

# Parental Investment and Child Development

Duc Thanh Nguyen

April 20, 2026

# Objectives

- Formulate a dynamic production function model for the various dimensions of human capital with endogenous parental investments to examine dynamic complementarities and interactions among different inputs in forming child cognitive skills, social-emotional skills and health.
- Use a maximum likelihood approach by the minorization-maximization algorithm to estimate the distribution of the latent factors and dynamic CES production functions of human capital.

## Contributions

- This research is the first attempt to analyze the determinants and interactions of three important dimensions of human capital: cognitive skills, noncognitive skills and health.
- It provides rare evidence in developing country settings on parental investments and child development
- It uses a latent factor approach to identify the latent, unobserved factors instead of (noisy) proxy variables to correct measurement error problems and explore the endogeneity of investments.

## Model: Dynamics of Skill Formation

- The framework builds on the dynamic factor models of Cunha et al. (2010), Cunha and Heckman (2007), Attanasio et al. (2017), Agostinelli and Wiswall (2016) and Aucejo and James (2021):

$$\Theta_{i,t}^k = f(\Theta_{i,t-1}^C, \Theta_{i,t-1}^{NC}, \Theta_{i,t-1}^H, l_{i,t}, P_i^C, P_i^{NC}, X_{i,t}, A_t^k, v_{i,t}, \varepsilon_{i,t}^k) \quad (1)$$

$$k \in \{C, NC, H\}$$

- I use a Constant Elasticity of Substitution (CES) production function

$$\Theta_{i,t}^k = [\gamma_{1t}^k (\Theta_{i,t-1}^C)^{\rho^{tk}} + \gamma_{2t}^k (\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^k (\Theta_{i,t-1}^H)^{\rho^{tk}} + \gamma_{4t}^k (l_{i,t})^{\rho^{tk}} + \gamma_{5t}^k (P_i^C)^{\rho^{tk}} + \gamma_{6t}^k (P_i^{NC})^{\rho^{tk}}]^{1/\rho^{tk}} e^{X'_{i,t} \delta_t^k + A_t^k + \mu^k v_{i,t} + \varepsilon_{i,t}^k} \quad (2)$$

Where  $\gamma_{1t}^k + \gamma_{2t}^k + \gamma_{3t}^k + \gamma_{4t}^k + \gamma_{5t}^k + \gamma_{6t}^k = 1$ .

# Measurement System

- The latent factors ( $\Theta_{i,t}^C$ ,  $\Theta_{i,t}^{NC}$ ,  $\Theta_{i,t}^H$ ,  $\Theta_{i,t-1}^C$ ,  $\Theta_{i,t-1}^{NC}$ ,  $\Theta_{i,t-1}^H$ ,  $P_i^C$ ,  $P_i^{NC}$ , and  $I_{i,t}$ ) are not directly measured, the factor model approach is used to extract these unobserved variables from a large set of observed data.
- Define the natural log of the factors as  $\theta_{i,t}^k = \ln(\Theta_{i,t}^k)$ ,  $\theta_{i,t-1}^k = \ln(\Theta_{i,t-1}^k)$ ,  $\mathcal{P}_i^C = \ln(P_i^C)$ ,  $\mathcal{P}_i^{NC} = \ln(P_i^{NC})$  and  $\mathcal{I}_{i,t} = \ln(I_{i,t})$
- The observed measures proxy the natural log of the factors.

## Measurement System

- There are two types of the observed measures, continuous and binary measures.
- The continuous measures are described by:

$$T_{i,j,\tau}^k = \alpha_{j,\tau}^k + \beta_{j,\tau}^k \theta_{i,\tau}^k + u_{i,j,\tau}^k \quad (3)$$

- The binary measures are described by:

$$T_{i,j,\tau}^k = \mathbb{1}[\alpha_{j,\tau}^k + \beta_{j,\tau}^k \theta_{i,\tau}^k + u_{i,j,\tau}^k > 0] \quad (4)$$

where:

- $T_{i,j,\tau}^k$ : the  $j$ th measure relating to latent factor  $k$  for individual  $i$  at time  $\tau$ ,  $\tau \in \{t-1, t\}$ .
- $\theta_{i,\tau}^k$ : unobserved/latent factors:  $\theta_{i,t}^k$ ,  $\theta_{i,t-1}^k$   $\mathcal{P}_i^C$ ,  $\mathcal{P}_i^{NC}$  and  $\mathcal{I}_{i,t}$
- $u_{i,h,t}$ : error terms.

