Intergenerational economic mobility : estimation and identification

Arnaud Lefranc
Université de Cergy-Pontoise, THEMA

Canazei IT 6 Winter School, January 2011

Introduction

- Focus of the talk : empirical assessment of intergenerational economic mobility (nuts and bolts mostly)
- Framework : Galtonian intergenerational earnings regression model

$$Y_i = \beta X_i + e_i$$

  where
  - Y and X denote the log of child's and father's earnings
  - β is the intergenerational elasticity (IGE) in earnings.
- Four questions :
  1. How to interpret β
  2. How to estimate β
  3. How to identify the structural determinants of β
  4. How to change β

Outline

Theoretical model

- Simplified version of Solon (2004) inspired by Becker and Tomes (1979)
- Notations
  - Y_{it} : earnings of generation t in family i
  - C : consumption
  - I_{it-1} : investment by generation t in the human capital of generation t
- Human capital accumulation :

$$H_{it} = \theta \log(I_{it-1}) + e_{it}$$

  where e_{it} denotes individual ability
- (Mincerian) Earnings function :

$$logY_{it} = \mu + pH_{it-1}$$

  where p denotes the returns to human capital

Family choice and the IGE

- Generation $t-1$ has Cobb-Douglas preferences over own consumption and child's income
$$U_{it-1} = C_{it-1}^{1-\alpha} Y_{it}^{\alpha}$$
- Optimal human capital investment :
$$I_{it-1} = \frac{\alpha\theta p}{1-\alpha(1-\theta p)} Y_{it-1}$$
- Implied intergenerational earnings transmission
$$\log Y_{it} = constant + \theta p \log Y_{it-1} + pe_{it}$$
- Intergenerational transmission of ability : and follows an AR(1) process :
$$e_{it} = \delta + \lambda e_{it-1} + \nu_{it}$$
- Earnings follow an AR(1) process with AR(1) disturbance, which imply the reduced form IGE :
$$\beta = \frac{\theta p + \lambda}{1 + \theta p \lambda}$$

Implications

- No structural interpretation of β
- Descriptive appeal : reduced form estimate of intergenerational transmission (Solon : omnibus measure).
- Normative appeal ? vaguely ... speed of regression to the mean
- Model also identifies key aspects of intergenerational transmission : educational transmission (θ), labor market (p), "mechanical" transmission of ability (λ)

Estimation issues

- Ideally, estimating the IGE requires the linked observation of both parents and children's permanent earnings. This is not available in most countries.
- Three main issues in usual data sets:
  - Classical measurement error
  - Life-cycle bias
  - Lack of direct information on parental income

Measurement error I

- Problem : we only observe father's current earnings $\tilde{X}_t$
  - assume $\tilde{X}_t = X + u_t$ and measurement error u_t is of the classical type
  - standard attenuation bias (Griliches, 1986):

$$plim \beta_{OLS} = \beta \frac{V(X)}{V(X) + V(u)} = \beta\lambda < \beta$$

- Solution 1 : averaging income over multiple periods
  - Replace $\tilde{X}_t$ by $\bar{X} = \frac{1}{T}\sum_1^T \tilde{X}_t$
  - we get a lower attenuation factor

$$\lambda = \frac{V(X)}{V(X) + V(u)/T}$$

  - Solon (1992), Mazumder (2002)
- Solution 2 : IV (see below)

Measurement error II

- Problem : life-cycle bias
  - earnings growth-rate is positively correlated with earnings levels
  - assumption of classical measurement error is invalid
  - letting a and a' denote the age at which father and child's income are observe, assume instead :

$$\begin{aligned} \tilde{X}_a &= \mu_a X \\ \tilde{Y}_{a'} &= \gamma_{a'} Y \end{aligned}$$

