Life's two lotteries: modeling the effects of genes and environments in human capital formation'

Marina Aguiar Palma,[1] Sjoerd van Alten, [2] Titus Galama, [3] Maarten Lindeboom, [4] and Soraya Roman [5]

[1]Vrije Universiteit Amsterdam and FGV-Rio de Janeiro m.aguiar.palma@vu.nl
[2]Vrije Universiteit Amsterdam
[3]University of Southern California and Vrije Universiteit Amsterdam
[4]Vrije Universiteit Amsterdam
[5]FGV-Rio de Janeiro

Motivation

Evidence that observed differences in human-capital outcomes (health, cognition, socio-emotional abilites etc) between socio-economic groups start early and grow over the lifetime of individuals

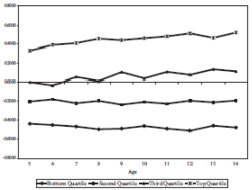


Figure: Math test scores and family income quintiles. Heckman (2007)

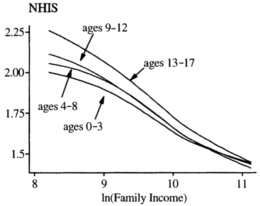


Figure: Reported Health and family income for children and adults. Case et al. (2002)

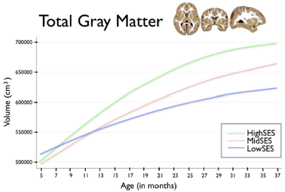


Figure: Gray matter and SES status by age. Hanson et al. (2013)

Motivation

How to explain this stylised facts.

- Existence of sensitive periods where children are especially sensitive to stimuli (negative or positive)
- Children of low-SES background being more sensitive to shocks and experiencing negative shocks more often.
- Skills beget skills (self- and cross-productivity)
- Current skills raise the productivity of investments in later skills (dynamic complementarity)

Further, skills and investments are latent traits that we do not fully observe

Literature and Contribution

An influential literature in Economics has a class of models of childhood development using a dynamic-factor models of child cognition, socio-emotional abilities, and health.

Cunha, Heckman, & Schennach, 2010; Attanasio, Meghir, Nix, & Salvati,2017; Attanasio, Meghir, & Nix, 2020; Del Bono, Kinsler, & Pavan, 2022; Agostinelli & Winswall (2023)

- The literature has recognised the importance of genetic endowments but has, for lack of better data, used birth outcome variables as proxies.
- One exception (Ronda et al forthcoming)

Our contribution:

- Add genetic endowments to the model of production function od child health and socio-emotional abilities
- Genetic endowments are measured by multiple PGIs

Data

The Dutch Lifelines cohort (N=167,000) of the population of the Northern Netherlands between 2007 and 2014. 63,836 participants were genotyped as part of the UGLI sample.

- Restrict the sample to children with observed socio-emotional (CBCL), and health (antropometrics) measures around age 8 (N= 1,139)
- For a subsample that is genotyped we calculated the EA PGI and the birthweight PGI (N= 434)

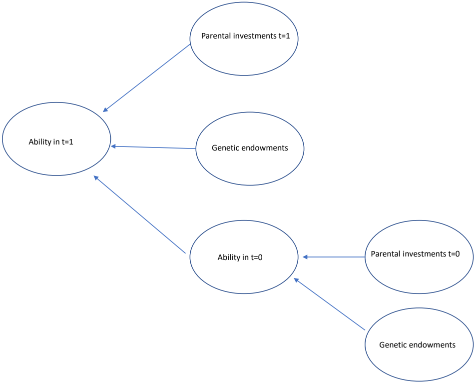
Model

Log-linear production function of child skills:

$$\ln \theta^s_{t_1} = \gamma^s_0 + \gamma^s_1 \ln i_{t_1} + \gamma^s_2 \ln \theta_{t_0} + \gamma^s_3 \ln \theta_{PH} + \gamma^s_4 BWPGI + \gamma^s_5 EAPGI + \boldsymbol{\gamma_6 X_{t_1}} + u^s_{t_1} \quad (1)$$

- $s = [h, se]$: child's health, external socio-emotional socio-emotional skills.
- i_{t_1}: parental investments.
- θ_{t_0}: child's initial skills
- $\boldsymbol{X}_{t_1}$ Controls

