Empirical Methods for Policy Evaluation II

Matteo Bobba

Toulouse School of Economics (TSE)

TSE PhD Program
Fall 2021

Part 4: Dynamic Latent Factor Models & RCTs

- Context
  - **A randomized early childhood intervention in Colombia**
- Methods
  - Cunha, Heckman and Shennach (ECMA, 2010)
- Application
  - Attanasio et al. (AER, 2020)

Human Capital and The Importance of the Early Years

- Human capital formation is a complex process
  - Human capital is multi-dimensional (cognitive, non-cognitive, health,...)
  - Skill formation is a dynamic process
  - The components of human capital interact both within and across periods
  - Both genetic and environmental factors are important inputs

- The early years is a particularly salient period for policy
  - Human capital is malleable (and vulnerable)
  - Dynamic complementarities ("skills beget skills")

- Well-designed and well-targeted interventions in the early years can partially compensate for exposure to adverse environments (and have long-lasting impacts)
  - Perry School experiment (Anderson, 2008; Heckman et al., 2010, 2011, 2013)
  - Jamaica Study (Walker et al., 1990, 2011; Gertler et al., 2013)

The Intervention

- Attanasio et al (2014) adapted the Jamaica curriculum to the Colombian context
  - Targeted to children 12-24 months living in families targeted by a cash transfer program
  - Weekly home visits of one hour from trained women from the community
  - Micronutrients supplementation
  - Cheap and scalable intervention
- Stratified-clustered RCT design
  - 1,429 children in 96 municipalities in central Colombia
  - Municipalities were randomized into 4 groups
    1. Stimulation
    2. Supplementation (micronutrients)
    3. Stimulation+Supplementation
    4. Control

Measurement Tools

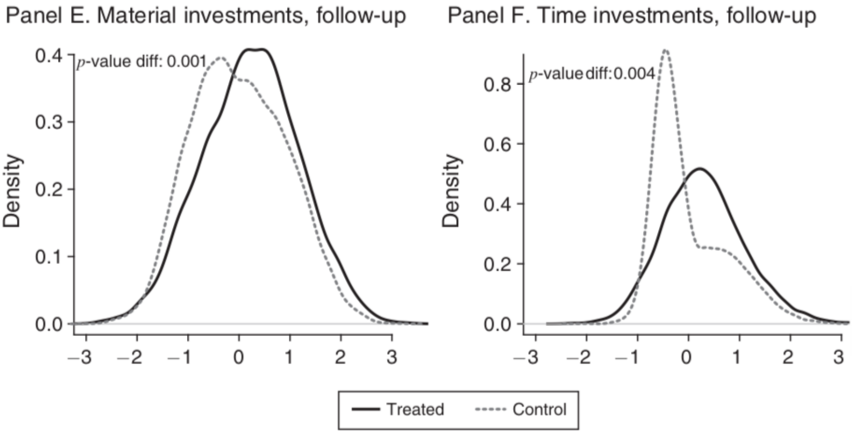
- A battery of measurement on children development
  - Motor and cognitive skills: Bailey test
  - Socio-emotional skills: bates temperament
  - Language development: Mc-Arthur-Bates
  - Height, weight, hemoglobin, and morbidity
  - Food intakes
  - Child care arrangements and time use
- Mothers and families
- Home visitor
- Qualitative instruments

Experimental Results

- Stimulation improved cognitive scores (adjusted for age, sex, testers, and baseline levels of outcomes) by 0.26 of a standard deviation (P=0.002)
- Stimulation also increased receptive language by 0.22 of a standard deviation (P=0.032)
- Micronutrient supplementation had no significant effect on any outcome and there was no interaction between the interventions
- No intervention affected height, weight, or hemoglobin levels

A First Hint at Mechanisms: Parental Investment

Figure: Kernel Density of Latent Factors



Takeaways from the Experimental Evaluation

- The experiment demonstrates impacts on child development
- What are the possible channels through which the intervention affected outcomes?
  - The stimulation provided during the home visits may be a new input in the development of the child
  - The intervention could generate impacts on child development through an increase in parental investments
  - The intervention could also have affected maternal cognitive or, more plausibly, socio-emotional skills

Part 4: Dynamic Latent Factor Models & RCTs

- Context
  - A randomized early childhood intervention in Colombia
- Methods
  - **Cunha, Heckman and Shennach (ECMA, 2010)**
- Application
  - Attanasio et al. (AER, 2020)

The Production Function of Human Capital

$$\boldsymbol{H}_{t+1} = g_t(\boldsymbol{H}_t, \boldsymbol{X}_t, \boldsymbol{I}_{t+1}, u_{t+1})$$

