

The Democratization of Education, its Apparent Paradox, a Plausible Explanation, and its Empirical Proof

Louis-André Vallet (CNRS)

Observatoire Sociologique du Changement
UMR 7049 CNRS & Sciences Po, Paris

Laboratoire Interdisciplinaire d'Evaluation
des Politiques Publiques, Sciences Po, Paris

Putting together papers published at two moments...

- SELZ M., VALLET L.-A., 2006. – "La démocratisation de l'enseignement et son paradoxe apparent", *Données sociales. La société française (Édition 2006)*, Paris, INSEE, pp. 101-107.
- VALLET L.-A., SELZ M., 2007. - "Évolution historique de l'inégalité des chances devant l'école : des méthodes et des résultats revisités", *Éducation et Formations*, 74, pp. 65-74.
- ICHOU M., VALLET L.-A., 2012. - "Performances scolaires, orientation et inégalités sociales d'éducation. Évolution en France en quatre décennies", *Éducation et Formations*, 82, pp. 9-18.
- ICHOU M., VALLET L.-A., 2013. - "Academic Achievement, Tracking Decisions, and Their Relative Contribution to Educational Inequalities: Change over Four Decades in France" in JACKSON M. (ed.), *Determined to Succeed? Performance versus Choice in Educational Attainment*, Stanford, Stanford University Press, pp. 116-148.

RESEARCH QUESTIONS

This presentation analyzes temporal trends in the intrinsic association between social origin and educational attainment (in the sense of highest degree obtained) in French society.

* First, it aims at *establishing stylised facts* about temporal dynamics in inequality of educational opportunity according to social origin for French society.

* Second, the presentation also especially aims at *highlighting that unconditional and conditional analyses conducted on the same data may reveal remarkably opposite trends*.

* Third, the presentation aims at *reconciling these different results*, especially by *relying on tentative hypotheses developed by Pierre Bourdieu & Jean-Claude Passeron in 1970 in France as well as Robert D. Mare in 1981 in the US*.

* Fourth, it aims at *establishing an empirical proof of these statements* on the basis of two longitudinal datasets that are nearly 40 years apart.

DATA

All the analyses I will present are based on a *representative sample composed of 515,591 cases* (coming from the French Labour Force Surveys for the years 1984, 1987, 1990, 1993, 1996, 1999 and 2001).

This is a representative sample of French-born men and women who belong to *19 three-year birth cohorts, from the oldest (1920-1922) to the youngest (1974-1976)*.

For each birth cohort, the *occupational class of the father in 11 categories* is cross-classified with the highest diploma obtained by the respondent.

This *education variable comprises 7 categories*, from ‘no diploma at all’ to ‘a tertiary degree of at least three years after baccalauréat’.

So, *the basic material for the whole analysis is a set of 19 (for birth cohorts) 11 x 7 contingency tables*.

Occupational class of the father

- 1. – Agriculteurs exploitants (PCS 10) ;
- 2. – Artisans ; Commerçants et assimilés (PCS 21 et 22) ;
- 3. – Chefs d'entreprise de 10 salariés ou plus ; Professions libérales (PCS 23 et 31) ;
- 4. – Cadres de la fonction publique ; Professions de l'information, des arts et des spectacles ; Cadres administratifs et commerciaux d'entreprises ; Ingénieurs et cadres techniques d'entreprises (PCS 33, 35, 37 et 38) ;
- 5. – Professeurs, professions scientifiques ; Instituteurs et assimilés (PCS 34 et 42) ;
- 6. – Professions intermédiaires de la santé et du travail social ; Clergé, religieux ; Professions intermédiaires administratives de la fonction publique ; Professions intermédiaires administratives et commerciales des entreprises ; Techniciens (PCS 43, 44, 45, 46 et 47) ;
- 7. – Employés civils et agents de service de la fonction publique ; Policiers et militaires ; Employés administratifs d'entreprises ; Employés de commerce ; Personnels des services directs aux particuliers (PCS 52, 53, 54, 55 et 56) ;
- 8. – Contremaîtres, agents de maîtrise (PCS 48) ;
- 9. – Ouvriers qualifiés de type industriel ; Ouvriers qualifiés de type artisanal ; Ouvriers qualifiés de la manutention, du magasinage et du transport (PCS 62, 63 et 65) ;
- 10. – Chauffeurs ; Ouvriers non qualifiés de type industriel ; Ouvriers non qualifiés de type artisanal (PCS 64, 67 et 68) ;
- 11. – Ouvriers agricoles (PCS 69).

