Did the apple fall far from the tree?

Dynamics of education and labor market inequality with ability uncertainty and family-informed priors

Andrew Proctor

Stockholm School of Economics

August 17, 2021

Summary

Aim:

- Examine the effect of uncertainty and learning about ability on intergenerational correlations in education and labor market outcomes, when children use family signals to form initial beliefs about their ability.

Approach:

- Use a cohort study and registry data to estimate a dynamic discrete choice model of education and employment, incorporating:
  - Multidimensional skills and ability endowments
  - Uncertainty and learning about ability, starting from a family-driven prior

Introduction

Intergenerational persistence

- Understanding the sources of intergenerational correlations in education and income is a central question in economics
  - Often focused on decomposing persistence into genetic and environmental components, i.e. **"Nature" vs. "Nurture"**
  - Twin and family studies in behavioral genetics and economics suggest that genetic factors account for $\approx$ 40-60% of variation in cognitive ability (*Smith-Woolley et al 2018*).
  - Lay perception about heritability similar at 58% in survey by *Willoughby et al (2018)*
- Standard theory models of intergenerational persistence since *Becker and Tomes (1979)* predict correlations arise due to:
  1. a direct affect of correlated ability on education and income,
  2. an endogenous human capital investment decision, which depends on parental income (and hence ability)

Introduction

Integrating family priors with DDC labor models

- But substantial evidence that **ability uncertainty is important for human capital decisions** (e.g. Stinebrickner & Stinebrickner 2012, Arcidiacono et al 2017) and that **family matters for beliefs** (Chowdry et al 2010, Dizon-Ross 2019)
- What are rational household beliefs about child ability when it's not observable?
  - Given that much of ability is heritable, ability of family members are valuable initial signals of child ability.
  - As "children" learn, through education and labor market experience, the role of family-driven priors should diminish (similar to Farber & Gibbon 1996, Altonji & Pierret 2001)
  - But while uncertainty is being resolved, RE Bayesian ability beliefs are biased towards parental/family signal.

Data

- Combine two general sources of Swedish data:
  - **Evaluation through Follow-Up (ETF) Study** (1992 cohort):
    - A multiple cohort, longitudinal study that provides rich data on educatonal achievement, beliefs, and goals for children (~10,000, nationally representative), with household interviews at multiple points.
    - For 1992 cohort, interviewed at 6 (*Mellanstadiet*), grade 9 (*Högstadiet*), and grade 12 (*Gymnasieskola*)
  - **Administrative "registry" data** for the entire population:
    - *Education:* schooling applications and enrollment, grade and credit completion data, financial aid.
    - *Ability:* cognitive and non-cognitive ability measures from military draft assesments.
    - *Employment:* occupation and income, unemployment, etc.
    - *Family linkages:* family members linked using multigenerational register.

General strategy

- Estimate a DDC model of education and employment, from lower secondary school to early career
  - General structure similar to *Keane and Wolpin*
  - Extends model to include *Arcidiacono et al (2017)* uncertainty about ability endowments, with Bayesian beliefs and correlated learning.
  - Major points of departure from *Arcidiacono et al (2017)*:
    - Individuals look to markers of ability for immediate family (parents, elder siblings) to form initial beliefs about their own ability.
    - Several employment sectors (similar to *Sullivan 2010*), so ability beliefs formed over education-occupation paths that aren't just vertically differentiated.

Outline of choices

Individuals (and their families) make a series of decisions about education and employment.

- In schooling:
  - Compulsory schooling: Household effort (frequency of child study and parental help)
  - Upper secondary: Whether or not to enter/complete studies, choice of university prep or vocational tracks.
  - After high school, whether to enter/complete:
    - Technical and vocation training
    - STEM or non-STEM Bachelors
  - After university, whether to enter/complete graduate degree.
- In the labor market (full or part time):
  - Whether to look for a job in different occupational sectors
  - Whether to switch jobs
- Individuals can:
  - Move from schooling to work or work to schooling
  - Participate in both school and work at the same time
  - Stay home (participate in neither)

Educational performance in lower secondary education

Achievement Test scores (grades 3,6,8,9) are a function of:

- Child and Parental Effort (from ETF), E_{it}
  - How often/how many hours do you study each week
  - How often do you get help at home with school work?
- Age, age_{it}
- (Unobserved) Ability, A_{it}
  - *Note: for all processes, parameter on ability is constrained to 1 in initial periods*
- Idiosyncratic shocks, $\epsilon_{it} \sim N(0, \sigma^2_{it})$

$$G_{it} = \gamma + \gamma_1 E_{it} + \gamma_2 age_{it} + \gamma_3 A_{it} + \epsilon_{it}$$

Educational performance in upper secondary education

▸ Upper secondary schooling details

GPA is a function of:

- Child and Parental Effort
- Age
- Ability
- Idiosyncratic shocks

$$G_{it} = \gamma + \gamma_1 E_{it} + \gamma_2 age_{it} + \gamma_3 A_{it} + \epsilon_{it}$$

Educational performance in higher education

Credit completion rate is a function of:

- ***Degree type*** (STEM or non-STEM), m
- Labor market participation (worked part/full-time), L_{it}
- Age
- Ability
- Idiosyncratic shocks

$$G_{imt} = \gamma_m + \gamma_1 L_{it} + \gamma_2 age_{it} + \gamma_3 A_{imt} + \epsilon_{it}$$

Employment overview

- Two types of employment: full-time or part-time
- 6 Occupational sectors, ℓ:
  - Elementary occupations (ISCO Skill Level 1)
  - Clerical, service and sales workers (Skill Level 2)
  - Craft and trades, skilled agricultural workers (Skill Level 2)
  - Non-STEM Technicians and Associate Professionals (Skill Level 3)
  - STEM Technicians and Associate Professionals (Skill Level 3)
  - Professional or managerial workers (Skill Level 4)
- Workers face a job offer probability that varies by state & type

Wage process

Log wage is a function of:

- Observable characteristics, X_{ilt}
  - Highest educational degree
  - Full-time/part-time employment status
  - Occupational experience
  - Ability measures
  - Immigrant status
  - Degree type (for skilled occupations)
- Occupation-specific ability, A_{il}
- Sector-specific time dummies, δ_{lt}
- Idiosyncratic shocks, $e_{ilt} \sim N(0, \sigma_l^2)$

$$w_{ilt} = \delta_{lt} + \gamma_1 X_{ilt} + \gamma_3 A_{il} + \epsilon_{ilt}$$

Ability

- Ability, A_i is a multi-dimensional vector consisting of ability in each educational stage and employment sector.
  - Assumed to be distributed multivariate normal, with unconstrained covariance matrix, Δ
- Individuals assumed to be rational, rely on Bayesian updating about beliefs.
  - Individuals are assumed not to know their own ability (*...how far away did the apple fall?*),
  - But for formation of initial beliefs, they're assumed to know the expected ability of people sharing their observable characteristics.

Example: ability in university

From the assumed university grade process:

$$G_{imt} = \gamma_m + \gamma_1 L_{it} + \gamma_2 age_{it} + A_{imt} + \epsilon_{it}$$
$$\hat{A}_{imt} + \hat{\epsilon}_{it} = G_{mt} - (\hat{\gamma}_m + \hat{\gamma}_1 L_{it} + \hat{\gamma}_2 age_{it})$$

- Think of $\hat{A}_{imt} + \hat{\epsilon}_{it}$ like a composite residual.

Without additional info, unconditional (prior) expectation is: $\mathbb{E}[A_{it}] = 0$

- But initial belief can be made better by conditioning on signals of ability, z_{it}
  - High school achievement: test scores and GPA
  - Assessed ability from draft
  - Own labor market outcomes (if working)
  - Parent and older sibling signals: education, occupation and wages, male relatives' ability scores from draft

Since we assume: $\mathbb{E}[A_{it} + \epsilon_{it}) \mid z_{it}] = \mathbb{E}[A_{it} \mid z_{it}]$

- Can predict ability conditional on observable signals via auxiliary regression.

Belief updating

In periods after the first in a given pursuit, individuals receive a signal of ability from their own performance:

- Compound error for previous period $\left[G_{m,t-1} - (\hat{\gamma}_{sm} + \hat{\gamma}_1 L_{i,t-1} + \hat{\gamma}_2 age_{i,t-1})\right]$ is an unbiased signal of A_{it}
- Update according to Bayes rule for prior belief and MVN distributional assumption.

Utility

- Individuals are forward looking and choose the sequence of education (j) and labor market (k) decisions (d_{it}) that maximizes the present value of expected lifetime utility.

$$V_t \equiv \max \mathbb{E} \left[\sum_{t=1}^{T} \beta^{t-1} \sum_j \sum_k (u_{jk}(Z_{it})) + \epsilon_{ijkt}) \mathbb{1}\{d_{it} = (j,k)\} \right]$$

- Where:
  - $Z_{it} = (Z_{1it}, Z_{2it})$ denotes the variables that affect the utility of schooling and work respectively.
  - β is the discount rate

Components of utility

- In both schooling and work components:
  - Demographics
  - Ability measures
  - Controls for previous choice (switching costs)
- Unique Aspects for schooling or work:
  - Schooling:
    - Expected ability in schooling option j
    - *For higher education:* financial resources (student aid, parental consumption transfers)
  - Work: Expected log wages in sector k
- Utility weights are also allowed to vary by latent class, producing type-specific unobserved heterogeneity.
- The home sector is set as the reference sector, hence utility of this option is normalized to 0.