  - μ and γ increase with age
  - OLS on current earnings provide biased estimates :

$$plim \beta_{OLS} = \beta \frac{\mu_a}{\gamma_{a'}}$$

- Solution 1 : rule of thumb of Haider and Solon = age 40 ($\mu_{40} = \gamma_{40} \simeq 1$)
- Solution 2 : net out age effects from the earnings model

TSIV estimation

- Problem : father's earnings are not observed but only some characteristics of the father
- Solution : Two-samples instrumental variables (TSIV)
  - First-step estimation : predict father's earnings on the basis of fathers characteristics, using a sample of the fathers
  - Second-step estimation : regress child's earnings on predicted father's earnings
- Properties :
  - as good as IV (Angrist and Krueger, 1995) and depends on the properties of the instrument
  - "solves" the classical measurement error issue
  - most papers use "instruments" that have an independent (positive) effect on the explained variable : positive bias but small in practice (Björklund and Jäntti, 1997)

Objective and motivations

Objective :

- assess changes in intergenerational earnings mobility in France across cohorts
- for this : estimate cohort-specific IGEs for cohorts born between 1931 and 1975
- IGE are derived from TSIV estimations and attention is limited to father-son pairs

Motivations :

- descriptive.
- analytic : How is IGE affected by changes in the overall economic environment ?
  Two major changes :
  - large rise in acess to education
  - fall in earnings inequality
- comparative : recent series of paper on changes in IGE,
  e.g US : Fertig (2003), Mayer and Lopoo (2004, 2005), Hertz (2006), Solon and Lee (2007), Aaronson and Mazumder (2008); UK : Blanden *et al.* 2002, Ermisch and Nicoletti (2005); etc

Main model

- The cohort-specific IGE are derived from the following equation

$$Y_{ict} = \alpha_t + \beta_c \hat{X}_{ic} + g(age_{ict}) \times \hat{X}_{ic} + f_c(age_{ict}) + e_{ict}$$

  - i, c, t are indices for individual, cohort and date
  - f_c and g : 4^{th} order polynomial functions;
  - $\hat{X}_{ic}$ is predicted father's earnings at age 40
  - age is normalized to zero at age 40
    $\Rightarrow \beta_c$ denotes the IGE at age 40 for all cohorts

Auxiliary (first-step) model

- The prediction of father's earnings is based on father's education
- The model allows for heterogeneity by cohorts in the effect of education and heterogeneity by education in age-earnings profiles :

$$X_{ict} = \alpha_t + \sum_j \gamma_c^j Educ_{ic}^j + f_c(age_{ict}, Educ_{ic}) + e_i$$

  - $Educ_{ic}^j$ is a set of education dummies
  - $f_c(age_{ict}, Educ_{ic})$ is a fourth polynomial in age, specific to each level of education and that varies across father's cohorts
  - age_{ict} is centered at age 40
- The equation is estimated on the sample that is representative of the father's cohorts
- $$\hat{X}_{ict} = \sum_j \hat{\gamma}_c^j Educ_{ic}^j$$

Data

The Education, Training and Occupation survey

- collected by the French National Statistical Institute
- 6 waves between 1964 and 2003.
- sample sizes btw/ 15000 and 25000
- **respondent's characteristics** :
  - annual earnings in previous year, # of month worked full- and part-time
  - family characteristics : number of children
- **father's characteristics** (waves 1970-2003)
  - education : highest degree - 6 groups (none, primary, general lower secondary, vocational lower secondary, upper secondary, higher education).

Data (continued)

Samples selection rules

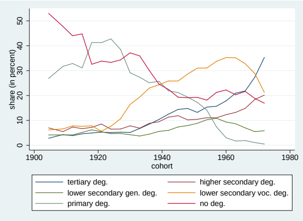
- **children samples**
  - male head of household, aged 28 to 50 years old at survey date
  - grouped into nine 5-years cohorts : [1931-1935] to [1971-1975]
  - exclude self-employed children as well as children whose father was self-employed
- **auxiliary samples ("pseudo-fathers")**
  - male heads of household, aged 25 to 60 years old at survey date, who report at least one child, and are not self-employed