Education variable (DIPL)

- 1. – Aucun diplôme ou diplôme non déclaré (codes 71 et blanc) ;
- 2. – Certificat d'études primaires (code 70) ;
- 3. – BEPC sans diplôme technique ou professionnel (code 60) ;
- 4. – CAP, BEP avec ou sans BEPC (codes 50 et 51) ;
- 5. – Baccalauréat ou diplôme de niveau équivalent : baccalauréat général et diplôme technique secondaire ; baccalauréat général seul ; baccalauréat de technicien, baccalauréat technologique, baccalauréat professionnel ou brevet professionnel ; BEI, BEC, BEA (codes 40, 41, 42 et 43) ;
- 6. – Diplôme d'une ou deux années après le baccalauréat : diplôme de 1^{er} cycle universitaire ; BTS, DUT ; diplôme paramédical ou social avec ou sans baccalauréat général (codes 30, 31, 32 et 33) ;
- 7. – Diplôme d'au moins trois années après le baccalauréat : diplôme de 2^{ème} ou 3^{ème} cycle universitaire ; diplôme de grande école, diplôme d'ingénieur (codes 10 et 11).

Making the dataset: Observational design

Enquêtes Génération	EMPLOI 1984	EMPLOI 1987	EMPLOI 1990	EMPLOI 1993	EMPLOI 1996	EMPLOI 1999	EMPLOI 2001 (sauf tiers sortant)
1920-1922	62-64 ans						
1923-1925	59-61 ans	62-64 ans					
1926-1928	56-58 ans	59-61 ans	62-64 ans				
1929-1931	53-55 ans	56-58 ans	59-61 ans	62-64 ans			
1932-1934	50-52 ans	53-55 ans	56-58 ans	59-61 ans	62-64 ans		
1935-1937	47-49 ans	50-52 ans	53-55 ans	56-58 ans	59-61 ans	62-64 ans	
1938-1940	44-46 ans	47-49 ans	50-52 ans	53-55 ans	56-58 ans	59-61 ans	61-63 ans
1941-1943	41-43 ans	44-46 ans	47-49 ans	50-52 ans	53-55 ans	56-58 ans	58-60 ans
1944-1946	38-40 ans	41-43 ans	44-46 ans	47-49 ans	50-52 ans	53-55 ans	55-57 ans
1947-1949	35-37 ans	38-40 ans	41-43 ans	44-46 ans	47-49 ans	50-52 ans	52-54 ans
1950-1952	32-34 ans	35-37 ans	38-40 ans	41-43 ans	44-46 ans	47-49 ans	49-51 ans
1953-1955	29-31 ans	32-34 ans	35-37 ans	38-40 ans	41-43 ans	44-46 ans	46-48 ans
1956-1958	26-28 ans	29-31 ans	32-34 ans	35-37 ans	38-40 ans	41-43 ans	43-45 ans
1959-1961		26-28 ans	29-31 ans	32-34 ans	35-37 ans	38-40 ans	40-42 ans
1962-1964			26-28 ans	29-31 ans	32-34 ans	35-37 ans	37-39 ans
1965-1967				26-28 ans	29-31 ans	32-34 ans	34-36 ans
1968-1970					26-28 ans	29-31 ans	31-33 ans
1971-1973						26-28 ans	28-30 ans
1974-1976							25-27 ans

STATISTICAL MODEL

To establish stylised facts about temporal trends in the association between social origin and education, I rely on the *Unidiff or log-multiplicative layer effect model* that very significantly improves on the temporally constant association model in each analysis.

In the context of such a model, the logged odds ratio that measures the association between social origins i and i' , and education categories j and j' , in birth cohort g , is equal to:

$$\text{Log}\left(\frac{m_{ijg}/m_{ij'g}}{m_{i'jg}/m_{i'j'g}}\right) = \beta_g (\psi_{ij} + \psi_{i'j'} - \psi_{ij'} - \psi_{i'j})$$

i.e., it is the product of an expression that only depends on the social origins and education categories considered (the structure of the association) by a cohort-specific parameter (its strength in a given cohort).

