Factor Models

B. Nonlinear: based on Cunha, Heckman and Schennach, Econometrica, 2010 (CHS) and Attanasio, Meghir and Nix, 2017 NBER WP

Costas Meghir

February 27, 2018

The setup

- We will consider the estimation of a production function of the form

$$\theta^c_{it+1} = [\gamma^c_1(\theta^c_{it})^\rho + \gamma^c_2(\theta^N_{it})^\rho + \gamma^c_3(\theta^I_{it})^\rho]^{\frac{1}{\rho}} + \eta^c_{it}$$

  where the elasticity of substitution is $\sigma = \frac{1}{1-\rho}$, $\rho < 1$
- We will start with the assumption that $E(\eta^c_{it}|\theta_i) = 0$
- We will take the case where there are at least three measurements per factor
- We will start with all measurements being continuous

The measurement system

- As before we assume that we possess a set of measurements (demeaned) such that

$$m_{iks}^{j} = a_{ks}^{j}\theta_{iks}^{j} + \varepsilon_{iks}^{j}$$

- With enough measurements it is possible to allow for some measurements to reflect more than one factor but we leave this for now.
- Assuming measurement errors are independent of the factors and of each other we can estimate the factor loadings, subject to the suitable normalization restrictions
- Rescaling the measurements m_{iks}^{j}/a_{ks}^{j} we are now ready to apply the Kotlarski theorem
- As we know the theorem states that we can identify the joint distribution of the θ and the ε nonparametrically

Implementing the approach

The following material has been developed with Emily Nix

- I now describe a flexible parametric approach making the problem simpler
- I assume that the measurement errors are $\varepsilon \sim N(0,\Omega)$, with Ω diagonal
- Now the next step is to assume a flexible parametric form for the distribution of the factors $g(\theta)$
- If we assume normality we will be imposing linearity ($\rho = 1$)
- We thus assume

$$g(\theta) = \lambda f_1(\theta|\mu_1,\Sigma_1) + (1-\lambda) f_2(\theta|\mu_2,\Sigma_2) \quad 0 \leq \lambda \leq 0.5$$

  where $f_k(\mu_k,\Sigma_k)$ is the normal distribution with mean μ_k and covariance matrix Σ_k
- Note the identification (and inference problem) when $\lambda = 0$.

Normalizations

- Usually it is irrelevant whether we set the mean of the log factors to zero or not
- However, when there are dynamics over more than one period this may cause bias in the substitution elasticity
- Moreover it is interesting to follow the developmental growth of children
- If the same test is being administered every time this is an easy problem to solve
  1. Assume that the mean of the measurement remains invariant across age
  2. Then any growth of the measurement is due to the growth of the factor
  3. This identifies the mean at each age
- Also observing one of the measures at all ages, allows us to keep the scale the same from period to period

Implementing the approach

- With this assumption it means that the measurements also follow a mixture of normals
- Each component of the mixture is

$$p_k(m) = \int f_\varepsilon(\varepsilon) f_k(\theta|\mu_k, \Sigma_k) d\theta \quad k = 1, 2$$

where $\varepsilon_i = m_i - a'\theta_i$ and

$$p(m) = \lambda p_1(m) + (1-\lambda) p_2(m)$$

- In what follows we write the measurement equations in stacked form (having removed means)

$$m_i = A'\theta_i + \varepsilon_i$$

with A being $n_m \times n_f$ and n_m is the total number of measurements and n_f is the number of factors

Implementing the approach

- Specifically each component is

$$p_j(m) = \frac{1}{(2\pi)^{k/2}|A'\Sigma_j A+\Omega|^{1/2}} exp[-0.5(m_i - A'\mu_j)'(A'\Sigma_j A+\Omega)^{-1}(m_i - A'\mu_j)$$

- We can thus obtain a likelihood function for the observables

$$logL_i = log[\lambda L_i^1 + (1-\lambda)L_i^2]$$

with the sample likelihood being $logL = \sum_{i=1}^{N} logL_i$

Comments

- This is a typical mixtures likelihood function exactly like the one we find when we integrate unobserved heterogeneity from economic models
- We can generalize this for greater flexibility by adding elements to the mixture
- Indeed we know we can do this because the distribution of the factors is nonparametrically identified
- So we can view this process as a nonparametric procedure
- We still have a very restrictive process for measurement error