Matching children and fathers cohorts

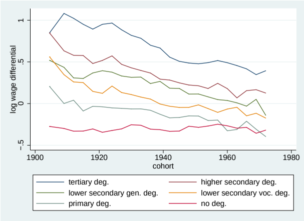
- for each cohort I know the distribution of their fathers' birth cohort
  $\Rightarrow$ cohort matching is based on this distribution

First-step results and trends

Returns to education

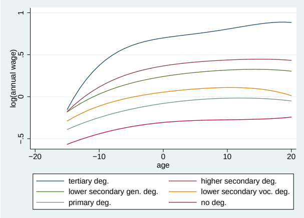


Distribution of education



Trends in inequality

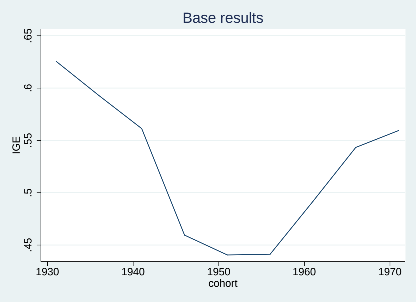
Age-earnings profiles



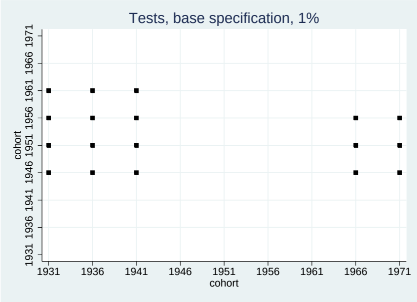
Main results

Figure: Cohort specific IGEs and tests

A- cohort specific IGEs



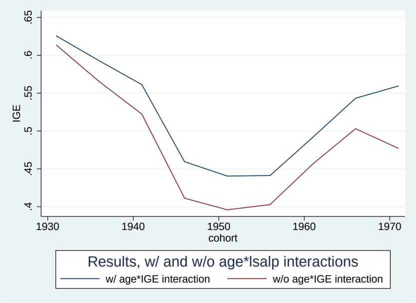
B- Unilateral 1%tests



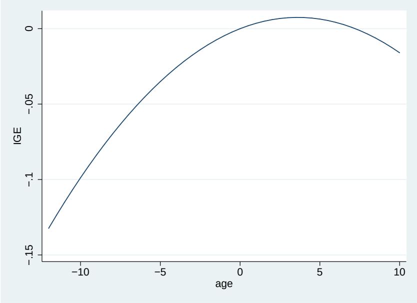
Note: in panel B, squares indicate that the estimated IGE for the row cohort is significantly lower than the IGE for the column cohort.

Influence of lifecycle biases

A- IGE w/ and w/o age interactions



B- interaction effect age*X



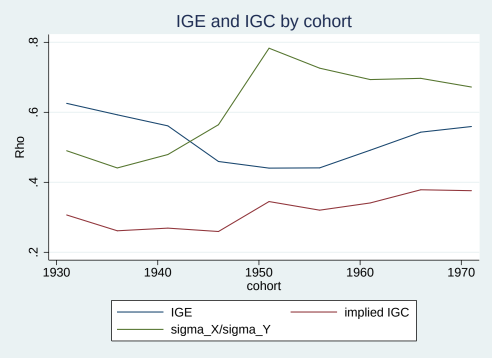
Correlation vs. elasticity

- Intergenerational (positional) mobility can also be measured the intergenerational correlation coefficient (IGC) ρ
- Link between the IGC and the IGE :

$$\beta = \rho \frac{\sigma_Y}{\sigma_X}$$

- If $\frac{\sigma_Y}{\sigma_X}$ falls due to a reduction in inequality among children, the IGE will "mechanically" fall as well.