The model in a simple picture



System of measurements to latent traits

$$
\begin{aligned}
M^{\boldsymbol{\Theta}_{t_1}} &= \beta^{\boldsymbol{\Theta}_{t_1}} + \lambda^{\boldsymbol{\Theta}_{t_1}} \ln(\boldsymbol{\Theta}_{t_1}) + \epsilon_{\boldsymbol{\Theta}_{t_1}} \\
M^{\theta_{t_0}} &= \beta^{\theta_{t_0}} + \lambda^{\theta_{t_0}} \ln(\theta_{t_0}) + \epsilon_{\theta_{t_0}} \\
M^{i_{t_1}} &= \beta^{i_{t_1}} + \lambda^{i_{t_1}} \ln(i_{t_1}) + \epsilon_{i_{t_1}} \\
M^{\boldsymbol{\Omega}} &= \beta^{\boldsymbol{\Omega}} + \lambda^{\boldsymbol{\Omega}} \ln(\boldsymbol{\Omega}) + \epsilon_{\boldsymbol{\Omega}} \\
M^{\mathbf{X}_{t_1}} &= \mathbf{X}_{t_1} \\
M^{\mathbf{PGI}_{t_1}} &= \mathbf{PGI}_{t_1}
\end{aligned}
$$

For identification, we assume: errors are orthogonal to latent variables, factor loading of the first measurement of each latent variable equal to one. Age invariant measures for latent traits measured repeatedly.

$$\epsilon \sim N(0, \Sigma^{\epsilon})$$

Estimation: Step 1

Estimation procedure is based on Attanasio et. al (2020)

We begin by standardizing all measurements, and anthropometric measures by WHO standardized weight-for-age and height-for-age z-scores.

We assume that the joint distribution of measures, $f(\tilde{m})$, follows a mixture of normal distribution
We use an Expectation Maximisation algorithm to estimate the means and variance-covariance matrices of $f(\tilde{m})$: $\tilde{\mu}^1, \tilde{\Sigma}^1, \tilde{\mu}^2, \tilde{\Sigma}^2, \tilde{\tau}$.

Estimation: Step 2

We estimate the distribution of latent variables $f(\ln \Psi)$, which is defined by the means and covariance-variance matrices of each mixture component, given by $\mu^1, \Sigma^1, \mu^2, \Sigma^2, \tau$, factor loadings matrix Λ and the covariance matrix of the distribution of errors Σ^ϵ. Note that the latter two are assumed to be the same between pre and post ChCC cohorts so that all differences between cohorts arise from differences in the distribution of latent variables.

We estimate the following system using minimum distance estimation:

$$\tau = E[\tilde{\tau}]$$

$$\Lambda\mu^1 + A = E[\tilde{\mu}^1]$$
$$\Lambda\mu^2 + A = E[\tilde{\mu}^2]$$
$$\Lambda'\Sigma^1\Lambda + \Sigma^\epsilon = E[\tilde{\Sigma}^1]$$
$$\Lambda'\Sigma^2\Lambda + \Sigma^\epsilon = E[\tilde{\Sigma}^2]$$

Synthetic Dataset: Step 3

We begin by generating a synthetic dataset of all latent traits and controls using the parameters from Step 2.

We estimate a production function of child skills using OLS.

As we have a multi-step procedure all standard errors are calculated using a bootstrapping procedure.

In a accompanying paper we prove, via Monte-Carlo exercises, that our estimator is able to recuperate all parameters of production function.

Percentage of information per measure of latent variables

	Measures	Perc. Information
Abilities at birth	Gestation in weeks	0.705
	Weight at birth	0.735
	Height at birth	0.546
Parental Investments	Sports startage	0.001
	Sweets frequency	0.003
	Television time	0.004
	Active now	0.543
	Active age 4	0.639
	Preschool use	0.000
	Computer time	0.003
	Reading time	0.003
	Outside time	0.051
Health at age 8	Weight for age	0.754
	Height for age	0.661
Socio-emotional External Abilities	(-)Delinquent Behavior	0.668
	(-)Aggresive conduct	0.638
Parental Health	Weight	0.283
	Height	0.269

Source: Own elaboration based on EM estimation

Production functions

	Health	Ext. Socio Emotional
	(1)	(2)
Investment	0.105	0.362
	(0.081, 0.123)	(0.020, 0.718)
Abilities at birth	0.053	0.056
	(0.029, 0.069)	(0.035, 0.075)
Parental health	0.929	−0.122
	(0.891, 1.010)	(−0.071, −0.193)
EA PGI	0.026	0.031
	(−0.079, 0.181)	(0.023, 0.037)
Birthweight PGI	0.027	−0.006
	(0.016, 0.036)	(−0.018, 0.009)

Note: 90% bootstrapped confidence interval in parenthesis. 30 replications. Controls are child's gender and the number of children in the household.

Conclusions

- Genetic endowments matters for socio-emotional abilities and health at age 8
- Evidence on the importance of using different PGIs for different abilities
- Effects of genes are small relative to parental investments.

Next Steps

- Model more time periods
- Improve measures of parental investments
- Model parental directly, then we can control for the fact that parents react to shocks to the production function
- Add interaction terms of investment*genes, or abilities*genes
- Use model to simulate the effects of difference policies, shocks to parental investments, parental income etc.