As a consequence, *temporal dynamics in the association between social origin and education will be entirely captured in the set of 19 β_g parameters* corresponding to the 19 birth cohorts. And the figures I will present will simply depict the dynamics of these estimated parameters.

THREE DIFFERENT ANALYSES

The basic material (a 19 x 11 x 7 three-way contingency table) can be analysed in three different ways.

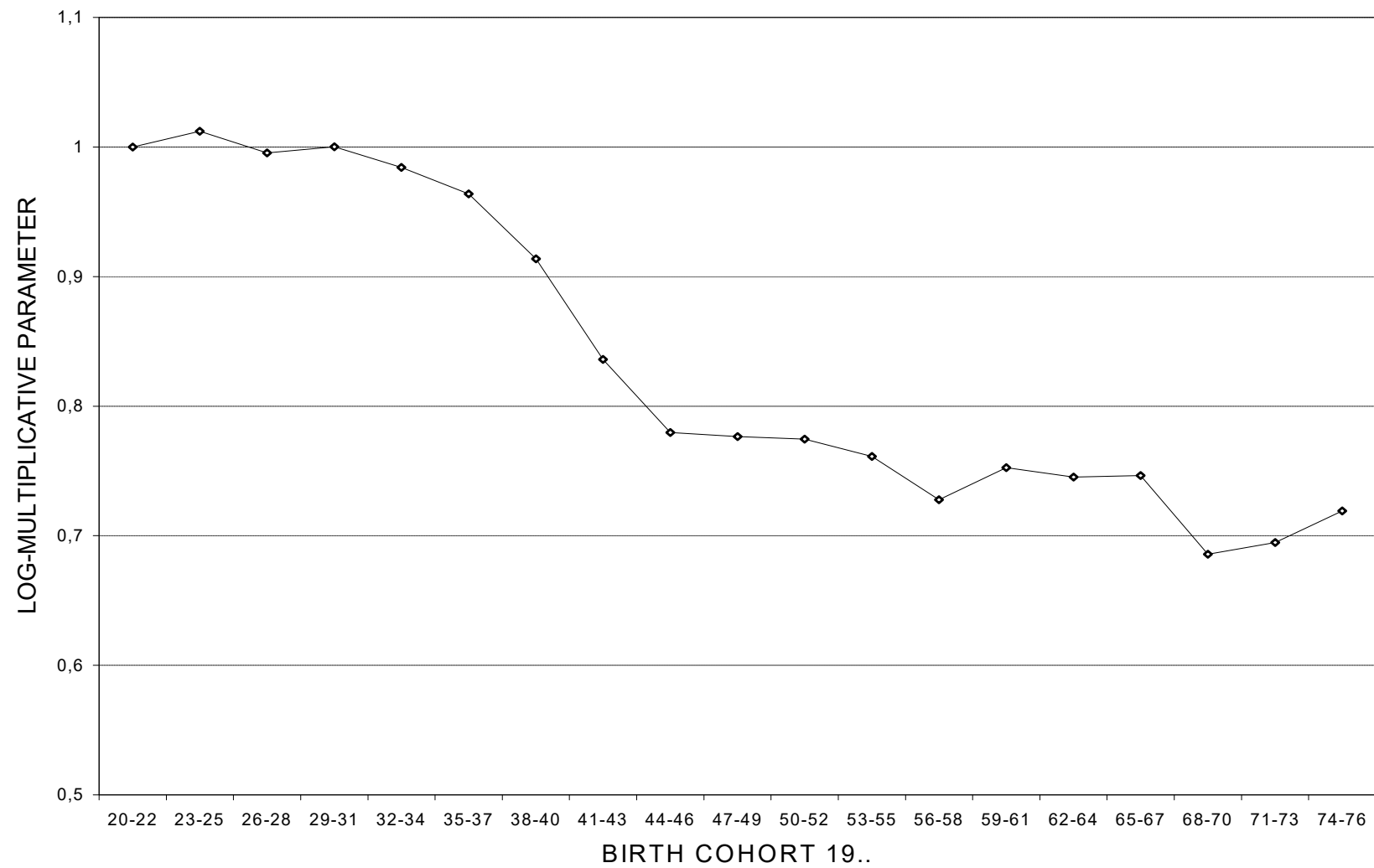
FIRST ANALYSIS

The first analysis is an *unconditional* one in the sense that it considers the whole population in each birth cohort. It is also based on the *whole distribution of the education variable* (7 categories). In other words, this is just an analysis of the 19 11 x 7 contingency tables at once.

