Factor Models

A. Linear

Costas Meghir

February 27, 2018

Introduction

- Child development is about the intergenerational transmission of poverty and human capital development.
- It matters because poverty is associated with permanent cognitive and health deficits with longer term effects.
- Understanding HC development involves understanding:
  1. The human capital production function from an early age and the role of investments
  2. what may be the optimal structure of investments
  3. How parents make decisions and what is their information set
  4. How effective policy can be and how it interacts with parental decisions
  5. It is an interdisciplinary field that needs to draw from economics, psychology and measurement, nutrition and neurology and even genetics.

Outline

1. Estimating human capital production functions
2. Models of the family and investing in children
3. Field Experiments on child development and interpreting them.

Human Capital production functions: Background

- Suppose we wish to estimate a human capital production function
- Fundamentally the problem we need to face is that neither output nor input is directly observable.
- We tend to have measures that reveal indirectly inputs and outputs in the human capital production function
- From this perspective this is a (complex) measurement error problem: inputs and output is measured with error

Background

- The flip side of this is that all the measures we observe are substantive in themselves
- However they may be generated by a small set of latent factors
- For example how many factors underlie asset prices? (King, Sentana, and Wadhwani, 1994 Econometrica)
- How many dimensions does intelligence have?
- We will work with the first interpretation: rather than investigating the appropriate number of dimensions we will assume them known and discuss estimation of HC production functions

Linear Production Functions

Formulating, Identifying and Estimating the Technology of Cognitive and noncognitive Skill Formation by Cunha and Heckman JHR, 2008

- Define two dimensions of skill: cognitive skill (θ^c) and non-cognitive skill (θ^N)
- Now consider a production function describing skill accumulation

$$\theta^c_{it+1} = \mu^c + \gamma^c_1 \theta^c_{it} + \gamma^c_2 \theta^N_{it} + \gamma^c_3 \theta^I_{it} + \eta^c_{it}$$

$$\theta^N_{it+1} = \mu^N + \gamma^N_1 \theta^c_{it} + \gamma^N_2 \theta^N_{it} + \gamma^N_3 \theta^I_{it} + \eta^N_{it}$$

- The term η^j_{it} is an unobserved random factor that does not correspond to any measurement.
- For now we will assume it is independent of the latent factors θ^j_{it} (exogeneity)
- The model above implies that past skill levels and investments θ^I_{it} produce future cognitive and non-cognitive skills
- It is designed to reflect the technological constraints that govern human capital policy

Linear Production Functions

- The point is that we cannot measure the θ directly (without error).
- Instead we have a set of measurements in the form of test scores, behavioral problem index and HOME score (Home Observation Measurement of the Environment)
- It turn out that we need a minimum of two measurements per factor if all measures are independent of each other.
- Take the system of measurement equations (and henceforth drop the time index t unless needed)

$$m_{ik}^{j} = a_{0k}^{j} + a_{1k}^{j}\theta_{i}^{j} + \varepsilon_{ik}^{j} \quad j = c, N, I \;\; k = 1, 2$$

- These equations relate what we actually observe (m) to the latent factors (θ).
- The ultimate aim is to estimate the joint distribution of θ from which we can then deduce $E(\theta_{it+1}^{j} | \theta_{it})$, $j = c, N$

Identification

- At this point we need the following normalisations:
  - Since we do not observe the latent factors we need to normalize one of the factor loadings per factor to 1(e.g. $a_1^j = 1$)
  - We also assume that $E(\theta_i^j) = 0$ and hence $E(m_{ik}^j) = a_{0k}^j$. Hence forth we will assume measurements are zero mean.

Measurement Error

- In its simplest form the measurement errors are assumed iid (classical)
- In general we can allow for dependence of errors across measurements (depending on how many measurements we have)
- Importantly we do not need to specify the distribution of the measurement errors

Kotlarski Theorem

- The key theorem in this literature is the Kotlarski theorem (1967 Pacific Journal of Mathematics):
- Under the measurement error structure $Y = \theta + \varepsilon$ with at least two measures per latent factor and iid errors the distribution of the latent factors and of the measurement errors is identified.
- This powerful result is at the heart of all the results. The important point is that all we need is independence of the measurement errors.
- Everything else is nonparametrically identified.