Implementing the approach: The EM algorithm

- An effective way of maximizing this likelihood is to use the EM algorithm. The key references are
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, B, 39, 1–38.
- Peter Arcidiacono and John Bailey Jones (2003) Finite Mixture Distributions, Sequential Likelihood and the EM algorithm, Econometrica, Vol. 71, No. 3 (May, 2003), 933–946
- The key idea is to use Bayes theorem to create a sequential algorithm. The notes below draw from Arcidiacono and Jones

The conditional probability of a type

- Think of λ_k as being the probability of being a particular type k.
- Consider first the following relationship:

$$\lambda_k p_k(m_i|\beta) = pr(k|m_i,\beta)f(m_i|\beta)$$

  where $f(m_i|\beta) \equiv L_i$ is the marginal distribution of the measurements and where $pr(k|m_i)$ is the probability of type being k given we observe data m_i and β are the unknown parameters (either type specific or common) .

- Hence

$$pr(k|m_i, \beta) = \frac{\lambda_k p_k(m_i|\beta_k)}{f(m_i|\beta)}$$

The first order conditions

- Now consider the first order conditions for a generalized version of our loglikelihood: $\Sigma_{i=1}^{N} logL_i = log\left[\Sigma_{k=1}^{K} \lambda_k p_k(m_i|\beta)\right]$

$$\frac{\partial loL_i}{\partial \beta} = \Sigma_{i=1}^{N}\left[\Sigma_k \frac{\lambda_k p_k}{L_i p_k}\frac{\partial p_k(m_i|\beta)}{\partial \beta}\right] = \Sigma_{i=1}^{N}\left[\Sigma_k pr(k|m_i,\beta)\frac{\partial logp_k(m_i|\beta)}{\partial \beta}\right] = 0$$

In the above we have multiplied and divided by p_k to get the log derivative.

- Now consider the following sequential algorithm:

1. Initialize parameters
2. Compute $pr(k|m_i,\beta)$ for each observation (expectation step)
3. Solve first order conditions conditional on $pr(k|m_i,\beta)$.
4. Estimate $\hat{\lambda}_k = \frac{1}{N}\Sigma_{i=1}^{N} pr(k|m_i,\hat{\beta})$
5. Go to 2 and start again with the updated parameters - continue until convergence

Binary / Discrete measurements

- Suppose a subset of measurements are discrete.
- We then think of the measurement system as a latent itself. For example we think of the binary variable $y_i = \mathbf{1}(m_i > 0)$
- Then set up the likelihood in terms of the underlying continuous latent measures and then integrate over the relevant range.
- Hence the likelihood for observation i becomes

$$logL_i = log[\int_R (\lambda L_i^1 + (1-\lambda)L_i^2)\, dm^d]$$

  where m^d denotes those measurements that are discrete and over which we are integrating and R denotes the appropriate range for m. For example if for that observation the binary variable was 1 then the appropriate range of m is $m > 0$.
- Note that with many discrete measurements this can be come a high dimensional integral, which can be very slow to compute.
- Also note that we need to put the intercepts back in place - we cannot preestimate them.

Identifying the Production function

- Once we have estimated the joint distribution of factors we can now estimate the production function
- One approach may be to draw a large random sample of factors from the joint distribution and then use nonlinear least squares to fit the conditional mean
- We can then use the parametric bootstrap (drawing parameters from the asymptotic distribution of the factor model) to estimate the standard errors.

- The production function we are estimating is
  $\theta^c_{it+1} = [\gamma^c_1(\theta^c_{it})^\rho + \gamma^c_2(\theta^N_{it})^\rho + \gamma^c_3(\theta^I_{it})^\rho]^{\frac{1}{\rho}} + \eta^c_{it}$
- The issue of endogeneity arises at this point, when we need to define the conditional distribution of the output given the inputs
- More generally the error η may be nonseparable: for example
  $\theta^c_{it+1} = [\gamma^c_1(\theta^c_{it})^\rho + \gamma^c_2(\theta^N_{it})^\rho + \gamma^c_3(\theta^I_{it})^\rho + \eta^c_{it}]^{\frac{1}{\rho}}$

- Now suppose that $\eta_{it}^{c} = \pi + u_{it}^{c}$ with u being independent of all factors
- Suppose we have at least three adult outcomes m^{A}. We suppose they have measurement equations

$$m_{ik}^{A} = a_{k}^{A}\theta_{it+1}^{c} + \delta\pi + \varepsilon_{ilt}^{c}$$

- Then we need to identify an additional factor π. If there is a non-cognitive skill production function we would add the non-cognitive factor as well in the measurement equations.
- Effectively we need to find some measurement that reflects the omitted factor. In CHS the idea is that nothing in childhood reveals that but it "surfaces" directly in adult measurements.