This analysis reveals an *unevenly declining association* between social origin and education, that is to say, *decreasing inequality of educational opportunity*.

Change was especially marked from birth cohort 1938-1940 to birth cohort 1944-1946, then it largely levelled off, and finally reappeared in birth cohort 1968-1970.

FIRST ANALYSIS:
UNCONDITIONAL, WHOLE DISTRIBUTION OF THE EDUCATION VARIABLE



SECOND ANALYSIS

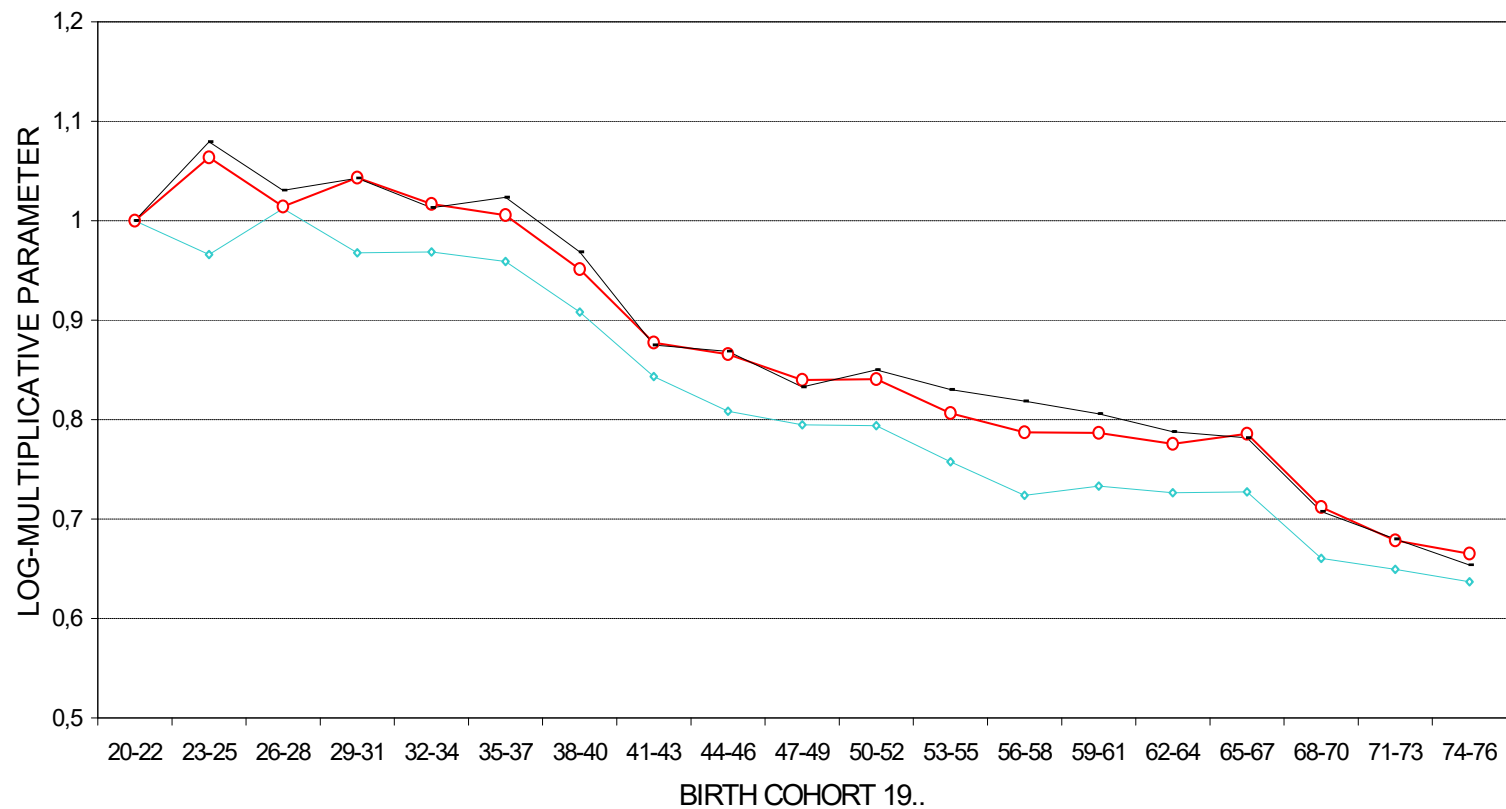
The second analysis is also an *unconditional* one in the sense that it again considers the whole population in each birth cohort. But it adopts a *dichotomous representation of the education variable* based on a threshold category. In other words, this is just an analysis of 19 11 x 2 contingency tables.