- $\boldsymbol{H}_t = \{\theta_{c,t}, \theta_{s,t}, \theta_{h,t}\}$ is human capital (including cognition, socio-emotional, and health)
- $\boldsymbol{X}_t = \{\theta_{m,t}, \theta_{f,t}, \theta_{r,t}\}$ are parental skills and other background variables (including mother, father and other caregivers)
- $\boldsymbol{I}_{t+1} = \{\theta_{M,t}, \theta_{T,t}\}$ are parental investments (including materials and time) between t and $t+1$
- Stocks of current period skills produce next period skills
- Stocks of current skills affect the current period productivity of investments

$$\frac{\partial^2 g_t}{\partial I_t \partial \boldsymbol{H}_t}$$

Measurement System

- For each $k \in \{c, n, m, f, ...\}$ and period t we have measures such that:

$$m_{k,t,j} = \alpha_{k,t,j}\theta_{k,t} + \epsilon_{k,t,j}$$

- $m_{k,t,j}$ is the j^{th} measurement of skill $\theta_{k,t}$
- $\alpha_{k,t,j}$ is a factor loading
- $\epsilon_{k,t,j}$ is the measurement error contained in $m_{k,t,j}$

- With two independent measures per factor, the distribution of the unobserved factor and the measurement error can be identified non-parametrically up to a change of location
    - To set the factor scale: $\alpha_{k,t,1} = 1, \, \forall t = 1, ...T$
    - To set the factor location: $E(\theta_{k,t}) = 0, \, \forall t = 1, ...T$

Identification: Factor Loadings

- Assume that the $\epsilon_{k,t,j}$ are uncorrelated across t, then

$$\mathrm{Cov}(m_{k,t,1}, m_{k,t+1,1}) = \mathrm{Cov}(\theta_{k,t}, \theta_{k,t+1})$$
$$\mathrm{Cov}(m_{k,t,2}, m_{k,t+1,1}) = \alpha_{k,t,2}\mathrm{Cov}(\theta_{k,t}, \theta_{k,t+1})$$

- Hence, we can identify the factor loading

$$\frac{\mathrm{Cov}(m_{k,t,1}, m_{k,t+1,1})}{\mathrm{Cov}(m_{k,t,2}, m_{k,t+1,1})} = \alpha_{k,t,2}$$

- With more than two measures per factor, this approach can be generalized to allow for serial correlation in the $\epsilon_{k,t,j}$
    - Need $2M + 1$ measures to identify factor loadings for M factors

Identification: Distribution of Latent Variables

- Once $\alpha_{k,t,j}$ are identified, we can re-write the measurement system as

$$\frac{m_{k,t,j}}{\alpha_{k,t,j}} = \theta_{k,t} + \frac{\epsilon_{k,t,j}}{\alpha_{k,t,j}}$$

- The signal is common to multiple measurements, but the noise is not
- To extract the noise from the signal, assume that the $\epsilon_{k,t,j}$:
    1. Are uncorrelated across measures conditional on θ
    2. Are independent from factors

- Then we can identify the joint distribution of $\{\theta_{k,t}\}_{t=1}^{T}$

- Cunha et al (2010) also discuss identification under the more general case with unknown mapping between measures and factors and non-separable measurement errors

The Investment Problem

- Parents choose investments by solving

$$\max_{\{C_t, I_{t+1}\}} U(C_t, \boldsymbol{H}_{t+1})$$
$$s.t.\ C_t + \boldsymbol{P}_t^I I_{t+1} = Y_t$$
$$\boldsymbol{H}_{t+1} = g_t(\boldsymbol{H}_t, \boldsymbol{X}_t, \boldsymbol{I}_{t+1}, u_{t+1})$$

- Investment could be in commodities or in time
- Could easily add parental labor supply decisions
- Could easily add other inter-temporal dimensions beyond HK

Endogeneity of Investment

- Parents choose investment in reaction to shocks, and so $E(u_{t+1}|I_{t+1}) \neq 0$
- Don't assume the parents know the production function, hence derive a reduced-form of the investment function

$$\boldsymbol{I}_{t+1} = f_t(Y_t, \boldsymbol{P}_t, \boldsymbol{X}_t, e_{t+1})$$

- Notice that Y_t and $\boldsymbol{P}_t$ are excluded from the production function
- One can estimate the investment equation jointly with the production function
- Or use a control function approach to estimate the role of investments in the production function
  - Plug estimated residuals $\hat{e}_{t+1}$ as additional covariates in the production function, so that $E(u_{t+1}|I_{t+1}, \hat{e}_{t+1}) = 0$