The Main Idea

- The key idea about identification is as follows. It is a deconvolution problem.
- Take a one factor model and assume known factor loadings.
- The data provides the unconditional distribution of observables $f(m)$.
- This can be written as

$$f(m) = \int f_m(m|\theta)g(\theta)d\theta$$

- The measurement equations together with the distribution of measurement errors define $f_m(m|\theta)$.
- Effectively the theorem states that if we can write $f_m(m|\theta) = f_{1m}(m_1|\theta) \times f_{2m}(m_2|\theta)$ then we can identify nonparametrically the form of f_{1m}, f_{2m} and the object of interest $g(\theta)$.

Estimation

- The problem is not quite the from we want it because of the factor loadings
- However redefining the measurements as m^j_{ik}/a^j_{1k} they are now in the form of the Kotlarski theorem.
- Thus estimation takes place in two stages:

1. The factor loadings are estimated a^j_{1k} first.
2. The joint distribution of the latent factors is then estimated

Estimating the Factor Loadings

- Write the measurement system as

$$M = \Theta \times A' + E$$

  where

  - M is $N \times (p \times k)$ where p is the number of measures per factor, k is the number of latent factors and N is the number of observations
  - Θ is the matrix of latent factors. Its dimensions is $N \times k$
  - A is the matrix of factor loadings. It is $(p \times k) \times k$
  - E is the matrix of all measurement errors

- Embedded in A are the normalization restrictions (so some values are already set to one) and the fact that each measurement corresponds to one factor (this can be relaxed). The latter are zero restrictions.

Estimating the Factor Loadings

- This formulation implies

$$V(M) = AV(\Theta)A' + V(E)$$

  where $V(X)$ is the covariance matrix of X.
- From the data we know $V(M)$ (we can estimate it). These are $\frac{1}{2}(pk)(pk+1)$ known moments
- Before any zero restrictions and normalizations the unknowns are

$$k^2 p + \frac{1}{2}k(k+1) + \frac{1}{2}pk(pk+1)$$

- However, with the normalizations, the zero restrictions and the iid assumptions of the ε the system is overidentified.
- Solving this system subject to the restrictions provides the factor loadings and the covariance matrix of the factor loadings.
- Note that this step does not use any distributional assumptions

Example

- As an example consider a two factor model with three measurements for each factor

$$
\begin{aligned}
m_{1i} &= \theta_{1i} + u_{1i} \\
m_{2i} &= a_2 \theta_{1i} + u_{2i} \\
m_{3i} &= a_3 \theta_{1i} + u_{3i}
\end{aligned}
$$

$$
\begin{aligned}
m_{4i} &= \theta_{2i} + u_{4i} \\
m_{5i} &= b_2 \theta_{2i} + u_{5i} \\
m_{6i} &= b_3 \theta_{2i} + u_{6i}
\end{aligned}
$$

- Assume $E(u_{ki} u_{sj}) = 1(k = s)\sigma_k^2$
- Note the normalization restrictions and the zero restrictions
- We have 21 distinct observed moments and 13 unknowns: the variance matrix of of θ (3 elements), $a_2,\ a_3,\ b_2,\ b_3$ and $\sigma_k^2,\ k = 1, ..., 6$

Example

$$
\begin{gathered}
V(m_1) = V(\theta_1) + \sigma_1^2 \\
C(m_1, m_2) = a_2 V(\theta_1) \\
C(m_1, m_3) = a_3 V(\theta_1) \\
C(m_2, m_3) = a_2 a_3 V(\theta_1)
\end{gathered}
$$

- Hence $a_2 = C(m_2, m_3)/C(m_1, m_3)$; $V(\theta_1) = C(m_1, m_1)/a_2$ and so on.
- Note the scope of nonclassical measurement error since the system is overidentified

Non-Classical Measurement error

- Cunha and Heckman extend their model for non-classical measurement error:
  - Assume that there is one measurement (say the first one) whose errors are independent from all others (across measurements and time periods)
  - Otherwise all other errors can be arbitrarily dependent
  - We still impose that measurements are independent of the actual factors - this is crucial.
- This relaxation of restrictions still allows for at least two independent instruments - such an assumption is at the heart of this approach

Estimation

- Now Suppose we assumed that the latent factors are jointly normal.
- This is consistent with the linear production function, although there are other distributions that have linear conditional expectations
- Under joint normality our job is finished because knowing the joint distribution defines the conditional mean
- Under linearity even without normality we can use an IV strategy as well. This is the approach in Cunha and Heckman (JHR)

Estimation

- Take the structural equations

$$\theta^c_{it+1} = \gamma^c_1 \theta^c_{it} + \gamma^c_2 \theta^N_{it} + \gamma^c_3 \theta^I_{it} + \eta^c_{it}$$

- Use the measurement equations to replace the latent factor by a proxy:

$$\theta^j_{it} = m^j_{i1t} - \varepsilon^j_{i1t}$$

- Note that if all measurements are rescaled by dividing by the factor loadings the coefficients will be scaled in the same way whatever measurement we use as a proxy.