# Parental Investments

$$\begin{aligned} \ln l_{i,t} = & \alpha_{1,t} + \alpha_{2,t} \ln \Theta_{i,t-1}^C + \alpha_{3,t} \ln \Theta_{i,t-1}^{NC} + C \alpha_{4,t} \ln \Theta_{i,t-1}^H \\ & + \alpha_{5,t} \ln P_i^C + \alpha_{6,t} \ln P_i^{NC} + \alpha_{7,t} X_{i,t} + \alpha_{8,t} Z_{i,t} + v_{i,t} \end{aligned} \quad (5)$$

where:

- $X_{i,t}$ : child gender, urban/rural residence and the number of siblings.
- $Z_{i,t}$ : instrumental variables including the log of wealth index reflecting parental resources, household economic shocks and the log of regional prices.
- $v_{i,t}$ : an error term.

# Estimation

- Measurement System:

$$T_{i,j,\tau}^k = \alpha_{j,\tau}^k + \beta_{j,\tau}^k \theta_{i,t-1}^k + u_{i,j,\tau}^k$$

$$T_{i,j,\tau}^k = \mathbb{1}[\alpha_{j,\tau}^k + \beta_{j,\tau}^k \theta_{i,t-1}^k + u_{i,j,\tau}^k > 0]$$
(6)

- Parental Investments:

$$\ln l_{i,t} = \alpha_{1,t} + \alpha_{2,t} \ln \Theta_{i,t-1}^C + \alpha_{3,t} \ln \Theta_{i,t-1}^{NC} + C \alpha_{4,t} \ln \Theta_{i,t-1}^H$$

$$+ \alpha_{5,t} \ln P_i^C + \alpha_{6,t} \ln P_i^{NC} + \alpha_{7,t} X_{i,t} + \alpha_{8,t} Z_{i,t} + v_{i,t}$$
(7)

- Production functions:

$$\Theta_{i,t}^k = [\gamma_{1t}^k (\Theta_{i,t-1}^C)^{\rho^{tk}} + \gamma_{2t}^k (\Theta_{i,t-1}^{NC})^{\rho^{tk}} + \gamma_{3t}^k (\Theta_{i,t-1}^H)^{\rho^{tk}} + \gamma_{4t}^k (l_{i,t})^{\rho^{tk}}$$

$$+ \gamma_{5t}^k (P_i^C)^{\rho^{tk}} + \gamma_{6t}^k (P_i^{NC})^{\rho^{tk}}]^{1/\rho^{tk}} e^{X'_{i,t} \delta_t^k + A_t^k + \mu^k v_{i,t} + \varepsilon_{i,t}^k}$$
(8)

# Estimation

- The distributions of the log factors,  $f(\theta)$ ,  $f(\mathcal{P})$ ,  $f(\mathcal{I})$ , are assumed to be distributed as a mixture of two normals.
- Model estimation consists of two steps.
  - First step: estimate the measurement system to recover the parameters  $\beta_{j,\tau}^C$ ,  $\beta_{j,\tau}^{NC}$ ,  $\beta_{j,\tau}^H$ ,  $\beta_j^{PC}$ ,  $\beta_j^{PNC}$ ,  $\beta_{j,t}^I$ ,  $\alpha_j^C$ ,  $\alpha_k^{NC}$ ,  $\alpha_j^{PC}$ ,  $\alpha_j^{PNC}$ ,  $\alpha_{j,t}^I$  and the latent factor distributions
  - Second step: use the estimated parameters from the first step to take individual-specific draws and use these draws as observable data to estimate investment and production functions.