Correlation vs. elasticity (ctd)



Comments : initial fall in the IGE may be driven by the fall in cross-section earnings inequality among children but the subsequent rise reflects an decrease in positional mobility

Assessing the contribution of education

- Let H_{ic} denote the human capital of child i in cohort c.
- Consider the following simplified intergenerational transmission model with two channels (education and residual) :

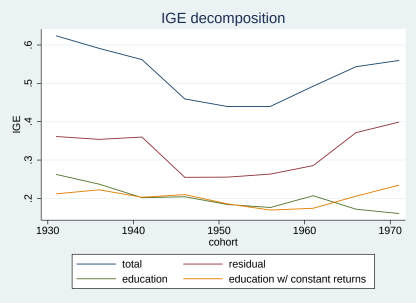
$$\begin{aligned} Y_{ic} &= \beta_c^1 H_{ic} + \beta_c^2 X_{ic} + e_{ic} \\ H_{ic} &= \gamma_c X_{ic} + u_{ic} \end{aligned}$$

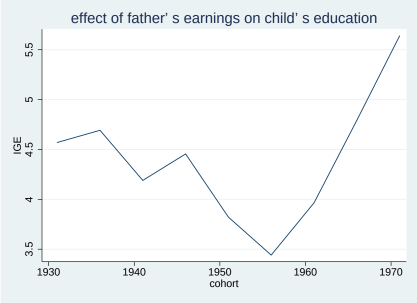
  - father's income can influence both child's education and the residual
- The reduced form IGE is given by :

$$\beta_c = \beta_c^1 \gamma_c + \beta_c^2$$

  - three components : the residual impact of father's earnings, the association between father's earnings and education and the returns to education.
  - Separating these three effects is key to assessing the contribution of educational systems to changes in the IGE

Assessing the contribution of education





Results :

- large share of residual effect
- the contribution of education decreases but mostly due to a fall in the returns to education

Main results

- Intergenerational mobility, as measured by the IGE falls and then rises again
- The fall in the IGE largely reflects the reduction in earnings inequality that occurred in the 1970's together with slightly higher educational mobility for cohorts born in the 1950's
- The subsequent rise in the IGE is driven by a rise in the positional association between fathers and children as well as a reduction in the mobility that occurs through the education system (massification without equalization of opportunities).
- This negative trends in mobility is to some extent hidden by historically low levels of cross sectional earnings inequality. However, as inequality is currently rising, IGE should be expected to rise even more. Bad news for equality of opportunity!

(More) Decomposition on the basis of observable covariates (1)

Decomposition procedure (from Bowles & Gintis, JEP 2002). Objective is to isolate the relevant dimensions for the transmission of advantage

- Assume two observable covariates C_1 and C_2. All variables are standardized.
- The regression equation :

$$Y = \beta_{YX} X + \beta_{YC_1} C_1 + \beta_{YC_2} C_2 + e$$

  where β denotes the partial regression coefficient...
- implies the following decomposition :

$$\rho = \beta_{YX} + \rho_{XC_1} \beta_{YC_1} + \rho_{XC_2} \beta_{YC_2}$$

  where ρ_{XC} denotes the correlation btw X and C.
- IGC = direct residual effect + effect mediated through C_1 + effect mediated through C_2

(More) Decomposition on the basis of observable covariates (2)

Application : Moods, Jonsson, Bihagen (2011)

- Population data from administrative register matched across generations and with results to tests taken at compulsory military enlistment
- Covariates : cognitive ability (reasoning, verbal, technical), non-cognitive ability (extroversion, stability, focus, independance), physical ability
- Overall, covariates explain 37% of intergenerational income correlation (cognitive=.20, non-cognitive=.13)
- Adding individual own education, leaves an unexplained part of 43%.
- Parental income still has a significant and sizable effect after controlling for a large set of individual and parental covariates
- Main limitation : descriptive and non-causal
- See also Blanden, Gregg, Macmillan (2007)

Nature vs. Nurture (Björklund, Jäntti, Solon (2005))