Estimation

- Estimation can be performed in two-steps
  1. Estimate the joint distribution of all latent factors
  2. Draws from such a distribution to estimate the investment equation as well as the production function with control function terms
- Inference can be done by using appropriate bootstrapping
  - This takes into account both estimation error at each stage and simulation error
  - Block-bootstrap with cluster at the municipality-level

Part 4: Dynamic Latent Factor Models & RCTs

- Context
  - A randomized early childhood intervention in Colombia
- Methods
  - Cunha, Heckman and Shennach (ECMA, 2010)
- Application
  - **Attanasio et al. (AER, 2020)**

The Intervention and HK Accumulation

- We can view the intervention as potentially affecting outcomes through three different channels
  1. Total factor productivity
  2. The productivity of the individual inputs
  3. The distribution of inputs (and in particular parental investment)
- Attanasio et al (2020) jointly estimate the distribution of inputs and the production function itself to test for these different hypotheses

Measurement equations

$$m^{\theta}_{j,d} = \mu^{k}_{j,d} + \alpha^{k}_{j} \ln \theta^{k}_{d} + \epsilon^{k}_{j}$$

- Treated and control children are allowed to have different factor distributions
- Latent factors are distributed as a mixture of two normal distributions
- Measurement errors terms are jointly normal
- Likelihood function for each group writes as

$$L_i(m) = \int_{\theta} f(m_i|\theta)[\tau\phi_A(\theta) + (1-\tau)\phi_b(\theta)]d\theta$$

- Each component is just a normal distribution with an analytical expression, where $f(m_i|\theta)$ is the density of the measurement system

The Production Function for Human Capital

- They consider two production functions
  1. One for the production of cognitive skills ($k = C$)
  2. One for the production of non-cognitive skills ($k = S$)

- Cobb-Douglas specification for the technology of skill formation

$$\ln(\theta^k_{i,d,t+1}) = A^k_d + \gamma^k_{1,d}\ln(\theta^C_{i,t}) + \gamma^k_{2,d}\ln(\theta^S_{i,t}) + \gamma^k_{3,d}\ln(P^C_{i,t}) + \gamma^k_{4,d}\ln(P^S_{i,t}) \\ + \gamma^k_{5,d}\ln(I^M_{i,t+1}) + \gamma^k_{6,d}\ln(I^T_{i,t+1}) + \gamma^k_{7,d}\ln(n_{i,t}) + \eta^k_{i,t+1}$$

- Use the estimates of the measurement system to predict Bartlett factor scores for each individual in the data: $\hat{\theta}^k_{i,t}, \hat{\theta}^k_{i,t}$
- All parameters are skill-specific and they are allowed to depend on treatment status d

Parental Investments

- Same log-linear specification for investment equation ($\tau = \{M, T\}$)

$$\ln(I^{\tau}_{i,t+1}) = \lambda^{\tau}_{0,d} + \lambda^{\tau}_{1,d}\ln(\theta^{C}_{i,t}) + \lambda^{\tau}_{2,d}\ln(\theta^{S}_{i,t}) + \lambda^{\tau}_{3,d}\ln(P^{C}_{i}) + \lambda^{\tau}_{4,d}\ln(P^{S}_{i}) + \lambda^{\tau}_{5,d}\ln(n_{i,t}) + \lambda^{\tau}_{6,d}Z_{i,t} + u_{i,t+1}$$

- Z_t are instruments (the average log price of toys, the average log price of food items in the municipality of residence, and past maternal exposure to conflict)
- All parameters are allowed to vary with the treatment, d
- Control function approach to account for the correlation between shocks and investments
  - Include residuals $\hat{u}_{i,t+1}$ as additional regressor in the production function

Evidence Supporting the Exclusion Restrictions

TABLE 3—BALANCE TEST FOR THE INSTRUMENTAL VARIABLES

	log toy price	log food price	Conflict
Mother's cognitive skill	0.021	0.007	−0.004
	(0.009)	(0.007)	(0.003)
Mother's socio-emotional skill	−0.006	−0.014	0.005
	(0.008)	(0.006)	(0.005)
Mother is married	0.013	0.030	0.017
	(0.016)	(0.012)	(0.008)
Wealth index	−0.003	0.002	0.002
	(0.007)	(0.006)	(0.004)
Terrorism	0.015	−0.012	0.005
	(0.014)	(0.013)	(0.009)
Constant	8.025	8.062	0.049
	(0.028)	(0.017)	(0.007)
Observations	1,010	1,023	1,023
R^2	0.011	0.021	0.017
F-statistic	1.656	2.147	1.438
F-test p-value	0.154	0.0671	0.219

Notes: All right-hand-side variables measured at baseline. Asymptotic standard errors in parentheses allowing for clustering at the municipality level. *Terrorism* is the number of terrorist attacks between conception of child and baseline. *Conflict* is the number of conflicts against civil population divided by the municipality population (in thousands) when the mother herself was a child.

