The Economics and the Econometrics of Human Development: Dynamic Models
Part IIIA

James J. Heckman
University of Chicago

CeMMAP Masterclass
University College of London
June 23, 2016
A Bare-Bones Model of Parenting as Investment
To focus ideas, I present a simple model of family investment and skill development based on Cunha (2007) and Cunha and Heckman (2007).
The Problem of the Parent
Life is assumed to last four periods:

- Two periods as a passive child who makes no economic decisions (and whose consumption is ignored) but who receives investment in the form of goods
- Two periods as a parent.

When the parent dies, she is replaced by the generation of her grandchild.

Denote by $\theta_1$ the initial capability level of a child drawn from the distribution $J(\theta_1)$.

The evolution of child skills depends on parental investments in the first and second period, $I_1$ and $I_2$. 
The productivity of parental investment depends on parental human capital, $\theta_{P,t}$.

(For notational simplicity, we set $\theta_{P,t} = \theta_P$.) Equate scalar human capital with skill for both parents and children.

Denoting by $\theta_3$ the human capital of the child when he reaches adulthood.

Recursive substitution of the technology of skill formation using a CES specification gives the following representation:

$$\theta_3 = \delta_2 \left[ \theta_1, \theta_P, \left( \gamma (l_1)^\phi + (1 - \gamma) (l_2)^\phi \right)^\frac{\rho}{\phi} \right], \quad (1)$$

for $0 < \rho \leq 1$, $\phi \leq 1$ and $0 \leq \gamma \leq 1$.

$\gamma$ is a skill multiplier.
How to Get Simple Representation
Consider the following parameterization of the stage-specific production functions:

\[ \theta_{t+1} = \delta_t \left\{ \gamma_{1,t} \phi_t + \gamma_{2,t} \phi_t + \gamma_{3,t} \phi_t \right\}^{\frac{\rho_t}{\phi_t}} \]

with \( 0 < \gamma_{1,t}, \gamma_{2,t}, \text{ and } \gamma_{3,t}; \rho_t \leq 1; \phi_t \leq 1; \text{ and } \sum_{k=1}^{3} \gamma_{k,t} = 1. \)
Substitute recursively. If $T = 2$, $\rho_1 = \rho_2 = 1$, $\delta_1 = 1$, and $\phi_1 = \phi_2 = \phi \leq 1$, skills at adulthood, $\theta_3 = \theta_{T+1}$, can be expressed as

$$\theta_3 = \delta_2 \left[ \gamma_{1, 2} \gamma_{1, 1} \theta_1^\phi + \gamma_{1, 2} \gamma_{2, 1} I_1^\phi + \gamma_{2, 2} I_2^\phi + \left( \gamma_{3, 2} + \gamma_{1, 2} \gamma_{3, 1} \right) \theta_P^\phi \right] \frac{1}{\phi}.$$

"Multiplier"
• **Multiplier:** \( \gamma = \gamma_{1,2} \gamma_{2,1} \).

• Arises from the conjunction of self-productivity \((\gamma_{2,1} \neq 0)\) and the productivity of investment \((\gamma_{1,2} \neq 0)\).

• Self-productivity joined with the productivity of investment generates dynamic complementarity.

• \( \gamma_{2,1} \) characterizes how much of the investment in period \( t = 1 \) propagates into skills at adulthood, \( \theta_3 \).

• The parameter \( \phi \) captures the substitutability/complementarity of investments.
• If \( \phi = 1 \), investments at different periods are (almost) perfect substitutes.

• They are perfect one-to-one substitutes if \( \gamma_{1,2} = \gamma_{2,2} \), in which case the timing of investment in skills does not matter for the developmental process.

• This is the only circumstance in which collapsing childhood into one period as in Becker–Tomes is without loss of generality.
• Polar opposite case: $\theta_3 = \delta_2 (\theta_1, \theta_P, \min (I_1, I_2))$.
• Closer to the empirical truth than perfect substitution.
• Complementarity has a dual face.
• Early investment is essential but ineffective unless later investments are also made.
• In this extreme case, there is no possibility of remediation.
• If parents are poor and unable to borrow against the future earnings of their children and, as a result, $I_1$ is low, there is no amount of investment at a later age, $I_2$, that can compensate for early neglect.
• The parameters of the technology determine whether early and later investments are complements or substitutes.

• “Direct” complementarity for Equation (1) holds if $\rho > \phi$, whereas substitutability holds otherwise.

• Another definition of complementarity in the literature distinguishes (in the case of $\rho = 1$) whether $\phi > 0$ (*gross* substitutes; the elasticity of substitution is greater than 1) or $\phi < 0$ (*gross* complements; the elasticity of substitution is less than 1).

• Cobb-Douglas ($\phi = 0$) is the boundary case.
• Given \( \rho \), the smaller \( \phi \), the harder it is to remediate low levels of early investment \( I_1 \) by increasing later investments.

• At the same time, the stronger the complementarity (the lower \( \phi \)).

• The more important it is to follow high volumes of early investments with high volumes of late investments to achieve high levels of production of adult human capital.
Parent’s Problem
The parent allocates resources across household consumption in both periods of the child’s life, \( c_1 \) and \( c_2 \); early and late investments, \( l_1 \) and \( l_2 \); and bequests, \( b' \).

Assets at the end of the first period, period \( a \), may be constrained to be non-negative.

Bequests are received when entering adulthood and may be positive or negative.

The state variables for the parent are her initial wealth, \( b \); human capital level, \( \theta_P \); and the initial skill level of the child, \( \theta_1 \).

Human capital is rewarded in the labor market according to the wage rate, \( w \).

Economy is characterized by one risk-free asset with return \( r \).
• $u(\cdot)$: parental utility function
• $\beta$: discount factor
• $\nu$: parental altruism given by the weight assigned to the utility of future generations
• $\theta'_1$: uncertain initial endowment of the child’s child
• Goal of the parent: optimize

$$V(\theta_P, b, \theta_1) = \max_{c_1, c_2, l_1, l_2} \left\{ u(c_1) + \beta u(c_2) + \beta^2 \nu \mathbb{E}[V(\theta_3, b', \theta'_1)] \right\}$$

(2)

• subject to (1), (3) and (4)
• Denote parental financial assets by $a$
• Parental labor market productivity grows at exogenous rate $g$
• One can represent the stage-of-childhood-specific budget constraints:

\[ c_1 + l_1 + \frac{a}{(1 + r)} = w\theta_P + b \]  \hspace{1cm} (3)

and

\[ c_2 + l_2 + \frac{b'}{(1 + r)} = w(1 + g)\theta_P + a \]  \hspace{1cm} (4)
• Allow for the possibility of borrowing constraints
  • $a \geq a$ (intragenerational)
  • $b' \geq 0$ (intergenerational)
If no intra- and intergenerational credit constraints are assumed, a key property of the Becker and Tomes (1986) model persists in this framework.

There is no role for initial financial wealth $b$, parental income, parental utility, or the magnitude of parental altruism $\nu$ (above zero) in determining the optimal level of investment because parents can borrow freely in the market to finance the wealth-maximizing level of investment.
• Even if the altruism parameter is zero \((\nu = 0)\), if the parents can make binding commitments, selfish parents \((\nu = 0)\) will still invest in the child, as long as the economic return in doing so is positive.

• However, even in this setup, returns to parental investments depend on parental skills, \(\theta_P\), as they affect the productivity of investments.

• The returns to investments are higher for children of parents with higher \(\theta_P\).
• These children receive higher levels of investment.
• This is a type of market failure due to the “accident of birth” that induces a correlation of human capital and earnings across generations even in the absence of financial market imperfections.
• The initial condition \( \theta_1 \) also affects investments.
• It creates a second channel of intergenerational dependence due to the “accident of birth” if it is genetically related to parental endowments, as considerable evidence suggests.
• When there are no intra-period constraints,

\[
\frac{l_1}{l_2} = \left[ \frac{\gamma}{(1 - \gamma)(1 + r)} \right]^{\frac{1}{1 - \phi}}.
\]

(5)

• \(\frac{l_1}{l_2}\) ↑ as \(\gamma\) ↑, \(\phi\) ↑, and \(r\) ↓.
Figure 1: The Ratio of Early to Late Investment in Human Capital as a Function of the Skill Multiplier for Different Values of Complementarity

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. Let $t$ denote the present value as of period “3” of the future prices of one efficiency unit of human capital:

$$t = P_W w = 3 z_w (1 + u)^w$$

The parents solve

$$\max_\mu \frac{1}{1+u} \left[ tk + e \right]$$

subject to the budget constraint

$$L_1 + L_2 (1+u) + e (1+u)^2 = P$$

and the technology of skill formation:

$$k = h L_1 + (1 - \phi) L_2$$

for $0 < \phi < 1$ and $\phi = 0$ (Cobb-Douglas) and $\phi = 0.5$ (Leontief) and for values of the skill multiplier between 0 and 0.9.

(Assumes $r = 0$)
• Imperfect credit markets create another channel of intergenerational dependence.

• One possible constraint is the impossibility of borrowing against the child’s future earnings (Becker and Tomes, 1986).
• Because $b' \geq 0$, parental wealth matters in this model when this constraint binds.

• Children coming from constrained families will have lower early and late investments.
• However, even with $b' \geq 0$, the ratio of early to late investment is not affected.
• Suppose constraints (3) and (4) bind separately.
• Suppose parental utility is given by
  \[ u(c) = \frac{c^\lambda - 1}{\lambda}. \]
• \( \lambda = 1 \) corresponds to perfect intertemporal substitutability.
The ratio of early to late investment is then

\[
\frac{l_1}{l_2} = \left( \frac{\gamma}{(1 - \gamma)(1 + r)} \right)^{\frac{1}{1 - \phi}}
\]

unconstrained ratio

\[\frac{l_1}{l_2} \uparrow \text{ as } \gamma \uparrow, \phi \uparrow, \text{ and } r \downarrow\]

\[\frac{\beta(1 + r)}{1 - \phi} \left( \frac{c_1}{c_2} \right)^{\frac{1 - \lambda}{1 - \phi}} = 1 \text{ if unconstrained, } <1 \text{ if constrained (}a > a\text{ binds)}\]

(6)
• If early parental income is low compared to later life income, or if $\lambda$ is small, the level and timing of family resources will influence the parental investment.

• Estimates from Cunha et al. (2010) suggest that $1/(1 - \phi) = 3$.

• Estimate of $\lambda \in [-3, -1.5]$ (Attanasio and Browning, 1995).

• $(1 - \lambda)/(1 - \phi) \in [0.83, 1.3]$.

• Notice that even if $\lambda = 1$, parents may hit constraints on the level of investment if future resources are of insufficient magnitude.

• This constraint could be very harmful to a child if it binds in a critical period of development and the complementarity parameter $\phi$ is low so that later life remediation is ineffective.
The Presence of Constraints is Not Synonymous with Low Levels of Investment
• However, the presence of constraints is not necessarily synonymous with a low level of investment.
• For a given family, a binding constraint implies that the investments are lower than the unconstrained optimum.
• Whether a family is constrained, however, is uninformative on how that family compares with others in terms of the effective level of investments provided.
Caucutt and Lochner (2012) use a variant of the model of Cunha (2007, 2013) to investigate the role of income transfers and credit constraints in the early years.

They find that a large proportion of young parents are credit constrained (up to 68% among college graduates) but that reducing borrowing constraints is effective in promoting skills only for the children in the generation in which they are relaxed.
Introducing Income Uncertainty
• Cunha (2007, 2013): overlapping generations model with stochastic innovations to parental income.

• If $g$ is stochastic on the interval $[-1, \infty)$, so parents face uncertain income growth, constraints play a dual role.

• First, as before, if the constraints bind, they reduce investments in the constrained periods.

• Second, because future income is uncertain, so is the likelihood of binding future constraints.
● Absent full insurance markets, consumption and investments in children are less than optimal, even if the parent is not currently constrained but expects to be constrained in the future with a probability greater than zero.
See Appendix on **The Problem of the Parent**
(based on Cunha, 2007; revised 2016)

[Link]
• Under this scenario, young parents who just entered the labor force accumulate more assets than they would in the absence of possible future constraints to ensure against bad future shocks.

• This implies a reduction in household consumption and investments in child human capital.
Recent Extensions of the Basic Model
• The recent literature has moved beyond the simple models just discussed.
• Most assume parental altruism.
• Some are explicitly paternalistic.
• They all feature investment in *goods*.
• Only recently has parental time been analyzed as an explicit input to child quality.
• Most models analyze how child investment depends on parental skills.
• Until recently, most studies considered the self-productivity of skills.
• Some recent papers ignore this feature, despite the empirical evidence that supports it.
• Most analyses assume that parents know the technology of skill formation, as well as the skills of their children, in making investment decisions.

• Some of the recent literature also ignores intergenerational transfers.

• Some papers consider extreme credit constraints that do not permit any borrowing (or lending), even within a lifetime of a generation, much less inter-generational transfers.

• Virtually the entire literature focuses on single-child models, exogenous fertility, and exogenous mating decisions.

• Most models are for single-parent families, for which the characteristics of the spouse are irrelevant.
First, with the exception of Cunha and Heckman (2008a) and Cunha et al. (2010), as well as very recent papers by Attanasio and Meghir, human capital is treated as a scalar.
Second, in many, but not all recent models, investments are also treated as scalars.
Third, families usually have more than one child.
Fourth, the models in the literature ignore the interaction of parents and children in the process of development.
Fifth, fertility is taken as exogenous.
• Do not take too literally models of credit constraints interacting with dynamic complementarity that take fertility as exogenously determined.
• Child’s development is influenced by the environment outside his family: day care, kindergarten, school, and neighborhood.

• In addition, the effectiveness of policies is determined in part by parental responses to them.

• Policies that complement rather than substitute for family investments will have greater impacts and lower costs.
Credit Constraints and the Effects of Family Income
The Effects of Family Income
The literature is unanimous in establishing that families with higher levels of long-run (or permanent) income on average invest more in their children and have children with greater skills.

The literature is much less clear in distinguishing the effect of income by source or in distinguishing pure income effects from substitution effects induced by changing wages and prices (including child-care subsidies or educational incentive payments).
Levels of permanent income are highly correlated with family background factors such as parental education and maternal ability, which, when statistically controlled for, largely eliminate the gaps across income classes.
The literature sometimes interprets this conditioning as reflecting parenting and parental investments, but it could arise from any or all of the panoply of correlates of permanent income associated with parental preferences and skills.
• Income operates through altruism or paternalism?
Effects of Borrowing Constraints
The literature also analyzes the effect of borrowing constraints on child outcomes.
Restrictions in Lending Markets for College Education
Belley and Lochner update the NLSY79 analysis of Carneiro and Heckman (2002) using NLSY97 data and claim that credit constraints seem to bind predominantly among less able poor children.

However, their analysis shows that, across all ability groups, college enrollment increased in 1997 compared to 1979.

The increases are more substantial for more affluent, low-ability children.
Figure 2: College Attendance by AFQT and Family Income Quartiles (1979)

Source: Belley and Lochner (2007).
Figure 3: College Attendance by AFQT and Family Income Quartiles (1997)

Source: Belley and Lochner (2007).
Figure 4: College Attendance by AFQT and Family Income Quartiles (1979 and 1997 on one graph)

Source: Belley and Lochner (2007).
The Timing of Income, Dynamic Complementarity, and Credit Constraints
• The interaction of dynamic complementarity and lifetime liquidity constraints motivates a recent literature.

• Dahl and Lochner (2012) investigate how credit constraints affect test scores of children in early adolescence.

• They exploit the policy variation in the Earned Income Tax Credit (EITC) as an exogenous instrument for the effect of income on child outcomes.
Studies on the Role of Income on Children’s Outcomes and on Credit Constraints

Link to Appendix
Lessons from the Literature on Family Income and Credit Constraints
• The literature on credit constraints and family income shows that higher levels of parental resources, broadly defined, promote child outcomes.
• However, a clear separation of parental resources into pure income flows, parental environmental variables, and parental investment has not yet been done.
• **Premature** to advocate income transfer policies alone as effective policies for promoting child development.
• Poverty (measured by income is not (necessarily) a good measure of child disadvantage).
The literature establishes the first-order importance of child ability for college going, irrespective of family income levels. More advantaged families with less able children send their children to college at greater rates than less advantaged families, but the literature does not establish the existence of market imperfections or any basis for intervention in credit markets. The observed empirical regularity may result from the exercise of parental preferences. Recent work shows that the returns to college for less able children are low, if not negative.
The literature that presents more formal econometric analyses of the importance of credit market restrictions on educational attainment shows little evidence for them.
The analysis of Caucutt and Lochner (2012) is an exception. They *calibrate* that a substantial fraction of the population is constrained due to the interaction of dynamic complementarity, the receipt of income, and the imperfection of lending markets. Much further research is required before definitive policy conclusions can be drawn on the empirical importance of the timing of receipt of income over the life cycle for child outcomes.
Recent work on credit constraints and human capital accumulation

“Inequality in Human Capital and Endogenous Credit Constraints” by Rong Hai and James J. Heckman (2016)

Link
Structural Models
See Appendix on **Overview of Structural Models of Parental Investments**

[Link]
Most studies of the role of income transfer programs discussed earlier do not investigate the interactions of public policy interventions and family investments.

To do so, some authors have estimated fully specified structural models and use them to study the effect of various types of policy experiments.
Overview of Policy Implications of Structural Models
Four main facts:

- First, subsidies to parental investments are more cost-effective in improving adult outcomes of children such as schooling attainment or earnings, when provided in the early stages of life (Caucutt and Lochner, 2012; Cunha, 2007; Cunha and Heckman, 2007).
• Second, financial investment subsidies have stronger effects for families who are already engaging in complementary investments.

• Targeted public investments and targeted transfers restricted to child-related goods that guarantee minimum investment amounts to every child increase the level of investments received by the children of the least-active parents (Caucutt and Lochner, 2012; Del Boca et al., 2014a; Lochner and Monge-Naranjo, 2016).

• Lee and Seshadri (2016) provide evidence on the importance of targeted education subsidies for increasing the educational expenditures of poor families.
• Third, time-allocation decisions are affected by transfers.
• Del Boca et al. (2014a) show that unrestricted transfers increase the time parents spend with their children through a wealth effect.
• The increase in child quality is minimal.
• Lee and Seshadri (2016) show how this effect is especially strong for parents without college education, whereas, in their model, public transfers negatively affect time spent with children for college-educated parents.

• Fourth, targeted conditional transfers (on a child’s ability improvements) are more cost-effective than pure income transfers to achieve any child outcome.
The Implications of Dynamic Complementarity for Investments across Children with Different Initial Endowments
• The average family usually has more than one child, and society allocates public investments across multiple children.
• The problem of intra-child allocations is sometimes formulated as a problem in fairness.
• CES representation of parental utility $V$ is often used:

$$V = \left( \sum_{k=1}^{N} \omega_k V_k^\sigma \right)^{\frac{1}{\sigma}}. \quad (7)$$
The Rawlsian version of maximal inequality aversion is obtained when $\sigma \to -\infty$, so utilities are perfect complements, and parents are concerned only with the maximization of the minimum outcome across children.
In a two-child version of the one-period-of-childhood model analyzed by Becker and Tomes (1979, 1986), under complementarity between initial endowment and investment, the optimal policy when $\sigma = 1$ is to invest less in the initially disadvantaged child.

Under substitutability, it is optimal to invest more in the disadvantaged child.
• Story richer when we consider a multiperiod model with dynamic complementarity.

• *Investing relatively more in initially disadvantaged young children can be efficient even when the $\omega_k$ are equal and $\sigma = 1$.*

• This is true even if there is complementarity in each period of the life cycle.
Dynamic complementarity is a force promoting compensating early stage investments.

In a multiperiod model at stage $t$

$$\theta_{t+1} = f^{(t)}(\theta_t, l_t),$$

Even if there is complementarity at all stages, so $f^{(t)}_1(\cdot) > 0$ (where $(\cdot)$ denotes the argument of the function), output-maximizing investments can be compensating.
• If $f_{12}^{(1)}(\cdot) < 0$, but $f_{12}^{(2)}(\cdot) > 0$, it is *always* efficient to invest relatively more in the initially disadvantaged child in the first period.
• It can also be productively efficient to invest in the disadvantaged child if $f_{12}^{(1)}(\cdot) > 0$, when initial endowments and investments are complements.
See Appendix on **Targeting Relatively More Investment Toward Disadvantaged Children**

[Link](#)
• Intuition.
• Suppose that there is increasing complementarity.
• In this case, the stock of skills in the second period has a greater effect on the productivity of investments than it does in the first period \( f_{12}^{(2)}(\cdot) > f_{12}^{(1)}(\cdot) \).
• First-period investments bolster the stock of second-period skills and prepare disadvantaged children to make productive use of them in the second period.
• This effect is stronger when \( f_{12}^{(2)}(\cdot) \) is larger.
Another force promoting greater initial investment in the disadvantaged child is diminishing self-productivity of skills in the first period \( f_{11}^{(1)}(\cdot) < 0 \).

The greater the diminishing returns to investment for the better-endowed child, the lower the benefits of early advantage.

Diminishing productivity of the stock of second-period skills \( f_{11}^{(2)}(\cdot) < 0 \) operates in the same fashion to limit the effects of any initial advantage.
The smaller the effect of the initial stock of skills on the productivity of investment in the first period \( f_{12}^{(1)}(\cdot) \), the weaker is the disequalizing force of complementarity toward promoting investment in the initially advantaged child.
● Summarizing:

1 The more concave are the technologies in terms of stocks of skills (the more they exhibit decreasing returns in the stocks of skills), the more favorable is the case for investing in more disadvantaged children.

2 The stronger is second-period complementarity \( f_{12}^{(2)}(\cdot) \), the stronger is the case for investing more in the initially advantaged child to build skill stocks to take advantage of this opportunity.

● The weaker is the first-period complementarity \( f_{12}^{(1)}(\cdot) \), the less offsetting is the disequalizing effect of complementarity coupled with initial advantage.
• In general, even when investment is greater in the first period for the disadvantaged child, it is optimal for second-period investment to be greater for the initially advantaged child.
• It is generally not efficient to make the disadvantaged child whole in the first period.
• Greater second-period complementarity then kicks in to promote disequalizing second-period investments.
Dynamic complementarity generates the following relationship:
Figure 5: Returns to a Unit Value Invested

Source: Heckman (2008)
See Appendix on **Some Evidence from Simulations on Why Dynamic Complementarity is a Force for Promoting Early Investment in Disadvantaged Children**

[Link](#)
• A dynamic state-space model with constraints and family investment decisions is the natural econometric framework for operationalizing the model.

• Discuss in Part IIIB.
Towards a More General Model of Parent-Child Interactions
• The productivity of any investment or parental stimulus is influenced by the child’s response to it.

• Parents and children can have different goals.

• For example, the child can be more shortsighted than the parent (Akabayashi, 2006b) or have different values for leisure and future human capital (Cosconati, 2013a).

• The parent may act as a principal whose goal is to maximize the effort from an agent—their child.
Economic Models of Parenting and Scaffolding
García and Heckman
Research Question

- How do child-parent interactions shape the formation of early character and cognitive skills?
- How child matures in this process?
1. Model parent-child interactions
   - Early-childhood, dynamic game
   - Learning and skill development
   - Technology depends on parental and child investment

2. Structurally estimate the model
   - Use data from the Infant Health and Development Program (IHDP)
     - Parent and child efforts
     - Character and cognitive skills
     - Parental expectations
Literature Overview

- Skill formation with passive children:
  - Cunha and Heckman (2008b); Cunha and Schennach (2010), Heckman et al. (2010, 2013b)
    - Lots of results on skills and investment
      - Self-productivity (of skills)
      - Dynamic complementarity (of investment)
      - Time without the mother hurts skill development
- Specific features of child-parent interactions:
  - Akabayashi (2006a); Cosconati (2013b); Lizzeri and Siniscalchi (2008a)
    - Child-maltreatment (theory only)
    - Optimal parenting styles (theory and estimation)
    - Parental guidance (theory only)
### Table 1: Models of Parent-Child Interaction

(“✓” means present; “X” means absent)

<table>
<thead>
<tr>
<th>Model</th>
<th>Parental Monetary Investments</th>
<th>Discordance in Preferences between Parent and Child</th>
<th>Multiple Children</th>
<th>Parental Incentives for Child Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosconati (2013a)</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Akabayashi (2006b)</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Hao et al. (2008)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lizzeri and Siniscalchi (2008b)</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

a Difference in discount factors.  
b Differences in utility functions.  
c Restrictions on leisure.  
d Differences in knowledge about proper task execution.  
e Time investments in the child.  
f Monetary investments in the child.  
g The authors analyze the effects of time investments.  
h Implications from the model are empirically tested.
Table 1 (continued): Models of Parent-Child Interaction (“✓” means present; “X” means absent)

<table>
<thead>
<tr>
<th>Model</th>
<th>Effort Produces Greater Skill</th>
<th>Parental Learning about Child Quality</th>
<th>Parental Actions Facilitate Acquisition of Information</th>
<th>Parental Beliefs Can Diverge from Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosconati (2013a)</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Akabayashi (2006b)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hao et al. (2008)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lizzeri and Siniscalchi (2008b)</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
</tr>
</tbody>
</table>

a Difference in discount factors.  
b Differences in utility functions.  
c Restrictions on leisure.  
d Differences in knowledge about proper task execution.  
e Time investments in the child.  
f Monetary investments in the child.  
g The authors analyze the effects of time investments.  
h Implications from the model are empirically tested.
Table 1 (continued): Models of Parent-Child Interaction (“✓” means present; “X” means absent)

<table>
<thead>
<tr>
<th>Child’s Effort Observable</th>
<th>Child’s Human Capital Observable</th>
<th>Model Is Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosconati (2013a)</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Akabayashi (2006b)</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Hao et al. (2008)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lizzeri and Siniscalchi (2008b)</td>
<td>✓</td>
<td>X</td>
</tr>
</tbody>
</table>

\(a\) Difference in discount factors.  
\(b\) Differences in utility functions.  
\(c\) Restrictions on leisure.  
\(d\) Differences in knowledge about proper task execution.  
\(e\) Time investments in the child.  
\(f\) Monetary investments in the child.  
\(g\) The authors analyze the effects of time investments.  
\(h\) Implications from the model are empirically tested.
Our Contribution to the Literature

- Model and estimate parent-child interactions to understand how:
  - Skill forms
  - Child matures
Parent

- The parent has target skills and target investments
- Target skills evolve according to a predetermined skills production function
- The parent knows the technology of skill formation
- Parent observes a noisy skill realization – function of true skill and child’s effort
- Parent chooses investment to minimize deviations from target
Parent’s Problem

\[ V_t(y_t) = \min_{u_t, \ldots, u_T} \mathbb{E} \left[ \sum_{\tau=t}^{T} \beta^\tau \left( \tilde{\theta}_\tau^' Q \tilde{\theta}_\tau + \tilde{u}_\tau^' R \tilde{u}_\tau \right) + \tilde{\theta}_T^' Q_f \tilde{\theta}_T \mid y_t \right] \] (9)

- \( Q, Q_f, R \geq 0 \) are weighting matrices
- we define

\[
\begin{align*}
\tilde{\theta}_{t+1} &= \tilde{A}\tilde{\theta}_t + \tilde{B}\tilde{u}_t + \tilde{F}\tilde{a}_t + \omega_t \\
\theta_{t+1} &= \Phi(\theta_t, u_t, a_t); \quad \overline{\theta}_{t+1} = \Phi(\overline{\theta}_t, \overline{u}_t, \overline{a}_t) \\
y_t &= C\tilde{\theta}_t + B\tilde{a}_t + v_t \\
\tilde{x}_t &= x_t - \overline{x}_{t} \text{ for } x_t = \theta, u_t, a_t
\end{align*}
\] (10)
• Enjoys parental investment and dislikes effort
• Her strategy could be “à la Cournot” or “à la Stackelberg”
• Uncertain on her ability and perfectly observes the rest of the components
Child’s Problem

\[ J_t = \max_{a_t, \ldots, a_T} \lambda_t J^1_t(y_t) + (1 - \lambda_t) J^2_t(y_t) \]

\[ J^1_t(y_t) = \mathbb{E} \left[ \sum_{\tau = t}^T \beta^\tau \left( -\tilde{a}_\tau + \tilde{u}_\tau \tilde{a}_\tau \right) | y_t \right] \]

\[ J^2_t(y_t) = \mathbb{E} \left[ \sum_{\tau = t}^T \beta^\tau \left( -\tilde{a}_\tau + R^P_t (\tilde{a}_\tau)^2 \right) | y_t \right] \]
Identification

- Extend Cunha and Heckman (2008) to include child inputs
- Linear technology, repeated measures are instrumental variables for measurement error
- Second moments from simultaneous parent and child optimization imply,

\[ \lambda_t = \frac{2M_t \text{cov}(a_t, u_t)}{\text{var}(u_t) + 2M_t \text{cov}(a_t, u_t)} \]  (11)

- \( M_t \) is a function of technology parameters and preference parameters \( R, \beta \)
- Property: \( \frac{\partial}{\partial R} \lambda_t < 0 \)
- Technology identified \( \implies \lambda \) identified up to \( R, \beta \)
• Technology: instrumental variables regression or MLE under distributional assumption
• Requires repeated measures of child’s cognitive and non-cognitive skills as well as child and maternal effort over short stages of child development
• \( \lambda \): Given technology and preferences, plug-in estimator
• Inference: bootstrap the estimators, test for non-zero technology coefficients and positive \( \lambda_t \)
Estimation Details

Link
Results

Child’s Maturity over Time

R = 0.00

Econ and Ecom of Hum Dev
Results

Child’s Maturity over Time

R = 0.05

Age of Child

Child’s Passivity

Lambda

10% Critical Value

Econ and Ecom of Hum Dev
A Simulation Exercise Based on Akabayashi (2006)

[Link]
Link to

“Measuring Early Investments in Children”

by Jennifer Culhane, Flávio Cunha, Irma Elo, and Zoe Pham (2016)
END
Appendix: The Problem of the Parent
The Problem of the Parent
• The parent is assumed to be the decision-maker in the household.
• The child passively accepts investment.
• The consumption of the child is not modeled.
• The problem solved by the parent depends on the age of the child.
• When the child is between ages 1 and $T - 1$, he only receives investments and cannot work.
• When the child reaches age $T$, the parent may invest a minimum level or something beyond that minimum.