- What is the engine of intergenerational transmission : genetic transmission or quality of the environment provided by the parents ? Distinct from the previous question.
- Three factors variance decomposition in individual earnings : genes G, family environment E, idiosyncratic component U

$$Y_i = gG_i + eE_i + uU_i$$

- Identification of g, e and u can be achieved from :
  - the correlation of Y across different sibling types
  - some reasonable assumptions on the covariance in G and E for different sibling types
  - example : monozygotic twins have perfectly correlated G and E, while dizygotic twins have their genes in common
- Main result for earnings : $g^2 \in [.15, .3]$
- Main limitation : sensitive to the assumptions about genes and environment correlation

IV estimates

- Idea : use some exogenous variation in parental attributes and look at the effect on the next generation
- Problem : good instruments are not easy to find
- Parental income
  - Instruments : job loss (Oreopoulos, Page, Stevens), union membership (Shea)
  - Result : sizable effect of parental income losses associated with job loss
  - Problems : quality of the instrument and identification of the channels through which parental income influence child's outcomes
- Parental education
  - Instruments : change in compulsory school laws (global (e.g. Chevalier) or local (Black et al.)), accidents (Maurin & McNally)
  - Usually positive and significant effect but through what channel ?

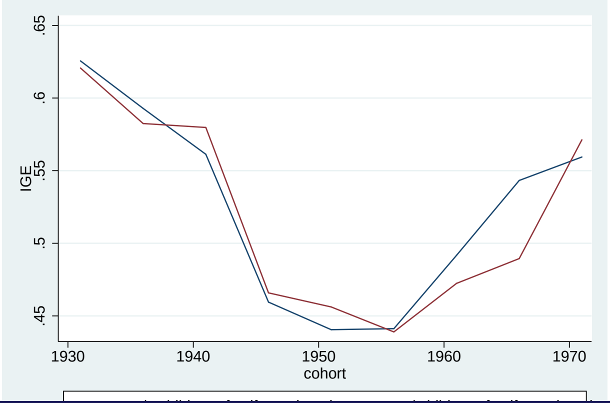
Concluding remarks

- Much progress over the last 10 years in the descriptive assessment of intergenerational earnings mobility
- Significant advances in our understanding of the underlying determinants of intergenerational persistence
- Still the evidence remains rather limited on the impact of policies aimed at fostering intergenerational mobility
  - Parental attributes are often hard to change which make direct intervention highly relevant
  - Cross-country evidence on the relationship between IGE and aggregate public policy (e.g. Ichino, Karabarbounis, Moretti, 2009)
  - Microeconomic evidence on the mobility effect of policy reforms : Pekkarinen, Uusitalo, Kerr (JPubE 2009) - school tracking reform in Finland; Dumas, Lefranc (2011) and Bingley, Westergard-Nielsen (2011) - effect of preschool and day care on intergenerational mobility

Including the children of self-employed fathers

- we don't observe their father's earnings
- artificially assign the same earnings as wage-earners with similar education

Figure: IGE by cohort w/ and w/o the children of self-employed



TSIV estimation (contd)

Properties of the TSIV estimator

- Case 1 : no independent effect of the instrument : $Y = \beta X + u$

$$plim\beta_{TSIV} = \frac{cov(Y, P_Z X)}{V(P_Z X)} = \beta + \frac{cov(u, P_Z X)}{V(P_Z X)} = \beta$$

- Case 2 : independent effect of the instrument : $Y = \lambda X + \gamma Z + u$

$$\begin{aligned} plim\beta_{OLS} &= \lambda + \gamma \frac{cov(X, Z)}{V(X)} \\ plim\beta_{TSIV} &= \lambda + \gamma \frac{cov(P_Z X, Z)}{V(P_Z X)} \\ plim(\beta_{OLS} - \beta_{TSIV}) &= \gamma \frac{cov(X, Z)}{V(X)} (1 - \frac{V(X)}{V(P_Z X)}) < 0 \end{aligned}$$