Ability and Selection; Human Capital; Returns to Schooling

Álvaro A. Novo
ISEG

Labor Economics
Lisbon, 2016

Parte I

Ability, Sorting, Selection

1 Gibbons-Katz Model

Technology

Two ability levels: $\eta \in \eta_H, \eta_L$

Two industries: A and B

Two time periods

Output: y_{ij} in industry i by worker of ability η_j. Output in industry A is more sensitive to ability than in industry B.

$y_{AH} > y_{BH} > y_{BL} > y_{AL}$

Complete information

High-ability (low-ability) workers would be employed in industry A (B) and earn high (low) wages.

Symmetric but imperfect information

First period

Noisy (non-manipulable) ex-ante signal $s \in s', s''$ where $s' > s''$

A higher signal leads to higher posterior probability $p(s)$ that worker is of high ability:

$1 > p(s') > p(s'')$ where $p(s) = \Pr(\eta = \eta_H)$

We assume that the signal is accurate enough that the following conditions on expected productivity hold:

$p(s')y_{AH} + [1 - p(s')]y_{AL} > p(s')y_{BH} + [1 - p(s')]y_{BL}$, and

$p(s'')y_{AH} + [1 - p(s'')]y_{AL} < p(s'')y_{BH} + [1 - p(s'')]y_{BL}$

Thus, productive efficiency dictates that high-signal (low-signal) workers begin their employment in industry A (B).

Second period

First-period output perfectly reveals ability to everyboby so that second period is played under perfect information.

Contracts

Risk-neutral workers. So w.o.l.o.g., compensation contracts are for one period with wages paid in advance equal to expected output (marginal product).

Wage profiles and mobility

First period

High-signal workers are employed in industry A and earn the wage

$$w_{1A} = p(s')y_{AH} + [1 - p(s')]y_{AL},$$

while low-signal workers are employed in industry B and earn the wage

$$w_{1B} = p(s'')y_{BH} + [1 - p(s'')]y_{BL},$$

Note that $w_{1A} > w_{1B}$.

Second period

High-ability workers are employed in industry A and earn the wage $w_{2A} = y_{AH}$, while low-wage workers are employed in industry B and earn the wage $w_{2B} = y_{BL}$.

Implications

- **Persistent inter-industry wage differentials**:

$$w_{tA} > w_{tB}, \quad \forall\, t$$

- **Wage gap grows with experience**:

$$w_{2A} - w_{2B} > w_{1A} - w_{1B}$$

  – Workers who move from A to B experience $w_{2B} - w_{1A} < 0$

  – Workers who move from B to A experience $w_{2A} - w_{1B} > 0$

TABLE 1

Industry wage differentials from cross-section and first-differenced regressions

January 1984 and 1986 CPS displaced workers survey plant closing sub-sample

Industry	(1) Cross-section[a]	(2) First-differenced[b]
Mining	0·510	0·429
	(0·043)	(0·051)
Primary Metals	0·223	0·262
	(0·053)	(0·055)
Fabricated Metals	0·177	0·196
	(0·049)	(0·052)
Machinery, except Electrical	0·273	0·248
	(0·040)	(0·042)
Electrical Machinery	0·131	0·083
	(0·048)	(0·047)
Transportation Equipment	0·274	0·272
	(0·045)	(0·047)
Lumber, Furniture	0·045	0·069
	(0·045)	(0·051)
Other Durables	0·164	0·091
	(0·047)	(0·046)
Food	0·195	0·170
	(0·045)	(0·048)
Textiles, Apparel	−0·005	0·053
	(0·040)	(0·045)
Paper, Printing	0·146	0·076
	(0·050)	(0·051)
Chemicals, Petroleum	0·267	0·186
	(0·044)	(0·045)
Transportation	0·329	0·130
	(0·041)	(0·044)
Utilities	0·262	0·285
	(0·064)	(0·057)
Wholesale Trade	0·154	0·085
	(0·036)	(0·034)
Retail Trade	—	—
FIRE	0·263	0·162
	(0·052)	(0·040)
Business, Professional Services	0·217	0·045
	(0·038)	(0·036)
Personal Services	0·013	−0·008
	(0·043)	(0·039)
Other Services	0·067	−0·036
	(0·046)	(0·036)
R^2	0·451	0·131
n	2576	2576