• If the parent invests the minimum amount, the child does not attend college but becomes a high school graduate and works full time.

• If the parent invests any amount beyond the minimum, the child attends school (college) full-time.

• At the end of the period, he becomes a college graduate.
The Problem When the Child Is Between 1 and $T - 1$ Years Old
- Parental labor supply is assumed to be perfectly inelastic.
- At each age $t$ of the child, the parent is subject to productivity innovations $\varepsilon_t$, corresponding to labor market uncertainty.
- The shocks $\varepsilon_t$ are independently and identically distributed across parents.
The shocks follow a first-order Markov process:

\[ \ln \varepsilon_{t+1} = \rho \varepsilon \ln \varepsilon_t + \sigma \eta_{\varepsilon, t}. \] (12)
• Parents are assumed to have positive earnings.
• Productivity innovations are restricted so that there exists $\varepsilon_{\text{min}}$ with the property that $\varepsilon_t \geq \varepsilon_{\text{min}} > 0$ for any $t = T + 1, \ldots, 2T$.
• The labor income of the parent is $wh\varepsilon_t$, where $w$ is the efficiency wage and $r$ is the risk-free discount rate.
• Innovations in wages and labor market uncertainty are missing in BTS.
• The level of capability of the parent, \( h \), is the outcome of investment decisions made by the grandparent.

• In similar fashion, the level of skill of the child when an adult, \( h' \), will also be the consequence of investments made by the parent, and satisfies \( h' = \theta_{T+1} \).

• Defining \( s_t \) as the stock of savings of the parent at age \( t \), the individual state variables for the parents of children who are between 1 and \( T - 1 \) years old is \((h, \theta_t, s_t, \varepsilon_t, t)\).
• Given the state variables, the parent chooses household consumption \( c_t \), savings \( s_{t+1} \), and investments \( I_t \) in the cognitive skill of the child.

• The savings of the parents are in a risk-free asset which pays a rate of interest \( r \).

• \( p \) denotes the price of the investment goods in cognitive skill.

• Following Laitner (1992), the parents cannot leave debts to their children and have negative net worth, so savings are subject to the lower bound equal to \( \frac{-wh_{\text{min}}}{(1+r)} \) (the “natural” borrowing limit).
\[ V(t, h, \theta_t, s_t, \varepsilon_t) \] is the value function of the parent of a child at age \( t \), \( 1 \leq t \leq T - 1 \). The problem of the parent is:

\[
V(t, h, \theta_t, s_t, \varepsilon_t) = \max_{c_t, l_t, s_{t+1}} \{ u(c_t) + \beta \mathbb{E}[V(t + 1, h, \theta_{t+1}, s_{t+1}, \varepsilon_{t+1}) | \varepsilon_t] \}.
\]
Subject to:

\[ c_t + p_l + s_{t+1} = wh^\varepsilon_t + (1 + r) s_t \quad (13) \]

\[ s_{t+1} \geq - (wh^\varepsilon_{\text{min}}) , \quad l_t, c_t \geq 0 \quad (14) \]

And the technology for capability formation
• Associating multiplier $\mu_t$ to the borrowing constraint in stage $t$, the optimal conditions for consumption and investments are given by:

$$u_c(c_t) = \beta(1 + r)E[V_s(t + 1, h, \theta_{t+1}, s_{t+1}, \varepsilon_{t+1}) | \varepsilon_t] + \mu_t$$  \hspace{1cm} (15)$$

$$\beta E \left[ \frac{\partial \theta_{t+1}}{\partial I_t} V_\theta(t + 1, h, \theta_{t+1}, s_{t+1}, \varepsilon_{t+1}) | \varepsilon_t \right]$$

$$= \beta(1 + r)pE[V_s(t + 1, h, \theta_{t+1}, s_{t+1}, \varepsilon_{t+1}) | \varepsilon_t] + \mu_t$$  \hspace{1cm} (16)$$
- This implies that the marginal utility of investments is equated to the marginal utility of consumption and to the marginal utility of future wealth.
- Whenever the constraint binds ($\mu_t \geq 0$), consumption and investment will be reduced as the agent would like to borrow more than ($wh_{\ell_{min}}$), but she is constrained.
Suppose now that the agent is not constrained in period \( t \).

Using the envelope condition for assets we can rewrite the optimal condition for investment and consumption making clear the dependence on expected future constraints:

\[
\beta \mathbb{E}_t \left[ \frac{\partial \theta_{t+1}}{\partial I_t} V_{\theta} (t + 1, h, \theta_{t+1}, s_{t+1}, \varepsilon_{t+1}) | \varepsilon_t \right] = p u_c (c_t) 
\]

(17)

\[
= \beta (1 + r) p \mathbb{E}_t [V_s (t + 1, h, \theta_{t+1}, s_{t+1}, \varepsilon_{t+1}) | \varepsilon_t]
\]

(18)

\[
= [\beta (1 + r)]^2 p \mathbb{E}_t [\mathbb{E}_{t+1} [V_s (t + 2, h, \theta_{t+2}, s_{t+2}, \varepsilon_{t+2}) | \varepsilon_{t+1}] + \mu_{t+1} | \varepsilon_t]
\]

(19)

\[
+ \mathbb{E}_t [\mu_{t+1} | s_{t+1} = -wh\varepsilon_{\min}] P(s_{t+1}^* < -wh\varepsilon_{\min})
\]
Where $s_{t+1}^*$ represents the optimal unconstrained amount of savings from stage $t+1$ to stage $t+2$ and

$$P(s_{t+1}^* < -wh\varepsilon_{min}) = P\left(\varepsilon_{t+1}wh - c_{t+1}^*(\varepsilon_{t+1}wh) - pl_{t+1}^*(\varepsilon_{t+1}wh)\right)$$

$$\quad \quad < -wh\varepsilon_{min} - (1 + r)s_t$$

(20)
• \( c_{t+1}^* \) and \( l_{t+1}^* \) represent the optimal unconstrained levels of consumption and investments in period \( t+1 \) which depend on the realization of income.

• Even when the parent is not constrained in period \( t \), the expectation of future constraints reduces current consumption and investments levels.

• The fear of hitting the constraint in the future induces a precautionary motive for savings which reduces current investments and consumption.
The Problem When the Child Is $T$ Years Old: Go to College or Not?
• Consider the decision to go to college (made by the parent).
• When the child reaches age $T$, the parent decides to invest the minimum amount, $I$, or something beyond that amount.
• The parent uses the relevant information to make that decision, which is contained in the vector of state variables $(h, \theta_t, s_t, \varepsilon_t, n_t)$. 

• Let \( \kappa \) be tuition cost. The parent’s problem can be stated as:

\[
V (T, h, \theta_T, s_T, \varepsilon_T) = \max_{c_T, l_T, s'_1} \{ u(c_T) + \beta \mathbb{E} [V (1, h', \theta'_1, s'_1, \varepsilon'_1)] \}.
\]
Subject to:

\[ c_T + s'_1 + pI = wh \varepsilon_T + w \theta_T + (1 + r) s_T \text{ if } I_T = I \]

\[ c_T + s'_1 + (pI_T + \kappa) = wh \varepsilon_T + (1 + r) s_T \text{ if } I_T > I \]

\[ s_T \geq 0 \]

And the technology for the production of skills
The budget constraint (21) states that a child who receives the minimum amount of investments $I$ works full time.

Refer to this child as a high school graduate.

Note that the high-school-graduate child’s earnings are pooled with the rest of the parental resources.

Abstract from productivity shocks for the child before he reaches adulthood.

If the parent decides to invest any amount above the minimum, so that $I_T > I$, then the parent must pay the variable cost of the investment, which is $p$ by unit, plus a fixed cost, $\varphi$—college tuition.
• A child who receives more than the minimum amount of investment does not work.
• This is described by the budget constraint (22).
• Note that equation (23) embodies the notion that the parent faces lifetime liquidity constraints.
• The parent dies and cannot leave debts to the child.
• Following Cunha (2007), one can establish a steady state general equilibrium.

• Firms producing final output under constant returns to scale.

• Also a child investment good is produced.

• Cunha (2007) establishes a stochastic general equilibrium for the steady state, extending Laitner to include human capital.
Return to main text
Appendix: Targeting Relatively More Investment Toward Disadvantaged Children Can Be Socially Efficient
Introduction
• Analyze the problem of investing in children with different initial endowments assuming that children are weighted equally ($\omega_k = 1$ for all $k$). Parents only care about total output.

• No social justice concerns.

• Families are assumed to only care about productivity.

• Consider the following two-stage model of childhood investment:

$$\theta_3 = f^{(2)}(\theta_2, l_2)$$  \hspace{1cm} (24)

$$\theta_2 = f^{(1)}(\theta_1, l_1)$$  \hspace{1cm} (25)

• $\theta_3$ represents the level of skill at the beginning of adulthood.
• The functions are assumed to be strictly concave in $l_2$ and $l_1$, respectively, and twice differentiable.
• Concavity in $\theta_2$ or $\theta_1$ is not required for an optimum, although it plays a role in signing terms in the comparative statics exercise below.
• The assumptions made below imply that all inputs are normal.
• Total resources are $E$.
• The price of input $i$ is $p_i$.
• There are two children: $A$ and $B$.
• Their initial endowments are $\theta_1^A$ and $\theta_1^B$, respectively.
• $\theta_1^A = \gamma \theta_1^B$ and consider how, from a position of initial equality ($\theta_1^A = \theta_1^B$ or $\gamma = 1$), raising the initial endowment of $A$ affects Benthamite allocations of investment goods between $A$ and $B$.

• Denote investment in the first period for child $A$ by $I_1^A$ and in the second period by $I_2^A$. $I_1^B$ and $I_2^B$ are defined analogously for child $B$. 
One Period of Childhood: Version of the Problem as in Becker and Tomes (1986)
• Parents (or social planners) seek to maximize the aggregate of adult skills ($\theta_2$):

$$\theta_2^A + \theta_2^B$$

subject to  

$$E = p_1(I_1^A + I_1^B).$$

• First order condition is:

$$F.O.C.: f_2^{(1)}(\gamma \theta_1^B, I_1^A) = f_2^{(1)}(\theta_1^B, I_1^B).$$
\[
\text{sign} \left( \frac{\partial I_1^A}{\partial \gamma} \right) = \text{sign} \left( f_{12}^{(1)}(\cdot) \right)_{\gamma=1}.
\]

- \( f_{12}^{(1)}(\cdot) \) is the value of \( f^{(12)} \) in the neighborhood of \((\cdot)\).
- Parents (social planners) invest more in the disadvantaged if inputs are substitutes with initial endowments and they invest less if they are complements.
- These are direct complements and substitutes.
In the multiperiod setting it is still optimal to invest more in the child with the lower initial endowment if $f_{12}^{(1)}(\cdot) < 0$ even though $f_{12}^{(2)}(\cdot) > 0$.

Pattern consistent with the evidence on the evolution of complementarity at later stages in the life cycle: $f_{12}^{(1)}(\cdot) < f_{12}^{(2)}(\cdot)$.

However, targeting relatively more investment to the initially more disadvantaged child can still be efficient if $0 \leq f_{12}^{(1)}(\cdot) \leq f_{12}^{(2)}(\cdot)$. 
To establish this suppose that parents (or social planners) seek to maximize

$$\theta_3^A + \theta_3^B,$$

Subject to

$$E = p_1(l_1^A + l_1^B) + p_2(l_2^A + l_2^B).$$
First Order Conditions:

\[ f_1^{(2)} \left( f^{(1)} \left( \theta^A_1, I^A_1 \right), I^A_1 \right) f_2^{(1)} \left( \theta^A_1, I^A_1 \right) = \lambda p_1 \]

\[ f_2^{(2)} \left( f^{(1)} \left( \theta^A_1, I^A_1 \right), I^A_2 \right) = \lambda p_2 \]

\[ f_1^{(2)} \left( f^{(1)} \left( \theta^B_1, I^B_1 \right), I^B_2 \right) f_2^{(1)} \left( \theta^B_1, I^B_1 \right) = \lambda p_1 \]

\[ f_2^{(2)} \left( f^{(1)} \left( \theta^B_1, I^B_1 \right), I^B_2 \right) = \lambda p_2 \]

\[ p_1 \left( I^A_1 + I^B_1 \right) + p_2 \left( I^A_2 + I^B_2 \right) = E. \]
Consider an enhancement of the endowment of $A$ in the neighborhood of initial equality ($\theta_1^A = \theta_1^B$).

As before, let $\theta_1^A = \gamma \theta_1^B$.

Perturb $\gamma$ from position $\gamma = 1$.

Take total differentials of the system of first order conditions:
\[
\begin{align*}
\left\{ f^{(2)}_{11}(\cdot)\left[f^{(1)}_2(\cdot)\right]^2 + f^{(2)}_1(\cdot)f^{(1)}_{22}(\cdot) \right\} dl^A_{1} + \left\{ f^{(2)}_{12}(\cdot)f^{(1)}_2(\cdot) \right\} dl^A_{2} + \theta^B_1 \left[ f^{(2)}_{11}(\cdot)f^{(1)}_1(\cdot)f^{(1)}_2(\cdot) + f^{(2)}_1(\cdot)f^{(1)}_{21}(\cdot) \right] d\gamma \\
= (d\lambda)p_1 + \lambda dp_1
\end{align*}
\]

\[
\begin{align*}
\left\{ f^{(2)}_{21}(\cdot)f^{(1)}_2(\cdot) \right\} dl^A_{1} + \left\{ f^{(2)}_{22}(\cdot) \right\} dl^A_{2} + \theta^B_1 \left\{ f^{(2)}_{21}(\cdot)f^{(1)}_1(\cdot) \right\} d\gamma = (d\lambda)p_2 + \lambda dp_2
\end{align*}
\]

\[
\begin{align*}
\left\{ f^{(2)}_{11}(\cdot)\left[f^{(1)}_2(\cdot)\right]^2 + f^{(2)}_1(\cdot)f^{(1)}_{21}(\cdot)\theta^B_1 \right\} dl^B_{1} + \left\{ f^{(2)}_{12}(\cdot)f^{(1)}_2(\cdot) \right\} dl^B_{2} = (d\lambda)p_1 + \lambda dp_1
\end{align*}
\]

\[
\begin{align*}
\left\{ f^{(2)}_{21}(\cdot)f^{(1)}_2(\cdot) \right\} dl^B_{1} + \left\{ f^{(2)}_{22}(\cdot) \right\} dl^B_{2} = (d\lambda)p_2 + p_2 d\lambda
\end{align*}
\]

\[
\begin{align*}
-dE + p_1 dl^A_{1} + p_2 dl^A_{2} + p_1 dl^B_{1} + p_2 dl^B_{2} + l^A_1 dp_2 + l^A_2 dp_1 + l^B_1 dp_1 + l^B_2 dp_2 = 0.
\end{align*}
\]
A Three-Stage Analysis
• Fruitful to analyze the problem in three stages.
• In the first stage, we consider, for a single agent, how as $\gamma \uparrow$, the allocation of a fixed bundle of resources between investment in the first period and investment in the second period is affected.
• Then in the second stage consider how, as $\gamma \uparrow$, the productivity of expenditure changes and how resources are allocated across $A$ and $B$.
• Clearly, resources shift to wherever they become more productive.
• In the third stage, consider how an increase in resources is allocated between the first and the second periods.

• Sometimes convenient to use fictitious child A specific prices ($p_1^A$ and $p_2^A$) and child B specific prices $p_1^B$ and $p_2^B$. 
• Expenditures on child A and child B:

\[ E_A = p_1 l_1^A + p_2 l_2^A \]
\[ E_B = p_1 l_1^B + p_2 l_2^B. \]

• Maximize each of \( \theta_3^A \) and \( \theta_3^B \) separately subject to \( E_A \) and \( E_B \).

• Then allocate \( E_A \) and \( E_B \) to equalize marginal productivity of expenditure across A and B.
• Assume concavity of the production functions in terms of $\theta_1$ or $\theta_2$.
• This allows us to use standard results from consumer theory.
• The “$d\gamma$” terms act like income-compensated price changes.
• They do not affect total resources $E$. Assuming interior solutions, $\gamma \uparrow$ is like an income-compensated change in the (child-specific) input prices $p_1$ and $p_2$. 
The effect of $\gamma \uparrow$ on the allocation of investments across periods holding $E_A$ fixed
Consider the effect of an increase in $\gamma$ on the allocation of period one and period two investment of child $A$ while $E_A$ is fixed.

We consider the allocation of $E_A$ and $E_B$ across $A$ and $B$ later.

The displacement system derived from the first order conditions:

$$
\begin{bmatrix}
  c & d & -p_1 \\
  d & e & -p_2 \\
  -p_1 & -p_2 & 0
\end{bmatrix}
\begin{bmatrix}
  dl_1^A \\
  dl_2^A \\
  d\lambda
\end{bmatrix}
= 
\begin{bmatrix}
  \lambda dp_1 - \theta_1^B (\text{Term 1}) d\gamma \\
  \lambda dp_2 - \theta_1^B (\text{Term 2}) d\gamma \\
  0
\end{bmatrix}.
$$

(26)
• The income compensated own price changes are negative.
• Cross effects can be shown to be positive under the conditions specified below.
• \(|M| > 0\) from the assumption of a regular optimum.
• To simplify the notation, suppress the “·” notation. We can sign

\[
c = \left[ f_{11}^{(2)} [f_{2}^{(1)}]^2 + f_{1}^{(2)} f_{22}^{(1)} \right] \leq 0
\]

if period 2 production is concave in \(\theta_2\) and period 1 production is concave in \(l_1\).
• We assume that all marginal products are strictly positive unless otherwise noted.

• But $c$ might still be negative if period 2 production is convex in $\theta_2$ ($f_{11}^{(2)} > 0$) provided $f_1^{(2)} f_2^{(1)}$ is sufficiently negative.

$$d = f_{12}^{(2)} f_2^{(1)} > 0$$ if there is second period complementarity and

$$e = f_{22}^{(2)} < 0$$ from concavity in $l_2$.  


• In displacement system (26),

\[
\text{Term 1} \equiv \begin{bmatrix} f_{11}^{(2)} f_1^{(1)} f_2^{(1)} + f_1^{(2)} f_{21}^{(1)} \end{bmatrix}
\]

may be of either sign.
• The second grouping of terms in Term 1 is positive under first period complementarity.
• It is negative under substitutability.
• The first grouping is negative under concavity of $f^{(2)}$ in $\theta_2$.
• Under second period complementarity ($f^{(2)}_{21} > 0$) and

$$\text{Term 2} = \begin{bmatrix} f^{(2)}_{21} & f^{(1)}_1 \end{bmatrix} \geq 0.$$
• The change associated with Term 1 alone is opposite in sign to the change in the income-constant price of $I_1^A$ which is negative.

• Similarly, a change associated with Term 2 alone is opposite in sign from a change in the price of $I_2^A$. 
Using Cramer’s Rule,

\[
\frac{\partial I_1^A}{\partial \gamma} = \frac{\begin{vmatrix}
-\theta_1^B \text{(Term 1)} & d & -p_1 \\
-\theta_1^B \text{(Term 2)} & e & -p_2 \\
0 & -p_2 & 0
\end{vmatrix}}{|M|} \theta_1^B.
\]

\[
= \left( \frac{(\text{Term 1}) p_2^2 - (\text{Term 2}) p_1 p_2}{|M|} \right) \theta_1^B.
\]
• Focus on the numerator since the denominator is positive.
• Substitute out for $p_1$ and $p_2$ using the first order conditions (26).
• The numerator can be written as

\[
\theta_1^B \left[ \frac{1}{\lambda^2} \right] \left\{ \left[ f_{11}^{(2)} f_{1}^{(1)} f_{2}^{(1)} + f_{1}^{(2)} f_{21}^{(1)} \right] [f_{2}^{(2)}]^2 - \left[ f_{21}^{(2)} f_{1}^{(1)} f_{1}^{(2)} f_{2}^{(1)} f_{2}^{(2)} \right] \right\}.
\]
Focusing further on the term in braces (which is multiplied by a positive term), we obtain

\[\left\{f_{11}f_1^{(1)}f_2^{(1)}\left[f_2^{(2)}\right]^2 + f_1^{(2)}f_{21}^{(1)}\left[f_2^{(2)}\right]^2 - f_{21}^{(2)}\left[f_1^{(2)}\right]^2 f_2^{(2)}f_1^{(1)}f_2^{(1)}\right\} .\]
\[ f(1) = f_2^{(1)} \left( f_2^{(2)} \right)^2 f_1^{(2)} \]

\[ = f_2^{(1)} \left( f_2^{(2)} \right)^2 f_1^{(1)} \left[ \frac{f_1^{(2)}}{f_1^{(2)}} \left( f_1^{(1)} \right) + \frac{f_2^{(1)}}{f_2^{(1)}} - \frac{f_2^{(2)}}{f_2^{(2)}} f_1^{(2)} \right] \]

\[ = f_2^{(1)} \left( f_2^{(2)} \right)^2 f_1^{(1)} \left[ \frac{\partial \ln f_1^{(2)}}{\partial \theta^A_2} \right] + \left( \frac{\partial \ln f_2^{(1)}}{\partial \theta^A_1} \right) - \left( \frac{\partial \ln f_2^{(2)}}{\partial \theta^A_2} \right) \]

- Diminishing marginal productivity of \( \theta_2 \)
- Effect of \( \theta_1^A \) on marginal productivity of \( l_1^A \) (under complementarity)
- Effect of \( \theta_2^A \) on marginal productivity of \( l_2^A \) (under complementarity)

- Note that \( f_1^{(1)} = \frac{\partial \theta^A_2}{\partial \theta^A_1} \).
- This is the marginal self productivity of \( \theta_1 \).
Thus the term in brackets is:

\[
\frac{\partial \ln f_1^{(2)}}{\partial \theta_1^A} + \frac{\partial \ln f_2^{(1)}}{\partial \theta_1^A} - \frac{\partial \ln f_2^{(2)}}{\partial \theta_1^A}
\]

The effect of \( \theta_1^A \) on the marginal productivity of \( \theta_2^A \)
The effect of \( \theta_1^A \) on the marginal productivity of \( I_1^A \)
The effect of \( \theta_1^A \) on the marginal productivity of \( I_2^A \)

\[
= \frac{\partial}{\partial \theta_1^A} \left[ \ln f_1^{(2)} + \ln f_2^{(1)} - \ln f_2^{(2)} \right]
\]

(27)
• Consider the three effects inside the bracket going from left to right.
• The first term is the effect of $\theta^A_1$ on the marginal product of $\theta^A_2$ in period 2 production.
• From concavity (in terms of $\theta^A_2$), this term is negative.
• Diminishing returns is a force toward investing less in the first period.
• This term reflects how first period stocks of skills augment second period stocks of skills.
• If, example, \( f_1^{(1)} = 0 \) (so \( \frac{\partial \theta^A_2}{\partial \theta^A_1} = 0 \)), this term is zero.

• This could occur if there is 100% depreciation of skills or if there is a threshold value of \( \theta_1 \) beyond which increases in \( \theta_1 \) do not affect \( \theta_2 \) and the agent is at or beyond the threshold.

