Estimating the technology of cognitive and noncognitive skill formation

Flavio Cunha, James Heckman, Susanne M. Schennach (Econometrica 2010)

Moritz Mendel

Department of Economics
University of Bonn

Topics in Applied Microeconomics

November 29, 2018

Overview

Introduction

- Skill formation is modelled as a multistage technology during childhood that determines adulthood outcomes
- (Non-)Cognitive Skill formation in a period depends on several key factors
  - Both cognitive and non-cognitive Skills in the previous period
  - Parental investment in skill formation during that period
  - Parental skills
- Key challenges
  - Identify skills and parental investment
  - Estimate the key parameters in the skill formation technology

Model

- T periods of childhood and A periods of adulthood
- Technology function $\theta_{k,t+1} = f_s(\theta_t, I_t, \theta_p, \eta_{k,t})$
- Adulthood outcomes $Q = g(\theta_{T+1})$
- Both cognitive and noncognitive skills matter for skill formation and adulthood outcomes $\theta_t = (\theta_{C,t}, \theta_{N,t})$
- Throughout the discussion the control vector X will be kept implicit

Model

- The model is built such as to accomodate for a number of empirical observations in the human capital literature
- The notions of critical periods $\frac{\partial \ f_{k,t}}{\partial I_{k,t}} = 0$ if $t \neq t'$ and of dynamic complementarity $\frac{\partial^2 f_{k,t}}{\partial \theta_{l,t} \partial I_{k,t}}$ are accomodated for in the structure of the model
- The paramteric assumptions the paper makes at a later stage make these notions more explicit

Identification of latent factors

- Skills and Investment are not observed but only related factors such as test scores
- Factor Analysis can be used to identify latent factors subject to important normalizations
- Consider a measurement $Z_{a,k,t,j}$ where a denotes the type k the factor and j the specific measurement
- As a simplifying assumption the paper considers $Z_{a,k,t,j} = \mu_{a,k,t,j} + \alpha_{a,k,t,j}\theta_{k,t} + \epsilon_{a,k,t,j}$
- Factor Analysis can be used to identify latent factors subject to important normalizations
- The ϵ are assumed to be uncorrelated over j and are assumed to have expectation 0
- The authors set the scale by assuming $\alpha_{a,k,t,1} = 1$ and normalize further $E[\theta_{k,t}] = 0, E[I_{k,t}] = 0$

Identification of latent factors

- Identification of the other factor loadings $\alpha_{a,k,t,2}$ can be accomplished via the covariances of the observed variables.

$$Cov(Z_{1,C,t,1}, Z_{1,C,t+1,1}) = Cov(\theta_{C,t}, \theta_{C,t+1})$$
$$\mathrm{Cov}(\mathrm{Z}_{1,C,t,2}, Z_{1,C,t+1,1}) = \alpha_{1,C,t,2} Cov(\theta_{C,t}, \theta_{C,t+1})$$

- Identification of the distribution of all the unobserved factors follows from a theoretical argument
- The intuition is that the signal is common to different measurements but not the noise
- The normalizing assumptions assumptions the made are sufficient to obtain a unique distribution of the factors

Identification of the technology function

- The shocks η and the technology function can not be identified together without normalizing assumption
- One potential assumption is additive seperability of $f(.)$ in η
- Another way is to normalize η such that is conditionally uniformly distributed. The technology function can then be written as $f_{k,s}(\xi_t, \eta_{k,t}) = G^{-1}_{\eta_{k,t}}(\eta \mid \xi_t)$

Anchoring

- Factors are anchored to an adulthood outcome such as to set them to an interpretable metric
- Denote adulthood outcomes by $Z_{4,j}$
- Assume the following model
  $Z_{4,j} = \mu_{4,j} + \alpha_{4,j,C}\theta_{C,T+1} + \alpha_{4,j,N}\theta_{N,T+1} + \epsilon_{4,j}$
- The anchor would then look as follows
  $g_{C,j}(\theta_{C,t+1}) = \mu_{4,j} + \alpha_{4,C,j}\theta_{C,T+1}$
- Identification can be shown along with the identification of the production function under the assumption that g is monotone

Notation

- ω denotes a specific individual and X represents a set of control variables
- $\xi_t \in \theta_{C,t} x \theta_{N,t} x \theta_{C,p} x \theta_{N,p} x I_{C,t} x I_{N,t}$
- Potential Outcomes: $Y_t(\omega, \xi_t) = \mu_{\xi_t}(X) + U_{\xi_t}$