Estimation

- OLS will not work because of measurement error: ε and the proxies m are correlated.
- Use the other measurement (m^j_{i2t}) as an instrument. The measurements errors must be independent
- This also gives another perspective to the numerous regressions including such proxies
- The idea of using two independent measures to solve the measurement error problem in linear and nonlinear models goes back some time (e.g. Hausman, J., H. Ichimura, W. Newey, and J. Powell, 1991, Measurement errors in polynomial regression models, Journal of Econometrics 50, 271-295).

Endogenous inputs

- Up to now we have assumed the inputs to be exogenous.
- Suppose the inputs are endogenous.
- In this context this means that there are unobserved inputs that do not correspond to any measurements (otherwise we could have included further latent factors in the same way.
- In Cunha and Heckman they suggest a fixed effects approach

$$\eta_{it}^{j} = \alpha^{j} \lambda_i + u_{it}^{j}$$

- With variation over time we can now take first differences of the proxies

$$\Delta m_{i1t+1}^{c} = \gamma_1^c \Delta m_{i1t}^{c} + \gamma_2^c \Delta m_{i1t}^{N} + \gamma_3^c \Delta m_{i1t}^{I} + \Delta u_{it}^{c}$$

- Use as instruments the first differences of the other proxy (Δm_{i2t}^{j})
- Notice that we can now use the residuals in levels to also estimate the distribution of the fixed effect $\lambda$and the factor loadings, based on the Kotlarski theorem.

Anchoring

- Even after estimating the production function it is hard to interpret the results if we do not know what the units of measurement refer to.
- A useful approach is to anchor the results to some outcome of interest, such as earnings
- So consider the regression of earnings on the latent factors

$$Y_i = a_1 \theta_{it}^c + a_2 \theta_{it}^N + a_3 \theta_{it}^I + v_{it}$$

-

  The coefficients on the latent factors transform the latent factor units to the units of the left hand side; So we can use these to transform the coefficients of the other equations

Using Factor Scores to estimate linear factor models

From Heckman, Pinto and Savalyev, American Economic Review 2010

- It is also possible to predict factor scores for each individual and use these for estimation or descriptive statistics.
- Consider the measurement equation written for one observation

$$M_i = A\theta_i + E_i$$

where θ_i is the $k \times 1$ vector of latent factors for the ith individual

Using Factor Scores to estimate linear factor models

From Heckman, Pinto and Savalyev, American Economic Review 2010

- We need a prediction $\hat{\theta}_i$ such that $E(\hat{(}\theta_i) = \theta_i)$
- Consider Some matrix $\mathscr{L}$ such that $\mathscr{L}A = I$
- A solution is given by Barttlet's (1937) estimator where we minimize the MSE of the prediction $\mathscr{L}M_i$ subject to the restriction $\mathscr{L}A = I$.
- the optimal is obtained by regressing the measures on the factor loadings by GLS where the weight matrix is given by the variance of E_i

$$\hat{\theta}_i = (A'Var(E)^{-1}A)^{-1}A'Var(E)^{-1}M_i$$

  where $\mathscr{L} = (A'Var(E)^{-1}A)^{-1}A'Var(E)^{-1}$.
- Bartlett, M. S. (1937) The statistical conception of mental factors, British Journal of Psychology

Using Factor Scores to estimate linear factor models

From Heckman, Pinto and Savalyev, American Economic Review 2010

- This procedure leads to an unbiased prediction of the latent factors for each individual
- However, these factor scores include estimation error, which is not orthogonal to the predicted latent factor will cause attenuation error.
- We can use the structure of the measurement error to correct for the bias