• If \( \theta^A_2 \) has a low or zero productivity in second period production, this term is small or zero.
• The second term is the effect of increasing $\theta_1^A$ on augmenting the productivity of first period investment in producing $\theta_2^A$.
• This is the term that drives the analysis in a one period model of childhood.
• The third term is the effect of increasing $\theta_1^A$ on augmenting the productivity of second period investment.

• Again, if there is no self-productivity ($\frac{\partial \theta_2^A}{\partial \theta_1^A} = 0$), this term is zero.

• Greater complementarity with later stages in the life cycle is a force toward investing less in the first period.
• Thus, in the absence of self-productivity \( f_1^1 = \frac{\partial \theta^A}{\partial \theta^1} = 0 \), the effect is driven solely by the second term.

• Under complementarity, the sign of the effect is positive.
Thus,

$$\frac{\partial I_1^A}{\partial \gamma} < 0$$

- if (a) $f^{(2)}$ concave in $\theta^A_2$, $f^{(1)}_{21} < 0$, $f^{(2)}_{21} > 0$
- and/or (b) $f^{(2)}$ is concave in $\theta_2$ and $\frac{\partial \ln f^{(1)}_2}{\partial \theta_1} < f^{(1)}_1 \frac{\partial \ln f^{(2)}_2}{\partial \theta_2}$, or
- if there are other configurations so that the term in brackets in (27) is positive.
• Because of the budget constraint it follows that

\[
\frac{\partial I_2^A}{\partial \gamma} > 0 \quad \text{if} \quad \frac{\partial I_1^A}{\partial \gamma} < 0.
\]

• The effects are offsetting.

• This is an analysis for allocation of investment \textit{within} the life cycle of child A.
Stage 2
The Effects on Productivity: Allocation Between $A$ and $B$
Let $\lambda_A$ be the productivity of expenditure on $A$.

- $\lambda_B$ is defined analogously for $B$.
- If, as $\gamma \uparrow$, $\lambda_A \uparrow$, it is optimal to allocate to $A (E_A \uparrow)$.
- If $\lambda_A \downarrow$ it is optimal to allocate less to $A (E_A \downarrow)$.
- The sign of this relationship hinges on the sign of Term 1 as we now show.
\[
\frac{\partial \lambda_A}{\partial \gamma} = \begin{vmatrix}
  c & d & (-\text{Term 1}) \\
  d & e & (-\text{Term 2}) \\
  -p_1 & -p_2 & 0 \\
\end{vmatrix}
\begin{vmatrix}
  M \\
\end{vmatrix}
\theta_1^B
\]
• Collecting terms and using the first order conditions (26), using

\[ p_1 = \frac{1}{\lambda} f_1^{(2)} f_2^{(1)} \quad \text{and} \quad p_2 = \frac{1}{\lambda} f_2^{(2)} \]
\[
\frac{\partial \lambda_A}{\partial \gamma} = \frac{\theta_1^B}{\lambda} \left( \frac{1}{|M|} \right) \left[ \underbrace{(\text{Term 1})[f_2^{(2)} d - f_1^{(2)} f_2^{(1)} e]}_{Q_1} - \underbrace{(\text{Term 2}) [f_2^{(2)} c - df_1^{(2)} f_2^{(1)}]}_{Q_2} \right],
\]

- Remember: \( d > 0; \ e < 0; \ c < 0. \)
\[ Q_1 = f_2^{(1)} \left[ f_2^{(2)} f_{12}^{(2)} - f_1^{(2)} f_{22}^{(2)} \right] > 0 \]

\[ \text{and} \]

\[ Q_2 = f_2^{(2)} f_{11}^{(2)} [f_2^{(1)}]^2 + f_2^{(2)} f_1^{(2)} f_{22}^{(1)} - f_{12}^{(2)} f_2^{(1)} f_1^{(2)} f_2^{(1)} < 0. \]
Thus

\[
\frac{\partial \lambda}{\partial \gamma} = \frac{\theta_1^B}{\lambda} \frac{1}{|M|} \left[ (\text{Term 1})(Q_1) + (\text{Term 2})Q_2 \right].
\]
So if Term 1 (+), then \( \frac{\partial \lambda A}{\partial \gamma} > 0 \).

This is a sufficient condition.

In this case, as \( \gamma \uparrow \) it is efficient to allocate *more* to \( A(E_A \uparrow) \).
• If Term 1 is sufficiently negative, it is optimal to allocate less to \( A(E_A) \downarrow \).

• Recall that a sufficient condition for Term 1 to be negative is that \( f_{21}^{(1)} < 0 \).

• But even if \( f_{21}^{(1)} > 0 \), if there is sufficiently strong diminishing returns in \( \theta_1(f_{11}^2 < 0) \), the optimal response of an increase in \( \gamma \) is to reduce \( I_1^A \) (i.e. to favor the disadvantaged child).
Stage 3:
Allocation of Changes in Endowments over Periods
• From standard results in consumer theory,

\[
\frac{\partial I_A^1}{\partial E_A} = \frac{(-1)}{|M|} \begin{vmatrix} d & -p_1 \\ e & -p_2 \end{vmatrix} = \frac{dp_2 - p_1 e}{|M|} = \frac{(f_2^{(1)})}{\lambda |M|} \left[ f_{12}^{(2)} f_2^{(2)} - f_1^{(2)} f_{22}^{(2)} \right] \geq 0.
\]

• \( d > 0, \ e < 0. \)

• Recall we assume \( f_{12}^{(2)} > 0. \)

• From concavity it follows that \( f_{22}^{(2)} < 0. \)

• Thus \( \frac{\partial I_A^1}{\partial E_A} > 0. \)
\[
\frac{\partial l_2^A}{\partial E_A} = \begin{vmatrix} c & -p_1 \\ d & -p_2 \end{vmatrix} = \frac{1}{\lambda |M|} \left\{ \left( f_2^{(1)} \right)^{(2)} f_1^{(2)} f_{12} - \left( f_2^{(2)} \right) \left[ f_{11}^{(2)} \left( f_2^{(1)} \right)^{(2)} + f_1^{(2)} f_{22}^{(1)} \right] \right\}
\]

- This expression is also positive.
- Thus inputs are normal under our assumptions.
- For the case \( p_1 = p_2 = 1 \) (which we can assume with no loss of generality).
\[ \frac{\partial I_1^A}{\partial E_A} = \frac{f_{12} f_2^{(1)} - f_2^{(2)}}{|M|} \]

\[ \frac{\partial I_2^A}{\partial E_A} = \frac{f_{11}^{(2)} (f_2^{(1)})^2 - f_1^{(2)} f_2^{(1)} + f_{12}^{(2)} f_2^{(1)}}{|M|} \]
Observe that $\frac{\partial I_A}{\partial E_A}$ is larger

(a) the greater the second period complementarity ($f_{12}^{(2)}$) (so that $I_A^1$ has greater productivity in producing final output $\theta_A^3$),

(b) the larger $f_2^{(1)} = \frac{\partial \theta_A^2}{\partial I_A^1}$ (so that $I_A^1$ is more productive in producing the intermediate product $\theta_A^2$);

(c) the more rapidly the decline in the productivity of $I_A^2$.

Intuitively, relatively more is allocated to first period investment the more productive is the first period investment.
Putting It All Together
• The second step is the key one.
• It determines the allocation of expenditure across children in response to an increase in endowment ($\gamma \uparrow$).
• The greater the decline in self productivity with increases in $\theta_1$ (the more negative $f_{11}^{(2)}$), the more likely it is that more resources are devoted to the less advantaged child.
• This negative effect is amplified by greater productivity of $\theta_1$ in period 1 ($f_{1}^{(1)}$) and greater productivity of $l_1$ in period 1.
● These effects are reinforced if there is substitutability between $\theta_1$ and $l_1(f_{21}^{(1)} < 0)$.

● If $f_{21}^{(1)}$ is positive, the redistributive effect is attenuated.

● This offsetting effect is weaker the smaller the productivity of $\theta_2$ in period 2 production.
• The first step explores substitution effects arising from the change in $\gamma$.
• The third step explores income effects across periods arising from transfers across children.
• The other steps determine the allocation of investment across periods for each child.
• The analysis of the third step for each child informs us that resources are differentially allocated to the more productive period.
• The analysis of the first step makes a similar claim but investigates how changes in $\gamma$ affect the relative productivity of investment in each period.
In Section 208 below, we establish that if first period investment \((I_1)\) and initial endowment \((\theta_1)\) are substitutes, \((f_{12}^{(1)} < 0)\), but \(\theta_2\) is complementary with second period investments \((f_{12}^{(2)} > 0)\), first period investments are greater for the more disadvantaged child.
But even if \( f^{(1)}_{12} > 0 \), greater first period investment in the initially disadvantaged child may be optimal.

This is more likely (ceteris paribus).
(a) the more steeply diminishing is the productivity of second period skills \( (f_{22}^{(2)}) \);  
(b) the greater the self productivity of the stock of skills in the first period \( (f_{1}^{(1)} = \frac{\partial \theta^2}{\partial \theta_1}) \);  
(c) the smaller first period complementarity \( (f_{21}^{(1)}) \) relative to second period complementarity and absolutely  
(d) the more rapidly diminishing the marginal productivity of \( \theta_1 (f_{11}^{(1)}) \);  
(e) the greater the second period complementarity \( (f_{12}^{(2)}) \);  
(f) the greater the first period productivity of investment \( (f_{2}^{(1)}) \) and  
(g) the more rapidly diminishing the productivity of second period investment \( (f_{22}^{(2)}) \).
• Roughly speaking, the more concave are the technologies in terms of stocks of skills, the more favorable is the case for investing relatively more in the disadvantaged child.

• The greater the second period complementarity \( (f_{12}^{(2)}) \), the greater the case for investing more in the initially disadvantaged child to allow the child to benefit from greater second period complementarity of the stock of skills with second period investment.
• In general, even when investment is greater in the first period for the disadvantaged child, second period investment is greater for the initially advantaged child.

• It is generally not efficient to make the initially disadvantaged child whole as it enters the second period when the effect of greater second period complementarity kicks in.
Appendix

- Direct proofs of some additional propositions
Proof that $f_{12}^{(1)} < 0$ is sufficient for $\frac{\partial f_A}{\partial \gamma} < 0$. 
Consider the bordered Hessian displacement system associated for the problem for both children treated together:

\[
\begin{bmatrix}
  c & d & 0 & 0 & -p_1 \\
  d & e & 0 & 0 & -p_2 \\
  0 & 0 & c & d & -p_1 \\
  0 & 0 & d & e & -p_2 \\
  -p_1 & -p_2 & -p_1 & -p_2 & 0
\end{bmatrix}
\begin{bmatrix}
  dl_1^A \\
  dl_2^A \\
  dl_1^B \\
  dl_2^B \\
  d\lambda
\end{bmatrix}
= 
\begin{bmatrix}
  \lambda dp_1 - \theta_A^1 (\text{Term 1}) d\gamma \\
  \lambda dp_2 - \theta_B^1 (\text{Term 2}) d\gamma \\
  \lambda dp_1 \\
  \lambda dp_2 \\
  -dE + \sum_{j \in \{A,B\}} \sum_{l \in \{1,2\}} l^j \cdot dp_l
\end{bmatrix}
\]
where as before

\[ c = \left[ f_{11}^{(2)} f_{2}^{(1)} \right]^2 + f_{1}^{(2)} f_{22}^{(1)} \leq 0 \]
• if period 2 production is concave in $\theta_2$ and period 1 production is concave in $I_1$.
• But it might also arise if period 1 production is convex in $\theta_2$.

\[ d = f^{(2)}_{12} f^{(1)}_2 > 0 \]  
if there is second period complementarity

\[ e = f^{(2)}_{22} < 0 \]  
from concavity in $I_2$. 
Recall that

\[ T_1 \equiv \text{Term } 1 \equiv \left[ f_{11}^{(2)} f_1^{(1)} f_2^{(1)} + f_1^{(2)} f_{21}^{(1)} \right] \]

may be of either sign.
• The second grouping of terms in Term 1 is positive under complementarity in the first period; negative under substitutability.

• The first grouping is negative under concavity of \( f^{(2)} \) in \( \theta_2 \) (but it might be positive if there are increasing returns).

• Under second period complementarity \( (f_{21}^{(2)} > 0) \)

\[
T_2 \equiv \text{Term 2} = \begin{bmatrix} f_{21}^{(2)} & f_1^{(1)} \end{bmatrix} \geq 0.
\]
• Let $H$ be the bordered Hessian associated with displacement system (28) and let $|H|$ be the determinant of the Hessian.
• $|H| > 0$ under the assumption of a regular optimum.
Then the income-compensated effect of a change in $p_2^A$ on $I_1^A$ is

$$\frac{\partial I_1^A}{\partial p_2^A} = \lambda \begin{vmatrix} d & 0 & 0 & -p_1 \\ 0 & c & d & -p_1 \\ 0 & d & e & -p_2 \\ 0 & -p_1 & -p_2 & 0 \end{vmatrix} / |H|$$

$$= \lambda d \begin{vmatrix} c & d & -p_1 \\ d & e & -p_2 \\ -p_1 & -p_2 & 0 \end{vmatrix} / |H|. \quad (29)$$
• The numerator of (29) is negative from the sufficiency conditions for an optimum for the two stage budgeting problem for $A$ and from second period dynamic complementarity ($d > 0$).
• Hence both inputs are Hicks-compensated cross substitutes:

\[ \frac{\partial I_1^A}{\partial p_2^A} < 0. \]

• and from symmetry

\[ \frac{\partial I_1^A}{\partial p_2^A} = \frac{\partial I_2^A}{\partial p_1^A} < 0. \]
Collecting results,

let \( S_{ij} = \frac{\partial I_i^A}{\partial p_j^A} \quad i, j \in \{1, 2\} \)

\[
\frac{\partial I_1^A}{\partial \gamma} = - \left\{ [S_{11}] [\text{Term 1} (-)] + [S_{12}] [\text{Term 2} (+)] \right\} d\gamma \tag{30}
\]

\[
\frac{\partial I_2^A}{\partial \gamma} = - \left\{ [S_{12}] [\text{Term 1} (-)] + [S_{22}] [\text{Term 2} (+)] \right\} d\gamma. \tag{31}
\]
• If Term 1 is sufficiently negative, which could happen even if $f_{21}^{(1)}(\cdot) > 0$, then

$$\frac{\partial I_1^A}{\partial \gamma} < 0.$$ 

• (Term 1 would be negative if $f_{21}^{(1)} < 0$) and possibly even

$$\frac{\partial I_2^A}{\partial \gamma} < 0.$$
• Term 1 positive \( \Rightarrow \frac{\partial I^A_1}{\partial \gamma} > 0 \) and \( \frac{\partial I^B_1}{\partial \gamma} < 0 \).
• Thus it may be efficient to allocate more to the less endowed, even in both periods.
• We can say something stronger.
• If $f_{12}^{(1)} < 0$, but $f_{12}^{(2)} > 0$, then as $\gamma \uparrow$, $I_A \downarrow$ and the term in braces in (30) is positive.
• To prove this define $T_1 = \text{Term 1}$ and $T_2 = \text{Term 2}$ and notice that

$$\frac{\partial I^A}{\partial \gamma} = \begin{vmatrix} -T_1 \theta_1^B & d & 0 & 0 & -p_1 \\ -T_2 \theta_1^B & e & 0 & 0 & -p_2 \\ 0 & 0 & c & d & -p_1 \\ 0 & 0 & d & e & -p_2 \\ 0 & -p_2 & -p_1 & -p_2 & 0 \end{vmatrix} = \frac{|N|}{|H|} \theta_1^B,$$
where

\[ N = \begin{cases} 
- T_1 e & \begin{vmatrix} c & d & -p_1 \\ d & e & -p_2 \\ -p_2 & -p_2 & 0 \end{vmatrix} \\
- T_1 e & \begin{vmatrix} 0 & 0 & -p_2 \\ c & d & -p_1 \\ d & e & -p_2 \end{vmatrix} \\
\hline
\hline
|M| > 0 
\end{cases} \]

\[ + T_2 d & \begin{vmatrix} c & d & -p_1 \\ d & e & -p_2 \\ -p_1 & -p_2 & 0 \end{vmatrix} \]

\[ + T_2 p_2 & \begin{vmatrix} 0 & 0 & -p_1 \\ c & d & -p_1 \\ d & e & -p_2 \end{vmatrix} \]

\[ \theta_1^B \]
\[ |N| = \begin{vmatrix} (-T_1 e + T_2 d)|m| - (T_1)p_2 (-p_2) & c & d \\ d & e \end{vmatrix} + T_2 p_2 (-p_1) & c & d \\ d & e \end{vmatrix} \theta_1^B \\
\text{if } T_1 < 0 \]

\(-\text{if } T_1 < 0\)
• Thus it follows as a sufficient condition that

\[ |N| < 0 \text{ if } \left( -T_1 e + T_2 d \right) < 0. \]

• Writing out \((-T_1 e + T_2 d),\)

\[
(-T_1 e + T_2 d) = -f_{11}^{(2)} f_1^{(1)} f_2^{(1)} f_{22}^{(2)} - f_1^{(2)} f_{21}^{(1)} f_{22}^{(2)} + f_{21}^{(2)} f_1^{(1)} f_{12}^{(2)} f_2^{(1)},
\]
and collecting the first and the last terms:

\[-f_1^{(1)} f_2^{(1)} \left[ f_{11}^{(2)} f_{22}^{(2)} - [f_{12}^{(2)}] \right] - f_1^{(2)} f_2^{(1)} f_2^{(2)}\]

\( (+) \) by concavity

\(- \) if \( f_2^{(1)} < 0 \)
so

\[ (-T_1 e + T_2 d) < 0 \quad \text{if} \quad f_{21}^{(1)} < 0, \]

and hence

\[ |N| < 0 \quad \text{if} \quad f_{21}^{(1)} < 0, \]
• so

\[ \frac{\partial I_{1}^{A}}{\partial \gamma} < 0 \] if \( f_{21}^{(1)} < 0 \).

• Notice, however, that even if \( f_{21}^{(1)}(\cdot) > 0 \), it is possible that

\[ \frac{\partial I_{1}^{A}}{\partial \gamma} < 0. \]
• (See the second term in equation (32).)
• Notice that the more negative $f^{(2)}_{22}$ (i.e., the more sharply are the diminishing returns to $l^A_2$ in period 2), the more negative is $\frac{\partial l^A_1}{\partial \gamma}$. 
• The intuition for this offsetting effect is that as second period investments become less effective, then it is more productive to invest relatively more in the first period.

• Concavity in terms of $\theta_2$ is not strictly required.
Next consider

\[ \frac{\partial I_2^A}{\partial \gamma} = \begin{vmatrix} c & -T_1 \theta^B_1 & 0 & 0 & -p_1 \\ d & -T_2 \theta^B_1 & 0 & 0 & -p_2 \\ 0 & 0 & c & d & -p_1 \\ 0 & 0 & d & e & -p_2 \\ -p_1 & 0 & -p_1 & -p_2 & 0 \end{vmatrix} \]

\[ |H| \]

\[ = \frac{\tilde{N}}{|H|} \theta^B_1 \]
\[ \tilde{N} = T_1 \begin{vmatrix} d & 0 & 0 & -p_2 \\ 0 & c & d & -p_1 \\ 0 & d & e & -p_2 \\ -p_1 & -p_1 & -p_2 & 0 \end{vmatrix} \]

\[ -T_2 \begin{vmatrix} c & 0 & 0 & -p_1 \\ 0 & c & d & -p_1 \\ 0 & d & e & -p_2 \\ -p_1 & -p_1 & -p_2 & 0 \end{vmatrix} \]
\[ T_1 \begin{bmatrix}
  d & c & d & -p_1 \\
  d & d & e & -p_2 \\
 -p_1 & -p_2 & 0 & 0 \\
 -p_1 & -p_2 & 0 & 0 \\
\end{bmatrix}
+ p_1 \begin{bmatrix}
  0 & 0 & -p_2 \\
  c & d & -p_1 \\
  d & e & -p_2 \\
\end{bmatrix} \]
\[- T_2 \begin{bmatrix}
c & d & -p_1 \\
d & e & -p_2 \\
-p_1 & -p_2 & 0
\end{bmatrix} + p_1 \begin{bmatrix}
0 & 0 & -p_1 \\
c & d & -p_1 \\
d & e & -p_2
\end{bmatrix}\]
\[ = (T_1d - T_2c) \begin{vmatrix} c & d & -p_1 \\ d & e & -p_2 \\ -p_1 & -p_2 & 0 \end{vmatrix} \]

\[ - T_1p_1p_2 \begin{vmatrix} c & d \\ d & e \end{vmatrix} + T_2(p_1)^2 \begin{vmatrix} c & d \\ d & e \end{vmatrix} \]

(-) if \( T_1 < 0 \)

(+)

- Focus on the term \((T_1d - T_2c)\)
\[
\left( T_1 d - T_2 c \right) - \left( T_1 p_1 - T_2 p_2 \right) = \left( \begin{array}{ccc} c & d & -p_1 \\ -p_1 & -p_2 & 0 \end{array} \right) - \left( \begin{array}{ccc} T_1 p_1 & -p_1 \\ -p_2 & T_2 p_2 \end{array} \right)
\]
Observe that

\[
(T_1 d - T_2 c) = \left[ f_{11}^{(2)} f_1^{(1)} f_2^{(1)} + f_{12}^{(1)} f_1^2 \right] f_{12}^{(2)} f_2^{(1)}

- f_{21}^{(2)} f_1^{(1)} \left[ f_{11}^{(2)} (f_2^{(1)})^2 + f_1^{(2)} f_{22}^{(1)} \right]
\]
\[
= f_{11}^{(2)} f_{1}^{(1)} f_{2}^{(1)} f_{12}^{(2)} f_{2}^{(1)} \\
+ f_{12}^{(1)} f_{1}^{(2)} f_{12}^{(2)} f_{2}^{(1)} \\
- f_{21}^{(2)} f_{1}^{(1)} f_{11}^{(2)} \left( f_{2}^{(1)} \right)^2 \\
- f_{21}^{(2)} f_{1}^{(1)} f_{1}^{(2)} f_{12}^{(1)} \\
= f_{12}^{(2)} f_{1}^{(2)} \\
\left[ f_{12}^{(1)} f_{12}^{(1)} - f_{1}^{(1)} f_{22}^{(1)} \right] \\
\text{(+) }
\overbrace{T_3}^{T_3} \\
\]

- and the last term is positive \((T_3 > 0)\), if in the period 1 production function \(f_{12}^{(1)} > 0\) (first period complementarity).
This is a sufficient condition for

\[ \frac{\partial I^A_{2}}{\partial \gamma} > 0. \]
• Notice that when Term 1 \((T_1)\) is negative, then \(T_3\) can be negative.\(^1\)

• Thus, it is possible that the efficient policy redistributes to the less endowed in period 1 but to the more endowed in period 2.
• It is also possible that as $\gamma \uparrow$, it is socially efficient to invest in the disadvantaged child in both periods, although this seems unlikely.

• In general, it is not efficient to make the initially disadvantaged child whole by the start of the second period, and second period complementarity reinforces starting of second period discrepancies.
Return to main text
Appendix: Some Evidence from Simulations on Why Dynamic Complementarity is a Force Toward Targeting Disadvantaged Children in the Early Years
• Dynamic complementarity is a force toward equalization of early stage investments even in the absence of family inequality aversion.

• To illustrate the mechanism underlying this claim, suppose that, for each child $k$, the outcome of interest for parents are children’s earnings $E_k$ and that they are a function of children’s adult human capital determined by “genes” $(\theta_{1,k})$ and early ($I_{1,k}$) and late ($I_{2,k}$) parental investments.
\[ E_k = wf^2(\theta_2,k, l_2,k) = f^2 \left( \gamma_2 \phi_2^{\theta_2,k} + (1 - \gamma_2) l_2^{\phi_2,k} \right)^{\rho_2 \phi_2} \] (33)

with

\[ \theta_2,k = f^1(\theta_1,k, l_1,k) = f^1 \left( \gamma_1 \phi_1^{\theta_1,k} + (1 - \gamma_1) l_1^{\phi_1,k} \right)^{\rho_1 \phi_1} \] (34)

- Where \( w \) is the payment to skill corresponding to one unit of human capital which is determined by equilibrium in the factor markets.

- Since \( w \) is common across families and siblings we assume that the measurement of human capital is chosen so that \( w = 1 \).
• The budget constraint faced by the parents with total resources $R^e$ is:

$$p_1 \sum_{k=1}^{n} I_1 + p_2 \sum_{k=1}^{n} I_2 = R^e. \quad (35)$$

• Consider the case of a parent with two children $i$ and $j$.

• We show that even in the absence of inequality aversion, the shape of the technology, and in particular the presence of decreasing returns in at least one of the two periods, might induce parents to follow a compensating strategy devoting more resources to the less endowed child, say $j$ ($\theta_{1,i} > \theta_{1,j}$).
• As a measure of parental compensation with respect to initial inequality we define the parameter $\tau$ as:

$$\tau \equiv \left( \frac{E_i}{E_j} \right) / \left( \frac{\theta_i}{\theta_j} \right),$$  \hspace{1cm} (36)

• Which captures how much earnings differences are inflated compared to initial endowment differences.

• If $\tau = 1$, the parents perfectly translate “genetic” differences into earnings.

• In results from a simulation exercise, Figure 6 shows that earnings differences are dampened compared to differences in initial endowments whenever $\rho_1 < 1$. 
Notes: The parental preference parameters used in the simulation are $\sigma = 1$ and $\omega_i = \omega_j = 0.5$. Total resources are $R^e = 4$. The technology of skill formation parameters, capturing increasing complementarity between skills and investments over time, are: $\gamma_1 = \gamma_2 = 0.5$, $\phi_1 = 0.6$, $\phi_2 = -0.5$, $\rho_2 = 1$. The parameter $\rho_1$ defines the degree of homogeneity of the first period technology. We vary the value of $\rho_1$ over the range $[0.1, 1]$. Child $i$ has a skill endowment of 5 while child $j$ of 1.
• We also consider the how changes in $\rho_1$ affect parental behavior in Figures 7, 8, and 9.
• Figure 7 shows the ratio of early ($I_1$) to late ($I_2$) investments.
Figure 7: Ratio Early to Late Investments

Notes: The solid line refers to the most endowed child, the dashed line to the least endowed child. The parameters used are as in Figure 6.
This ratio is always higher for the less endowed child $j$ whenever $\rho_1$ is smaller than one.