Parameters of Interest

- Consider two different values $\xi_t'$and ξ_t''
- By assuming that the η are independent of ξ we get that $Y_{\xi_t'} - Y_{\xi_t''} \perp \xi_t \mid X$
- Thus we can identify the ATE on skill formation of skill k in $t+1$ from moving from ξ_t' to ξ_t'' as $ATE(\xi_t', \xi_t'', t+1, k) = \int_{\eta_{k,t}} (f(\xi_t'', \eta_{k,t}) - f(\xi_t', \eta_{k,t}))\phi_t(\eta_{k,t}) d\eta_{k,t}$

Accounting for the Endogeneity of Parental Investment

- Assuming that the η are independent of the inputs would imply that parental decisions do not depend upon unobserved opportunities
- This is unlikely to be true
- We can use adulthood outcomes to include unobserved time invariant heterogeneity into the production function
- One can include such heterogeneity by writing:
  $Z_{4,j} = \alpha_{4,C,j}\theta_{C,T+1} + \alpha_{4,N,j}\theta_{N,T+1} + \alpha_{4,\pi,j}\pi + \epsilon_{4,j}$
- The π is identified by similar arguments as the other factors
- A more realistic assumption on potential outcomes
  $Y_{\xi_t'} - Y_{\xi_t''} \perp \xi_t \mid X, \pi$

Policy Relevant Treatment effects

Policy Relevant Treatment effects

From the discussion of the heterogeneity we can now identify average treatment effects conditional on π . Since we have that $\theta_{k,t} = f_t(\theta_{t-1}, I_{t-1}, \theta_p, \eta_{k,t-1})$ we can substitute for skill levels such as to obtain a production function that only depends on shocks, Investment, innate ability, parental environment and shocks. We can now define a theory parental Investment and thereby define policy relevant treatment effects.

Data

- 2207 firstborn children from the NLSY9 sample
- Starting in 1986 the children of the female respondents have been assesed every 2 years
- There is data available on cognitive and noncognitive skills and investment

Practical Estimation

The discussion above would technically allow us to identify the model non-parametrically. That would however be computationally demanding and require a lot of data. The authors do instead invoke parametric assumptions such as to estimate the production function.

Parametric assumptions and Estimation

$$\theta_{k,t+1} = [\gamma_{s,k,1}\theta_{C,t}^{\phi_{s,k}} + \gamma_{s,k,2}\theta_{N,t}^{\phi_{s,k}} + \gamma_{s,k,3}\theta_{C,t}^{\phi_{s,k}} + \gamma_{s,k,4}\theta_{C,P}^{\phi_{s,k}} + \gamma_{s,k,5}\theta_{N,P}^{\phi_{s,k}}]^{\frac{1}{\phi_{s,k}}} e^{\eta_{k,t+1}} \tag{1}$$

$$\eta_{k,t} \sim N(0, \delta_{\eta,s}) \tag{2}$$

$$\gamma_{s,k,l} \geq 0 \tag{3}$$

$$\sum_{l=1}^{5} \gamma_{s,k,l} = 1 \tag{4}$$

Practical estimation

$$Z_{a,k,t,j} = \mu_{a,k,t,j} + \alpha_{a,k,t,j} ln(\theta_{k,t}) + \epsilon_{a,k,t,j} \tag{5}$$

$$\epsilon_{a,k,t,j} \sim N(0, \wedge_t) \tag{6}$$

Practical Estimation

If we now define a likelihood for the whole θ vector we get a full likelihood for the observed data.

Likelihood

$\int_{\times\epsilon} p(\theta) p_{\times\epsilon}(Z = z \mid \theta) d\epsilon$

Instead of maximizing the original likelihood which would require solving a large number of integrals of nonlinear functions one can use the formula on the last page to compute the likelihood of Z as implied by the distribution of θ recursively. Therefore the authors use a technique that is related to the Kahlman filter. In order to illustrate the algorithm consider the following identity

$$P(z) = P(z_1) \prod_{t=1}^{T} P(z_{t+1} | z^t) \tag{7}$$

Filtering

- The Filtering technique helps us to perform the following steps:
  1. $p(\theta_{t-1}^{t-1}) \longrightarrow p(\theta_t \mid z^{t-1})$
  2. $p(\theta_t^{t-1}) \longrightarrow p(\theta_t \mid z^t)$
- With the Likelihood function we can then get $p(z_t \mid z^{t-1})$

Results

- The model is estimated assuming two stages of development
- Controlling for endogeneity seems to be of first order importance
- There is cross-productivity of non-cognitive skills on cognitive skills but not vice versa
- Investment is more productive for cognitive skills at the earlier period. That is not the case for non-cognitive skills

Conclusion

- The paper introduces a model of human capital accumulation that can be non-parametrically identified
- The empirical section does however produce biased results which is due to the fact that the CES-Production Function already assumes a scale and location. According to Wistwall and Agistonelli setting a fixed location for θ introduces a bias