts besides head</td>
<td>0.71 0.51 0.67 0.48</td>
<td>0.15 0.15 0.14 0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.25) (0.28) (0.21) (0.26)</td>
<td>(0.26) (0.26) (0.26) (0.25)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis.
Beliefs about the technology of skill formation

Table 3
Maternal Beliefs about the Technology of Skill Formation

<table>
<thead>
<tr>
<th></th>
<th>25th percentile</th>
<th>Median</th>
<th>75th percentile</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\psi,0}$</td>
<td>-0.015</td>
<td>0.101</td>
<td>0.236</td>
<td>0.115</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\mu_{\psi,1}$</td>
<td>0.077</td>
<td>0.296</td>
<td>0.554</td>
<td>0.365</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.016)</td>
<td>(0.022)</td>
<td>(0.016)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>$\mu_{\psi,2}$</td>
<td>0.065</td>
<td>0.166</td>
<td>0.285</td>
<td>0.192</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\mu_{\psi,3}$</td>
<td>-0.008</td>
<td>0.094</td>
<td>0.335</td>
<td>0.190</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.010)</td>
<td>(0.024)</td>
<td>(0.020)</td>
<td>(0.051)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis.
### Sensitivity analysis

#### Table 4

**Alternative Definition of Scenarios and Maternal Beliefs**

<table>
<thead>
<tr>
<th>Regressors</th>
<th>$\mu_{\psi,0}$</th>
<th>$\mu_{\psi,1}$</th>
<th>$\mu_{\psi,2}$</th>
<th>$\mu_{\psi,3}$</th>
<th>(p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (baseline)</td>
<td>0.018</td>
<td>0.147</td>
<td>0.112</td>
<td>0.070</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.043)</td>
<td>(0.022)</td>
<td>(0.062)</td>
<td>-</td>
</tr>
<tr>
<td>Dummy for alternative</td>
<td>0.067</td>
<td>-0.027</td>
<td>-0.032</td>
<td>-0.081</td>
<td>1.080</td>
</tr>
<tr>
<td>scenario #1</td>
<td>(0.037)</td>
<td>(0.094)</td>
<td>(0.048)</td>
<td>(0.136)</td>
<td>(0.364)</td>
</tr>
<tr>
<td>Dummy for alternative</td>
<td>0.280</td>
<td>0.469</td>
<td>0.175</td>
<td>0.424</td>
<td>33.910</td>
</tr>
<tr>
<td>scenario #2</td>
<td>(0.028)</td>
<td>(0.071)</td>
<td>(0.037)</td>
<td>(0.103)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Dummy for alternative</td>
<td>0.206</td>
<td>0.027</td>
<td>0.051</td>
<td>0.091</td>
<td>6.750</td>
</tr>
<tr>
<td>scenario #3</td>
<td>(0.041)</td>
<td>(0.104)</td>
<td>(0.054)</td>
<td>(0.152)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis, except in the F-test column where we report p-values.
## Preferences

<table>
<thead>
<tr>
<th></th>
<th>25th percentile</th>
<th>Median</th>
<th>75th percentile</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{i,1}$</td>
<td>0.0261 (0.0004)</td>
<td>0.0312 (0.0002)</td>
<td>0.0400 (0.0007)</td>
<td>0.0313 (0.0004)</td>
<td>0.0002 (0.0000)</td>
</tr>
<tr>
<td>$\alpha_{i,2}$</td>
<td>0.0669 (0.0005)</td>
<td>0.0777 (0.0008)</td>
<td>0.0942 (0.0007)</td>
<td>0.0795 (0.0005)</td>
<td>0.0003 (0.0000)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis.