Figure 8 shows that the less endowed child receives a higher amount of early investment whenever the period 1 technology exhibits substitutability between skills (initial endowments) and investments (i.e. when $\rho_1 < \phi_1$).
Notes: The solid line refers to the most endowed child, the dashed line to the least endowed child. The parameters used are as in Figure 6.
- Figure 9 shows that the most endowed child always receives a higher level of late investment.
Figure 9: Levels of Late Investments

Notes: The solid line refers to the most endowed child, the dashed line to the least endowed child. The parameters used are as in Figure 6.
• Late investments are an increasing function of $\rho_1$ for the more endowed child while they are decreasing in $\rho_1$ for the less endowed child.

• As $\rho_1$ decreases the less endowed child receives a higher level of early investments and a level of late investments which is increasingly closer to the one of his more endowed brother.

• This explains why earnings tend to be equalized as $\rho_1$ decreases.
• We conclude that if the technology of skill formation is defined over more than one period, parents might exhibit compensating behavior in investments in children’s human capital even in absence of inequality aversion.

• In particular, less endowed children receive a higher level of early investment than their more endowed siblings if the technology of skill formation exhibits substitutability between initial (genetic) endowments and the level of early investments.
Return to text
Appendix:
Overview of Structural Models of Parental Investments
Table 2: Structural Models of Parental Investments
(“✓” means present; “X” means absent)

<table>
<thead>
<tr>
<th></th>
<th>OLG Model</th>
<th>Dynastic Links</th>
<th>Explicit Models of Parental Preferences, Altruism (a) or Paternalism (p)</th>
<th>Model Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cunha and Heckman (2007)</td>
<td>✓</td>
<td>A, B, C</td>
<td>✓ (a)</td>
<td>X</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>✓</td>
<td>A, B, C</td>
<td>✓ (a)</td>
<td>✓</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>✓</td>
<td>B, C</td>
<td>✓ (a)</td>
<td>✓</td>
</tr>
<tr>
<td>Del Boca et al. (2014b)</td>
<td>X</td>
<td>X</td>
<td>✓ (p)</td>
<td>✓</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>✓</td>
<td>A, C</td>
<td>✓ (p)</td>
<td>✓</td>
</tr>
<tr>
<td>Cunha et al. (2013)</td>
<td>X</td>
<td>X</td>
<td>✓ (p)</td>
<td>✓</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>X</td>
<td>X</td>
<td>✓ (p)</td>
<td>✓</td>
</tr>
<tr>
<td>Lee and Seshadri (2016)</td>
<td>✓</td>
<td>A, B, C</td>
<td>✓ (a)</td>
<td>✓</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>X</td>
<td>A, B, C</td>
<td>✓ (a)</td>
<td>✓</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>X</td>
<td>A, B, C</td>
<td>✓ (a)</td>
<td>✓</td>
</tr>
</tbody>
</table>

A: Through parental skills, B: Through asset transfers, C: Through genes (initial conditions), D: Natural borrowing limit, E: Limits can be more stringent than natural limit.
Table 2: Structural Models of Parental Investments (“✓” means present; “X” means absent)

<table>
<thead>
<tr>
<th>Study</th>
<th>Parental Goods Investment</th>
<th>Parental Time Investment</th>
<th>Technology Depends on Parental Skill</th>
<th>Self-productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cunha and Heckman (2007)</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Del Boca et al. (2014b)</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Cunha et al. (2013)</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lee and Seshadri (2016)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Restuccia and Urrutia (2004)</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

A: Through parental skills, B: Through asset transfers, C: Through genes (initial conditions), D: Natural borrowing limit, E: Limits can be more stringent than natural limit.
Table 2: Structural Models of Parental Investments (“✓” means present; “X” means absent)

<table>
<thead>
<tr>
<th>Study</th>
<th>Parental Learning About Technology</th>
<th>Bequests</th>
<th>Intrigenerational Borrowing</th>
<th>Multiple Skills of Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cunha and Heckman (2007)</td>
<td>X</td>
<td>✓</td>
<td>✓&lt;sup&gt;E&lt;/sup&gt;</td>
<td>X</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>X</td>
<td>✓</td>
<td>✓&lt;sup&gt;D&lt;/sup&gt;</td>
<td>X</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>X</td>
<td>X</td>
<td>✓&lt;sup&gt;E&lt;/sup&gt;</td>
<td>X</td>
</tr>
<tr>
<td>Del Boca et al. (2014b)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cunha et al. (2013)</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lee and Seshadri (2016)</td>
<td>X</td>
<td>✓</td>
<td>✓&lt;sup&gt;D&lt;/sup&gt;</td>
<td>X</td>
</tr>
<tr>
<td>Restuccia and Urrutia (2004)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

A: Through parental skills, B: Through asset transfers, C: Through genes (initial conditions), D: Natural borrowing limit, E: Limits can be more stringent than natural limit.
Table 2: Structural Models of Parental Investment ("✓" means present; "X" means absent)

<table>
<thead>
<tr>
<th>Multichild Families (Preferences for Equity vs. Efficiency)</th>
<th>Endogenous Fertility Decisions</th>
<th>Multiple Parents</th>
<th>Endogenous Mating Decisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cunha and Heckman (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Del Boca et al. (2014b)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cunha et al. (2013)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lee and Seshadri (2016)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Restuccia and Urrutia (2004)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

A: Through parental skills, B: Through asset transfers, C: Through genes (initial conditions), D: Natural borrowing limit, E: Limits can be more stringent than natural limit.
Return to main text
Inequality in Human Capital and Endogenous Credit Constraints (2016)
Rong Hai and James J. Heckman
• A dynamic model of schooling and working
• Agents subject to uninsured human capital risks, and face restrictions on their borrowing possibilities
## Table 3: Structural Models on Educational Choices and Credit Constraints

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Human Capital Investment</th>
<th>Labor Supply</th>
<th>GSL</th>
<th>Private Loan Limit</th>
<th>CRRA Risk Aversion</th>
<th>Parental Influence</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keane and Wolpin (2001) Education and work experience</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Borrowing limits not observed, proxied by a function of age and human capital; the parameters of the borrowing limit are estimated</td>
<td>Estimate $\gamma = 0.4826$</td>
<td>Parental transfers is a function of parental education and individuals choices</td>
<td>NLSY79 (1979-1992)</td>
</tr>
<tr>
<td>Navarro (2011)</td>
<td>Education</td>
<td>No</td>
<td>No</td>
<td>Models an endogenous natural borrowing limit based on education and inelastic labor supply, due to borrowers' limited repayment ability in the presence of uninsurable wage risk; however, in estimation, the borrowing limit is set equal to the lowest level of assets observed in samples in each period independently of the model</td>
<td>Estimate $\gamma = 0.82$</td>
<td>No</td>
<td>NLSY79 and PSID</td>
</tr>
<tr>
<td>Lochner and Monge-Naranjo (2011) Education and work experience</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Endogenous credit limit based on borrowers’ cost of default (including temporary exclusion from credit market and wage garnishments), due to private lenders’ limited ability to punish default; parameters on the cost of default are calibrated outside the model</td>
<td>Set $\gamma = 2$</td>
<td>No</td>
<td>NLSY79 (1979-2006)</td>
</tr>
</tbody>
</table>

**Source:** NLSY: National Longitudinal Survey of Youth. BHPS: British Household Panel Survey.
### Table 3: Structural Models on Educational Choices and Credit Constraints Cont’d

<table>
<thead>
<tr>
<th>Model</th>
<th>Human Capital Investment</th>
<th>Labor Supply</th>
<th>GSL</th>
<th>Private Loan Limit</th>
<th>CRRA Risk Aversion</th>
<th>Parental Influence</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson (2013)</td>
<td>Education and work experience</td>
<td>Yes</td>
<td>Yes</td>
<td>Borrowing limits not observed; proxied by a function of age and human capital; the parameters of the borrowing limit proxy equation are estimated</td>
<td>Set $\gamma = 2$</td>
<td>Parental transfers is a function of parental income and child choices</td>
<td>NLSY97 (1997 to 2007)</td>
</tr>
<tr>
<td>Abbott, Gallipoli, Meghir, Violante (2016)</td>
<td>Education</td>
<td>Yes</td>
<td>Yes</td>
<td>No borrowing for high-school students; exogenous fixed debt limits for workers in the work-stage and college students whose parents are wealthy, respectively; borrowing limits are calibrated to match the fraction of households with zero or negative net worth and the aggregate private loans/GSL ratio, respectively</td>
<td>Set $\gamma = 2$</td>
<td>Parental transfers explicitly modeled</td>
<td>Multiple data, including NLSY79, NLSY97</td>
</tr>
<tr>
<td>Blundell, Costas Dias, Meghir, Shaw (2016)</td>
<td>Education and work experience</td>
<td>Yes</td>
<td>Yes</td>
<td>No borrowing permitted</td>
<td>Set $\gamma = 1.56$</td>
<td>Parental income and background factors affect youth’s psychic cost of schooling</td>
<td>BHPS (1991 to 2008)</td>
</tr>
<tr>
<td>This paper</td>
<td>Education and work experience</td>
<td>Yes</td>
<td>Yes</td>
<td>Model-determined natural borrowing limit based on education and labor supply decisions, due to borrowers’ limited repayment ability in the presence of uninsurable wage risk; no new auxiliary parameters for borrowing limit is added in estimation, unlike many previous papers</td>
<td>Set $\gamma = 2$</td>
<td>Parental transfer is a function of parental education and net worth, and individuals choices; parental education affects youth’s psychic cost of schooling</td>
<td>NLSY97 (1997-2013)</td>
</tr>
</tbody>
</table>

- Build model with natural borrowing limit
- Extend the existing literature by analyzing how cognitive and noncognitive ability affect choices through:
  (i) Psychic costs of working and schooling
  (ii) The technology of human capital production
  (iii) The discount factor
• Strong effects of adolescent endowments of cognitive and noncognitive ability on human capital development.
• Tuition costs and family transfers to children play important roles in explaining differences in life outcomes due to human capital investments.
Who Are the Credit Constrained?

- Credit constrained agents fall into two groups:
  (a) Those with poor initial endowments and family background who acquire little human capital and have low wage levels and low life cycle wage growth
  (b) The very able and those from good family backgrounds who have high levels of human capital, high wage levels, and high life cycle wage growth
Figure 10: Evolution of Average Natural Borrowing Limit by Ability Endowments

(a) Natural Borrowing Limit vs Cognitive Ability over Ages 17 to 50

(b) Natural Borrowing Limit vs Noncog. Ability over Ages 17 to 50
Kuhn Tucker Multiplier on the Borrowing Constraint
(This is constraint on intertemporal lending and borrowing)

\[ \lambda_{s,t}(c_t, s_{t+1}; \Omega_t) = \frac{\partial u_c(c_t; \Omega_t)}{\partial c_t} - \exp(-\rho(\theta_c, \theta_n)) \left( \frac{\partial E V_{t+1}}{\partial s_{t+1}} \right) \]  

(37)
Figure 11: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 21

Fraction Constrained ($\lambda_{s,t} > 0$) & $s_t$

Local polynomial smooth

Net Worth

Multiplier Positive

-50000 0 50000 100000

95% CI

kernel = epanechnikov, degree = 0, bandwidth = 592.01, pwidth = 888.01
Figure 11: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 21 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) & $\theta_c$

Local polynomial smooth

Cognitive Ability

Multiplier Positive

kernel = epanechnikov, degree = 0, bandwidth = .35, pwidth = .52
Figure 11: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 21 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) & $\theta_n$

Local polynomial smooth

Noncognitive Ability

Multiplier Positive

95% CI

kernel = epanechnikov, degree = 0, bandwidth = .39, pwidth = .59
Figure 11: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 21 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) & $F^\psi(e_t, k_t, \theta, \epsilon_{w,t})$

Local polynomial smooth

Human Capital

Multiplier Positive

$\begin{align*}
\text{Fraction Constrained } & (\lambda_{s,t} > 0) \\
& \& F^\psi(e_t, k_t, \theta, \epsilon_{w,t})
\end{align*}$

kernel = epanechnikov, degree = 0, bandwidth = .08, pwidth = .12
Figure 12: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 30

Fraction Constrained ($\lambda_{s,t} > 0$) vs $s_t$
Figure 12: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 30 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) vs $\theta_c$

Local polynomial smooth

Cognitive Ability

Multiplier Positive

95% CI

kernel = epanechnikov, degree = 0, bandwidth = .28, pwidth = .42
Figure 12: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 30 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) vs $\theta_n$

Local polynomial smooth
Figure 12: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 30 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) vs $F^\psi(e_t, k_t, \theta, \epsilon_{w,t})$

Local polynomial smooth

Human Capital

Multiplier Positive

95% CI

Ipoly smooth

kernel = epanechnikov, degree = 0, bandwidth = .09, pwidth = .13
Figure 13: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 40

Fraction Constrained ($\lambda_{s,t} > 0$) vs $s_t$

Local polynomial smooth

Net Worth

Multiplier Positive

95% CI  Ipoly smooth

kernel = epanechnikov, degree = 0, bandwidth = 1878.82, pwidth = 2818.23
Figure 13: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 40 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) vs $\theta_c$

Local polynomial smooth
Figure 13: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 40 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) vs $\theta_n$

Local polynomial smooth

Kernel = epanechnikov, degree = 0, bandwidth = .74, pwidth = 1.11
Figure 13: Multiplier $\lambda_{s,t}$ and Borrowing Constrained Youths $\lambda_{s,t} > 0$ at Age 40 Cont’d

Fraction Constrained ($\lambda_{s,t} > 0$) vs $F^\psi(e_t, k_t, \theta, \epsilon_{w,t})$

Local polynomial smooth

kernel = epanechnikov, degree = 0, bandwidth = .17, pwidth = .25
Figure 14: Borrowing Constrained ($\lambda_{s,t} > 0$) & Human Capital $F^\psi(e_t, k_t, \theta, \epsilon_{w,t})$

Age 21

Local polynomial smooth

Age 30

Local polynomial smooth

Age 40

Local polynomial smooth

kernel = epanechnikov, degree = 0, bandwidth = .08, pwidth = .12

kernel = epanechnikov, degree = 0, bandwidth = .09, pwidth = .13

kernel = epanechnikov, degree = 0, bandwidth = .17, pwidth = .25
Return to main text
Estimation Details
Mid to late 1980s
Eight sites across the US
Site-specific, stratified randomization
985 families
Low birth weight, premature infants
Randomized intervention lasted first 3 years of child’s life
Data collected frequently from 0-8 years, also 18 years
We do not make use of randomized treatment
Preparation of IHDP Data

- We use data from 5 ages: 1, 3, 5, 8, 18 years
- After attrition and non-response, sample size 833 of 985
- Partially-missing response (<20%): nonparametric imputation
- For each type of factor, measures chosen so that units are consistent over time
- Measures preserved by monotonic transforms, we use ranks
- Inference: 10,000 bootstrap samples
IHDP Factor Measures

- Cognitive measures: IQ, PPVT, and math examinations
- Character Skills: Maternally-reported antisocial attitudes, social withdrawal, and depressive behavior indices
  - Example: Doesn’t get along with other children? Not true, sometimes, often
- Child Effort/Compliance: Maternally-reported indices of rule-breaking behavior, aggression, and destruction indices
  - Example: Runs away from home? Not true, sometimes, often
- Maternal Effort/Investment: Learning material provision, time spent helping with reading/homework, and activity frequency/quality indices
  - Example: Read books, magazines together? Daily, weekly, etc.
## Cognitive Skills

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$t = 1$</th>
<th>$t = 3$</th>
<th>$t = 5$</th>
<th>$t = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Productivity</td>
<td>0.826</td>
<td>0.855</td>
<td>0.980</td>
<td>0.911</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.022)</td>
<td>(0.014)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Cross-Productivity</td>
<td>0.323</td>
<td>0.077</td>
<td>-0.006</td>
<td>-0.036</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.041)</td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Child’s Effort Productivity</td>
<td>-0.045</td>
<td>0.000</td>
<td>0.000</td>
<td>0.096</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.027)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Mother’s Effort Productivity</td>
<td>0.477</td>
<td>0.365</td>
<td>0.103</td>
<td>-0.071</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>
## Results

### Non-cognitive Skills

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$t = 1$</th>
<th>$t = 3$</th>
<th>$t = 5$</th>
<th>$t = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Productivity</td>
<td>0.929</td>
<td>0.417</td>
<td>0.904</td>
<td>0.199</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.022)</td>
<td>(0.016)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Cross-Productivity</td>
<td>0.275</td>
<td>-0.014</td>
<td>-0.010</td>
<td>0.163</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.054)</td>
<td>(0.017)</td>
<td>(0.026)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Child’s Effort Productivity</td>
<td>0.125</td>
<td>0.066</td>
<td>0.160</td>
<td>0.487</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.030)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Mother’s Effort Productivity</td>
<td>0.118</td>
<td>0.022</td>
<td>-0.033</td>
<td>0.043</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.014)</td>
<td>(0.019)</td>
<td>(0.026)</td>
<td>(0.027)</td>
</tr>
</tbody>
</table>
Return to main text
A Simulation Exercise Based on Akabayashi (2006)
In this document I simulate the model that Akabayashi (2006) develops in his paper *An Equilibrium Model of Child Maltreatment*.

The layout of the model is the following.

The child’s human capital development in each period follows a linear low of motion:

\[ h_{\tau+1} = (1 - \delta) h_{\tau} + \varphi s_{\tau} H^\gamma + \phi a_{\tau} \]  

for \( \tau = 1, \ldots, T \).

\( \delta \) is a human capital depreciation parameter, \( \varphi, \phi, \gamma \) are technology parameters, and \( H \) is the given and fixed parent’s human capital.

\( s_{\tau} \in [0, 1] \) and \( a_{\tau} \) are endogenous variables.
The parents are not able to observe either the true level of human capital of the children or the effort they make. Instead, at each time $\tau$, they observe an outcome variable, $y_\tau$, which evolves according to the following linear rule:

$$y_\tau = h_\tau + a_\tau + \nu_\tau$$

(39)

where $\nu_\tau \sim N(0, \sigma^2_{\nu_\tau})$ for $\tau = 1, \ldots, T$.

The author lets $\sigma^2_{\nu_\tau} \equiv \frac{K}{s_\tau}$ because more time spent with the child reduces the uncertainty of the observation error.

(38) is interpreted as the state equation and (39) as the observation equation.

The author postulates a linear incentive schedule.
The service, $d_{\tau}$, is as follows:

$$d_{\tau} = (s_{\tau} + b_{\tau} \mathbb{E}[a_{\tau} | l_{\tau}]) H^{\gamma}$$

for $\tau = 1, \ldots, T$ where $b_{\tau}$ is defined as the slope of the incentive schedule and is an endogenous decision of the parent.

- $l_t \equiv \{y_t, \ldots, y_1\}$ is the information set at $t$.
- Suppose that a parent picks a relatively high $b_{\tau}$ and that the observation of his daughter’s performance, $y_{\tau}$, deviates from his human capital forecast, $\hat{h}_{\tau}$, by a lot such that $y_{\tau} - \hat{h}_{\tau}$ is very negative.
- Then, $d_{\tau}$ is relatively low and the child suffers from a low service, which the author interprets as abuse.
• The child’s utility function is
\[
\max\{a_\tau\}_{\tau=1}^T \mathbb{E} \left[ u \left( \sum_{\tau=t}^{T} \frac{1}{1+\rho_{ct}} \tau^{-t} (d_\tau - v(a_\tau)) + \frac{1}{1+\rho_{ct}} \tau^{-t+1} Bh_{T+1} \right) \big| I_{t-1} \right] \tag{41}
\]
subject to (38) and (40) and given \{s_\tau, b_\tau\}.

• The parent’s utility function is
\[
\max\{b_\tau, s_\tau\}_{\tau=1}^T \mathbb{E} \left[ U \left( \sum_{\tau=t}^{T} \frac{1}{1+\rho_{pt}} \tau^{-t} [c_\tau + \alpha u(\cdot, a_\tau)] \right) \big| I_{t-1} \right] \tag{42}
\]
subject to \(c_\tau = \pi(1 - s_\tau)H^{\gamma}\), (38), (39), (40) and to the child’s optimal decision rule, and where \(c_\tau\) is consumption at \(\tau\), and \(\pi\) efficiency unit wage.

• The Nash Equilibrium of the model can be solved easily.
• The child decides the optimal level of effort to solve her utility maximization problem by taking parent’s choices, $\{b_\tau, s_\tau\}_{\tau=1}^T$, as given.

• The parent solves his utility maximization problem to choose the optimal level of time and incentive slope by taking the optimal decision rule of the kid as given.
• First, when the parent receives a new observation, \( y \), he updates his contemporaneous belief about the child’s current level of human capital by taking an average of the previous belief and the new observation weighted by the degree of uncertainty:

\[
\hat{h}^u = \hat{h} + \frac{A'(\hat{h})\sigma_h^2}{A'(\hat{h})^2\sigma_h^2 + \sigma_v^2} (y - A(\hat{h}) - C(\sigma_h^2))
\]  \( (43) \)

• Then, the parent uses the updated belief, \( \hat{h}^u \), and the human capital formation rule to forecast the child’s human capital level in next period:

\[
\hat{h}' = F(\hat{h}^u) + G(\sigma_h^2)
\]  \( (44) \)
Finally, the parent updates the uncertainty regarding the child’s human capital:

\[(\sigma_h^2)' = \Phi \sigma_h^2\]  

(45)

where
\begin{align*}
A(\hat{h}) &= \hat{h} + \phi \psi D_t(\hat{h}) \tag{46} \\
A'(\hat{h}) &= 1 + \phi \psi D_t'(\hat{h}) \tag{47} \\
C(\sigma_h^2) &= a + \frac{\psi}{R \alpha \sigma_h^2 - \psi} (a + \psi \phi D_t(H) - KR \alpha Q_t) \tag{48} \\
F(\hat{h}^u) &= (1 - \delta) \hat{h}^u + \phi^2 \psi D_t(\hat{h}^u) \tag{49} \\
G(\sigma_h^2) &= \phi \left[ a + \frac{\psi}{R \alpha \sigma_h^2 - \psi} (a + \psi \phi D_t(H) - KR \alpha Q_t) \right] + \psi s^* H^\gamma \tag{50} \\
\Phi &= \frac{F'(\hat{h}^u)^2 \sigma_v^2}{A'(\hat{h}^u)^2 \sigma_h^2 + \sigma_v^2} \tag{51} \\
F'(\hat{h}^u) &= (1 - \delta) + \phi^2 \psi D_t'(\hat{h}^u) \tag{52}
\end{align*}
Simulation Parameterization
I use the following parameters and functional forms to simulate the model:

- **Time Horizon:** $T = 80$.
- **Technology parameters:** $\phi = .7, \varphi = .01, \delta = .001, \gamma = 0.5$
- **Preference parameters:**
  
- $\psi = 7, \alpha = 7, K = 1.5, R = 2, \alpha = .9, B = 50$
- **Human Capital Parameters:** $h_1 = 100, H = 40000$.
- **Filter Initial Values:** $\hat{h}_1 = 200, \sigma_{h_1}^2 = 5000$.
- **Other parameters:** $K = 1.5, \pi = 2$.
- **Discount Rates:** $\rho_i(h) = \exp(-.02 \times h)$ for $i = p, c$. 
Simulation Results
Figure 15: log Incentive Slope and log Time Spent with the Child
Figure 16: Child’s Effort
Figure 18: Phi

![Graph of Phi over time](image-url)
Return to main text
Studies on the Role of Income on Children’s Outcomes and Credit Constraints
## Table 4a: Studies on the Role of Income on Children’s Outcomes

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Outcome Studied: Test Scores (T), Schooling (S)</th>
<th>Timing of Income (Developmental Stage of the Child at Which Income Effects are Studied)</th>
<th>Separate the Effect of Income from Changes in Labor Supply or Family Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carneiro and Heckman (2002)</td>
<td>NLSY79*</td>
<td>S</td>
<td>Early (E) and Late (L) College Enrollment</td>
</tr>
<tr>
<td>Belley and Lochner (2007)</td>
<td>NLSY79*, NLSY97*</td>
<td>S</td>
<td>L High school completion and College Enrollment</td>
</tr>
<tr>
<td>Dahl and Lochner (2012)</td>
<td>NLSY79*, C-NLSY79*</td>
<td>T</td>
<td>E Preadolescence (ages 8 to 14)</td>
</tr>
<tr>
<td>Duncan et al. (1998)</td>
<td>PSID*</td>
<td>S^d</td>
<td>E and L Childhood and Preadolescence (ages 0 to 15)</td>
</tr>
</tbody>
</table>
**Table 4a (cont.): Studies on the Role of Income on Children’s Outcomes**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Outcome Studied: Test Scores (T), Schooling (S)</th>
<th>Timing of Income (Developmental Stage of the Child at Which Income Effects are Studied)</th>
<th>Separate the Effect of Income from Changes in Labor Supply or Family Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2011)</td>
<td>T</td>
<td>E Early Childhood (ages 2 to 5)</td>
<td>X</td>
</tr>
<tr>
<td>Loken (2010)</td>
<td>S</td>
<td>E Childhood (ages 1 to 11)</td>
<td>✓ c</td>
</tr>
<tr>
<td>Loken et al. (2012)</td>
<td>S</td>
<td>E Childhood (ages 1 to 11)</td>
<td>X</td>
</tr>
<tr>
<td>Milligan and Stabile (2011)</td>
<td>T</td>
<td>E Childhood (ages 0 to 10)</td>
<td>X</td>
</tr>
<tr>
<td>Carneiro et al. (2013)</td>
<td>S&lt;sup&gt;e&lt;/sup&gt;</td>
<td>E and L Childhood to Adolescence (ages 0 to 17)</td>
<td>X</td>
</tr>
</tbody>
</table>
**Table 4b: Studies on the Role of Income on Children’s Outcomes**

<table>
<thead>
<tr>
<th></th>
<th>Distinguishes the Effects of Contemporaneous vs. Permanent Income</th>
<th>Sources of Income Whose Effects are Studied</th>
<th>Instrument Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carneiro and Heckman (2002)</td>
<td>✓</td>
<td>Total family income</td>
<td>None</td>
</tr>
<tr>
<td>Belley and Lochner (2007)</td>
<td>X</td>
<td>Total family income</td>
<td>None</td>
</tr>
<tr>
<td>Dahl and Lochner (2012)</td>
<td>X&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Total family income</td>
<td>Policy variation in EITC eligibility</td>
</tr>
<tr>
<td>Duncan et al. (1998)</td>
<td>✓</td>
<td>Total family income</td>
<td>None</td>
</tr>
</tbody>
</table>
### Table 4b (continued): Studies on the Role of Income on Children’s Outcomes