## Estimation

- The likelihood of all the observed measures conditional on  $\theta_i$ :

$$L(T_i|\theta_i) = \prod_{k=1}^K \prod_{j=1}^J \prod_{\tau=1}^T f(T_{i,j,\tau}^k|\theta_i) \quad (9)$$

Where  $\theta_i = [\theta_{i,\tau}^k]'$  =  $[\theta_{i,\tau}^C, \theta_{i,\tau}^{NC}, \theta_{i,\tau}^H, \mathcal{P}_i^C, \mathcal{P}_i^{NC}, \mathcal{I}_{i,\tau}]'$

- The log-likelihood function:

$$\mathcal{L}(\Psi) = \sum_{i=1}^N \ln L(T_i) = \sum_{i=1}^N \ln \left( \int L(T_i|\theta_i) dF(\theta) \right) = \sum_{i=1}^N \ln \left( \int L(T_i|\theta_i) f(\theta) d\theta \right) \quad (10)$$

Where:

- $f(\theta) = \sum_{c=1}^2 \tau_c f(\theta|\mu_c, \Omega_c)$ , where  $\mu_c$ ,  $\Omega_c$  and  $\tau_c$  are the mean, covariance and the mixture probability of the two normals.
- $\Psi$  is all the parameters of the model,  $\Psi = \{\alpha, \beta, \sigma, \tau_c, \mu_c, \Omega_c\}$ .

# Estimation

- The log-likelihood function is estimated using the minorization-maximization algorithm.
- Given the unobservable nature of the factors, the log-likelihood function is constructed by integrating over the distributions of the unobservable factors.

## Data and variables

- Data is from Round 2 (at age 12) and Round 3 (at age 15) of the Young Lives survey for the Older cohort in Vietnam.

Latent factors		Observed variables
Child's cognitive skills - Round 2	$\theta_2^C$	<ol style="list-style-type: none"> <li>1. PPVT test</li> <li>2. Math Test</li> <li>3. Reading level</li> <li>4. Writing level</li> </ol>
Child's cognitive skills - Round 3	$\theta_3^C$	<ol style="list-style-type: none"> <li>1. PPVT test</li> <li>2. Math Test</li> <li>3. Cloze test</li> </ol>
Child's noncognitive skills - Round 2 and Round 3	$\theta_2^{NC},$ $\theta_3^{NC}$	<ol style="list-style-type: none"> <li>1. Self-esteem score</li> <li>2. Self-efficacy score</li> <li>3. Self-respect and inclusion score</li> </ol>
Child's health - Round 2 and Round 3	$\theta_2^H, \theta_3^H$	<ol style="list-style-type: none"> <li>1. Child height for age z-score</li> <li>2. Child weight</li> <li>3. How is child health?</li> </ol>

# Data and variables

Latent factors	$\mathcal{P}^C$	Observed variables
Parental cognitive skills	$\mathcal{P}^C$	<ol style="list-style-type: none"> <li>1. Mother's years of education</li> <li>2. Father's years of education</li> </ol>
Parental noncognitive skills	$\mathcal{P}^{NC}$	<ol style="list-style-type: none"> <li>1. Self-esteem score</li> <li>2. Self-efficacy score</li> <li>3. self-respect and inclusion score</li> </ol>
Parental Investments	$\mathcal{I}_3$	<ol style="list-style-type: none"> <li>1. Expenditure on the Young Lives child</li> <li>2. Number of hours studying outside school as a proxy for the time that parents dedicate to the child</li> <li>3. Quality of relationship between child and parents</li> </ol>

# Estimates: Determinants of Parental Investments

Variables	Parental investments
Child's cognitive skills, age 12	-0.024 (0.030)
Child's noncognitive skills, age 12	-0.102 (0.076)
Child's health, age 12	0.218*** (0.032)
Parental cognitive skills	0.402*** (0.057)
Parental noncognitive skills	0.172*** (0.066)
Wealth index	0.177*** (0.045)
<i>Observations</i>	<i>961</i>

*Note:* Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

- Selected covariates are reported.