[a] The dependent variable is log (pre-displacement weekly earnings). The reported estimates are the coefficient values for the pre-displacement industry dummy variables. The base industry is retail trade. The reported regression also includes eight pre-displacement occupation dummies, a spline function in previous tenure (with breaks at one, two, three, and six years), years of schooling, pre-displacement experience and its square, a marriage dummy, a female dummy, a non-white dummy, year of displacement dummies, three region dummies, and interactions of the female dummy with marriage and the experience variables.
[b] The dependent variable is log(post-displacement weekly earnings/pre-displacement weekly earnings). The reported estimates are the coefficient values for the difference between the post-displacement and pre-displacement industry dummy variables. The base industry is retail trade. The reported regression also includes eight occupation change dummy variables; three dummy variables for post-displacement employment in agriculture, construction, or public administration; experience and experience interacted with the female dummy variable; years since displacement, and year-of-displacement dummy variables. The numbers in parentheses are standard errors.

GIBBONS & KATZ UNMEASURED ABILITY

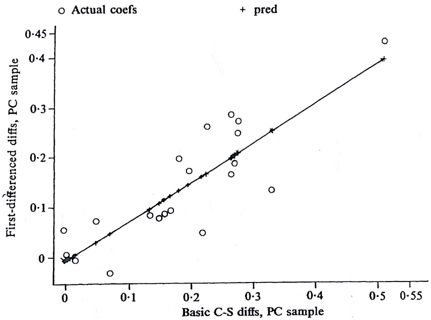

FIGURE 1

First-differenced vs. C – S ind diffs, PC sample

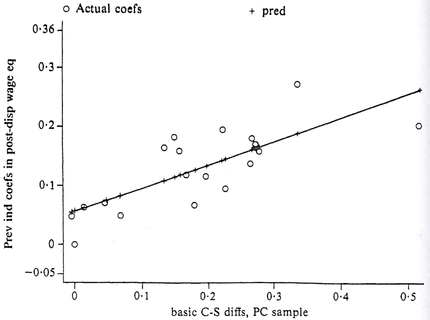

FIGURE 2

The transferability of industry diffs, PC sample

TABLE 2

The effect of pre-displacement industry on pre- and post-displacement wages

January 1984 and 1986 CPS displaced workers survey plant closing sub-sample

Pre-displacement industry	(1) Pre-displacement	(2) Post-displacement
Mining	0·510	0·208
	(0·043)	(0·053)
Primary Metals	0·223	0·099
	(0·053)	(0·066)
Fabricated Metals	0·177	0·070
	(0·049)	(0·061)
Machinery, except Electrical	0·273	0·162
	(0·040)	(0·049)
Electrical Machinery	0·131	0·168
	(0·048)	(0·060)
Transportation Equipment	0·274	0·168
	(0·045)	(0·056)
Lumber, Furniture	0·045	0·074
	(0·045)	(0·057)
Other Durables	0·164	0·121
	(0·047)	(0·058)
Food	0·195	0·119
	(0·045)	(0·055)
Textiles, Apparel	−0·005	0·049
	(0·040)	(0·050)
Paper, Printing	0·146	0·187
	(0·050)	(0·062)
Chemicals, Petroleum	0·267	0·174
	(0·044)	(0·054)
Transportation	0·329	0·278
	(0·041)	(0·051)
Utilities	0·262	0·141
	(0·064)	(0·080)
Wholesale Trade	0·154	0·162
	(0·036)	(0·044)
Retail Trade	—	
FIRE	0·263	0·183
	(0·052)	(0·065)
Business, Professional Services	0·217	0·199
	(0·038)	(0·048)
Personal Services	0·013	0·064
	(0·043)	(0·053)
Other Services	0·067	0·050
	(0·046)	(0·058)
R^2	0·451	0·327
n	2576	2576

The dependent variable in column (1) is log(pre-displacement weekly earnings). The dependent variable in column (2) is log(post-displacement weekly earnings). The reported estimates are the coefficient values for the pre-displacement industry dummy variables. The base industry is retail trade. The numbers in parentheses are standard errors. Each of the reported regressions includes eight pre-displacement occupation dummies, a spline function in previous tenure (with breaks at one, two, three, and six years), years of schooling, experience and its square, a marriage dummy, a female dummy, a nonwhite dummy, year of displacement dummies, three region dummies, and interactions of the female dummy with marrigae and the experience variables. The experience variables in column (1) use pre-displacement experience, while the experience variables in column (2) use current experience. The regression reported in column (2) also includes years since displacement.