Three threshold categories are independently considered:

- 'getting at least baccalauréat' vs 'getting less than baccalauréat', (columns 5 to 7 vs columns 1 to 4);
- 'getting at least a tertiary degree of 1 or 2 years after baccalauréat' vs 'getting less than that' (columns 6 & 7 vs columns 1 to 5);
- 'getting a tertiary degree of at least 3 years after baccalauréat' vs 'getting less than that' (column 7 vs columns 1 to 6).

The analysis reveals that, *in each case, the trend in the intrinsic association between social origin and dichotomised education is monotonically downwards*. Change is also more evenly distributed over birth cohorts than in the first analysis.

SECOND ANALYSIS:
UNCONDITIONAL, DICHOTOMOUS REPRESENTATION OF THE EDUCATION VARIABLE



- ◇— AT LEAST BACCALAUREAT vs LESS THAN THAT
- AT LEAST 1 OR 2 YEARS AFTER BACCALAUREAT vs LESS THAN THAT
- △— AT LEAST 3 YEARS AFTER BACCALAUREAT vs LESS THAN THAT

THIRD ANALYSIS

The third analysis is a *conditional* one in the sense that it considers progressively reduced subsets of the whole population in each birth cohort. It also applies a *sequential analysis based on four 'educational transitions'*.

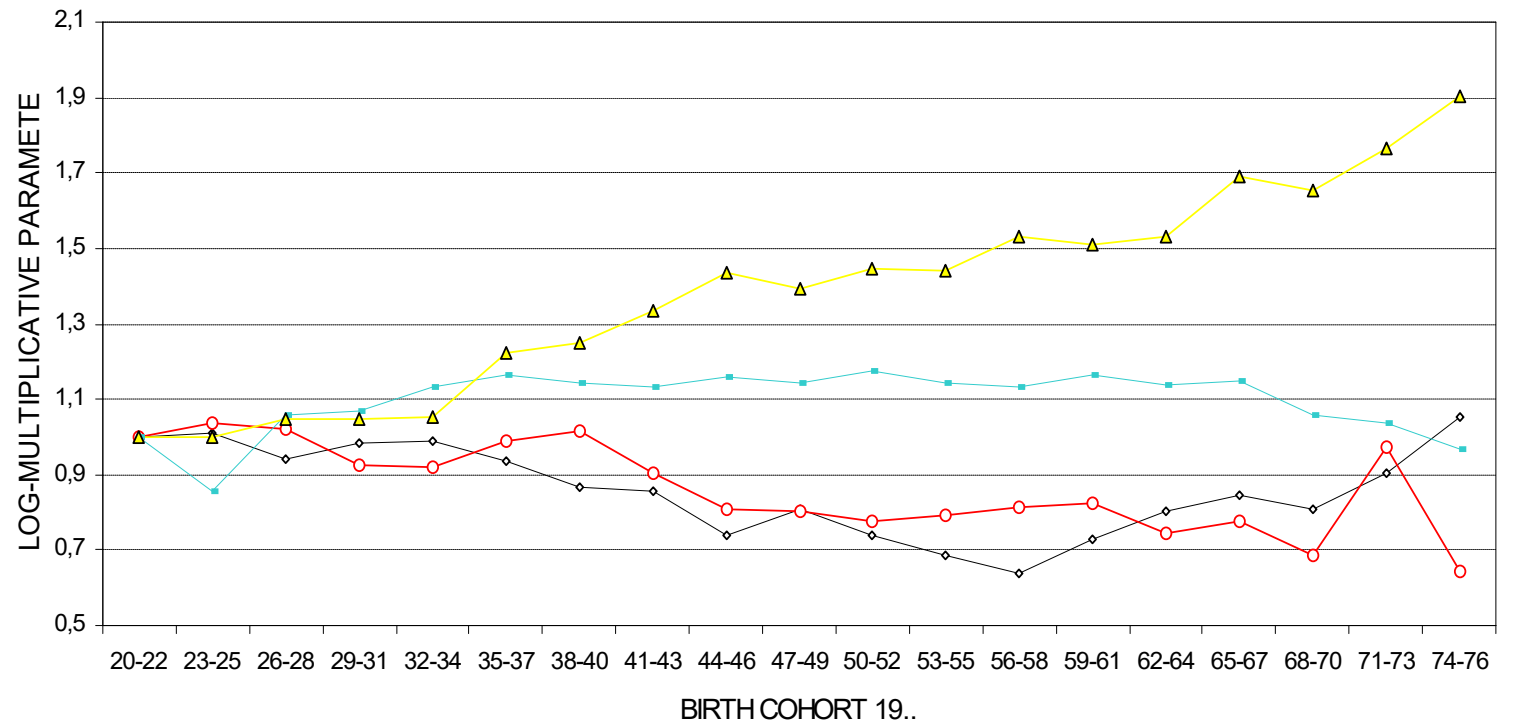
The result is that *temporal dynamics in inequality of educational opportunity are strikingly different in different transitions*.