<table>
<thead>
<tr>
<th>Distinguishes the Effects of Contemporaneous vs. Permanent Income</th>
<th>Sources of Income Whose Effects are Studied</th>
<th>Instrument Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2011)</td>
<td>Total family income</td>
<td>Random assignment to programs offering welfare transfers conditional on employment or education related activities, or full time work</td>
</tr>
<tr>
<td>Loken (2010)</td>
<td>Total family income</td>
<td>Oil discovery (inducing regional increase in wages)</td>
</tr>
<tr>
<td>Loken et al. (2012)</td>
<td>Total family income</td>
<td>Oil discovery (inducing regional increase in wages)</td>
</tr>
<tr>
<td>Milligan and Stabile (2011)</td>
<td>Child related tax benefits and income transfers</td>
<td>Variation in benefits eligibility</td>
</tr>
<tr>
<td>Carneiro et al. (2013)</td>
<td>Total family income</td>
<td>None</td>
</tr>
</tbody>
</table>
## Table 4c: Studies on the Role of Income on Children’s Outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Effect of Income on Human Capital Investments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carneiro and Heckman (2002)</td>
<td>Percentage of people constrained = weighted gap in educational outcome to highest income group: 5.1% are constrained in college enrollment (1.2% among low income, low ability, 0.2% low income high ability), 9% in completion of 2-year college (5.3% among low income, low ability, 0.3% low income high ability). No effect of timing of receipt of family income on child outcomes.</td>
</tr>
<tr>
<td>Belley and Lochner (2007)</td>
<td>High school completion: +8.4% for highest income quartile compared to lowest in 79, +6.7 in 97 cannot reject equal effect of income; college enrollment: +9.3% for highest income quartile compared to lowest in 79, +16 in 97, cannot reject equal effect of income.</td>
</tr>
<tr>
<td>Dahl and Lochner (2012)</td>
<td>$1,000 extra per year for 2 years: +6% of a standard deviation in math and reading combined PIAT score.</td>
</tr>
<tr>
<td>Duncan et al. (1998)</td>
<td>$10,000 increase in average (age 0-15) family income: +1.3 years of schooling in low income ($\leq 20,000) families, +0.13 in high income ones. Relevance of income is stronger in the early years (age 0-5): $10,000 increase in average (age 0-5) family income leads to extra 0.8 years of schooling in low income families, 0.1 in high income ones. Income at age 6-10 and 11-15: no significant effect. Similar results in a sibling differences model.</td>
</tr>
</tbody>
</table>

---

The relevance of income in early childhood (age 0-5) is stronger than in later years (age 6-15).
## Table 4c (continued): Studies on the Role of Income on Children’s Outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Effect of Income on Human Capital Investments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan et al. (2011)</td>
<td>$1,000 extra per year for 2 to 5 years: +6% of a standard deviation in child’s achievement score.</td>
</tr>
<tr>
<td>Loken (2010)</td>
<td>OLS: positive relationship of average (age 1-13) family income on children’s education, IV: no causal effect. Results are robust to different specification and splitting the sample by parental education.</td>
</tr>
<tr>
<td>Loken et al. (2012)</td>
<td>Non-linear IV (quadratic model): increase of $17,414, +0.74 years of education for children in poor families, +0.05 for children in rich families.</td>
</tr>
<tr>
<td>Milligan and Stabile (2011)</td>
<td>Low education mothers: positive effects of child benefits on cognitive outcomes for boys, on emotional outcomes for girls, weak on health. Results are non robust to the exclusion of Quebec.</td>
</tr>
<tr>
<td>Carneiro et al. (2013)</td>
<td>All outcomes: monotone and concave relationship with permanent income. £100,000 increase in permanent father’s earnings: +0.5 years of schooling. Timing of income: a balanced profile between early (age 0-5) and late childhood (age 6-11) is associated with the best outcomes; shifting income to adolescence is associated with better outcomes in dropping out of school, college attendance, earnings, IQ and teen pregnancy. Early and late childhood income are complements in determining schooling attainment, early and adolescent income are substitutes.</td>
</tr>
</tbody>
</table>
## Table 5a: Studies on Tests of Credit Constraints

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Outcome Studied: Schooling (S)</th>
<th>Timing of Income (Developmental Stage of the Child at Which Constraints are Studied)</th>
<th>Explicit Dynamic Model</th>
<th>Who is Affected by Constraints: Parent of the Agent (P), Agent / Child (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keane and Wolpin (2001)</td>
<td>NLSY79*</td>
<td>S</td>
<td>L College Enrollment</td>
<td>✓</td>
</tr>
<tr>
<td>Carneiro and Heckman (2002)</td>
<td>NLSY79*, C-NLSY79*</td>
<td>S</td>
<td>E and L College Enrollment, Completion, Delayed Entry</td>
<td>X</td>
</tr>
</tbody>
</table>
### Table 5a (continued): Studies on Tests of Credit Constraints

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Outcome Studied: Schooling (S)</th>
<th>Timing of Income (Developmental Stage of the Child at Which Constraints are Studied)</th>
<th>Explicit Dynamic Model</th>
<th>Who is Affected by Constraints: Parent of the Agent (P), Agent / Child (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameron and Taber (2004)</td>
<td>NLSY79*</td>
<td>S</td>
<td>X</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Adolescence and College Enrollment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>NLSY79*, C-NLSY79*</td>
<td>T^a</td>
<td>✓</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E and L Childhood and Adolescence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method to Test for Credit Constraints</td>
<td>Find Presence of Credit Constraints</td>
<td>Effect of Income or Constraints on Human Capital Investments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Keane and Wolpin (2001)               | YES                                 | Increase borrowing limit to $3,000 ($3 \times \text{max estimated})
|                                      | But irrelevant for schooling decisions | no change in mean highest grade completed; +0.2% in college enrollment; -0.2$ on mean hourly wage rate; increase in consumption and reduction in market hours; moderate reduction in parental transfers especially for the least educated parents. |
| Carneiro and Heckman (2002)           | YES                                 | (1) Conditioning on ability and family background factors, the role of income in determining schooling decisions is minimal. The strongest evidence is in the low ability group. The test is not robust to accounting for parental preferences and paternalism. Observed differences in attendance might be due to a consumption value of child’s schooling for parents. (2) There is no evidence of an independent effect on college enrollment of early or late income once permanent income is accounted for. (3) The claim that higher IV than OLS estimates of the Mincer coefficient implies credit constraints are incorrect: instruments used are invalid, the quality margin is ignored and self selection and comparative advantage can produce the result also in absence of financial constraints. |
|                                      | But affecting at most 8% of students |                                                          |
Table 5b (continued): Studies on Tests of Credit Constraints

<table>
<thead>
<tr>
<th>Method to Test for Credit Constraints</th>
<th>Find Presence of Credit Constraints</th>
<th>Effect of Income or Constraints on Human Capital Investments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameron and Taber (2004)</td>
<td>NO</td>
<td>Theoretical prediction: if borrowing constraints, IV estimates using direct costs of schooling higher than using opportunity costs. Data: IV estimates using the presence of a local college are smaller than the ones using foregone earnings. Regressions which interact college costs and characteristics potentially related to credit availability: no evidence of excess sensitivity to costs for potentially constrained sample. Structural model: almost 0% of the population is found to borrow at a rate higher than the market one.</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>YES</td>
<td>50% of young parents are constrained: high school dropouts (50%), high school graduates (38%), college dropouts (60%), college graduates (68%); and 12% of old parents are constrained. Families with college graduate parents benefit the most from a reduction in credit constraints.</td>
</tr>
</tbody>
</table>
Return to main text
Measuring Early Investments in Children

Jennifer Culhane, Flávio Cunha, Irma Elo, and Zoe Pham

Children’s Hospital of Philadelphia, Rice University, University of Pennsylvania, Rice University

June 5, 2016
A large body of research has demonstrated that investments in early childhood.

In this paper, we consider two forms of measurement of investments:

- Home Observation for the Measurement of the Environment (HOME).
- Language Environment Analysis (LENA).

HOME (or one of its many adaptations/versions) is widely used in many studies; LENA increasingly so.
Figure: PHD Study Data Collection Timeline

Wave 1 Begins:
Recruitment of primiparous women begins. First survey measure "prior" beliefs.

Wave 2 Begins:
First follow-up survey begins. We measure parental investments in children. Retention rate: 96%

Wave 3 Begins:
Second follow-up survey begins. We assess development and measure "posterior" beliefs

Wave 1 Ends:
Expected recruitment is 900 families.

Wave 2 Ends:
Expected number of families who complete first follow-up is around 760.

Wave 3 Ends:
Expected number of families who complete both follow-up surveys
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

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.

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

What about the green curve? It has the same slope as the black one but it is shifted to the right — hence the item with the green curve has the same discrimination parameter as the item with the black curve. The discrimination parameters $a_i$ are sometimes called slope parameters, just like the item difficulties are a.k.a. location parameters. The slope of the 2PL item response function at $b_i$ is equal to $a_i/b_i$. 

Jennifer Culhane, Flávio Cunha, Irma Elo, and Zoe Pham (Children's Hospital of Philadelphia, Rice University, University of Pennsylvania, Rice University)
Properties of an Informative IRT Scale
Properties of an Informative IRT Scale: IIF
Properties of an Informative IRT Scale: TCC

Informative Scale

Theta

Expected Score

-4 -3 -2 -1 0 1 2 3 4

0 2.26 8.05 15.4 22.6 29.8 37.3 42.9 45

Jennifer Culhane, Flávio Cunha, Irma Elo, and Zoe Pham (Children's Hospital of Philadelphia, Rice University, University of Pennsylvania, Rice University)
Properties of an Informative IRT Scale: IIF
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, there 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

0
-5
-10
1
5

IRT Properties of Full Scale HOME

Item Information Function - All Items

PHD Study
Test Characteristic Curve of the HOME Scale - All Items

PHD Study

Expected Score

Theta

0

-4

-3

-2

-1

0

1

2

3

4

13.7

31.3

37.9

40.3

44.4

IRT Properties of Full Scale HOME
IRT Properties of Full Scale HOME

Information Function of the HOME Scale - All Items

PHD Study

Test information

Standard error

Information

Theta

0

20

40

60

80

-4

-2

0

2

4

Standard Error

0

.2

.4

.6

.8

1
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:

$$h_1 = 1.0 + 0.5\theta - 0.25\theta^2 + \zeta \quad (1)$$

- 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
- $b_{1\text{ inf}}$: .48, .49, .5, .51, .52
- Kernel: epanechnikov, Bandwidth: 0.0014

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

- Density
- $b_{2\text{ inf}}$: -.27, -.26, -.25, -.24, -.23
- Kernel: epanechnikov, Bandwidth: 0.0016

**Coeff. on theta, Actual HOME**

- Density
- $b_{1\text{ uninf}}$: .43, .44, .45, .46, .47, .48
- Kernel: epanechnikov, Bandwidth: 0.0025

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

- Density
- $b_{2\text{ uninf}}$: -.36, -.34, -.32, -.3, -.28
- Kernel: epanechnikov, Bandwidth: 0.0031
Measuring Quality and Quantity of Time: LENA Pro

Jennifer Culhane, Flávio Cunha, Irma Elo, and Zoe Pham (Children's Hospital of Philadelphia, Rice University, University of Pennsylvania, Rice University)

Measuring Early Investments in Children

June 5, 2016
Figure 2
Pilot LENA Pro System Study in the HIPPY Program - HISD

Audio Environment per Hour

Adult Word Counts per Hour

Meaningful  Distant  TV  Noise  Silence

Conversation Turn Counts per Hour

Child Vocalization Counts per Hour

Jennifer Culhane, Flávio Cunha, Irma Elo, and Zoe Pham (Children's Hospital of Philadelphia, Rice University, University of Pennsylvania, Rice University)
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

Philadelphia Human Development Study
Conversation Turn Counts
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:

$$\ln \lambda_{i,j} = X_{i,j} \delta_j + \ln s_{i,j}$$  \hspace{1cm} (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).
Conditional on $\epsilon_i$, the probability of observing a count equal to:

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

where $\Pr (y_{i,j} \mid \epsilon_i) = \Pr (Y_{i,j} = y_{i,j} \mid \epsilon_i)$ 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:

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

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 $\frac{1}{\alpha}$.
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 $\sigma^2_{\mu}$.

We are interested in estimating the unobserved heterogeneity captured by $\mu_i$ across families.
Previous literature shows that early investments matter.

Difficult to simultaneously and precisely measure quantity and quality of parent-child interaction.

New forms of measurement will help us understand:

- What types of investment are more important for each dimension of human capital.
- How each type of investment is influenced by existing stock of human capital of the child.
- Whether each type is sensitive to opportunity costs, parental beliefs, and household characteristics.

Design new interventions that can directly or indirectly influence each type of investment.
The Economics and the Econometrics of Human Development: Testing and Operationalizing the Theory Part IIIB

James J. Heckman
University of Chicago

CeMMAP Masterclass
University College of London
June 23, 2016
Testing and Operationalizing the Theory
Decompose the $\theta_t$ vector into three subvectors:

$$\theta_t = (\theta_{C,t}, \theta_{N,t}, \theta_{H,t})$$  \hfill (1)$$

where

- $\theta_{C,t}$ is a vector of cognitive abilities (e.g., IQ) at age $t$,
- $\theta_{N,t}$ is a vector of noncognitive abilities (e.g., patience, self-control, temperament, risk aversion, discipline, and neuroticism) at age $t$, and
- $\theta_{H,t}$ is a vector of health stocks for mental and physical health at age $t$. 

• Functionings (task $j$) at age $t$:

$$Y_{j,t} = \psi_{j,t}(\theta_t, e_{j,t}, X_{j,t}), \quad j \in \{1, \ldots, J_t\} \quad \text{and} \quad t \in \{1, \ldots, 2T\}$$ (2)

• $Y_{j,t}$: outcome from activity $j$ at time $t$
• $\theta_t$ is the vector of skills at age $t$
• $X_{j,t}$ is a vector of purchased inputs that affect the functionings
• $e_{j,t}$ is effort in task
• $T$ is the length of childhood
• $T$ is the length of adulthood
• $2T$ is total lifetime
Effort: \( e_{j,t} \)

\[
e_{j,t} = \delta_j(\theta_t, A_t, X_{j,t}, R_{j,t}^a(I_{t-1}) | u).
\]  \hspace{1cm} (3)

- \( A_t \): environment
- \( R_{j,t}^a \): incentives
Estimating and Interpreting the Distribution of Skills, the Maps Between Skills and Functionings and the Technology of Skill Formation
• Low dimensional skills ("factors") generate a high dimensional set of functionings.
• Dimension and factor structures selected through a variety of methods.
• Exploratory factor analysis.
• Novel Bayesian procedures—avoid arbitrary methods in Exploratory Factor Analysis.
Cunha et al. (2010) present conditions under which the outcome equation and technology equation are non-parametrically identified.

They develop methods for accounting for the measurement error of inputs, anchoring estimated skills on adult outcomes (so that scales are defined in meaningful units), and accounting for the endogeneity of investments.
• Heckman et al. (2013) develop and apply simple and easily implemented least-squares estimators of linear factor models to estimate equations for outcomes.
Linear Factor Example
Link to “Formulating, Identifying”
(Cunha & Heckman 2008, JHR)
Nonparametric Factor Models Are Natural Frameworks for Estimating Skills and Determining Frontier Skill Sets
Estimating the Technology of Production of Cognitive and Noncognitive Skills
(builds on Cunha, Heckman, & Schennach, 2010)
Link to “Bayesian Exploratory Factor Analysis”
by Frühwirth-Schnatter et al. (JOE, 2014)
Skills as Determinants of Outcomes
Multiple Skills Shape Human Achievement Across a Variety of Dimensions
Estimating and Interpreting the Distribution of Skills, Maps Between Skills and Outcomes and the Technology of Skill Formation
Augmented Factor Models Are Natural Frameworks for Estimating Skills
Figure 1 from (Eisenhauer et al., 2014) plots the probability and the return of enrolling in college immediately after having graduated high school as a function of the deciles of scalar summaries of cognitive and noncognitive skills.
Figure 1: The Probability and Returns of College Enrollment by Endowments Levels

Source: Eisenhauer et al. (2014)

Note: College enrollment refers to the individuals who enroll in college immediately after having finished high school. Returns are expressed in units of millions of dollars. Higher deciles correspond to higher levels. See Eisenhauer et al. (2014) for greater details.
The return is calculated over a 65-year-long working life. Lifecycle earning profiles are simulated using the estimated parameters.

See Eisenhauer et al. (2014) for a precise description of the model, data, and computations.
See Appendix on Evidence on The Predictive Power of Cognitive and Socioemotional Traits

Link to Appendix
Similar Patterns for a Variety of Diverse Outcomes

A low dimensional vector of capabilities predicts a wide variety of outcomes

- Crime
- Wages
- Health
- Healthy behaviors (smoking, drug use)
- Trust
- Voting behavior
- Employment
- Participation in welfare
Technology of Skill Formation

\[ \theta_{k,t+1} = f_{s,k} (\theta_t, I_{k,t}, \theta_{P,t}) \tag{4} \]

for \( k \in \{C, N, H\} \), \( t \in \{1, 2, \ldots, T\} \).
• To estimate the technology of skill formation: Have to solve three problems.

1. Don’t observe \((\theta_t, I_{k,t}, \theta_P,t)\) directly, but have many measurements on it. **Measurement error in nonlinear systems.**

2. Don’t know which scale to use to measure components of \(\theta_t\). **Anchor test scores on adult outcomes.**

3. Investment \(I_{k,t}\) may be chosen by parents based on information that may be unobserved by the econometrician \((\eta_{k,t})\). **Endogeneity of investment.**
Estimates of Technologies of Skill Formation and Some Implications
Technology equations with shocks:

\[ \theta_{k,t+1} = \left[ \gamma_{t,k,1} \theta_{C,t}^{\phi_{t,k}} + \gamma_{t,k,2} \theta_{N,t}^{\phi_{t,k}} + \gamma_{t,k,3} l_{t}^{\phi_{t,k}} \right. \]

\[ + \gamma_{t,k,4} \theta_{C,P}^{\phi_{t,k}} + \gamma_{t,k,5} \theta_{N,P}^{\phi_{t,k}} \left. \right] \frac{\rho_{t,k}}{\phi_{t,k}} e^{\eta_{k,t}}, \tag{5} \]

where \( \gamma_{t,k,l} \geq 0 \) and \( \sum_{l=1}^{5} \gamma_{t,k,l} = 1, \ 0 \leq \rho_{t,k} \leq 1, \ \phi_{t,k} < 1, \ k \in \{ C, N \}, t \in \{ 1, 2 \}. \)

• Skills evolve over the life cycle
• Parental investments explain 34% of variance of educational attainment
• Self-productivity becomes stronger as children become older, for both cognitive and noncognitive skill formation (i.e., \( \frac{\partial \theta_{t+1}}{\partial \theta_t} \uparrow t \)).
• Strong cross effects (noncognitive skills foster cognitive investment)
• Complementarity between cognitive skills and investment becomes stronger as children become older. The elasticity of substitution for cognitive production is smaller in second stage production
• **Emerging dynamic complementarity.**
• It is more difficult to compensate for the effects of adverse environments on *cognitive* endowments at later ages than it is at earlier ages. This pattern of the estimates helps to explain the evidence on ineffective cognitive remediation strategies for disadvantaged adolescents reported in Cunha et al. (2006), Cunha (2007), and later papers.
• Complementarity between *noncognitive* skills and investments stays roughly constant over the life cycle.
• Suggests that later life investments should be more focused on promoting noncognitive—personality—skills.
• The evidence on which adolescent interventions are successful is consistent with this evidence.
The Implications of the Estimates for Design of Policy

- Targeted strategies
- Consider a policy for a social planner to optimize the stock of education in society.
- Assume (for simplicity) full control of investment (ignores parental responses)
- The bulk of the evidence in the child development literature shows **reinforcement** of investment by parents.
- No consideration of social fairness, equality of opportunity or equality of final outcomes—just efficiency.
- Yet with these estimates the optimal policy invests the most in the disadvantaged.
- As an empirical matter, social justice is enhanced by what is productively efficient.
Figure 2: Socially Optimal Early and Late **Levels** of Investment by Initial Skills

**Source:** Cunha et al. (2010). Optimal investments to maximize aggregate education in society.
Densities of Ratio of Early to Late Investments
Maximizing Aggregate Education Versus Minimizing Aggregate Crime

Ratio Early to Late Investment

Education
Crime

Figure 5

Econ and Ecom of Hum Dev
Review of the Literature

**Table 1: Skill Production Functions**

<table>
<thead>
<tr>
<th>Study</th>
<th>Cognitive</th>
<th>Non-cognitive</th>
<th>Health</th>
<th>Functional Form</th>
<th>Anchoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Todd and Wolpin (2003)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>Linear</td>
<td>X</td>
</tr>
<tr>
<td>Bernal and Keane (2010)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>Linear</td>
<td>X</td>
</tr>
<tr>
<td>Cunha and Heckman (2008)</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>Linear</td>
<td>✓</td>
</tr>
<tr>
<td>Cunha et al. (2010)</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>CES</td>
<td>✓</td>
</tr>
<tr>
<td>Todd and Wolpin (2007)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>Linear</td>
<td>X</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>CES</td>
<td>✓</td>
</tr>
<tr>
<td>Del Boca et al. (2014)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>Log-Linear</td>
<td>X</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>CES</td>
<td>✓</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>Linear</td>
<td>X</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/S</td>
<td>X</td>
</tr>
<tr>
<td>Bernal and Keane (2011)</td>
<td>✓</td>
<td></td>
<td>X</td>
<td>Linear</td>
<td>X</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>CES</td>
<td>X</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>CES</td>
<td>X</td>
</tr>
</tbody>
</table>
## Table 1 (continued): Skill Production Functions

<table>
<thead>
<tr>
<th>Study</th>
<th>Cognitive</th>
<th>Self Productivity</th>
<th>Health</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cognitive</td>
<td>Non-cognitive</td>
<td>Health</td>
</tr>
<tr>
<td>Todd and Wolpin (2003)</td>
<td>✓ - N/A</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bernal and Keane (2010)</td>
<td>✓ - N/A</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cunha and Heckman (2008)</td>
<td>0.977</td>
<td>0.884</td>
<td>X</td>
</tr>
<tr>
<td>Cunha et al. (2010)</td>
<td>0.487/0.902</td>
<td>0.649/0.868</td>
<td>X</td>
</tr>
<tr>
<td>Todd and Wolpin (2007)</td>
<td>0.21 - 0.34</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>0.735/0.799 /0.872</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Del Boca et al. (2014)</td>
<td>(0.14, 0.503)/(0.172, 0.922)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>✓ - N/A</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>✓ - N/A</td>
<td>N/A</td>
<td>X</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>N/S</td>
<td>N/S</td>
<td>X</td>
</tr>
<tr>
<td>Bernal and Keane (2011)</td>
<td>✓ - N/A</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>0.591/0.566</td>
<td>0.435/0.413</td>
<td>X</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>0.283/0.697/0.605</td>
<td>X</td>
<td>0.671/0.763/0.861/0.785 age5/age8/age12/age15</td>
</tr>
<tr>
<td></td>
<td>age8/age12/age15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 (continued): Skill Production Functions

<table>
<thead>
<tr>
<th></th>
<th>Cognitive</th>
<th>Non-cognitive</th>
<th>Health</th>
<th>Increasing Investments and Skill Complementarity over Time$^h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Todd and Wolpin (2003)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>U</td>
</tr>
<tr>
<td>Bernal and Keane (2010)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>U</td>
</tr>
<tr>
<td>Cunha and Heckman (2008)</td>
<td>0.003</td>
<td>0.028</td>
<td>X</td>
<td>U</td>
</tr>
<tr>
<td>Cunha et al. (2010)</td>
<td>0.000/0.008$^c$</td>
<td>0.083/0.011$^c$</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Todd and Wolpin (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>U</td>
</tr>
<tr>
<td>Cunha (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Del Boca et al. (2014)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Caucutt and Lochner (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Bernal (2008)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>U</td>
</tr>
<tr>
<td>Gayle et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/S</td>
</tr>
<tr>
<td>Bernal and Keane (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>U</td>
</tr>
<tr>
<td>Attanasio et al. (2015)</td>
<td>0.03/0.038 (not signif.)</td>
<td>0.11/0.122</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Meghir et al. (2015)</td>
<td>0.158/0.223/-0.054/0.018</td>
<td>X</td>
<td>-0.045/-0.054/0.061</td>
<td>✓</td>
</tr>
<tr>
<td>age5/age8/age12/age15 (ages 12 and 15 not signif.)</td>
<td></td>
<td></td>
<td>age 8/age12/age15 (ages 8 and 12 not signif.)</td>
<td></td>
</tr>
</tbody>
</table>
Bayesian Exploratory Factor Analysis
by Frühwirth-Schnatter et al. (JOE, 2014)
Recent revival: New challenges raised by the empirical research

- Profusion of high-dimensional data: How to aggregate them and extract meaningful information?
- How many latent factors should be retained?
- How to specify the structural relationships between observed measurements and latent variables?
This paper

Pick dimensions of space and factor loadings simultaneously
Motivation

**Dedicated Factor Model**

Factor model where each measurement loads on a single latent factor.

- Facilitates **interpretation**: Each latent factor is related to a well-defined subset of measurements.
  \[\Rightarrow\] cf. Thurstone’s (1947) simple structure.

- Achieves **identification**: Rules out most rotation problems that are usually solved by imposing restrictions on the factor loading matrix.
Contributions

Technical contributions

1. **Factor search procedure**: Sampling to produce identified models.
2. Sampling the latent factors and their correlation matrix in a dimension-varying model.
3. Block sampling of the correlation matrix jointly with the factors.
Model Specification

- Set of $Q$ measurements $M = (M_1, \ldots, M_Q)'$ related to covariates $X$ and to $K$ latent factors $\theta = (\theta_1, \ldots, \theta_K)'$:

$$M = \beta X + \alpha \theta + \varepsilon$$

with $K << Q$.

- Independence assumptions:

$$\theta \perp \perp X \perp \perp \varepsilon \qquad \varepsilon_i \perp \perp \varepsilon_j \quad \forall i \neq j$$

- Distributional assumptions:

$$\theta \sim \mathcal{N}(0, \Omega)$$
$$\varepsilon \sim \mathcal{N}(0, \Sigma)$$
$$\Sigma = \text{diag}(\sigma^2_1, \ldots, \sigma^2_Q)$$
• Covariance structure:

\[ \text{cov}(M \mid X) = \alpha \Omega \alpha' + \Sigma \]

• **Rotation problems!** Restrictions are required to achieve identification.
Each measurement is allowed to load on at most one latent factor.

The dedicated structure achieves identification.