Parte II

Human Capital Theory

2 Introduction to human capital

Human capital

Any inherent or acquired ability that produces labor market earnings or psychic income.[1]

Human capital investment

Any activity that augments earnings capacity (monetary plus psychic income) by augmenting abilities.

[1] The non-material satisfactions that accompany an occupation or economic activity.

Classic contributions to human capital theory and empirical work:

- A. Smith (1776)
- T.W. Schultz (AER, 1960, JPE, 1961)
- G. Becker – Human Capital (1964)
- J. Mincer – Schooling, Experience and Earnings (1974)
- Y. Ben Porath (1967, JPE)

The role of education was also addressed by A.M. Spence (1974) and J. Stiglitz (1975) on signalling/sorting alternative.

3 Becker-Rosen Model

Assumptions

- Individuals maximize lifetime present value of earnings
- No post-schooling investments
- Individuals have constant discount rates
- Foregone earnings only cost of schooling (no direct tution/fees)

Variation in schooling across individuals arise from differences in **"ability"** and differences in **"opportunity"** (discount rates/financing costs).

Each individual possesses a **human capital production function**:

$$ln(y_i) = h(s, a_i),$$

where a_i is **exogenous ability** of i, s is years of **schooling** and

$$h_s > 0,\ h_{ss} < 0 \text{ and } h_a > 0.$$

Thus,

$$y(s, a_i) = \exp\{h(s, a_i)\}.$$

Each individual has a constant **individual-specific discount rate** δ_i

At birth, the individual **picks schooling**, s, to solve

$$\max_s V(s) = \int_s^{\infty} y(s, a_i) e^{-\delta_i t} dt,$$

subject to

$$y(s, a_i) = e^{h(s, a_i)}.$$

Equivalently,

$$\max_s \frac{1}{\delta_i} y(s, a_i) e^{-\delta_i s}$$

The f.o.c. are:

$$y_s / y = \delta_i,$$

or

$$h_s(s, a_i) = \delta_i$$

That is, we have picked s in order to set the **marginal internal rate of return to schooling equal to the discount rate.**

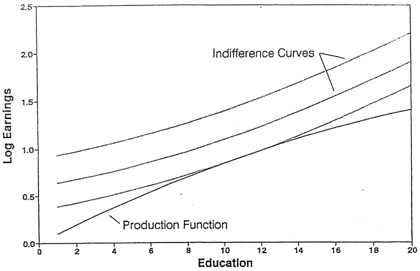

Fig. 3. Determination of optimum schooling.

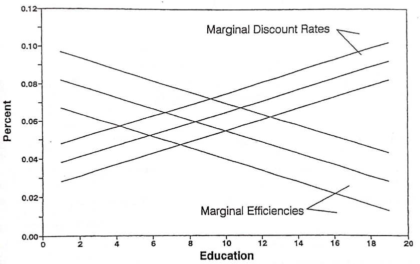

Fig. 4. Marginal benefit and marginal cost schedules for different individuals.

Key empirical issue

Ability bias in estimates of economic return to schooling.

If schooling increases with ability, $ds/da > 0$, then assuming same discount rate and variation in unobserved a_i, the **sign of ability bias** depends on sign of h_{sa} – whether ability makes **schooling investment more or less productive** (positive or negative correlation).

The omitted variable bias

True regression: $y_i = \beta_1 + \beta_2 S_i + \beta_3 A_i + \epsilon_i$
Short regression: $y_t = \beta_1^* + \beta_2^* S_i + u_i$

If we assume that ability and schooling are related:

$$A_i = a + \beta_{32} S_i + v_i,$$

then, by substitution:

$$\beta_2^* = \beta_2 + \beta_3 \times \beta_{32}.$$

We expect $\beta_3 > 0$ and $\beta_{32} > 0 \Rightarrow \boldsymbol{\beta_2^*}$ **is upward biased**.

Parte III
Empirical analysis of education and earnings

- **Issues** arising in estimating the “returns” to schooling
    1. Ability bias
    2. Measurement error in schooling
- **Instrumental variables** to generate a “natural experiment” in the quantity of schooling
    1. Angrist and Krueger (1991)
    2. Wald estimator
    3. Review of instrumental variables
    4. Problems with weakly correlated instruments

4 Issues in estimating the returns to schooling

Naive structural model

$$\log y_i = X_i\beta + \rho S_i + u_i,$$

$$S_i = Z_i\delta + v_i,$$

where y_i is **earnings** for i, S_i is **years of schooling**, X_i is a set of **control variables** and Z_i is a set of **schooling determinants**.