Correcting for measurement error

- Suppose we wish to estimate a regression of the form

$$y_i = \theta_i' \beta + \varepsilon_i$$

- The factor score method replace θ_i by $\hat{\theta}_i$ where $\hat{\theta}_i = \theta_i + v_i$ with v_i being the measurement error.
- By construction we have that $\Sigma_{\hat{\theta}} = \Sigma_\theta + \Sigma_v$ where Σ_c is the covariance matrix of c
- We have that $\hat{\beta} = (\hat{\theta}'\hat{\theta})^{-1}\hat{\theta} y_i$
- Now note that $Cov(\hat{\theta}, \theta) = \Sigma_\theta$. Hence

$$plim \hat{\theta} = \Sigma_{\hat{\theta}}^{-1} \Sigma_\theta \beta$$

- We already have an estimate of Σ_θ and of $\Sigma_{\hat{\theta}}$.
- So we can premultiply by an estimate of $\Sigma_\theta^{-1}\Sigma_{\hat{\theta}}$ to correct for the bias:

$$\tilde{\beta} = \hat{\Sigma}_\theta^{-1} \hat{\Sigma}_{\hat{\theta}} \hat{\beta}$$

Results from Cunha and Heckman

- The research question is:
  - How do skills get formed?
  - How persistent are initial conditions
  - What is the productivity of investments
  - How does the production function between cognitive and noncognitive skills differ

The Data

Table 1

Summary Dynamic Measurements: White Children NLSY/1979

	Age 6 and 7			Age 8 and 9			A
	Observations	Mean	Standard Error	Observations	Mean	Standard Error	C
Piat math[a]	753	−1.0376	0.5110	799	0.0423	0.6205	
Piat reading recognition[a]	751	−1.0654	0.4303	795	−0.0932	0.6543	
Antisocial score[a]	753	0.0732	0.9774	801	−0.0843	1.0641	
Anxious score[a]	778	0.1596	1.0016	813	−0.0539	1.0187	
Headstrong score[a]	780	0.0192	0.9882	813	−0.2127	1.0000	
Hyperactive score[a]	780	−0.0907	0.9673	815	−0.1213	1.0148	
Conflict score[a]	779	0.0177	0.9977	815	−0.0057	0.9935	
Number of books[b]	629	3.9173	0.3562	821	3.9220	0.3104	
Musical instrument[c]	628	0.4650	0.4992	821	0.4896	0.5002	
Newspaper[c]	629	0.5326	0.4993	821	0.5043	0.5003	
Child has special lessons[c]	627	0.5470	0.4982	820	0.7049	0.4564	
Child goes to museums[d]	628	2.2596	0.9095	821	2.3082	0.8286	
Child goes to theater[d]	630	1.8111	0.8312	820	1.8012	0.7532	
Natural logarithm of family income[e]	865	10.4915	1.3647	936	10.4494	1.5689	
Mother's highest grade completed[f]	1053	12.9620	2.2015	1053	12.9620	2.2015	
Mother's arithmetic reasoning score[g]	776	0.6050	1.0132	776	0.6050	1.0132	
Mother's word knowledge score[g]	776	0.5894	0.7666	776	0.5894	0.7666	
Mother's paragragh composition score[g]	776	0.5464	0.7311	776	0.5464	0.7311	
Mother's numerical operations score[g]	776	0.4945	0.8189	776	0.4945	0.8189	
Mother's coding speed score[g]	776	0.4554	0.8084	776	0.4554	0.8084	
Mother's mathematical knowledge score[g]	776	0.5297	1.0259	776	0.5297	1.0259	

a. The variables are standardized with mean zero and variance one across the entire CNLSY/79 sample.

Table 2

Unanchored Technology Equations:[a] *Measurement Error is Classical, Absence of Omitted Inputs Correlated with* θ_t *White Males, CNLSY/79*

	Noncognitive Skill $\left(\theta_{t+1}^N\right)$			Cognitive Skill $\left(\theta_{t+1}^C\right)$		
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Lagged noncognitive skill, $\left(\theta_t^N\right)$	0.884	0.884	0.884	0.028	0.028	0.028
	(0.021)	(0.021)	(0.021)	(0.013)	(0.013)	(0.013)
Lagged cognitive skill, $\left(\theta_t^C\right)$	0.003	0.003	0.003	0.977	0.977	0.977
	(0.013)	(0.012)	(0.013)	(0.038)	(0.038)	(0.038)
Parental investment, $\left(\theta_t^I\right)$	0.072	0.078	0.080	0.064	0.069	0.071
	(0.020)	(0.021)	(0.024)	(0.013)	(0.014)	(0.015)
Mother's education, S	0.004	0.004	0.004	0.003	0.003	0.003
	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)
Mother's cognitive skill, A	−0.006	−0.006	−0.006	0.025	0.025	0.025
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)

a. Let $\theta_t' = \left(\theta_t^N, \theta_t^C, \theta_t^I\right)$ 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:

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \psi_1^k S + \psi_2^k A + \eta_t^k.$$

In this table we show the estimated parameter values and standard errors (in parentheses) of $\gamma_1^k, \gamma_2^k, \gamma_3^k, \psi_1^k$, and ψ_2^k in Columns 1–6. In Columns 1 and 4, the parental investment factor is normalized on the log-family income equation. In Columns 2 and 5, the parental investment factor is normalized on trips to the museum. In Columns 3 and 6, we normalize the parental investment factor on trips to the theater.

Omitted Inputs

Table 7
Unanchored Technology Equations:[a] *Measurement Error is Classical, Allows for Omitted Input λ Correlated with θ_t, White Males, CNLSY/79*

Independent Variable	Noncognitive Skill $\left(\theta_{t+1}^{N}\right)$	Cognitive Skill $\left(\theta_{t+1}^{C}\right)$
Lagged noncognitive skill, $\left(\theta_t^N\right)$	0.8848	0.0276
	(0.021)	(0.013)
Lagged cognitive skill, $\left(\theta_t^C\right)$	0.0022	0.9891
	(0.013)	(0.039)
Parental investment, $\left(\theta_t^I\right)$	0.0797	0.0844
	(0.020)	(0.017)
Omitted correlated inputs, λ	0.2835	1.0000
	(0.134)	(normalized)

Discrete measurements

- In many cases the measurements are discrete
- In this case the simple covariance restrictions do not identify the factor loadings
- In this case we need to set up the problem accounting for such discreteness if we wish to estimate all factor loadings
- Under linearity however the problem is still quite simple; the IV strategy with two measurements will work so long as we have at least one continuous measurement
- This becomes much more complex with nonlinear models.

Discussion

- Why is this useful and where does it take us?
- The linear factor model is interesting but of limited usefulness from an economic point of view
- Embedded in this is the assumption that inputs in the production function are perfectly substitutable for each other.
- Hence we cannot answer questions relating to the complementarity/substitutability of investments

Material from Cunha, Heckman and Schennach

- Consider as an alternative the production function

$$\theta_{it+1}^{c} = [\gamma_1(\theta_{it}^{c})^{\rho} + \gamma_2^{c}(\theta_{it}^{N})^{\rho} + \gamma_3^{c}(\theta_{it}^{I})^{\rho}]^{\frac{1}{\rho}} + \eta_{it}^{c}$$

- Now we have an interesting testable hypothesis: $\rho = 1$.
- Take this in the context of child development and ask whether say parental time and child care or expenditures are perfectly substitutable
- And more interestingly: how substitutable are investments at different ages.
- Here the problem is much more complex: it is a nonlinear model with measurement errors.
- However, we know it is identifiable from the Kotlarski theorem, since the latter establishes the non-parametric identification of the distribution of θ: In general the conditional mean will be nonlinear.

A model for the timing of investments

- Consider the human capital production function

$$Q = [\tau_1 I_1^{\phi} + \tau_2 I_2^{\phi} + \tau_3 \theta_c^{\phi} + \tau_4 \theta_p^{\phi}]^{1/\phi}$$

$$Q = [\tau_1 I_1^{\phi} + \tau_2 I_2^{\phi} + \tau_3 \theta_{C,1}^{\phi} + \tau_4 \theta_{N,1}^{\phi} + \tau_5 \theta_{C,P}^{\phi} + \tau_6 \theta_{N,P}^{\phi}]^{1/\phi}$$

- Here we have (as an example) adult human capital depending on investments in two periods, initial endowments and parental characteristics
- Take lifetime discounted earnings to be $R(Q) = \sum_{t=3}^{A} \beta^t wQ$ with w being the constant wage (for simplicity)
- We can derive an optimal rule for parental investments
- The optimal investment rule then becomes

$$log(\frac{I_1}{I_2}) = \frac{1}{1-\phi}\left[log(\frac{\tau_1}{\tau_2}) - log(1+r)\right]$$

- Amount of total investment driven by cost of raising funds and altruistic link between families.

A model for the timing of investments