At least three dedicated measurements are required for each factor (Anderson and Rubin, 1956) → Maximum number of factors is $Q/3$.

$\Omega$ is restricted to be a correlation matrix $R$ to set the scale of the factors: $\text{diag}(R) = (1, \ldots, 1)'$. 
**Indicator matrix** for dedicated structure: $\Delta$, where each row indicates to which latent factor $k$ measurement $M_q$ is dedicated to:

$$\Delta_{M_q} = (0 \ldots 0 \ 1 \ 0 \ldots 0)$$

$k^{th}$ element
An Example with $Q = 12$ and $K = 4$

This model is identified:

- At least three measurements loading on each factor.
- Number of factors = Number of nonzero columns (3 here).
- Up to column-switching and sign-switching.
- Measurement $M_8$ is discarded.
### Existing methods for dimension-varying models

- Reversible Jump MCMC (Green, 1995)
- Split and Merge moves (Richardson & Green, 1997)
- Birth and Death moves (Stephens, 2000)
- Evolutionary search algorithm (Carvalho et al., 2008)
- Nonparametric Factor Analysis (Griffiths & Ghahramani, 2005; Paisley and Carin, 2009)
- Sparse Bayesian Infinite Factor Models (Bhattacharya & Dunson, 2011)

How to design a sampling scheme that secures the identification of the dedicated factor model and works in practice?
• Indicators: \( \Pr(\Delta M_q = e_k) = \tau_k \), for \( k = 0, \ldots, K \)

\[
\tau_{0k} \sim \text{Beta}(\kappa_0, \xi_0) \\
\tau_k = (1 - \tau_{0k}) \tau^*_k \\
\tau^* = (\tau^*_1, \ldots, \tau^*_K) \sim \text{Dir}(\kappa_1, \ldots, \kappa_K)
\]

• Idiosyncratic variances: \( \sigma^2_q \sim IG(c_0q, C_{0q}) \)

• Factor loadings: \( \alpha_q \mid \sigma^2_q, \Delta M_q \neq e_0 \sim \mathcal{N}(0, A_0 \sigma^2_q) \)

• Regression coefficients: \( \beta_q \sim \mathcal{N}(0, B_0) \)

• Factor correlation matrix: Inverse-Wishart in augmented model (MDA).
Intuition of the new sampler

• Based on the **Metropolis-Hastings algorithm** (Hastings, 1970).

• Use **intermediate Gibbs sweeps** to generate a proposal.

• Relax identifying restrictions temporarily to allow bigger moves, but accept **identified models** only.

• Approach inspired by the tempered transitions (Neal, 1996).
Factor Selection: Sampling Identified Models

Intermediate steps in augmented model

Gibbs moves

Gibbs moves in reverse order

Expand model

M-H step

MCMC iterations

$t - 1$ $t$ $t + 1$

reject M-H

accept M-H
Advantages of the approach

- Symmetry of the intermediate Gibbs sweeps guarantees the **detailed balance condition** of the Markov chain is verified and greatly simplifies computations.

- **Easy to implement:** Only unrestricted Gibbs sweeps are required to construct the algorithm.

- **Good performance** in practice in recovering the ‘true’ dedicated structure.
Empirical Application

- High-dimensional data on cognitive abilities, personality traits and physical health measurements.

- Application of BEFA: Exploratory analysis to unravel a small set of meaningful factors.

- Comparison to results obtained from classical approaches.
Data: The British Cohort Study

• Survey of all babies born (alive or dead) in one particular week of April 1970 in England, Scotland, Wales and Northern Ireland.

• Information drawn from different sweeps:
  - birth: background characteristics
  - age 10: cognitive, mental and physical health measurements

• Sample size: 2,080 males.
Data: The BCS — 131 items

Cognitive tests (cont.)
- Picture Language Comprehension Test
- Friendly Math Test
- Edinburgh Reading Test

British Ability Scales (BAS):
- BAS - Matrices
- BAS - Recall Digits
- BAS - Similarities
- BAS - Word Definition

Personality trait scales
- Rutter’s Behavioral Disorder Scale (19 continuous items)
- Conners Hyperactivity Scale (19 continuous items)
- Child Developmental Scale (53 continuous items)
- Locus of Control Scale (16 binary items)
- Self-Esteem Scale (12 binary items)

Physical health measures (continuous)
height, head circumference, weight, diastolic and systolic blood pressure
### Results from Classical Methods

#### Table 2: Classical Methods to Select the Number of Factors

<table>
<thead>
<tr>
<th>Method</th>
<th># of factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattell’s Scree Plot</td>
<td>5–6</td>
</tr>
<tr>
<td>Onatski Test</td>
<td>6</td>
</tr>
<tr>
<td>Optimal Coordinates</td>
<td>11</td>
</tr>
<tr>
<td>Velicer’s Rule</td>
<td>11</td>
</tr>
<tr>
<td>Kaiser’s Rule</td>
<td>25</td>
</tr>
<tr>
<td>Horn’s Parallel Analysis</td>
<td>(a) 15 or (b) 36</td>
</tr>
<tr>
<td>Bayesian Information Criterion</td>
<td>17</td>
</tr>
<tr>
<td>Akaike Information Criterion</td>
<td>47</td>
</tr>
</tbody>
</table>

(a) applies the criterion $\lambda > \bar{\lambda}$; (b) applies the criterion $\lambda > 0$, where $\lambda$ denotes the eigenvalues from the reduced correlation matrix.
Posterior Correlation Matrix of the Factors

Cognitive Ability
Behavioral Problems [M]
Anxiety [M]
Attention Problems–Hyperactivity [M]
Attention Problems [T]
Anxiety [T]
School Phobia [T]
Conduct Problems [T]
Motor Coordination Problems [T]
Depression [T]
Concentration Problems [T]
Positive Sense of Self [C]
Body Build

Posteriors of Estimated Correlations

-1.0
-0.5
0.0
0.5
1.0
Summary

- New approach for dedicated factor models with unknown dimension and unknown structure *a priori*.
- Simple to implement and good results in practice, cf. Monte Carlo simulations in paper.
- BEFA uncovers the rich structure of the BCS data: 13 factors are identified, their correlation matrix is estimated.
Return to text from
“Bayesian Exploratory Factor Analysis”
“Notes on: Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation”
by Flávio Cunha and James J. Heckman (2008)

*Journal of Human Resources*
Estimating the Technology of Production of Cognitive and Noncognitive Skills

(1) Analyze the evolution of both cognitive and noncognitive outcomes using the equation system

\[
\begin{bmatrix}
\theta_{t+1}^N \\
\theta_{t+1}^C
\end{bmatrix} =
\begin{bmatrix}
\gamma_1^N & \gamma_2^N \\
\gamma_1^C & \gamma_2^C
\end{bmatrix}
\begin{bmatrix}
\theta_t^N \\
\theta_t^C
\end{bmatrix} +
\begin{bmatrix}
\gamma_3^N \\
\gamma_3^C
\end{bmatrix}
\theta_t^I +
\begin{bmatrix}
\eta_t^N \\
\eta_t^C
\end{bmatrix},
\]

(6)

where \(\theta_t^I\) can be a vector.

- In more compact notation, \(\theta_{t+1} = A\theta_t + B\theta_t^I + \eta_t\), where
  \(A = \begin{bmatrix}
\gamma_1^N & \gamma_2^N \\
\gamma_1^C & \gamma_2^C
\end{bmatrix}\) and \(B = (\gamma_3^N, \gamma_3^N)'\).

- Allowing for stage-specific coefficients, this equation becomes
  \(\theta_{t+1} = A_t\theta_t + B_t\theta_t^I + \eta_t\)

(2) Define \(\eta'_t = (\eta_t^N, \eta_t^C)\)
Determine how stocks of cognitive and noncognitive skills at date \( t \) affect the stocks at date \( t + 1 \), examining both self productivity (the effects of \( \theta_t^N \) on \( \theta_{t+1}^N \), and \( \theta_t^C \) on \( \theta_{t+1}^C \)) and cross productivity (the effects of \( \theta_t^C \) on \( \theta_{t+1}^N \) and the effects of \( \theta_t^N \) on \( \theta_{t+1}^C \)) at each stage of the life cycle.
(4) Develop a dynamic factor model where we proxy $\theta'_t = (\theta^N_t, \theta^C_t, \theta^I_t)$ by vectors of measurements on skills and investments which can include test scores as well as outcome measures. In our analysis, test scores and parental inputs are indicators of the latent skills and latent investments. We account for measurement errors in output and input variables. We find substantial measurement errors in the proxies for parental investment and in the proxies for cognitive and noncognitive skills.

(5) Instead of imposing a particular index of parental input based on components of the home score, we estimate an index that best fits the data.
Instead of relying solely on exclusion restrictions to generate instruments to correct for measurement error in the proxies for $\theta_t$, and for endogeneity, we use covariance restrictions that exploit a feature of our data that there are many more measurements on $\theta_{t+1}$ and $\theta_t$ than the number of latent factors. This allows us to secure identification from cross equation restrictions using multiple indicator-multiple cause (MIMIC) (Jöreskog and Goldberger, 1975) and linear structural relationship (LISREL) (Jöreskog, Sörbom, and Magidson, 1979) models. When instruments are needed, they arise from the internal logic of the model developed in Cunha et al. (2006) and Cunha and Heckman (2007), using methods developed by Madansky (1964) and Pudney (1982).
Instead of relying on test scores as measures of output and change in output due to parental investments, we anchor the scale of the test scores using adult outcome measures: log earnings and the probability of high school graduation. We thus estimate the effect of parental investments on the adult earnings of the child and on the probability of high school graduation.
A Model for the Measurements

- Assume access to measurement systems that can be represented by a dynamic factor structure:

\[ Y_{j,t}^k = \mu_{j,t}^k + \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k, \text{ for } j \in \{1, \ldots, m_t^k\}, k \in \{C, N, I\} \]

- \(m_t^k\) is the number of measurements on cognitive skills, noncognitive skills, and investments in period \(t\); and where \(\theta_t^k\) is a dynamic factor for component \(k\), \(k \in \{C, N, I\}\).

- \(\text{Var}(\varepsilon_{j,t}^k) = \sigma_{k,j,t}^k\)

- Initial conditions of the process, \((\theta_1^C, \theta_1^N)\), which correspond to endowment of abilities.
• Multiple measurements of abilities in the first period of our data, we can identify the distribution of the latent initial conditions.

• We also identify the distribution of each $\theta_t = (\theta^c_t, \theta^N_t, \theta^I_t)$, as well as the dependence across $\theta_t$ and $\theta_{t'}, t \neq t'$.

• The $\mu_{j,t}^k$ and the $\alpha_{j,t}^k$ can depend on regressors which we keep implicit.
• As above, let $\theta^C_t$ denote the stock of cognitive skill of the agent in period $t$.
• We do not observe $\theta^C_t$ directly.
• Instead, we observe a vector of measurements, such as test scores, $Y^C_{j,t}$, for $j \in \{1, 2, \ldots, m^C_t\}$.
• Assume that:

$$Y^C_{j,t} = \mu^C_{j,t} + \alpha^C_{j,t} \theta^C_t + \varepsilon^C_{j,t}, \text{ for } j \in \{1, 2, \ldots, m^C_t\}$$  \hspace{1cm} (7)

and set $\alpha^C_{1,t} = 1$ for all $t$.
• Some normalization is needed to set the scale of the factors.
• The $\mu^C_{j,t}$ may depend on regressors.
• Similar equation for noncognitive skills at age $t$, relating $\theta_t^N$ to proxies for it:

$$Y_{j,t}^N = \mu_{j,t}^N + \alpha_{j,t}^N \theta_t^N + \varepsilon_{j,t}^N, \text{ for } j \in \{1, \ldots, m_t^N\} \quad (8)$$

and we normalize $\alpha_{1,t}^N = 1$ for all $t$.

• Measurement equations for investments, $\theta_t^I$:

$$Y_{j,t}^I = \mu_{j,t}^I + \alpha_{j,t}^I \theta_t^I + \varepsilon_{j,t}^I, \text{ for } j \in \{1, \ldots, m_t^I\} \quad (9)$$

and the factor loading $\alpha_{1,t}^I = 1$.

• The $\varepsilon$’s are measurement errors that account for the fallibility of our measures of latent skills and investments.
We analyze a linear law of motion for skills:

\[
\theta_{t+1}^k = \gamma_0^k + \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \eta_t^k
\]

for \( k \in \{C, N\} \) and \( t \in \{1, \ldots, T\} \), where the error term \( \eta_t^k \) is independent across agents and over time for the same agents, but \( \eta_t^C \) and \( \eta_t^N \) are freely correlated.

We assume that the \( \eta_t^k, k \in \{C, N\} \), are independent of \((\theta_1^C, \theta_1^N)\).

Can relax the independence assumption and allow for unobserved inputs (Cunha and Heckman, 2008).

The \( \gamma_1^k, l = 0, \ldots, 3 \) may depend on regressors which we keep implicit.
• We allow the components of $\theta_t$ to be freely correlated for any $t$ and with any vector $\theta_{t'}$, $t' \neq t$, and we can identify this dependence.

• Assume that any variables in the $\mu_{j,t}^k$ are independent of $\theta_t$, $\varepsilon_{j,t}^k$, and $\eta_t^k$ for $k \in \{C, N, I\}$ and $t \in \{1, \ldots, T\}$.

• We now establish conditions under which the technology parameters are identified.
Semiparametric Identification

- The goal of the analysis is to recover the joint distribution of \( \{ \theta^C_t, \theta^N_t, \theta^I_t \}_{t=1}^T \), the distributions of \( \{ \eta^k_t \}_{t=1}^T \) and \( \{ \varepsilon^k_{j,t} \}_{j=1}^{m^k_t} \) nonparametrically, as well as the parameters \( \{ \alpha^k_{j,t} \}_{j=1}^{m^k_t} \), \( \{ \gamma^k_{j,t} \}_{j=1}^{3} \) for \( k \in \{ C, N \} \), and for \( t \in \{ 1, \ldots, T \} \).
- Identification of the means of the measurements is straightforward under our assumptions.
1. Classical Measurement Error for the Case of Two Measurements Per Latent Factor:

\[ m_t^C = m_t^N = m_t^I = 2 \]

Assumption 1

\( \varepsilon_{j,t}^k \) is mean zero and independent across agents and over time for \( t \in \{1, \ldots, T\}; j \in \{1, 2\}; \) and \( k \in \{C, N, I\}; \)
Assumption 2

$\varepsilon_{j,t}^k$ is mean zero and independent of $(\theta_{C_t}^C, \theta_{N_t}^N, \theta_{I_t}^I)$ for all $t, \tau \in \{1, \ldots, T\}; j \in \{1, 2\};$ and $k \in \{C, N, I\};$

Assumption 3

$\varepsilon_{j,t}^k$ is mean zero and independent from $\varepsilon_{j,t}^l$ for $i, j \in \{1, 2\}$ and $i \neq j$ for $k = l$; otherwise $\varepsilon_{j,t}^k$ is mean zero and independent from $\varepsilon_{j,t}^l$ for $i, j \in \{1, 2\}; k \neq l, k, l \in \{C, N, I\}$, and $t \in \{1, \ldots, T\}$. 
Identification of the Factor Loadings

- Since we observe \( \left\{ \left[ Y_{j,t}^k\right]_{j=1}^2 \right\}_{t=1}^T \) for every person, we can compute \( \text{Cov}(Y_{1,t}^k, Y_{2,\tau}^l) \) from the data for all \( t, \tau \) and \( k, l \) pairs, where \( t, \tau \in \{1, \ldots, T\} \); \( k, l \in \{C, N, I\} \).
- Consider, for example, measurements on cognitive skills.
- Recall that \( \alpha_{1,t}^C = 1 \).
- We know the left hand side of each of the following equations:

\[
\begin{align*}
\text{Cov}(Y_{1,t}^C, Y_{1,t+1}^C) &= \text{Cov}(\theta_t^C, \theta_{t+1}^C), \quad (11) \\
\text{Cov}(Y_{2,t}^C, Y_{1,t}^C) &= \alpha_{2,t}^C \text{Cov}(\theta_t^C, \theta_{t+1}^C), \quad \text{and} \quad (12) \\
\text{Cov}(Y_{1,t}^C, Y_{2,t+1}^C) &= \alpha_{2,t+1}^C \text{Cov}(\theta_t^C, \theta_{t+1}^C). \quad (13)
\end{align*}
\]
• We can identify $\alpha_{2,t}^{C}$ by taking the ratio of (12) to (11) and $\alpha_{2,t+1}^{C}$ from the ratio of (13) to (11).

• Proceeding in the same fashion, we can identify $\alpha_{j,t}^{k}$ for $t \in \{1, \ldots, T\}$ and $j \in \{1, 2\}$, up to the normalizations $\alpha_{1,t}^{k} = 1$, $k \in \{C, N, I\}$.

• We assume that $\alpha_{2,t}^{k} \neq 0$ for $k \in \{C, N, I\}$ and $t \in \{1, \ldots, T\}$.

• If $\alpha_{2,t}^{k} = 0$, we would violate the condition that states that there are exactly $m_{t}^{k} = 2$ valid measurements for the factor $\theta_{t}^{k}$.
The Identification of the Joint Distribution of \( \{(\theta^C_t, \theta^N_t, \theta^I_t)\}_{t=1}^T \)

- Once the parameters \( \alpha^k_{1,t} \) and \( \alpha^k_{2,t} \) are identified (up to the normalization \( \alpha^k_{1,t} = 1 \)), we can rewrite (7), (8), and (9) as

\[
\frac{Y^k_{j,t}}{\alpha^k_{j,t}} = \frac{\mu^k_{j,t}}{\alpha^k_{j,t}} + \theta^k_t + \frac{\varepsilon^k_{j,t}}{\alpha^k_{j,t}}, \quad j \in \{1, 2\}
\]

for \( \alpha^k_{j,t} \neq 0, \ k \in \{C, N, I\}; \ t \in \{1, \ldots, T\} \).
Now, define

\[ Y_j = \left\{ \left( \frac{Y_{j,t}^C}{\alpha_{j,t}^C}, \frac{Y_{j,t}^N}{\alpha_{j,t}^N}, \frac{Y_{j,t}^I}{\alpha_{j,t}^I} \right) \right\}_{t=1}^T \]  
for \( j = 1, 2 \).

Similarly, define

\[ \varepsilon_j = \left\{ \left( \frac{\varepsilon_{j,t}^C}{\alpha_{j,t}^C}, \frac{\varepsilon_{j,t}^N}{\alpha_{j,t}^N}, \frac{\varepsilon_{j,t}^I}{\alpha_{j,t}^I} \right) \right\}_{t=1}^T \]  
for \( j = 1, 2 \),

and

\[ \mu_j = \left\{ \left( \frac{\mu_{j,t}^C}{\alpha_{j,t}^C}, \frac{\mu_{j,t}^N}{\alpha_{j,t}^N}, \frac{\mu_{j,t}^I}{\alpha_{j,t}^I} \right) \right\}_{t=1}^T \]  
for \( j = 1, 2 \).

Let \( \theta \) denote the vector of all factors in all time periods:

\[ \theta = \left\{ \left( \theta_{t}^C, \theta_{t}^N, \theta_{t}^I \right) \right\}_{t=1}^T. \]
• We rewrite the measurement equations as

\[ Y_1 = \mu_1 + \theta + \varepsilon_1, \]
\[ Y_2 = \mu_2 + \theta + \varepsilon_2. \]

• Under the assumption that measurement error is classical, we can apply Kotlarski’s Theorem (Kotlarski, 1967) and identify the joint distribution of \( \theta \) as well as the distributions of \( \varepsilon_1 \) and \( \varepsilon_2 \).

• Since \( \alpha_{j,t}^k \) is identified, it is possible to recover the distribution of \( \varepsilon_{j,t}^k \) for \( j \in \{1, 2, \ldots, m_t^k\} \); \( k \in \{C, N, I\} \) and \( t \in \{1, 2, \ldots, T\} \).
The Identification of the Technology Parameters Assuming Independence of $\eta$.

- Assume that $\eta^k_t$ is independent of $(\theta^c_t, \theta^N_t, \theta^I_t)$.
- Consider, for example, the law of motion for noncognitive skills,

$$\theta^N_{t+1} = \gamma^N_0 + \gamma^N_1 \theta^N_t + \gamma^N_2 \theta^C_t + \gamma^N_3 \theta^I_t + \eta^N_t$$

for $t \in \{1, \ldots, T\}$.
- Assume that $\eta^N_t$ is serially independent but possibly correlated with $\eta^C_t$. 

Econ and Ecom of Hum Dev
Define

\[ \tilde{Y}_{1,t+1}^N = Y_{1,t+1}^N - \mu_{1,t+1}^N \]
\[ \tilde{Y}_{1,t}^N = Y_{1,t}^N - \mu_{1,t}^N \]
\[ \tilde{Y}_{1,t}^C = Y_{1,t}^C - \mu_{1,t}^C \]
\[ \tilde{Y}_{1,t}^I = Y_{1,t}^I - \mu_{1,t}^I \]
• Substitute the measurement equations \( \tilde{Y}_{1,t+1}, \tilde{Y}_{1,t}, \tilde{Y}_{1,t}^C, \tilde{Y}_{1,t}^I \) for \( \theta_{t+1}^N, \theta_t^N, \theta_t^C, \theta_t^I \), respectively:

\[
\tilde{Y}_{1,t+1}^N = \gamma_0^N + \gamma_1^N \tilde{Y}_{1,t}^N + \gamma_2^N \tilde{Y}_{1,t}^C + \gamma_3^N \tilde{Y}_{1,t}^I \\
+ \left( \varepsilon_{1,t+1}^N - \gamma_1^N \varepsilon_{1,t}^N - \gamma_2^N \varepsilon_{1,t}^C - \gamma_3^N \varepsilon_{1,t}^I + \eta_{t+1}^N \right) (15)
\]

for \( t \in \{1, \ldots, T\} \).

• If we estimate (15) by least squares, we do not obtain consistent estimators of \( \gamma_k^N \) for \( k \in \{1, 2, 3\} \) because the regressors \( \tilde{Y}_{1,t}^N, \tilde{Y}_{1,t}^C, \tilde{Y}_{1,t}^I \) are correlated with the error term \( \omega_{t+1} \), where

\[
\omega_{t+1} = \varepsilon_{1,t+1}^N - \gamma_1^N \varepsilon_{1,t}^N - \gamma_2^N \varepsilon_{1,t}^C - \gamma_3^N \varepsilon_{1,t}^I + \eta_{t+1}^N.
\]
- Can instrument $\tilde{Y}_{1,t}^N$, $\tilde{Y}_{1,t}^C$, $\tilde{Y}_{1,t}^I$ using $\tilde{Y}_{2,t}^N$, $\tilde{Y}_{2,t}^C$, $\tilde{Y}_{2,t}^I$ as instruments by applying two-stage-least squares to recover the parameters $\gamma_k^N$ for $k = 1, 2, 3$.
- See Madansky (1964) or Pudney (1982) for the precise conditions on the factor loadings.
- Our instruments are “internal instruments” justified by the model.
- The suggested instruments are also independent of $\eta_t^N$ because of the assumed lack of serial correlation in $\eta_t^N$. 
• We can repeat the argument for different time periods.
• In this way, we can identify stage-specific technologies for each stage of the child’s life cycle.
• We can perform a parallel analysis for the cognitive skill equation.
2. Non-classical Measurement Error

We can replace Assumption 3 with the following assumption and still obtain full identification of the model.

**Assumption 4**

\[ \varepsilon_{1,t}^k \text{ is independent of } \varepsilon_{j,\tau}^l \text{ for } j \in \{2, \ldots, m_t^k\}; \ k, l \in \{C, N, I\} \text{ and } t, \tau \in \{1, 2, \ldots, T\}, \ m_t^k \geq 2. \ \varepsilon_{1,t}^k \text{ is independent of } \varepsilon_{1,\tau}^k, \text{ for } t \neq \tau. \]

Otherwise the \[ \varepsilon_{j,\tau}^l \text{, for } j \in \{2, \ldots, m_t^k\}; \ k, l \in \{C, N, I\} \text{ and } t, \tau \in \{1, 2, \ldots, T\} \] can be arbitrarily dependent.
• Proof of identification is as follows.
• Let \( Y_{j,t}^k = \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k \), for \( j \in \{1, \ldots, m_t^k\} \); \( t \in \{1, \ldots, T\} \) and \( k \in \{C, N, I\} \).
• Normalize \( \alpha_{1,t}^k = 1 \) for all \( k \in \{C, N, I\} \) and \( t \in \{1, \ldots, T\} \).
• Within a \( k \) system, for a fixed \( t \), we can compute \( \text{Cov} \left( Y_{j,t}^k, Y_{1,t}^k \right) = \alpha_{j,t}^k \text{Var} \left( \theta_t^k \right), j \in \{1, \ldots, m_t^k\} \).
• For temporally adjacent systems, we can compute

\[
\begin{align*}
\text{Cov} \left( Y_{1,t-1}^k, Y_{1,t}^k \right) &= \text{Cov} \left( \theta_{t-1}^k, \theta_t^k \right), \\
\text{Cov} \left( Y_{1,t-1}^k, Y_{j,t}^k \right) &= \alpha_{j,t}^k \text{Cov} \left( \theta_{t-1}^k, \theta_t^k \right), j \in \{2, \ldots, m_t^k\}.
\end{align*}
\]
• Hence we can identify $\alpha_{j,t}^k$, $j \in \{1, \ldots, m_t^k\}$; $t \in \{1, \ldots, T\}$; and $k \in \{C, N, I\}$ and thus $\text{Var}(\theta_t^k)$, $t \in \{1, \ldots, T\}$; $k \in \{C, N, I\}$.

• With these ingredients in hand, we can identify $\text{Var}(\varepsilon_{j,t}^k)$, $t \in \{1, \ldots, T\}$, as well as $\text{Cov}(\varepsilon_{j,t}^k, \varepsilon_{j',t}^l) = \text{Cov}(Y_{j,t}^k, Y_{j',t}^k) - \alpha_{j,t}^k \alpha_{j',t}^l \text{Var}(\theta_t^k)$,

since we know every ingredient on the right hand side of the preceding equation.