Some Z_i have to be **excluded from** X_i for identification of IV.

The coefficient ρ is the casual effect of schooling on earnings

It gives the expected percentage (log) gain in earnings if a randomly selected i was given an additional year of schooling.

$\hat{\rho}^{OLS}$ is consistent only if $E[u_i v_i] = 0$.

Most likely $E[u_i v_i] \neq 0$; $\hat{\rho}^{OLS}$ is biased.

- **Ability bias** – positive correlation of S_i and u_i
- **Measurement error** – likely negative correlation of S_i and u_i

Potential solutions

- **Intrumental variables (IV)** – Variables that affect schooling, but not earnings directly (conditional on schooling).
- **Controls for ability** – test scores, Griliches (1977, EMA)
- **Fixed effects** or **family effects** – Assume that residuals have restricted covariance structure and use either individual or same family repeated observations over time to difference out correlations between u_i and v_i.

Reference: David Card, *Handbook of Labor Economics*, Chapter 30.

5 Instrumental variables for schooling (Angrist and Krueger, 1991)

Natural experiment: interaction between compulsory schooling laws based on date of birth and school entrance rules.

Enter first grade (on average):
born 1st quarter: 6.07 years old; **born 4th quarter**: 6.45 years old.

Typically, legal drop out age, **16 or 17 years old**. Thus, those born in the 4th quarter attain legal drop out age with **less schooling**.

If **no direct effects of quarter of birth on earnings**, then quarter of birth is an **instrument** for years of schooling.

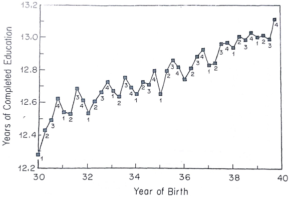

FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

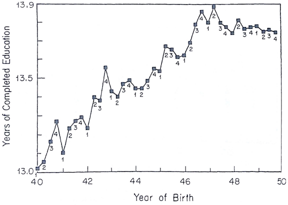

FIGURE II
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

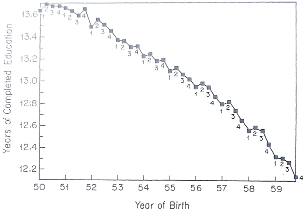

FIGURE III
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

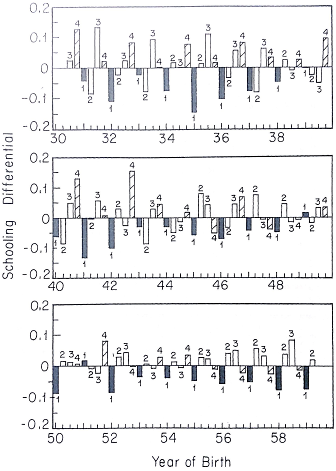

FIGURE IV
Season of Birth and Years of Schooling
Deviations from $MA(+2,-2)$

The Effect of Quarter of Birth on Various Educational Outcome Variables

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect[a] I	II	III	F-test[b] [P-value]
Total years of education	1930–1939	12.79	−0.124	−0.086	−0.015	24.9
			(0.017)	(0.017)	(0.016)	[0.0001]
	1940–1949	13.56	−0.085	−0.035	−0.017	18.6
			(0.012)	(0.012)	(0.011)	[0.0001]
High school graduate	1930–1939	0.77	−0.019	−0.020	−0.004	46.4
			(0.002)	(0.002)	(0.002)	[0.0001]
	1940–1949	0.86	−0.015	−0.012	−0.002	54.4
			(0.001)	(0.001)	(0.001)	[0.0001]
Years of educ. for high school graduates	1930–1939	13.99	−0.004	0.051	0.012	5.9
			(0.014)	(0.014)	(0.014)	[0.0006]
	1940–1949	14.28	0.005	0.043	−0.003	7.8
			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]
Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7
			(0.001)	(0.001)	(0.001)	[0.1599]
	1940–1949	0.11	0.000	0.004	0.001	3.9
			(0.001)	(0.001)	(0.001)	[0.0091]
Completed doctoral degree	1930–1939	0.03	0.002	0.003	0.000	2.9
			(0.001)	(0.001)	(0.001)	[0.0332]
	1940–1949	0.04	−0.002	0.001	−0.001	4.3
			(0.001)	(0.001)	(0.001)	[0.0050]

a. Standard errors are in parentheses. An $MA(+2, -2)$ trend term was subtracted from each dependent variable. The data set contains men from the 1980 Census, 5 percent Public Use Sample. Sample size is 312,718 for 1930–1939 cohort and is 457,181 for 1940–1949 cohort.

b. F-statistic is for a test of the hypothesis that the quarter-of-birth dummies jointly have no effect.