Estimation overview

Estimation details

- Solving the dynamic programming problem is simplified by using CCP estimation methods *(Hotz and Miller 1993)* rather than full solution methods.
- Following *Arcidiacono et al 2017*, permanent unobserved heterogeneity in preferences and ability is incorporated using latent types.
    - With unobserved types, sequential estimation of production and utility parameters fails because likelihood function is no longer additively separable between the two.
    - EM algorithm is applied to restore separability *(Arcidiacono and Jones 2003, Arcidiacono and Miller 2011)*

Counterfactual policy analysis

Consider the effect of different counterfactual experiments and policy changes:

- **Change information:** No uncertainty in ability
- **Change constraints:**
  - Change (*impose*) cost of education
  - Change household financial resources during education (eg move family from 1^{st} to 3^{rd} pctile)
- **Change market structure**:
  - Change tracking options in upper secondary school.

Comparing between benchmark and counterfactual equilibria the:

- distributional outcomes for education and employment
- intergenerational correlations in outcomes

... thank you!

Appendix

Details of upper secondary schooling

Back to educational performance specification

- About 90% of students enter upper secondary schooling after finishing compulsory (lower secondary) schooling.
- Upper secondary school divided into **3 types of tracks**, which at this time consisted of *(Mellander 2017)*:
  1. University prep "theory" tracks (3 or 4 years), e.g. Humanities, Sciences, or Social sciences
  2. Two-year "theory" tracks, with limited college access
  3. Vocational tracks (two years), e.g.:
     - car mechanic
     - operation of industrial processes;
     - forestry and farming.

Upper secondary school reforms

- **1995 Reform:**
  - Changed course content so that all tracks provided basic eligibility to university.
- **2011 Reform:**
  - Largely reversed 1995 Reform,
  - Increased apprenticeship component to vocational tracks,
  - Raised admission requirements for upper secondary schooling.

Recursive formulation of value function

Back to estimation overview

- The value function can be re-expressed as a Bellman equation:

$$V_t = \underbrace{u_{jk}(Z_{it}) + \epsilon_{ijkt}}_{\text{flow utility}} + \underbrace{\beta \mathbb{E}\left[V_{t+1}(Z_{i,t+1} \mid Z_{it}, d_{it} = (j,k))\right]}_{\text{continuation value}}$$

- Define the ex ante value function, $\bar{V}_t$ as the expected value of the value function at the beginning of time t, before ϵ_{ijkt} is revealed:

$$\bar{V}_t(Z_{i,t}) = \int V_t f(\epsilon) d\epsilon_t$$

- The expected discount PV of utility of a given choice at time, t, conditional on the history until t, can be expressed as:

$$v_{jkt}(Z_{it}) = u_{jk}(Z_{it}) + \beta \mathbb{E}\left[\bar{V}_{t+1}\left(Z_{i,t+1} \mid Z_{it}, d_{it} = (j,k)\right)\right]$$

Conditional choice probability

Assuming Type 1 GEV Errors, we get the following expression for the *conditional value function*, v_{jkt}:

$$v_{jkt}(Z_{it}) = u_{jk}(Z_{it}) + \beta\mathbb{E}[\ln(\sum_j \sum_k \exp(v_{jk,t+1}(Z_{i,t+1})) \mid Z_{it}, d_{it} = (j,k))]$$

The optimal choice at time t is then:

$$\delta_t(Z_{it}, \epsilon_{ijkt}) = \arg\max_d \left[v_{jkt}(Z_{it}) + \epsilon_{ijkt}\right]$$

Integrating out over the error distribution delivers the *conditional choice probability*: the probability of choice $d_{it} = (j,k)$ conditional on the state Z_{it}

$$p_t(d_t|Z_{it}) = \int \left\{\arg\max_d \left[v_{jkt}(Z_{it}) + \epsilon_{ijkt}\right] = d_t\right\} f(\epsilon_{ijkt})d\epsilon_{ijkt}$$

Sequential approach

With Type 1 GEV idiosyncratic shocks that are uncorrelated *and no type-specific unobserved heterogeneity*, the likelihood function is additively separable in choices and outcomes and can be estimated sequentially, as follows:

- **First stage:** estimate ...
  - parameters for educational performance and wage equations,
  - choice probabilities associated with choice alternatives
- **Second stage:** estimate ...
  - Flow utility parameters, taking first-stage as given

Restoring separability with type-specific unobserved heterogeneity

- Likelihood function is not additively separable in choices and outcomes (wages, grade performance) with type specific heterogeneity,
- To restore additive separability, apply **Expectation Maximization (EM) algorithm** to restore separability. Iteratively perform:
  - E-Step
    - Take as given parameters from outcome equations
    - Update posterior ability distribution from the observed outcome data
  - M-Step
    - Take as given posterior ability distribution from E-Step, maximize log-likelihood of the outcome equations