First transition – *'any diploma' vs 'no diploma at all'* (columns 2 to 7 vs column 1): inequality has decreased until cohort 1956-1958, then has grown up again.

Second transition – *'at least lower secondary or lower vocational diploma' vs 'only primary education certificate'* (columns 3 to 7 vs column 2): inequality has rather monotonically decreased from cohort 1938-1940.

Third transition – *'at least baccalauréat' vs 'only lower secondary or lower vocational diploma'* (columns 5 to 7 vs columns 3 & 4): inequality appears remarkably stable over birth cohorts (except in the early cohorts where it increases, and the late cohorts where it decreases).

THIRD ANALYSIS:
CONDITIONAL, SEQUENTIAL ANALYSIS OF FOUR EDUCATIONAL TRANSITIONS



- ◇— FIRST TRANSITION - ANY DIPLOMA vs NO DIPLOMA AT ALL
- SECOND TRANSITION - AT LEAST LOWER SECONDARY OR LOWER VOCATIONAL DIPLOMA vs ONLY PRIMARY EDUCATION CERTIFICATE
- THIRD TRANSITION - AT LEAST BACCALAUREAT vs ONLY LOWER SECONDARY OR LOWER VOCATIONAL DIPLOMA
- ▲— FOURTH TRANSITION - TERTIARY DEGREE vs ONLY BACCALAUREAT

Fourth transition – *‘tertiary degree’ vs ‘only baccalauréat’* (columns 6 & 7 vs column 5): inequality has continuously increased over birth cohorts.

So, as regards statistical association between social origin and education, *increasing temporal trend in the last transition is exactly the opposite of decreasing temporal trend in the general and unconditional analysis.*

A REMINDER ABOUT THE 'WANING EFFECTS' PATTERN

Trends are therefore markedly different in different transitions. But, of course, *the strength of the baseline association (i.e. the association in the first cohort) is also markedly different from one transition to another.*

As usually, we observe the *waning effects pattern*: the strength of the baseline association between social origin and success or failure is stronger in the first transitions, then progressively weakens in the subsequent ones.