# Estimates of Production Functions

	Cognitive skills, age 15	Noncognitive skills, age 15	Health, age 15
Child's cognitive skills at age 12	0.613*** (0.044)	0.001 (0.039)	0.064** (0.029)
Child's noncognitive skills at age 12	0.145*** (0.055)	0.339*** (0.077)	0.130** (0.053)
Child's health at age 12	0.042* (0.025)	0.065 (0.053)	0.670*** (0.073)
Parental Investments	0.261*** (0.075)	0.505*** (0.167)	0.346** (0.170)
Parental cognitive skills	0.017 (0.041)	-0.005 (0.122)	-0.184 (0.193)
Parental noncognitive skills	-0.079 (0.050)	0.096 (0.081)	-0.026 (0.065)
$A_t$	-0.010 (0.024)	0.211*** (0.031)	0.205*** (0.024)
Control Function (investment residuals)	-0.126* (0.068)	-0.132 (0.262)	-0.420** (0.179)
Complementarity( $\rho$ )	-0.168 (0.117)	-1.851** (0.757)	0.034 (0.327)
Elasticity of substitution	0.856*** (0.141)	0.351* (0.211)	1.036*** (0.195)
<i>Observations</i>		961	

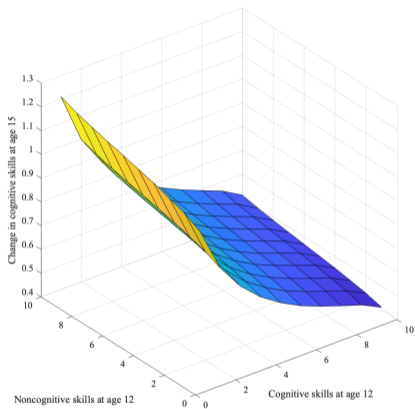
Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

# Estimates of Production Functions - Marginal Effects

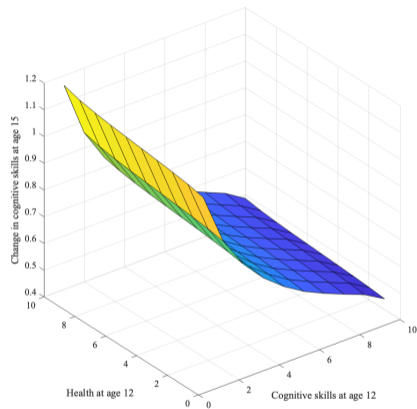
	Cognitive skills, age 15	Noncognitive skills, age 15	Health, age 15
Child's cognitive skills at age 12	0.614*** (0.043)	0.002 (0.039)	0.064** (0.029)
Child's noncognitive skills at age 12	0.145*** (0.055)	0.316*** (0.076)	0.130** (0.052)
Child's health at age 12	0.042* (0.025)	0.085 (0.062)	0.670*** (0.065)
Parental investments	0.261*** (0.075)	0.513*** (0.162)	0.346** (0.169)
Parental cognitive skills	0.017 (0.042)	-0.008 (0.123)	-0.184 (0.194)
Parental noncognitive skills	-0.079 (0.050)	0.093 (0.078)	-0.026 (0.067)
<i>Observations</i>		961	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

# Cognitive Skills: Self-productivity ( $\partial\Theta_t^C/\partial\Theta_{t-1}^C$ )

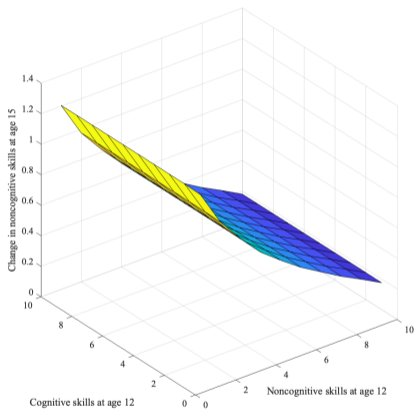


(a)