Results: Investment functions

Table 5—Estimates of the Material and Time Investment Equations

	Instruments: prices and conflict		Instruments: prices only
	Material investment	Time investment	Material investment
Intercept	−0.015 [−0.114, 0.078]	0.001 [−0.089, 0.089]	−0.013 [−0.11, 0.078]
Treatment	0.209 [0.038, 0.365]	0.318 [0.155, 0.48]	0.204 [0.037, 0.362]
log child's cognitive skill (t)	0.130 [0.016, 0.246]	0.068 [−0.044, 0.18]	0.132 [0.017, 0.25]
log child's socio-emotional skill (t)	−0.028 [−0.133, 0.087]	0.027 [−0.083, 0.145]	−0.030 [−0.131, 0.088]
log mother's cognitive skill	0.748 [0.582, 0.939]	0.349 [0.162, 0.498]	0.750 [0.583, 0.943]
log mother's socio-emotional skill	0.069 [−0.008, 0.139]	0.022 [−0.06, 0.108]	0.068 [−0.008, 0.139]
log number of children	−0.129 [−0.18, −0.077]	−0.128 [−0.186, −0.072]	−0.128 [−0.18, −0.078]
log toy price	−0.096 [−0.168, −0.027]	−0.020 [−0.085, 0.037]	−0.094 [−0.163, −0.026]
log food price	0.091 [0.006, 0.178]	0.042 [−0.026, 0.121]	0.091 [0.006, 0.178]
Maternal childhood exposure to conflict	−0.009 [−0.08, 0.063]	−0.089 [−0.139, −0.032]	

Results: Production function of cognitive skills

TABLE 6—ESTIMATES OF THE PRODUCTION FUNCTION FOR COGNITIVE SKILLS

Instruments:	OLS	IV			
		Prices, conflict	Prices, conflict, treatment	Prices	Prices and treatment
	(1)	(2)	(3)	(4)	(5)
Intercept	−0.018 [−0.094, 0.053]	−0.019 [−0.111, 0.079]	0.007 [−0.089, 0.104]	0.003 [−0.091, 0.072]	−0.009 [−0.083, 0.058]
Treatment	0.083 [−0.025, 0.192]	0.049 [−0.12, 0.391]		−0.028 [−0.186, 0.156]	
log child's cognitive skill (t)	0.675 [0.589, 0.77]	0.648 [0.544, 0.795]	0.638 [0.522, 0.747]	0.626 [0.525, 0.746]	0.631 [0.533, 0.739]
log child's socio-emotional skill (t)	0.001 [−0.091, 0.087]	0.012 [−0.098, 0.143]	0.019 [−0.094, 0.14]	0.015 [−0.084, 0.126]	0.02 [−0.079, 0.127]
log mother's cognitive skill	0.213 [0.089, 0.35]	−0.075 [−0.456, 0.5]	−0.173 [−0.538, 0.201]	−0.102 [−0.495, 0.291]	−0.094 [−0.45, 0.21]
log mother's socio-emotional skill	0.103 [0.031, 0.173]	0.084 [−0.03, 0.163]	0.063 [−0.035, 0.151]	0.06 [−0.019, 0.152]	0.074 [−0.026, 0.155]
log number of children	0.042 [−0.01, 0.092]	0.085 [−0.07, 0.154]	0.084 [0.011, 0.163]	0.089 [0.002, 0.176]	0.086 [0.023, 0.164]
log material investment	0.088 [0.016, 0.157]	0.594 [0.025, 1.179]	0.784 [0.204, 1.383]	0.542 [0.041, 0.996]	0.516 [0.195, 0.946]
log time investment	0.038 [−0.051, 0.129]	−0.171 [−1.198, 0.312]	−0.311 [−0.985, 0.217]		
Goodness-of-fit: Gap in output between treated and control					
Measured in the data	0.115	0.115	0.115	0.115	0.115
Predicted by the model	0.114	0.132	0.084	0.095	0.117