• By a similar argument, we can identify

$$\text{Cov}(\varepsilon_{j,t}^k, \varepsilon_{j',\tau}^l) = \text{Cov}(Y_{j,t}^k, Y_{j',\tau}^k) - \alpha_{j,t}^k \alpha_{j',\tau}^l \text{Cov}(\theta_t^k, \theta_{\tau}^l). \quad (17)$$
• We can rewrite the measurement equations as a system:

$$
Y_{j,t}^k \quad \frac{\alpha_{j,t}^k}{\alpha_{j,t}^k} = \mu_{j,t}^k + \theta_t^k + \frac{\varepsilon_{j,t}^k}{\alpha_{j,t}^k},
$$

\( j \in \{1, \ldots, m_t^k\}; \ t \in \{1, \ldots, T\}; \ k \in \{C, N, I\}. \)

• Applying Schennach (2004), we can identify the joint distribution of \((\theta_1^C, \ldots, \theta_T^C, \theta_1^N, \ldots, \theta_T^N, \theta_1^I, \ldots, \theta_T^I)\) as well as the joint distribution of \(\{\varepsilon_{j,t}^k\}, \ j \in \{1, \ldots, m_t^k\}; \ t \in \{1, \ldots, T\} \) and \(k \in \{C, N, I\}\) using multivariate deconvolution.
The Identification of the Technology with Correlated Omitted Inputs

- It is unrealistic to assume that omitted inputs are serially independent.
- Fortunately, we can relax this assumption.
- Assume now that $\eta^k_t$ is not independent of $\theta'_t = (\theta^C_t, \theta^N_t, \theta^I_t)$.
- Consider a model in which $\eta^k_t$ can be decomposed into two parts:

\[
\eta^N_t = \gamma^N_4 \lambda + \nu^N_t \quad \text{and} \quad \eta^C_t = \gamma^C_4 \lambda + \nu^C_t
\]

so that the equations of motion can be written as

\[
\begin{align*}
\theta^N_{t+1} &= \gamma^N_0 + \gamma^N_1 \theta^N_t + \gamma^N_2 \theta^C_t + \gamma^N_3 \theta^I_t + \gamma^N_4 \lambda + \nu^N_t, \\
\theta^C_{t+1} &= \gamma^C_0 + \gamma^C_1 \theta^N_t + \gamma^C_2 \theta^C_t + \gamma^C_3 \theta^I_t + \gamma^C_4 \lambda + \nu^C_t.
\end{align*}
\]

(18)  

(19)
In this section, we normalize $\gamma_4^N = 1$.

The term $\lambda$ is a time-invariant input permitted to be freely correlated with $\theta_t$.

We allow $\lambda$ to have a different impact on cognitive and non-cognitive skill accumulation.

Let $\nu_t = (\nu_t^N, \nu_t^C)$.

We make the following assumption.

**Assumption 5**

The error term $\nu_t$ is independent of $\theta_t$, $\lambda$, $\nu_\tau$, conditional on regressors for any $\tau \neq t$. 
• Under this assumption, we can identify both a stage-invariant technology and a stage-varying technology.
• We first analyze the stage-invariant case.
• Consider, for example, the law of motion for noncognitive skills.
• For any periods $t, t + 1$ we can compute the difference

$$
\begin{align*}
\theta_{t+1}^N - \theta_t^N &= \gamma_1^N (\theta_t^N - \theta_{t-1}^N) + \gamma_2^N (\theta_t^C - \theta_{t-1}^C) + \gamma_3^N (\theta_t^I - \theta_{t-1}^I) + \nu_t^N - \nu_{t-1}^N. \\
\end{align*}
$$

• We use the measurement equations for to proxy $\theta_{t+1}$ and $\theta_t$:

$$
\begin{align*}
\tilde{Y}_{1,t+1}^N - \tilde{Y}_{1,t}^N &= \gamma_1^N (\tilde{Y}_{1,t}^N - \tilde{Y}_{1,t-1}^N) + \gamma_2^N (\tilde{Y}_{1,t}^C - \tilde{Y}_{1,t-1}^C) + \gamma_3^N (\tilde{Y}_{1,t}^I - \tilde{Y}_{1,t-1}^I) + \nu_t^N - \nu_{t-1}^N \\
&+ \left\{ (\epsilon_{1,t+1}^N - \epsilon_{1,t}^N) - \gamma_1^N (\epsilon_{1,t}^N - \epsilon_{1,t-1}^N) - \gamma_2^N (\epsilon_{1,t}^C - \epsilon_{1,t-1}^C) - \gamma_3^N (\epsilon_{1,t}^I - \epsilon_{1,t-1}^I) \right\}.
\end{align*}
$$
• OLS applied to (21) does not produce consistent estimates of $\gamma_1^N$, $\gamma_2^N$, $\gamma_3^N$ and because the regressors $\left(\tilde{Y}_{1,t}^k - \tilde{Y}_{1,t-1}^k\right)$ are correlated with the error term $\omega$, where

$$\omega = \left(\varepsilon_{1,t+1}^N - \varepsilon_{1,t}^N\right) - \gamma_1^N \left(\varepsilon_{1,t}^N - \varepsilon_{1,t-1}^N\right) - \gamma_2^N \left(\varepsilon_{1,t}^C - \varepsilon_{1,t-1}^C\right) - \gamma_3^N \left(\varepsilon_{1,t}^I - \varepsilon_{1,t-1}^I\right).$$

• However, we can instrument $\left(\tilde{Y}_{1,t}^k - \tilde{Y}_{1,t-1}^k\right)$ using $\left\{\left(Y_{j,t-1}^k - Y_{j,t-2}^k\right)\right\}_{j=2}^{m_t^k}$ as the instruments.

• These instruments are valid because of the generalization of investment equation (9) in Cunha and Heckman (2007) to a $T$ period model.

• Using a two-stage least squares regression with these instruments allows us to recover the parameters $\gamma_1^N$, $\gamma_2^N$ and $\gamma_3^N$.

• We can identify $\gamma_0^N$ if we assume that $E(\lambda) = 0$. 
Following a parallel argument, we can identify $\gamma_0^N$, $\gamma_1^N$, $\gamma_2^N$ and $\gamma_3^N$ using the data on the evolution of cognitive test scores.

Next, define

$$\psi_{t+1}^k = \theta_{t+1}^k - (\gamma_0^k + \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I).$$

From the measurement equations, we know the joint distribution of $(\theta_{t+1}^k, \theta_t^N, \theta_t^C, \theta_t^I)$ for $k \in \{C, N\}$.

We have established how to obtain the parameter values $\gamma_0^N$, $\gamma_1^N$, $\gamma_2^N$ and $\gamma_3^N$.

Consequently, we know the distribution of $\psi_t^k$ for $k \in \{C, N\}$ and $t \in \{1, \ldots, T\}$. 
We have $2T$ equations:

\[
\begin{align*}
\psi_T^N &= \lambda + \nu_T^N, \\
\psi_T^{N-1} &= \lambda + \nu_T^{N-1}, \\
&\vdots \\
\psi_1^N &= \lambda + \nu_1^N \\
\psi_T^C &= \gamma_T^C \lambda + \nu_T^C, \\
\psi_T^{C-1} &= \gamma_T^{C-1} \lambda + \nu_T^{C-1}. \\
\end{align*}
\]

Under Assumption 5 we can apply Kotlarski’s Theorem to this system and obtain the distribution of $\lambda$ and $\nu_t$ for any $t$.

Note that we can identify the parameter $\gamma^C$ from the covariance:

\[
\text{Cov} \left( \psi_t^N, \psi_\tau^C \right) = \gamma^C \text{Var} (\lambda)
\]

for any $t, \tau \in \{1, \ldots, T\}$ since the variance of $\lambda$ is known.

This approach solves the problem raised by correlated omitted inputs for stage-invariant technologies.
• For the stage-varying case, a similar argument applies. Recall that the first period of life is \( t = 1 \).

• In place of equation (20), we can write

\[
\theta_{t+1}^N - \theta_t^N = \gamma_0, t - \gamma_0, t-1 + \gamma_1, t \theta_t^N - \gamma_1, t-1 \theta_{t-1}^N + \gamma_1, t \theta_t^C \\
- \gamma_2, t-1 \theta_{t-1}^C + \gamma_3, t \theta_t^I - \gamma_3, t-1 \theta_{t-1}^I + \nu_t^N - \nu_{t-1}^N.
\]

• Using the measurement equations to proxy \( \theta_{t+1} \) and \( \theta_t \), we obtain

\[
\tilde{Y}_{1,t+1}^N - \tilde{Y}_{1,t}^N = \gamma_0, t - \gamma_0, t-1 + \gamma_1, t \tilde{Y}_{1,t}^N - \gamma_1, t-1 \tilde{Y}_{1,t-1}^N + \gamma_2, t \tilde{Y}_{1,t}^C - \gamma_2, t-1 \tilde{Y}_{1,t-1}^C \\
+ \gamma_3, t \tilde{Y}_{1,t}^I - \gamma_3, t-1 \tilde{Y}_{1,t-1}^I + \nu_t^N - \nu_{t-1}^N \\
+ \left\{ \left( \epsilon_1, t - \epsilon_1, t-1 \right) - \left( \gamma_1, t \epsilon_1, t - \gamma_1, t-1 \epsilon_1, t-1 \right) \right\} - \left( \gamma_3, t \epsilon_1, t - \gamma_3, t-1 \epsilon_1, t-1 \right), \quad t \geq 2.
\]
• We can instrument $\tilde{Y}_{1,t}^k$, $\tilde{Y}_{1,t-1}^k$, $k \in \{C, N, I\}$, using 
$\{Y_{j,t-1}^k\}_{j=2}^{m_k^t}$, $k \in \{C, N, I\}$ and $l \geq 2$, as instruments.

• The validity of the instruments is based on the generalization of
investment equation (9) in Cunha and Heckman (2007),
discussed in our analysis of stage-invariant technologies.

• Thus we can identify the coefficients of (22) except for the
intercepts.

• We can identify relative intercepts $(\gamma^N_{0,t} - \gamma^N_{0,t-1})$,
$t \in \{2, \ldots, T\}$.

• With these intercepts in hand, we can identify the remaining
parameters by the preceding proof provided we have enough
proxies for each factor in each period.
We can set the scale of the factors by estimating their effects on log earnings for children when they become adults.

Let $Y$ be adult earnings.

We write

$$\ln Y = \mu_T + \delta_N \theta_T^N + \delta_C \theta_T^C + \varepsilon,$$

where $\varepsilon$ is not correlated with $\theta_T$ or $\varepsilon_{j,T}^k$.

Define

$$D = \begin{pmatrix} \delta_N & 0 \\ 0 & \delta_C \end{pmatrix}.$$
• Assume $\delta_N \neq 0$ and $\delta_C \neq 0$.

• For any given normalization of the test scores we can transform the $\theta_t$ to an earnings metric by multiplying equation system (6) by $D$:

$$D \theta_{t+1} = \left(DAD^{-1}\right) (D\theta_t) + (DB) \theta^l_t + (D \eta_t),$$  \hspace{1cm} (24)

and work with $D\theta_{t+1}$ and $D\theta_t$ in place of $\theta_{t+1}$ and $\theta_t$.

• The cross terms in $(DAD^{-1})$ are affected by this change of units but not the self-productivity terms.

• The relative magnitude of $\theta^l_t$ on the outcomes can be affected by this change in scale.
• We can use other anchors besides earnings.
• Cunha and Heckman (2008) report results from two anchors in this paper:
  (a) log earnings
  (b) the probability of graduating from high school.
• For the latter, we use a linear probability model.
Table 3: Unanchored Technology Equations: Measurement Error is Classical, Absence of Omitted Inputs Correlated with $\theta_t$, White Males, CNLSY/79

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Noncognitive Skill $\left(\theta^N_t\right)$</th>
<th>Cognitive Skill $\left(\theta^C_{t+1}\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Lagged Noncognitive Skill, $\left(\theta^N_t\right)$</td>
<td>0.884</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Lagged Cognitive Skill, $\left(\theta^C_t\right)$</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Parental Investment, $\left(\theta^I_t\right)$</td>
<td>0.072</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Maternal Education, $S$</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Maternal Cognitive Skill, $A$</td>
<td>-0.006</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>
Findings:

1. Both cognitive and noncognitive skills show strong persistence over time;

2. Noncognitive skills in one period affect the accumulation of next period cognitive skills, but cognitive skills in one period do not affect the accumulation of next period noncognitive skills;

3. The estimated parental investment factor affects noncognitive skills slightly more strongly than cognitive skills, but the differences are not statistically significant;

4. The mother’s ability affects the child’s cognitive ability but not noncognitive ability;

5. The mother’s education plays no role in affecting the evolution of ability after controlling for parental investments, and mother’s ability.
We contrast the OLS estimates of this model (presented in Table 9) with our measurement-error corrected versions in subsection 6 below.
The dynamic factors are statistically dependent.

Table 4 shows the evolution of the correlation patterns across the dynamic factors.

The correlation between cognitive and noncognitive skills is 0.18 at ages six and seven, and grows to around 0.28 at ages 12 and 13.

There is a strong contemporaneous correlation among noncognitive skill and the home investment.
Table 4: Contemporaneous Correlation Matrices: Measurement Error is Classical, Absence of Omitted Inputs Correlated with $\theta_t$, White Males, CNLSY/79

<table>
<thead>
<tr>
<th></th>
<th>Noncognitive</th>
<th>Cognitive</th>
<th>Investments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period 1</strong> – Children ages 6 and 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncognitive</td>
<td>1.0000</td>
<td>0.1892</td>
<td>0.3426</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.1892</td>
<td>1.0000</td>
<td>0.2921</td>
</tr>
<tr>
<td>Investments</td>
<td>0.3426</td>
<td>0.2921</td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>Period 2</strong> – Children ages 8 and 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncognitive</td>
<td>1.0000</td>
<td>0.2334</td>
<td>0.4065</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.2334</td>
<td>1.0000</td>
<td>0.3835</td>
</tr>
<tr>
<td>Investments</td>
<td>0.4065</td>
<td>0.3835</td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>Period 3</strong> – Children ages 10 and 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncognitive</td>
<td>1.0000</td>
<td>0.2643</td>
<td>0.4785</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.2643</td>
<td>1.0000</td>
<td>0.4892</td>
</tr>
<tr>
<td>Investments</td>
<td>0.4785</td>
<td>0.4892</td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>Period 4</strong> – Children ages 12 and 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncognitive</td>
<td>1.0000</td>
<td>0.2845</td>
<td>0.5511</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.2845</td>
<td>1.0000</td>
<td>0.6111</td>
</tr>
<tr>
<td>Investments</td>
<td>0.5511</td>
<td>0.6111</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
• The correlation starts off at 0.40 at ages six and seven and grows to 0.55 by ages 12 and 13.
• The same pattern is true for the correlation between cognitive skills and home investments.
• The correlation between these two variables goes from 0.38 at ages six and seven to 0.61 at ages 12 and 13.
Allowing for Non-Classical Measurement Error

Table 5: Unanchored Technology Equations\textsuperscript{a}: Measurement Error is Classical, Allows for Omitted Input $\lambda$ Correlated with $\theta_t$, White Males, CNLSY/79

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Noncognitive Skill $\left(\theta^N_{t+1}\right)$</th>
<th>Cognitive Skill $\left(\theta^C_{t+1}\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged Noncognitive Skill, $\left(\theta^N_t\right)$</td>
<td>0.8848</td>
<td>0.0276</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Lagged Cognitive Skill, $\left(\theta^C_t\right)$</td>
<td>0.0022</td>
<td>0.9891</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Parental Investment, $\left(\theta^I_t\right)$</td>
<td>0.0797</td>
<td>0.0844</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Omitted Correlated Inputs, $\lambda$</td>
<td>0.2835</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(normalized)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Let $\theta_{t+1}$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let $\lambda$ denote omitted inputs that are potentially correlated with $\theta_t$. The technology equations are:

$$1 = N + C + I + \lambda + \nu$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma$. Note that for identification purposes we normalize $C\gamma = 1$. Investment is normalized on family income.
(a) Let $\theta'_t = (\theta^N_t, \theta^C_t, \theta^I_t)$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let $\lambda$ denote omitted inputs that are potentially correlated with $\theta_t$. The technology equations are:

$$\theta^k_{t+1} = \gamma^k_1 \theta^N_t + \gamma^k_2 \theta^C_t + \gamma^k_3 \theta^I_t + \gamma^k_4 \lambda + \nu^k_t$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma^k_1$, $\gamma^k_2$, $\gamma^k_3$ and $\gamma^k_4$. Note that for identification purposes we normalize $\gamma^C_4 = 1$. Investment is normalized on family income.
Anchoring estimates of the factor scale using adult outcomes

- Table 6 reports estimates of the time-invariant technology that use the earnings data for persons age 23-28 to anchor the output of the production function in a log dollar metric.
- We initially assume that $\eta_t$ is serially uncorrelated and that measurement error is classical.
- We relax these assumptions below, when we report estimates of more general specifications.
- Our fitted earnings function is linear in age, and depends on the final level of the factors $\theta_{T+1}^C$ and $\theta_{T+1}^N$. 
Table 6: Anchored Technology Equations\textsuperscript{a}: Anchoring on Log Earnings and Graduation from High School Measurement Error is Non-Classical, No Omitted Inputs Correlated with $\theta_t$, White Males, CNLSY/79

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Noncognitive Skill</th>
<th>Cognitive Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Lagged Noncognitive Skill, $\left(\theta_t^N\right)$</td>
<td>0.8844</td>
<td>0.8843</td>
</tr>
<tr>
<td></td>
<td>(0.0210)</td>
<td>(0.0210)</td>
</tr>
<tr>
<td>Lagged Cognitive Skill, $\left(\theta_t^C\right)$</td>
<td>0.0084</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>(0.0364)</td>
<td>(0.0053)</td>
</tr>
<tr>
<td>Parental Investment, $\left(\theta_t^I\right)$</td>
<td>0.0101</td>
<td>0.0079</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>Maternal Education, $S$</td>
<td>0.0006</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>Maternal Cognitive Skill, $A$</td>
<td>-0.0008</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0007)</td>
</tr>
</tbody>
</table>
One problem that might arise in using log earnings as an anchor for this sample is that log earnings are observed for the children who are born to very young mothers, making it a very selected sample.

To check the robustness of these conclusions with regard to the log earnings anchor, we also use high school graduation for a person at least 19 years-old to anchor the parameters of the technology equations.

We model the probability of high school graduation as a linear probability equation.

It is interesting to note that in the metric of the probability of graduating from high school, the estimated parental investment factor affects cognitive skills more strongly than noncognitive skills.
• This is because cognitive skills receive higher weight in the high school graduation equation than in the log earnings equation.
• The relative strength of these effects is reversed across the two metrics.
• The choice of a metric is not innocuous.
Table 7: Unanchored Stage Specific Technology Equations$^a$:
Measurement Error is Classical, No Omitted Inputs Correlated with $\theta_t$, White Males, CNLSY/79

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Noncognitive Skill $\left(\theta_{t+1}^N\right)$</th>
<th>Cognitive Skill $\left(\theta_{t+1}^C\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Lagged Noncognitive Skill, $\left(\theta_t^N\right)$</td>
<td>0.9849</td>
<td>0.9383</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Lagged Cognitive Skill, $\left(\theta_t^C\right)$</td>
<td>0.0508</td>
<td>-0.0415</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Parental Investment, $\left(\theta_t^I\right)$</td>
<td>0.0533</td>
<td>0.1067</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Maternal Education, $S$</td>
<td>0.0034</td>
<td>-0.0028</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Maternal Cognitive Skill, $A$</td>
<td>0.0007</td>
<td>-0.0077</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
</tbody>
</table>

$^a$ Let $\theta_{tt}$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let $S$ denote mother’s education and $A$ denote mother’s cognitive ability. The technology equations are:

$$
\begin{align*}
\gamma_{t+1}^N &= \gamma_{t+1}^C + \psi_{t+1}^N + \eta_{t+1}^N \\
\gamma_{t+1}^C &= \gamma_{t+1}^C + \psi_{t+1}^C + \eta_{t+1}^C \\
\gamma_t^I &= \gamma_t^I + \psi_t^I + \eta_t^I \\
\end{align*}
$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_{t+1}^N$, $\gamma_{t+1}^C$, and $\gamma_t^I$. Stage 1 consists of the transition from ages 6-7 to ages 8-9. Stage 2 refers to the transition from ages 8-9 to 10-11. Stage 3 is the transition from ages 10-11 to 12-13.
Table 8: Anchored Stage Specific Technology Equations\textsuperscript{a}: Anchor: Log Earnings of the Child Between Ages 23-28 Measurement Error is Classical, No Omitted Inputs Correlated with $\theta_t$, White Males, CNLSY/79

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Noncognitive Skill ($\theta_{t+1}^N$)</th>
<th>Cognitive Skill ($\theta_{t+1}^C$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Stage 2</td>
<td>Stage 3</td>
</tr>
<tr>
<td>Lagged Noncognitive Skill, ($\theta_{t}^N$)</td>
<td>0.9849</td>
<td>0.9383</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Lagged Cognitive Skill, ($\theta_{t}^C$)</td>
<td>0.1442</td>
<td>-0.1259</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.115)</td>
</tr>
<tr>
<td>Parental Investment, ($\theta_{t}^I$)</td>
<td>0.0075</td>
<td>0.0149</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Maternal Education, $S$</td>
<td>0.0005</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Maternal Cognitive Skill, $A$</td>
<td>0.0001</td>
<td>-0.0011</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Let $\theta_{t}^N, \theta_{t}^C, \theta_{t}^I$ denote the noncognitive, cognitive and investment dynamic factors, respectively. Let $S$ denote mother’s education and $A$ denote mother’s cognitive ability. The technology equations are:

$$11 2 3 1 3 2 1 \gamma \gamma \gamma \gamma \psi \psi \eta ++ = + + +$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $12 31 , 2 , , k k k k$ and $\gamma \gamma \gamma \gamma$. Stage 1 consists of the transition from ages 6-7 to ages 8-9. Stage 2 refers to the transition from ages 8-9 to 10-11. Stage 3 is the transition from ages 10-11 to 12-13.
Estimating the Components of the Home Investment Dynamic Factor

Table 9: OLS Estimation of the Technology Equations: Measurement Error is Classical, Absence of Omitted Inputs Correlated with $\theta_t$, White Males, CNLSY/79, Cont’d

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Antisocial Score ($t + 1$)</th>
<th>PIAT Math ($t+1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antisocial Score, $t$</td>
<td>0.6431</td>
<td>0.0333</td>
</tr>
<tr>
<td></td>
<td>(0.0165)</td>
<td>(0.0096)</td>
</tr>
<tr>
<td>PIAT Math, $t$</td>
<td>0.0933</td>
<td>0.5909</td>
</tr>
<tr>
<td></td>
<td>(0.0317)</td>
<td>(0.0184)</td>
</tr>
<tr>
<td>HOME Score, $t$</td>
<td>0.0147</td>
<td>-0.0137</td>
</tr>
<tr>
<td></td>
<td>(0.0059)</td>
<td>(0.0034)</td>
</tr>
<tr>
<td>Maternal Education</td>
<td>0.0358</td>
<td>0.0208</td>
</tr>
<tr>
<td></td>
<td>(0.0091)</td>
<td>(0.0053)</td>
</tr>
<tr>
<td>Maternal ASVAB Arithmetics</td>
<td>-0.0254</td>
<td>0.0658</td>
</tr>
<tr>
<td></td>
<td>(0.0190)</td>
<td>(0.0110)</td>
</tr>
</tbody>
</table>
• Generally, OLS coefficients are downward-biased, showing much smaller self productivity, cross productivity and investment productivity effects.

• The estimated effect of the HOME score on the Math score is perverse.
Return to main text
Estimating the Technology of Production of Cognitive and Noncognitive Skills
(builds on Cunha, Heckman, & Schennach, 2010)
\[ \theta^C_{t+1} = f^C_t \left( \theta^C_t, \theta^N_t, I^C_t, \eta^C_t \right) \]
\[ \theta^N_{t+1} = f^N_t \left( \theta^C_t, \theta^N_t, I^N_t, \eta^N_t \right) \]
A Model for the Measurements
\[ Y_{j,t}^C, \quad Y_{j,t}^N \]
\[ Y_{j,t}^k = \mu_{j,t}^k + \alpha_{j,t}^k \theta_t^k + \varepsilon_{j,t}^k \text{ for } j = 1, \ldots, m_t^k, \ t = 1, \ldots, T, \text{ and } k = C, N \]
\[ \alpha_{1,t}^k = 1 \text{ and } \mu_{1,t}^k = 0 \text{ for } k = C, N \]
• Parental and schooling investments $I_t^k$ not directly observed.
• Observe a vector of measurements $X_{j,t}^k$.

$$X_{j,t}^k = \mu_{j,t}^X + \beta_{j,t}^k I_t^k + \varepsilon_{j,t}^l$$ for $j = 1, \ldots, m_t^k$, $t = 1, \ldots, T$, and $k = C, N$ (26)
Skill Measurements
\[ M_j^k = \mu_j^M + \delta_j^k \theta_j^M + \varepsilon_j^M \] (27)
Identification

- Identification of the factor loadings and of the joint distributions of the latent variables
• First establish identification of the factor loadings under the assumption that \( m^k_t \geq 3, \ k = C, N, IC, IN, MC, MN \).

• Without loss of generality, we focus on \( \alpha^C_{j,t} \) and note that similar expressions can be derived trivially for the loadings for the remaining latent factors.
Since we observe \( \{ Y_{j,t}^C \}_{j=1}^{m_t^C} \) for every person, we can compute from the data \( \text{Cov} (Y_{i,t}^C, Y_{j,t}^C) \) for all \( i, j \) pairs.

Let \( \text{Var}(\theta_t^C) = \sigma_{\theta_t^C}^2 \) denote the variance of \( \theta_t^C \), \( t = 1, 2, \ldots, T \), and note that it may change over time.