Table 5
Maternal Beliefs about the Technology of Skill Formation
This table shows the comparative statics of optimal investments in relation to preference and belief parameters. Each row shows what happens to investments as we move one parameter and fix the other parameters at the median value. In the last row, we replace the human capital at birth from the mean value to the value at the first percentile.
<table>
<thead>
<tr>
<th>Cases</th>
<th>Factual investment</th>
<th>Counterfactual investment</th>
<th>% Change</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\psi,2} = 0.267$</td>
<td>1.84</td>
<td>1.92</td>
<td>4.4%</td>
<td>10.3%</td>
</tr>
<tr>
<td>$\mu_{\psi,3} = 0.000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_{\psi,2} = 0.454$</td>
<td>1.84</td>
<td>2.05</td>
<td>11.7%</td>
<td>26.9%</td>
</tr>
<tr>
<td>$\mu_{\psi,3} = 0.000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Beliefs and Investments: Anthropology

- !Kung San in the Kalahari desert in Botswana and Namibia (e.g., Lee, 1979) vs. Ache in Paraguay (see Kaplan and Dove, 1987; Hill and Hurtado, 1996).
  - Both groups believe that the development of motor skills by children depends on parental encouragement and teaching.
  - Different environments lead both groups to behave in very different ways.
- Gusii in Kenya (see LeVine et al, 1994).
The argument that low subjective expectations about returns may affect investments has been recognized in developmental psychology for over 50 years (Hunt, 1961; Vygostky, 1978).

Huge empirical literature attempting to estimate what parents know about child developmental milestones (Epstein, 1979; Hess et al., 1980; Ninio, 1988; Mansbach and Greenbaum, 1999).
Expecting too little, too late however is not characteristic of teenagers’ knowledge in all areas of development. In fact, when we look at items about basic care, health and nutrition, and perceptual and motor development, we discover that their expectations are quite accurate. By contrast, when we look at how they view infant needs and abilities in the areas of mental development – cognitive, social, and language, it is here that we find teenagers attributing skills to babies many months too late. And, not surprisingly, our analyses show that it is the younger infant who is most likely viewed as a creature of physical needs and growth without corresponding mental activity.

This view of the infant is also evident in teenagers’ responses to the videotape measure. Mean ratings indicate that they can neither observe the signs of learning in babies nor recognize the appropriate activities by which adults support this learning.
Lynd and Lynd (1929, 1937) reported that working-class mothers ranked “strict obedience” as their most important childrearing goal more frequently than higher-SES mothers did. Many studies, conducted in the US in the 1990s or in other developed countries, replicate these findings.

Kohn (1963) argues that the stronger preferences towards socio-emotional skills by lower-SES mothers reflect those mothers’ forecasts for their children choosing occupations in which obedience and conformity have relatively higher returns.

This finding is also reported in Lareau’s ethnographic study *Unequal Childhood: “Natural Accomplishment of Growth” and “Concerted Cultivation.”*
Aizer and Stroud (2010) track the smoking habits of educated and non-educated pregnant women before and after the release of the 1964 Surgeon General Report on Smoking and Health.

Before the release of the report, educated and non-educated pregnant women smoked at roughly the same rates.

After the report, the smoking habits of educated women decreased immediately, and there was suddenly a ten-percentage point gap between pregnant women who were educated and non-educated in smoking.

Could the divergence of early investments in the last 20 years be the result of divergence in expectations? We don’t know, but it is possible that this is the case.
I presented research in which we aim to formulate a model of human development in which mothers have subjective expectations about a parameter of the technology of skill formation.

The model is useful to understand how maternal knowledge about the importance of investments in children affect investment choices.

Large body of literature in many fields suggest that beliefs may play an important role in determining familial investments in children.
At the same time, the literature suggests that these beliefs are endogenous.

Parents expectations about future occupations of children, or the skills that will be most important for their survival, determine parental beliefs about what skills children should learn, and what skills they believe are malleable.

So, if correct, this framework suggests that it may be difficult to change parental beliefs.

At the same time, research in economics shows that most educated parents react to information that improves children’s health.

And some home visitation programs have been very successful in positively affecting children’s health (but not all).

So, future research should aim to understand the process of belief formation.