Identification and Endogenous Investments

- The basic premise of the model is that all relevant heterogeneity is captured by the included factors
- It is thus important to make sure the specification is complete
- Beyond the issue of permanent omitted variables there is the question of temporal shocks
- Investments may respond to shocks:
  - A positive shock to child cognition may lead to more or less investment depending on complementarities and the structure of preferences
- Thus we also control for the possible correlation between TFP shocks and investment

Controlling for endogenous investments

- We view investments as being chosen as a function of child background and the economic environment
- Excluded instruments:
  - household wealth
  - Marital status
  - Male and female community level wages
- We also include all predetermined variables used in the production function.

Econometric approach to endogeneity

- We assume

$$ln\theta_{It} = \beta_c ln\theta_{c\,t} + \beta_h ln\theta_{h\,t} + \beta_{mc} ln\theta_{mc\,t} + \beta_{mh} ln\theta_{mh\,t} + \beta_I^{'} Z_t + v_t$$

where Z are the instruments

- The production function takes the form

$$ln\theta_{c\,t+1} = a + ln[\delta_c \theta_{c\,t}^{\rho} + \delta_h \theta_{h\,t}^{\rho} + \delta_{mc} \theta_{mc\,t}^{\rho} + \delta_{mh} \theta_{mh\,t}^{\rho} + \delta_I \theta_{It}^{\rho} + \delta_T \theta_{Tt}^{\rho}]^{\frac{1}{\rho}}$$

$$+\beta' x + \delta_1 v_I + \delta_2 v_T + u_t^{*}$$

- The measurements will now include the instruments as factors measured with no error.
- Once the investment equation is estimated we then include the residual in the production function as an extra regression
- This is a control function approach.

The Young Lives Survey

- The Young Lives Survey surveyed two cohorts of children over numerous periods of their childhood.
- It covers India, Vietnam, Peru and Ethiopia - We use data from India.
- There are two cohorts: One was first surveyed when the children were 1 and the other when they were 8.
- We use the young cohort. This is observed at ages 1, 5, 8 and 12

Descriptive Statistics

Household Characteristics	
Subject child is Male	0.54
Urban	0.24
Scheduled caste	0.18
Scheduled tribe	0.15
Hindu	0.88
Muslim	0.07
Number of children	1.89
	1.00
Number older siblings	0.69
	1.03
Household size	5.44
	2.36
Mother Characteristics	
Mother weight	46.39
	9.39
Mother years of school	3.62
	4.42
Mother's age	23.66
	4.35
Observations	1,910

Note: Standard deviations in italics.

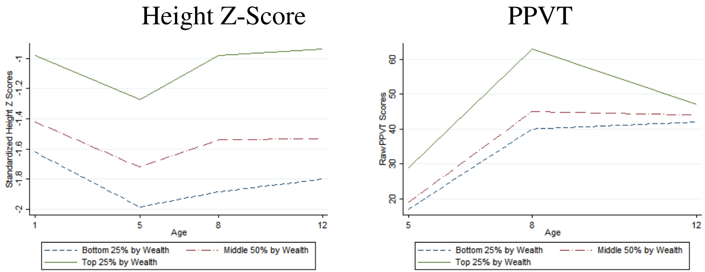
Descriptives Across Waves

	Age 1	Age 5	Age 8	Age 12
Child Characteristics				
Fraction stunted	0.31	0.36	0.30	0.29
Fraction underweight	0.32	0.45	0.46	
Fraction wasted		0.19	0.28	0.33
Height for age Z-score	-1.30	-1.66	-1.45	-1.45
	1.48	*0.99*	*1.04*	*1.03*
Raw score PPVT test		27.47	58.51	43.08
		21.10	*30.43*	*7.82*
Amount spent on books		3.48	8.98	13.00
		5.40	*13.02*	*16.97*
Household Economic Wellbeing				
Annual income		873.57	1407.98	1749.95
		1219.24	*2033.67*	*1841.78*
Wealth index	0.40	0.46	0.51	0.59
	0.20	*0.20*	*0.18*	*0.17*
Percent below $2/day		0.63	0.45	0.27
Child Work				
Daily hours chores		0.06	0.34	0.82
Daily hours family business		0.00	0.01	0.12
Daily hours paid work			0.01	0.05