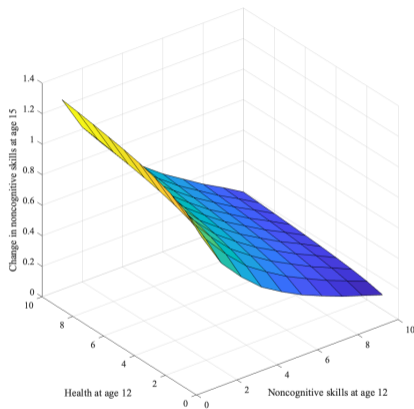


(b)

# Noncognitive Skills: Self-productivity ( $\partial\Theta_t^{NC} / \partial\Theta_{t-1}^{NC}$ )

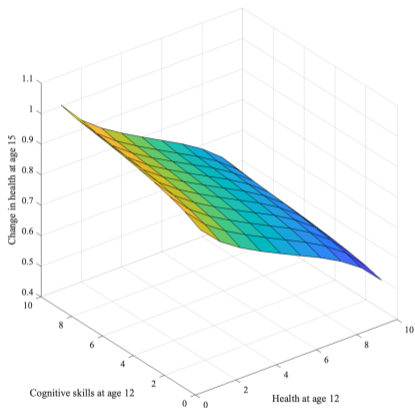


(a)

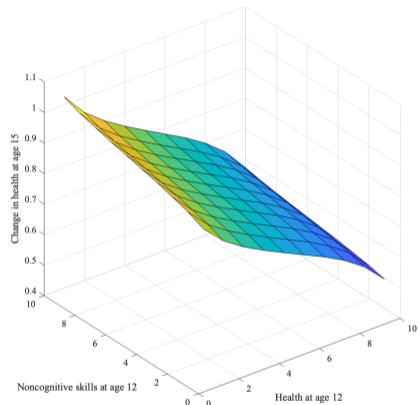


(b)

# Health: Self-productivity ( $\partial\Theta_t^H/\partial\Theta_{t-1}^H$ )



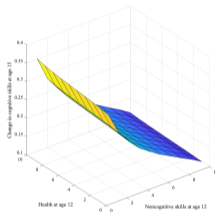
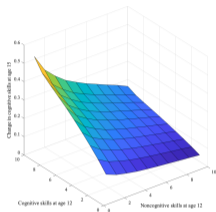
(a)



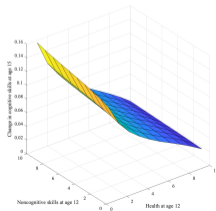
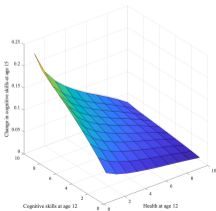
(b)

# Cognitive skill: Cross-productivity

- Cross-productivity from Noncognitive Skills to Cognitive Skills ( $\partial\theta_t^C/\partial\theta_{t-1}^{NC}$ )

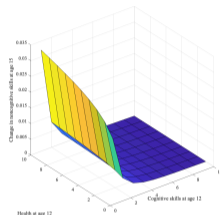
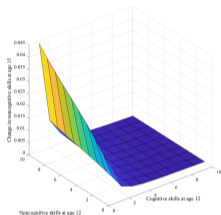


- Cognitive Skills: Cross-productivity from Health to Cognitive Skills ( $\partial\theta_t^C/\partial\theta_{t-1}^H$ )

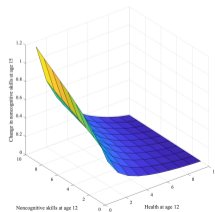
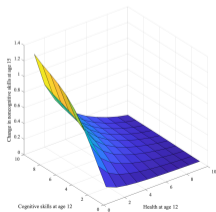


# Noncognitive skill: Cross-productivity

- Cross-productivity from Cognitive Skills to Noncognitive Skills ( $\partial\Theta_t^{NC}/\partial\Theta_{t-1}^C$ )

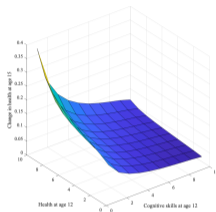
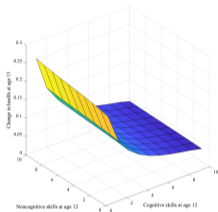