Results: Production function of socio-emotional skills

Table 7—Estimates of the Production Function for Socio-Emotional Skills

Instruments:	OLS	IV			
	(1)	Prices, conflict (2)	Prices, conflict (3)	Prices, conflict, treatment (4)	Prices, conflict, treatment (5)
Intercept	−0.009 [−0.08, 0.063]	−0.02 [−0.087, 0.064]	−0.006 [−0.084, 0.066]	−0.02 [−0.096, 0.058]	−0.022 [−0.089, 0.051]
Treatment	−0.011 [−0.124, 0.093]	−0.071 [−0.318, 0.166]	−0.116 [−0.31, 0.115]		
log child's cognitive skill (t)	0.106 [0.018, 0.192]	0.074 [−0.019, 0.21]	0.09 [−0.017, 0.194]	0.094 [0.003, 0.22]	0.099 [0.002, 0.193]
log child's socio-emotional skill (t)	0.522 [0.403, 0.659]	0.513 [0.374, 0.672]	0.499 [0.389, 0.663]	0.516 [0.387, 0.669]	0.512 [0.403, 0.656]
log mother's cognitive skill	−0.077 [−0.217, 0.049]	−0.146 [−0.443, 0.297]	−0.15 [−0.349, 0.084]	−0.018 [−0.315, 0.328]	−0.083 [−0.231, 0.078]
log mother's socio-emotional skill	0.037 [−0.058, 0.119]	0.035 [−0.062, 0.141]	0.048 [−0.053, 0.135]	0.049 [−0.054, 0.134]	0.043 [−0.042, 0.126]
log of number of children	0.099 [0.047, 0.153]	0.127 [0.017, 0.236]	0.133 [0.034, 0.223]	0.101 [0.028, 0.168]	0.101 [0.036, 0.163]
log material investment	0.154 [0.06, 0.256]	0.015 [−0.621, 0.428]		−0.144 [−0.688, 0.34]	
log time investment	0.109 [−0.006, 0.213]	0.487 [−0.177, 1.258]	0.549 [−0.022, 1.147]	0.448 [−0.125, 1.057]	0.324 [0.025, 0.691]
Goodness-of-fit: Gap in output between treated and control					
Measured in the data	0.087	0.087	0.087	0.087	0.087
Predicted by the model	0.057	0.080	0.050	0.103	0.098

Interpreting the impact of the intervention

- Cognitive development among children aged 3-4 is mainly determined by
  - Initial cognition (measured at ages 1-2)
  - Stimulation provided by older siblings
  - Material investments provided by parents
- Socio-emotional development is mainly determined by
  - Initial socio-emotional skills (measured at ages 1-2)
  - Stimulation provided by older siblings
  - Time investments provided by parents
- The main channel through which the intervention affects child development is by shifting parental investments
  - Measurement issue with (self-reported) time investments

Complementarities between Inputs

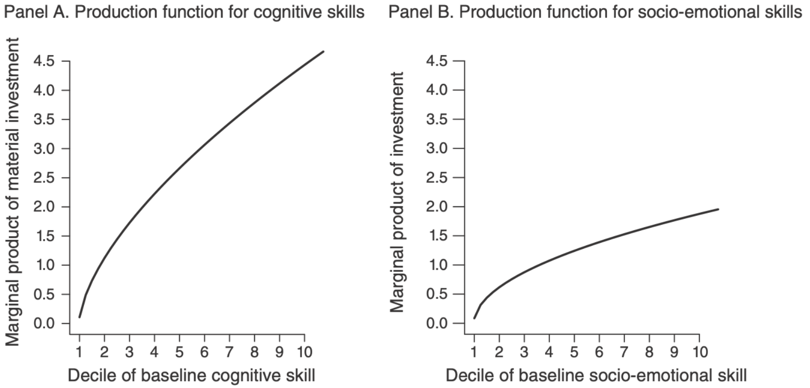


FIGURE 2. COMPLEMENTARITY BETWEEN INVESTMENTS AND BASELINE SKILLS

- Return to investment is higher for children with better initial conditions
- Also, mothers with higher levels of cognition invest more in their children
- Vicious cycles for social inequalities

Implications for longer-term outcomes

- Parental crowding-in of material and time resources suggest treatment group will diverge in the medium/long run
- But also possible fade-out due to mean reversion
- Additional data collection two years after the end of the intervention suggests partial fade out (and same parental investments as control group)
- How can we permanently improve parental practices ?

Wrapping up on Factor Models and RCTs

- How would the paper look like without the RCT?
- Model here provides a roadmap to interpret of how the intervention affected child development
- Sometimes you deliberately don't want to use the experimental variation as this is what you actually want to explain
- Need additional sources of variation beyond the experiment