Income and amount spent on books are annual amounts in the past 12 months in USD. At age 5, 1
age 8, 1USD≅49INR, and at age 12, 1USD≅62INR.

Wealth Gradients

Figure: Wealth Gradient in Height and in the Peabody Picture Vocabulary Test(PPVT)



- Year 12 PPVT is wrong in this graph

Model

$$\begin{aligned} \theta^{c}_{t+1} &= G\left(\theta^{c}_{t}, \theta^{h}_{t}, \theta^{I}_{t}, X_t\right) \\ \theta^{h}_{t+1} &= F\left(\theta^{c}_{t}, \theta^{h}_{t}, \theta^{I}_{t}, X_t\right) \end{aligned}$$

$$\theta_{ct+1} = \left[\delta_{ct}(\theta_{ct})^{\rho_t} + \delta_{ht}(\theta_{ht})^{\rho_t} + \delta_{cpt}(\theta_{cp})^{\rho_t} + \delta_{hpt}(\theta_{hp})^{\rho_t} + \delta_{It}(\theta_{It})^{\rho_t}\right]^{\frac{1}{\rho_t}} A_{ct}$$

$$\theta_{ht+1} = \left[\alpha_{ct}(\theta_{ct})^{\zeta_t} + \alpha_{ht}(\theta_{ht})^{\zeta_t} + \alpha_{cpt}(\theta_{cp})^{\zeta_t} + \alpha_{hpt}(\theta_{hp})^{\zeta_t} + \alpha_{It}(\theta_{It})^{\zeta_t}\right]^{\frac{1}{\zeta_t}} A_{ht}$$

where

$$\begin{aligned} A_{ct} &= exp\left(\delta_{0t} + \delta_{Xt}X_t + u_{ct}\right) \\ A_{ht} &= exp\left(\alpha_{0t} + \alpha_{Xt}X_t + u_{ht}\right) \end{aligned}$$

$$ln\theta_{It} = \gamma_0 + \gamma_{ct}\, ln\theta_{ct} + \gamma_{ht}\, ln\theta_{ht} + \gamma_{cpt}\, ln\theta_{cp} + \gamma_{hpt}\, ln\theta_{hp} + \gamma^{'}_{Xt}X_t + \gamma^{'}_{pt}\, lnp_{It} + \gamma_Y\, ln\theta_{Yt} + v_t \quad (1)$$

where v_t reflects random shocks, and θ_{Yt} represents parental resources, lnp_{It} represents log prices for child investment goods. All other variables are as defined in the production functions.

Controlling for the endogeneity of Investments

$$E\left(u_{ct} \mid Q_t, Z_t\right) = \kappa_c v_t$$

$$E\left(u_{ht} \mid Q_t, Z_t\right) = \kappa_h v_t$$

Signoal to noise ratio

$$s_j^{\ln \theta_{kt}} = \frac{(\lambda_{jkt})^2 \; Var(\ln \theta_{kt})}{(\lambda_{jkt})^2 \; Var(\ln \theta_{kt}) + Var(\varepsilon_{jkt})}$$

The information content of measures

	Age 1	Age 5	Age 8	Age 12
	Child Cognition			
PPVT		51%	33%	39%
Math			74%	68%
English				64%
Language				50%
EGRA (rasch)			52%	
CDA (rasch)		36%		
	Child Health			
Height Z-Score	55%	72%	60%	60%
Weight Z-Score	81%	73%	77%	
Weight in kg				64%
Health Status	8%	1%	5%	
	Investments			
Books		22%	21%	30%
Clothing		38%	31%	44%
Shoes		43%	38%	30%
Uniform		11%	15%	19%
Meals/day		2%	5%	2%
Food groups/day		7%	6%	1%
	Resources			
Income		69%	84%	82%
Wealth		38%	49%	49%
	Parental Cognition (fixed over age)			
Mother's education	79%			
Father's education	55%			
Literacy	40%			
	Parental Health (fixed over age)			
Mother's weight	73%			
Mother's height	11%			

PPVT: Peabody Picture Vocabulary Test, EGRA: Reading comprehension test, CDA: Cognitive Development Assessment. Books, clothing, shoes and uniform measured in monetary units.