Recall that \( \alpha_{1,t}^C = 1 \). Consequently, we obtain:

\[
\text{Cov} \left( Y_{1,t}^C, Y_{2,t}^C \right) = \alpha_{2,t}^C \sigma_{\theta_t}^2
\] (28)

\[
\text{Cov} \left( Y_{1,t}^C, Y_{3,t}^C \right) = \alpha_{3,t}^C \sigma_{\theta_t}^2
\] (29)

\[
\text{Cov} \left( Y_{2,t}^C, Y_{3,t}^C \right) = \alpha_{2,t}^C \alpha_{3,t}^C \sigma_{\theta_t}^2.
\] (30)
Taking ratios:

\[
\frac{\text{Cov} \left( Y_{2,t}, Y_{3,t}^C \right)}{\text{Cov} \left( Y_{1,t}, Y_{3,t}^C \right)} = \alpha_{2,t}^C
\]

\[
\frac{\text{Cov} \left( Y_{2,t}^C, Y_{3,t}^C \right)}{\text{Cov} \left( Y_{1,t}^C, Y_{2,t}^C \right)} = \alpha_{3,t}^C.
\]
• We can identify $\alpha_{2,t}^C$ and $\alpha_{3,t}^C$ from the ratios of covariances.
• Identify $\alpha_{j,t}^C$ for $j = 2, 3, \ldots, m_t^C$, $t = 1, \ldots, T$ up to the normalization $\alpha_{1,t}^C = 1$.
• Identify $\sigma_{\theta_t^C}^2$ for all $t = 1, 2, \ldots, T$.
• Once the parameters $\alpha_{1,t}^C, \alpha_{2,t}^C, \ldots, \alpha_{m_t^C,t}^C$ are identified (up to the normalization $\alpha_{1,t}^C = 1$), we can rewrite (25) as:

$$
\frac{Y_{j,t}^C}{\alpha_{j,t}^C} = \theta_t^C + \frac{\varepsilon_{j,t}^C}{\alpha_{j,t}^C}, \ j = 1, 2, \ldots, m_t^C.
$$

(31)
Can identify distribution of $\theta_t^C$ using results in Schennach (2004a,b)
\[ \theta = \left( \left\{ \theta^C_t \right\}_{t=1}^T, \left\{ \theta^N_t \right\}_{t=1}^T, \left\{ I^C_t \right\}_{t=1}^T, \left\{ I^N_t \right\}_{t=1}^T, \theta^C_M, \theta^N_M \right) \].
\[ \mathbf{Z}_1 = \left( \left\{ Y_{1,t}^C / \alpha_{1,t} \right\}_{t=1}^T, \left\{ Y_{1,t}^N / \alpha_{1,t} \right\}_{t=1}^T, \left\{ X_{1,t}^C / \beta_{1,t} \right\}_{t=1}^T, \left\{ X_{1,t}^N / \beta_{1,t} \right\}_{t=1}^T, M_1^C / \delta_1^C, M_1^N / \delta_1^N \right) \right)' \]

\[ \mathbf{Z}_2 = \left( \left\{ Y_{2,t}^C / \alpha_{2,t} \right\}_{t=1}^T, \left\{ Y_{2,t}^N / \alpha_{2,t} \right\}_{t=1}^T, \left\{ X_{2,t}^C / \beta_{2,t} \right\}_{t=1}^T, \left\{ X_{2,t}^N / \beta_{2,t} \right\}_{t=1}^T, M_2^C / \delta_2^C, M_2^N / \delta_2^N \right) \right)' . \]
These vectors consist, respectively, of the first and the second measurements for each factor. The corresponding measurement errors are

\[
\varepsilon_1 = \begin{pmatrix} 
\{\varepsilon_{1,t}/\alpha_{1,t}\}_{t=1}^T, \{\varepsilon_{1,t}/\alpha_{1,t}\}_{t=1}^T, \\
\{\varepsilon_{1,t}/\beta_{1,t}\}_{t=1}^T, \{\varepsilon_{1,t}/\beta_{1,t}\}_{t=1}^T, \\
\varepsilon_{M1}/\delta_{1} 
\end{pmatrix},
\]

\[
\varepsilon_2 = \begin{pmatrix} 
\{\varepsilon_{2,t}/\alpha_{2,t}\}_{t=1}^T, \{\varepsilon_{2,t}/\alpha_{2,t}\}_{t=1}^T, \\
\{\varepsilon_{2,t}/\beta_{2,t}\}_{t=1}^T, \{\varepsilon_{2,t}/\beta_{2,t}\}_{t=1}^T, \\
\varepsilon_{M2}/\delta_{2} 
\end{pmatrix}.
\]
 Distribution $\theta$ identified by the following Theorem
Theorem 1

Let $Z_1$, $Z_2$, $\theta$, $\varepsilon_1$, $\varepsilon_2$ be random variables taking values in $\mathbb{R}^L$ and related through

$$
Z_1 = \theta + \varepsilon_1
$$
$$
Z_2 = \theta + \varepsilon_2.
$$

If, for $k, k' = 1, \ldots, L$,

$$
E [\varepsilon_{1,k} | \theta_k, \varepsilon_{2,k}] = 0 \quad (32)
$$

$$
\varepsilon_{2,k} \perp (\theta, \varepsilon_{2,k'}) \text{ for } k \neq k' \quad (33)
$$

then the density of $\theta$ is given by:

$$
p_{\theta} (\theta) = (2\pi)^{-L} \int e^{-i\chi \theta} \frac{E \left[ e^{i\chi \cdot Z_2} \right]}{\prod_{k=1}^{L} E \left[ e^{i\chi_k Z_2,k} \right]} \prod_{k=1}^{L} \exp \left( \int_0^{\chi_k} \frac{E \left[ iZ_{1,k} e^{i\zeta_k Z_{2,k}} \right]}{E \left[ e^{i\zeta_k Z_{2,k}} \right]} d\zeta_k \right) d\chi, \quad (34)
$$

provided all the above expectations exist.
Note that $\theta$ can trivially include elements that are perfectly measured.

Identification of the value added technology function

$$\theta^k_{t+1} = f^k_t (\theta^C_t, \theta^N_t, I^k_t, \theta^C_m, \theta^N_m)$$

for $k = C, N$. 

(35)
Note that, even if $\theta$ were perfectly observed, we could not separately identify the distribution of $\eta^k_t$ and the function $f^k_t$ because, without further normalization, a change in the density of $\eta^k_t$ can be undone by a change in the function $f^k_t$.

One approach is to assume that (35) is additively separable in $\eta^k_t$.

Another way to avoid this ambiguity is to normalize $\eta_t$ to have a uniform density on $[0, 1]$.

As alternatives, any of the other normalizations suggested in Matzkin (2003) could be used.

To show that $f^k_t$ for $k = C$ or $N$ is nonparametrically identified, we note that, from the knowledge of $p_\theta (\theta)$, we can calculate, for any $\bar{\theta} \in \mathbb{R}$,

$$P \left[ \theta^k_{t+1} \leq \bar{\theta} | \theta^C_t, \theta^N_t, l^k_t \right] \equiv G \left( \bar{\theta} | \theta^C_t, \theta^N_t, l^k_t \right)$$
Next, we set
\[
  f_t^k (\theta_t^C, \theta_t^N, I_t^k, \eta_t) = G^{-1} (\eta^k_t | \theta_t^C, \theta_t^N, I_t^k)
\]
where \( G^{-1} (\eta^k_t | \theta_t^N, \theta_t^C, I_t^k) \) denotes the inverse of \( G (\tilde{\theta} | \theta_t^N, \theta_t^C, I_t^k) \) with respect to its first argument, i.e. the value \( \tilde{\theta} \) such that \( \eta^k_t = G (\tilde{\theta} | \theta_t^N, \theta_t^C, I_t^k) \).

By construction, this operation will produce a function \( f_t^k \) producing outcomes \( \theta_{t+1}^k \) with the appropriate distribution, thanks to the well-known property that a random variable is mapped into a uniformly distributed variable under the mapping defined by its own cdf.
The more traditional separable technology function with zero mean disturbance can also be handled by our analysis,

$$\theta_{t+1}^k = f_t^k (\theta_t^C, \theta_t^N, I_t^k) + \eta_t^k,$$

simply by defining

$$f_t^k (\theta_t^C, \theta_t^N, I_t^k) \equiv E [\theta_{t+1}^k | \theta_t^C, \theta_t^N, I_t^k],$$
The density of $\eta^k_t$ conditional on all variables can also be identified by

$$p_{\eta^k_t | \theta^C_t, \theta^N_t, l^k_t} (\eta^k_t | \theta^C_t, \theta^N_t, l^k_t) = p_{\theta^k_{t+1} | \theta^C_t, \theta^N_t, l^k_t} (\eta^k_t + E[\theta^k_{t+1} | \theta^C_t, \theta^N_t, l^k_t] | \theta^C_t, \theta^N_t, l^k_t)$$

since $p_{\theta^k_{t+1} | \theta^C_t, \theta^N_t, l^k_t}$ is known from $p_\theta (\theta)$. 

Econ and Ecom of Hum Dev
Anchoring
• Anchoring value added in meaningful cardinal outcomes
• It is common in the literature on child investment to measure skills in units of the test score selected to have factor loading of one.

\[ P \left[ D = 1 \mid \theta^C_T, \theta^N_T \right] = g \left( \theta^C_T, \theta^N_T \right), \]
Since the dummy \( D \) is perfectly observed, the corresponding element of the two repeated measurement vectors \( Z_1 \) and \( Z_2 \) are identical and equal to \( D \).

Theorem 1 then implies that the joint density of \( D, \theta_t^C \) and \( \theta_t^N \) is identified, thus making it possible to identify 
\[
P \left[ D = 1 \mid \theta_T^C, \theta_T^N \right].
\]

\( g^C (\theta_T^C) \) and \( g^N (\theta_T^N) \) from the function \( g (\theta_T^C, \theta_T^N) \), by integrating over one of the variables, e.g.,
\[ g^C (\theta_T^C) \equiv \int g (\theta_T^C, \theta_T^N) \, p_{\theta_N^T} (\theta_T^N) \, d\theta_T^N \]

\[ g^N (\theta_T^N) \equiv \int g (\theta_T^C, \theta_T^N) \, p_{\theta_C^T} (\theta_T^C) \, d\theta_T^C \]

The “anchored” skills, denoted by \( \tilde{\theta}_t^k \), are then defined as

\[ \tilde{\theta}_t^k = g^k (\theta_t^k) \]

for \( k = C, N \).
\[ \tilde{f}_t^k \left( \tilde{\theta}_t^C, \tilde{\theta}_t^N, l_t^k, \eta_t^k \right) \equiv g^k \left( f_t^k \left( [g^C]^{-1} \left( \tilde{\theta}_t^C \right), [g^N]^{-1} \left( \tilde{\theta}_t^N \right), l_t^k, \eta_t^k \right) \right), \]
\[
\tilde{f}_t^k \left( \tilde{\theta}_t^C, \tilde{\theta}_t^N, l_t^k, \eta_t^k \right) = f_t^k \left( g^C (\theta_t^C), g^N (\theta_t^N), l_t^k, \eta_t^k \right)
\]

\[
= g^k \left( f_t^k \left( \left[ g^C \right]^{-1} (g^C (\theta_t^C)), \left[ g^N \right]^{-1} (g^N (\theta_t^N)) \right), l_t^k, \eta_t^k \right)
\]

\[
= g^k \left( \theta_t^k + 1 \right) = \tilde{\theta}_{t+1}^k,
\]
Hence, $\tilde{f}_t^k$ provides the equation of evolution for the anchored skills $\tilde{\theta}_t^k$ that is consistent with the equation of evolution $f_t^k$ of the original skills $\theta_t^k$. 
More General Factor Measurement Models
The identification of a general nonlinear factor model

A factor model of the general form

\[ Y_j = a_j(\theta, \varepsilon_j) \text{ for } j = 1, \ldots, m, \]  

(36)
Theorem 2

The distribution of $\theta$ in Model (36) is identified, if the following conditions are satisfied

1. All $\varepsilon_j$ for $j = 1, \ldots, m$ are mutually independent and independent from $\theta$.
2. $p_{Y_1|Y_3}(Y_1|Y_3)$ and $p_{Y_1|\theta}(Y_1|\theta)$ each form a complete family of distributions (indexed by $Y_3$ and $\theta$, respectively).
3. The density $p_{Y_2|\theta}(Y_2|\theta)$ is bounded uniformly in $Y_2$ and $\theta$.
4. Whenever $\theta \neq \tilde{\theta}$, $p_{Y_2|\theta}(Y_2|\theta)$ and $p_{Y_2|\tilde{\theta}}(Y_2|\tilde{\theta})$ differ at least over a set of nonzero Lebesgue measure.
5. There exists a known functional $M$, mapping a density to a real number, that has the property that $M\left[p_{Y_1|\theta}(\cdot|\theta)\right] = \theta$. 
Proof of Theorem 2.

Since $m \geq 3$, we can use Theorem 1 in Hu and Schennach (2006) to prove that the distribution of $\theta$ is identified, after setting $x = Y_1$, $y = Y_2$, $z = Y_3$, and $x^* = \theta$. Assumption 1 herein implies that the Assumption 1 in Hu and Schennach (2006) holds. All remaining assumptions are directly assumed. Theorem 1 in Hu and Schennach (2006) then implies that the joint density of $Y_1$ and $\theta$ is identified, which provides the density of $\theta$, after integration over $Y_1$. 

\qed
Theorem 2 does not state that the distributions of the errors $\varepsilon_j$ or that the functions $a_j(\cdot, \cdot)$ are fully identified.

In fact, it is always possible to alter the distribution of $\varepsilon_j$ and the dependence of the function $a_j(\cdot, \cdot)$ on its second argument in ways that cancel each other out, as has often been noted in the nonseparable literature (e.g. Matzkin, 2003).
Further results on skills “anchoring”
• We assume throughout that the factor loadings $\alpha_{1t}^k$ have been estimated beforehand.

• We need to find the relationship between adult log income $\ln E$ and skills at the end of childhood $\theta^C_T, \theta^N_T$.

• The end result will be some normalization defined by some functions $g^C$ and $g^N$ or, for short:

$$g(\theta_t) \equiv \begin{pmatrix} g^C(\theta^C_t) \\ g^N(\theta^N_t) \end{pmatrix}$$

where $\theta_t \equiv \begin{pmatrix} \theta^C_t \\ \theta^N_t \\ I_t \end{pmatrix}$.
• Log linear separable case

\[ \ln E = \mu_T + \delta^C \theta^C_T + \delta^N \theta^N_T + \xi \]

where \( \xi \) is an error term.

• A linear errors-in-variable models can be estimated with standard 2SLS using \( \frac{Y_{j,T}^C}{\alpha_{jT}^C} \) for \( j = 2, \ldots, m^C_T \) as instruments for the regressor \( \frac{Y_{1,T}^N}{\alpha_{1T}^N} \) and using \( \frac{Y_{j,T}^N}{\alpha_{jT}^N} \) for \( j = 2, \ldots, m^N_T \) as instruments for the regressor \( \frac{Y_{1,T}^N}{\alpha_{1T}^N} \).

• Of course, any observation other than \( j = 1 \) can also be singled out as the regressor.

• One can also use the average \( \frac{1}{m^C_T/2} \sum_{j=1}^{m^C_T/2} \frac{Y_{j,T}^C}{\alpha_{jT}^C} \) as an instrument for \( \frac{1}{m^C_T/2} \sum_{j=m^C_T/2+1}^{m^C_T} \frac{Y_{j,T}^C}{\alpha_{jT}^C} \), etc.
Then, we set \( g^C(\theta^C_t) \equiv \exp(\delta^C \theta^C_T) \) and \( g^N(\theta^N_t) = \exp(\delta^N \theta^N_T) \).

We are using exponentials here so that, in the limiting case of Cobb-Douglas technology, we would get a product of the form \( \exp(\delta^C \theta^C_T) \exp(\delta^N \theta^N_T) \), which becomes the sum \( \delta^C \theta^C_T + \delta^N \theta^N_T \) after taking logs, as in the earlier log linear paper.
Nonlinear case with separable error
Assume

\[ E = g \left( \theta^C_T, \theta^N_T \right) + \xi \text{ with } \mathbb{E} \left[ \xi | \theta^C_T, \theta^N_T \right] = 0. \]

In this case, the function of interest could be

\[ g \left( \theta^C_T, \theta^N_T \right) \equiv \mathbb{E} \left[ E | \theta^C_T, \theta^N_T \right]. \]
If we specify \( g \left( \theta^C_T, \theta^N_T \right) \) parametrically as \( g \left( \theta^C_T, \theta^N_T; \delta \right) \), this can be estimated using moment conditions of the form

\[
\mathbb{E} \left[ (E - g \left( \theta^C_T, \theta^N_T; \delta \right)) \frac{\partial g \left( \theta^C_T, \theta^N_T; \delta \right)}{\partial \delta} \right] \equiv \mathbb{E} \left[ u \left( E, \theta_T; \delta \right) \right] = 0
\]
How is it possible to extract two anchors $g^C(\theta^C_T)$ and $g^N(\theta^N_T)$ from one function $g(\theta^C_T, \theta^N_T; \delta)$?

The most natural choice in this case is to average, e.g.:

$$g^C(\theta^C_T) = \int g(\theta^C_T, \theta^N_T, \delta) f(\theta^N_T) d\theta^N_T$$
• Another possible choice is to take \( f (\theta^N_T|\theta^C_T) \) instead of \( f (\theta^N_T) \), which would improve the predictive power of \( g^C (\theta^C_T) \) when used alone to predict \( E \).

• Alternatively, we can directly fit a separable function of the form \( g^C (\theta^C_T; \delta^C) + g^N (\theta^N_T; \delta^N) \) by solving

\[
\mathbb{E} \left[ (E - (g^C (\theta^C_T; \delta^C) + g^N (\theta^N_T; \delta^N))) \frac{\partial (g^C (\theta^C_T; \delta^C) + g^N (\theta^N_T; \delta^N))}{\partial (\delta^C, \delta^N)} \right] = 0.
\]
Nonlinear nonseparable case
Nonlinear weakly separable case
Anchoring the Factors in the Metric of Earnings
\[
\ln E_T = \mu_T + \delta_N \theta_T^N + \delta_C \theta_T^C + \varepsilon_T
\]  \hspace{1cm} (37)

Define

\[
D = \begin{pmatrix}
\delta_N & 0 \\
0 & \delta_C
\end{pmatrix}.
\]

We assume \(\delta_N \neq 0\) and \(\delta_C \neq 0\).
For any given normalization of the test scores we can transform the $\theta_t$ to an earnings metric by multiplying the technology equation by $D$:

$$D\theta_{t+1} = (DA_tD^{-1}) (D\theta_t) + (DB) I_t + (D\eta_t), \quad (38)$$
Further results on skills “anchoring”
We assume throughout that the factor loadings $\alpha^k_{1t}$ have been estimated beforehand.

$$
\begin{align*}
g (\theta_t) &\equiv \begin{pmatrix} g^C (\theta^C_t) \\ g^N (\theta^N_t) \\ I_t \end{pmatrix} \\
\text{where } \theta_t &\equiv \begin{pmatrix} \theta^C_t \\ \theta^N_t \\ \theta^I_t \end{pmatrix} \equiv \begin{pmatrix} \theta^C_t \\ \theta^N_t \\ I_t \end{pmatrix}.
\end{align*}
$$
Log linear separable case
Nonlinear case with separable error
Nonlinear nonseparable case
The CES Specification
Assuming that the error term $\eta_t^C$ enters the specification in the same way as the covariates (i.e., it’s a “random input”), we have:

$$g^C (\theta_{t+1}^C) = f^C (g (\theta_t^C), \eta_t) = \left( \gamma_1^C \left( g^C (\theta_t^C) \right)^\alpha + \gamma_2^C \left( g^N (\theta_t^N) \right)^\alpha + \gamma_3^C (I_t)^\alpha + \gamma_4^C \left( \eta_t^C \right)^\alpha \right)^{1/\alpha}$$
Computational details:

\[
\frac{\partial f_C \left( g \left( \theta_t \right), \eta_t^C \right)}{\partial \eta_t} = \left( \gamma_1^C \left( g \left( \theta_t^C \right) \right)^\alpha + \gamma_2^C \left( g^N \left( \theta_t^N \right) \right)^\alpha + \gamma_3^C \left( I_t \right)^\alpha + \gamma_4^C \left( \eta_t^C \right)^\alpha \right)^{(1/\alpha) - 1} \gamma_4^C \left( \eta_t^C \right)^\alpha - 1
\]

\[
\eta_t^C = \frac{1}{\gamma_4^C} \left( \left( g^C \left( \theta_t^{C+1} \right) \right)^\alpha - \gamma_1^C \left( g^C \left( \theta_t^C \right) \right)^\alpha - \gamma_2^C \left( g^N \left( \theta_t^N \right) \right)^\alpha - \gamma_3^C \left( I_t \right)^\alpha \right)^{1/\alpha}
\]
Return to main text
Evidence on The Predictive Power of Cognitive and Socioemotional Traits
Figure 3: Sorting into Schooling

The figure illustrates the distribution of cognitive and socio-emotional factors among different levels of schooling attainment. The x-axis represents the factors, and the y-axis shows the density of these factors. The lines represent different educational outcomes, including HS Drop, GED, HS Grad., Some College, and 4yr Coll. Grad., each with distinct densities across the factor spectrum. The graph visually demonstrates how sorting into schooling is influenced by these factors.
Figure 4: The Probability of Educational Decisions, by Endowment Levels, Dropping from Secondary School vs. Graduating

Figure 5: The Effect of Cognitive and Socio-emotional endowments on Physical Health at age 40 (PCS-12)

Figure 6: The Effect of Cognitive and Socio-emotional endowments on Ever Participated in Welfare (1996-2006)

Figure 7: The Effect of Cognitive and Socio-emotional endowments on Trusting People (2008)

Figure 8: The Probability and Returns of College Enrollment by Endowments Levels

Source: Eisenhauer et al. (2014)
Note: Early college enrollment refer to the individuals who enroll in college immediately after having finished high school. Returns are expressed in units of millions of dollars.
Figure 9: 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.
Source: Heckman et al. (2006).
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.

Source: Heckman et al. (2006).
Figure 11: Probability of being a high school dropout by age 30 (males)

Source: Heckman et al. (2006).
Figure 12: Probability of being a 4-year college graduate by age 30 (males)

Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

Source: Heckman et al. (2006).
**Figure 13: Probability of daily smoking by age 18 (males)**

Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).

Source: Heckman et al. (2006).
Figure 14: Mean log wages by age 30 (males)

Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (50 draws).

Source: Heckman et al. (2006).
Figure 15: Average Treatment Effect of Graduating from a Four-Year College by Outcome

A. (log)Wages

B. PV Wages

Figure 15: Average Treatment Effect of Graduating from a Four-Year College by Outcome, Cont’d

C. Smoking

D. Health Limits Work

Figure 16: The Effect of Cognitive and Socioemotional Endowments

A. (log)Wages

B. PV Wages

Notes: For each of the four outcomes, we present three figures that study the impact of cognitive and socioemotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socioemotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socioemotional endowments. Notice that we define as “decile 1” the decile with the lowest values of endowments and “decile 10” as the decile with the highest levels of endowments. The bottom left figure displays the average levels of endowment across deciles of cognitive endowments. The bottom right figure mimics the structure of the left-hand side figure but now for the socioemotional endowment.

Source: Heckman, Humphries, and Veramendi (2016)
Figure 16: The Effect of Cognitive and Socioemotional Endowments, Cont’d

C. Health Limits Work

D. Smoking

Notes: For each of the four outcomes, we present three figures that study the impact of cognitive and socioemotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socioemotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socioemotional endowments. Notice that we define as “decile 1” the decile with the lowest values of endowments and “decile 10” as the decile with the highest levels of endowments. The bottom left figure displays the average levels of endowment across deciles of cognitive endowments. The bottom right figure mimics the structure of the left-hand side figure but now for the socioemotional endowment.

Source: Heckman, Humphries, and Veramendi (2016) Econ and Ecom of Hum Dev
Figure 17: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels are Highlighted Using Bold Letters)

A. Dropping from HS vs. Graduating from HS

B. HS Dropout vs. Getting a GED

Notes: For each of the four educational choices, we present three figures that study the probability of that specific educational choice. Final schooling levels do not allow for further options. For each pair of schooling levels \( j \) and \( j + 1 \), the first subfigure (top) presents \( \text{Prob}(D_j = 0|d^C, d^{SE}) \) where \( d^C \) and \( d^{SE} \) denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. \( \text{Prob}(D_j = 0|d^C, d^{SE}) \) is computed for those who reach the decision node involving a decision between levels \( j \) and \( j + 1 \). The bottom left subfigures present \( \text{Prob}(D_j = 0|d^C) \) where the socioemotional factor is integrated out. The bars in these figures display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving levels \( j \) and \( j + 1 \). The bottom right subfigures present \( \text{Prob}(D_j = 0|d^{SE}) \) for a given decile of socioemotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels \( j \) and \( j + 1 \).

Source: Heckman, Humphries, and Veramendi (2016)
Figure 17: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels are Highlighted Using Bold Letters), Cont’d

C. HS Graduate vs. College Enrollment  
D. Some College vs. 4-year college degree

Notes: For each of the four educational choices, we present three figures that study the probability of that specific educational choice. Final schooling levels do not allow for further options. For each pair of schooling levels \( j \) and \( j + 1 \), the first subfigure (top) presents \( \text{Prob}(D_j = 0 | d^C, d^{SE}) \) where \( d^C \) and \( d^{SE} \) denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. \( \text{Prob}(D_j = 0 | d^C, d^{SE}) \) is computed for those who reach the decision node involving a decision between levels \( j \) and \( j + 1 \). The bottom left subfigures present \( \text{Prob}(D_j = 0 | d^C) \) where the socioemotional factor is integrated out. The bars in these figures display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving levels \( j \) and \( j + 1 \). The bottom right subfigures present \( \text{Prob}(D_j = 0 | d^{SE}) \) for a given decile of socioemotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels \( j \) and \( j + 1 \).
Figure 17: The Effect of Cognitive and Socioemotional Endowments on Work Limited

Source: Heckman et al. (2016).
Figure 17: Average Treatment Effect of Education on Present Value of Wages, by Decision Node and Endowment Levels (cont.)

D. Some College vs. 4-year college degree

Source: Heckman et al. (2016).
Figure 18: Average Treatment Effect of Education on Smoking, by Decision Node and Endowment Levels

A. Dropping from HS vs. Graduating from HS

Source: Heckman et al. (2016).
Figure 19: Cognitive and Socio-Emotional Factors: Physical Health, Males
Figure 20: The Effect of Cognitive and Socio-Emotional Endowments on Mental Health at Age 40
Figure 21: Probability of Being a Four-Year College Graduate by Age 30

ii. By Decile of Cognitive Factor

iii. By Decile of Personality
Figure 22: Mean Log Wages by Age 30 (males)

Notes: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (50 draws).
**Figure 23**: Probability of Daily Smoking by Age 18 (Males)

**Notes**: The data are simulated from the estimates of the model and our NLSY79 sample. We use the standard convention that higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws).
Figure 24: Cognitive and Socio-Emotional Factors: Probability of Graduating from Secondary School, Males
Figure 25: The Effect of Cognitive and Socio-Emotional Endowments on Probability of White-Collar Occupation
Figure 26: The Effect of Cognitive and Socio-Emotional Endowments on Smoking
Figure 27: The Effect of Cognitive and Socio-Emotional Endowments on Heavy Drinking During Adulthood
Figure 28: The Effect of Cognitive and Socio-Emotional Endowments on Pearlin’s “Personal Mastery Scale”: Sense of Self-Mastery
Figure 29: The Effect of Cognitive and Socio-Emotional Endowments on Trusting People (2008)
Figure 30: The Effect of Cognitive and Socio-Emotional Endowments on Ever Divorced