Percentage of Age Group Enrolled in School by Birthday and Legal Dropout Age[a]

Date of birth	Type of state law[b]: School-leaving age: 16 (1)	School-leaving age: 17 or 18 (2)	Column (1) − (2)
	Percent enrolled April 1, 1960		
1. Jan 1–Mar 31, 1944	87.6	91.0	−3.4
(age 16)	(0.6)	(0.9)	(1.1)
2. Apr 1–Dec 31, 1944	92.1	91.6	0.5
(age 15)	(0.3)	(0.5)	(0.6)
3. Within-state diff.	−4.5	−0.6	−4.0
(row 1 − row 2)	(0.7)	(1.0)	(1.2)
	Percent enrolled April 1, 1970		
4. Jan 1–Mar 31, 1954	94.2	95.8	−1.6
(age 16)	(0.3)	(0.5)	(0.6)
5. Apr 1–Dec 31, 1954	96.1	95.7	0.4
(age 15)	(0.1)	(0.3)	(0.3)
6. Within-state diff.	−1.9	0.1	−2.0
(row 1 − row 2)	(0.3)	(0.6)	(0.6)
	Percent enrolled April 1, 1980		
7. Jan 1–Mar 31, 1964	95.0	96.2	−1.2
(age 16)	(0.1)	(0.2)	(0.2)
8. Apr 1–Dec 31, 1964	97.0	97.7	−0.7
(age 15)	(0.1)	(0.1)	(0.1)
9. Within-state diff.	−2.0	−1.5	0.5
(row 1 − row 2)	(0.1)	(0.2)	(0.3)

a. Standard errors are in parentheses.

b. Data set used to compute rows 1–3 is the 1960 Census, 1 percent Public Use Sample; data set used to compute rows 4–6 is 1970 Census, 1 percent State Public Use Sample (15 percent form); data set used to compute rows 7–9 is the 1980 Census, 5 percent Public Use Sample. Each sample contains both boys and girls. Sample sizes are 4,153 for row 1; 12,512 for row 2; 7,758 for row 4; 24,636 for row 5; 42,740 for row 7; and 131,020 for row 8.

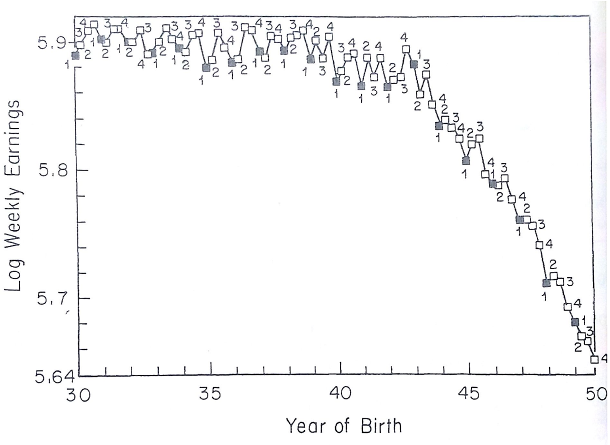

FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

Panel A: Wald Estimates for 1970 Census—Men Born 1920–1929[a]

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.1484	5.1574	−0.00898 (0.00301)
Education	11.3996	11.5252	−0.1256 (0.0155)
Wald est. of return to education			0.0715 (0.0219)
OLS return to education[b]			0.0801 (0.0004)

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110 (0.00274)
Education	12.6881	12.7969	−0.1088 (0.0132)
Wald est. of return to education			0.1020 (0.0239)
OLS return to education			0.0709 (0.0003)

a. The sample size is 247,199 in Panel A, and 327,509 in Panel B. Each sample consists of males born in the United States who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5 percent sample, and the 1970 Census sample is from the State, County, and Neighborhoods 1 percent samples.

b. The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

Wald estimator

Ratio of differences between **treatment and control** group in **dependent and explanatory** variables to calculate a slope coefficient (here the return to education).

6 Using twins to control for genetic ability and family environment

This is an alternative to the IV approach.

The key idea: some of the unobserved differences that bias a cross-sectional comparison of education and earnings are reduced or eliminated within families.

A “pure family effects” is plausible for identical twins, because identical twins share genetics and almost always share the same family background environment.

Thanks