Table: The Coefficients of the Investment Equations

	Age 5	Age 8	Age 12
	Child human capital		
Cognition		0.113 [0.03,0.17]	0.09 [−0.01,0.13]
Health	0.095 [0.05,0.13]	0.012 [−0.01,0.07]	0.051 [0,0.13]
Gender	−0.013 [−0.1,0.04]	−0.026 [−0.12,0.05]	0.056 [−0.04,0.16]
	Parental human capital		
Parental Cognition	0.01 [−0.06,0.07]	0.004 [−0.02,0.1]	−0.02 [−0.05,0.04]
Parental Health	−0.013 [−0.05,0.04]	−0.01 [−0.06,0.03]	−0.032 [−0.11,0.01]

Table: The Coefficients of the Investment Equations

	Age 5	Age 8	Age 12
		Prices	
Price Clothes	−0.063	0.031	0.099
	[−0.15,0.02]	[−0.12,0.11]	[−0.02,0.23]
Price Notebook	−0.383	−0.196	−0.231
	[−0.53,−0.23]	[−0.37,−0.05]	[−0.34,−0.13]
Price Mebendazol	0.047	−0.156	0.011
	[0.01,0.1]	[−0.26,−0.11]	[−0.04,0.05]
Price Food	−0.082	−0.328	−0.256
	[−0.28,0.2]	[−0.66,0.04]	[−0.47,−0.09]
		Household Characteristics	
Resources	0.457	0.644	0.587
	[0.3,0.59]	[0.42,0.75]	[0.41,0.68]
Older Siblings	0.039	0.058	−0.032
	[−0.02,0.1]	[−0.01,0.11]	[−0.09,0.03]
Number of Children	−0.096	−0.041	−0.049
	[−0.14,−0.04]	[−0.09,0.01]	[−0.11,0.01]
Urban	0.349	0.103	0.066
	[0.2,0.54]	[0,0.28]	[−0.09,0.23]
Hindu	−0.005	−0.01	0
	[−0.02,0.01]	[−0.02,0.01]	[−0.01,0.01]
Muslim	−0.105	−0.247	−0.02
	[−0.33,0.07]	[−0.4,−0.05]	[−0.27,0.15]
Mother's Age	0.014	−0.141	0.016
	[−0.12,0.14]	[−0.26,0.08]	[−0.1,0.2]
Scheduled Caste	−0.074	−0.066	−0.269
	[−0.17,0.05]	[−0.21,0.1]	[−0.42,−0.11]
Scheduled Tribe	0.073	−0.106	−0.254
	[−0.05,0.2]	[−0.23,0.05]	[−0.4,−0.1]
BC Caste	−0.034	0.061	−0.169
	[−0.15,0.08]	[−0.06,0.22]	[−0.3,−0.05]
Prices and Income (P-values)	0	0	0
Prices (P-values)	0	.005	.001

Production of Cognitive Skills and Health

	Cognition			Health		
Age	5	8	12	5	8	12
			Lagged Skills			
Cognition		0.29	0.6		−0.02	−0.03
		[0.22,0.43]	[0.53,0.67]		[−0.07,0.01]	[−0.06,0.03]
Health	0.18	0.15	0.02	0.69	0.82	0.92
	[0.11,0.25]	[0.1,0.19]	[−0.01,0.07]	[0.64,0.75]	[0.76,0.87]	[0.85,0.98]
			Investment and Parental Skills			
Investment	0.47	0.65	0.19	0.1	0.12	0.04
	[0.31,0.56]	[0.47,0.75]	[0.07,0.29]	[0.01,0.2]	[0.06,0.21]	[−0.06,0.1]
Parent Cog	0.32	−0.01	0.17	0.01	0.03	0.04
	[0.25,0.39]	[−0.1,0.06]	[0.13,0.21]	[−0.05,0.06]	[0,0.07]	[−0.01,0.06]
Parent Health	0.03	−0.09	0.03	0.2	0.05	0.04
	[−0.02,0.1]	[−0.14,−0.02]	[0,0.07]	[0.15,0.28]	[0.02,0.08]	[0.02,0.09]