- Cross-productivity from Health to Noncognitive Skills ( $\partial\Theta_t^{NC}/\partial\Theta_{t-1}^H$ )

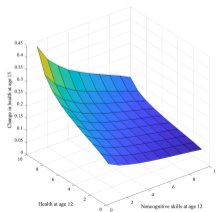
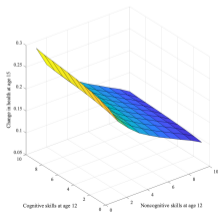


# Health: Cross-productivity

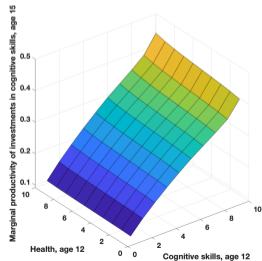
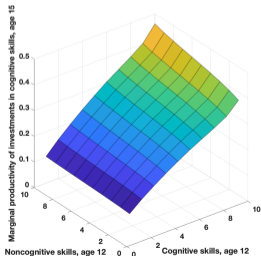
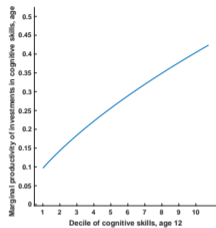
- Cross-productivity from Cognitive Skills to Health ( $\partial\Theta_t^H/\partial\Theta_{t-1}^C$ )



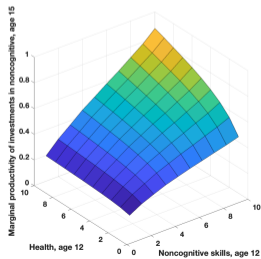
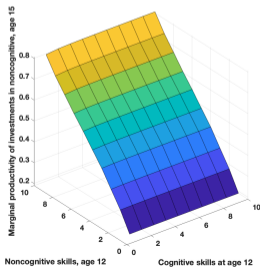
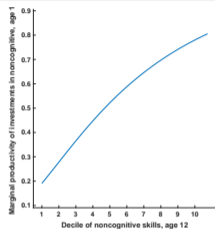
- Cross-productivity from Noncognitive Skills to Health ( $\partial\Theta_t^H/\partial\Theta_{t-1}^{NC}$ )



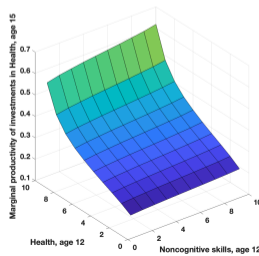
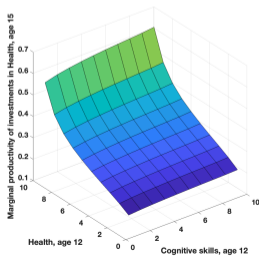
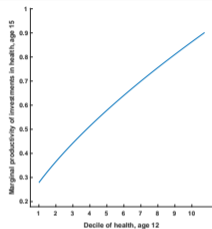
# Dynamic Complementarity between Investments and Cognitive Skills



# Dynamic Complementarity between Investments and Noncognitive Skills



# Dynamic Complementarity between Investments and Health



## Conclusions

- Self-effects are present and strong in the production of all human capital dimensions.
- There is the existence of cross-productivity from noncognitive skills and health to cognitive skills and from cognitive and noncognitive skills to health.
- Parental investments strongly and directly affect the accumulation of skills and health
- There are strong dynamic complementarities between parental investments and three dimensions of human capital.

## Conclusions

- The self-effects, cross-effects and dynamic complementarities together become a dynamic multiplier effect mechanism of human capital accumulation.
- Parental skills and resources are positively associated with parental investments.
- These effects could lead to substantially different growth rate of human capital and substantial increases in inequality.
- These results indicate the importance of interventions by boosting investments at early ages that can alter child development path, especially for disadvantaged children.
- Lack of parental investments can seriously hinder the development of a child.

Thank you!