Production of Cognitive Skills and Health

	Cognition			Health		
Age	5	8	12	5	8	12
			Demographic Characteristics			
Num Child	0 [−0.02,0.01]	−0.01 [−0.03,0.02]	−0.03 [−0.05,−0.01]	0.01 [−0.01,0.02]	0 [−0.02,0]	0 [−0.01,0.01]
Older Sibs	0.01 [−0.01,0.03]	−0.01 [−0.03,0]	0 [−0.01,0.02]	−0.03 [−0.05,−0.01]	0 [−0.01,0.01]	0.01 [0,0.02]
Gender	0.01 [−0.01,0.02]	0.04 [0.02,0.05]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.01 [−0.01,0.01]	0.02 [0.01,0.02]
Urban	−0.01 [−0.01,0]	−0.03 [−0.03,−0.01]	−0.01 [−0.02,0]	0 [−0.01,0]	0.01 [0,0.01]	0 [−0.01,0]
Hindu	−0.01 [−0.03,0]	−0.01 [−0.03,0.01]	0.03 [0.02,0.05]	0.01 [−0.01,0.02]	0 [−0.01,0]	0 [−0.01,0.01]
Muslim	0 [0,0]	0 [0,0]	−0.01 [−0.01,0]	0 [0,0]	0 [0,0]	0 [0,0]
Mother Age	0.01 [0,0.03]	0.01 [0,0.03]	−0.01 [−0.02,0.01]	0 [−0.01,0.02]	0 [−0.01,0.01]	−0.02 [−0.02,0]
Sched Caste	0 [−0.02,0]	0.02 [0.01,0.03]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0]	0 [−0.01,0.01]
Sched Tribe	0.06 [0.04,0.08]	−0.01 [−0.02,−0.01]	−0.01 [−0.01,0]	0.01 [0.01,0.02]	−0.01 [−0.02,−0.01]	0 [0,0.01]
BC Caste	−0.02 [−0.04,−0.01]	0 [−0.01,0.02]	0 [−0.01,0.01]	−0.02 [−0.03,−0.01]	0.01 [0,0.02]	0 [−0.01,0]
			Production function structure and test of exogeneity for investment			
(ρ, ζ)	−0.11 [−0.37,−0.01]	−0.06 [−0.19,0.07]	0.28 [0,0.35]	−0.03 [−0.22,0.03]	0.23 [0.05,0.36]	−0.2 [−0.2,0.18]
Subst. Elast	0.9 [0.73,0.99]	0.95 [0.84,1.07]	1.39 [1,1.54]	0.97 [0.82,1.04]	1.31 [1.05,1.56]	0.83 [0.83,1.23]
Log TFP	−0.03 [−0.09,0.04]	0.03 [−0.06,0.05]	0.03 [−0.03,0.07]	0.03 [0,0.08]	−0.02 [−0.04,0.03]	0 [−0.03,0.02]
Inv. Res	−0.39 [−0.54,−0.12]	−0.77 [−0.9,−0.53]	−0.21 [−0.31,−0.05]	−0.08 [−0.24,0.08]	−0.06 [−0.17,0.06]	0.01 [−0.1,0.1]

Marginal Produt of Investment

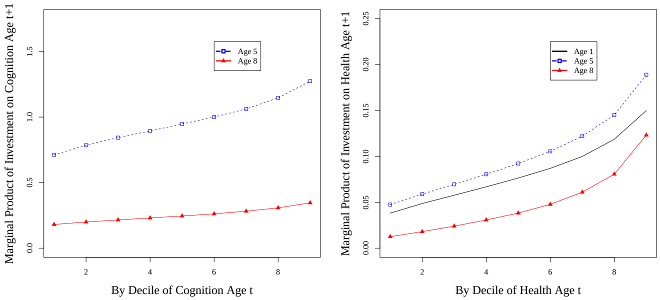


Figure: Marginal Product of Investment on Health and Cognition

Note: The y-axis represents the impact on the outcome in question, in standard deviation units, of increasing cognition or health by one standard deviation.