More precisely, the logistic contrast (or logged odds ratio) between the extreme categories in each transition, i.e. children of teachers on one hand, and children of agricultural workers on the other hand, is:

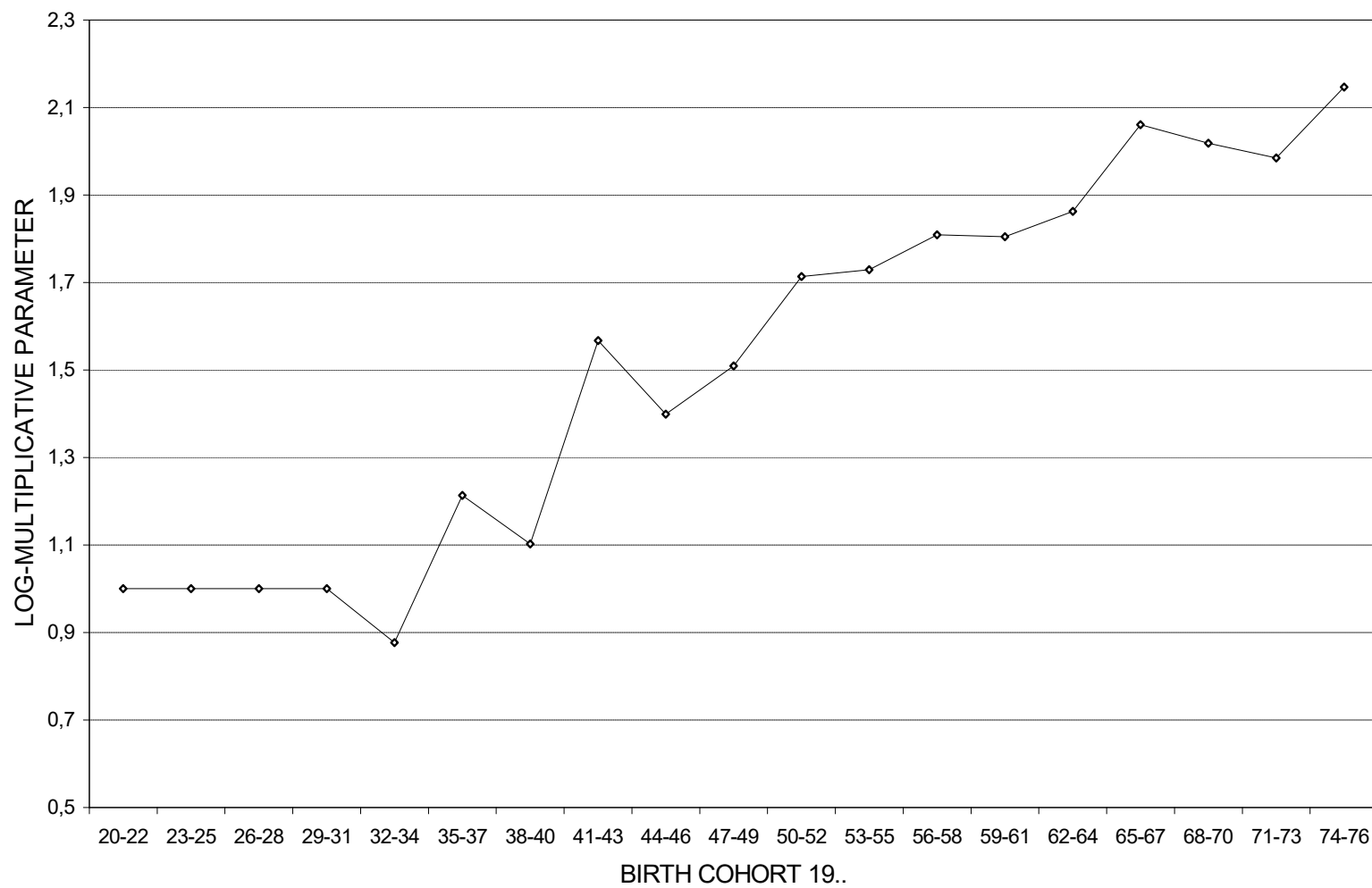
- 3.036 in the *'first transition'*,
- and 3.853 in the *second transition*,
- but 2.562 in the *third transition*,
- and 0.992 in the *fourth transition*.

THIRD ANALYSIS – ADDENDUM

Does the increasing trend in the last transition only result from increasing heterogeneity of the baccalauréat diploma per se, coming for instance from the creation of technological baccalauréat in 1968, then vocational baccalauréat in 1985?

The answer is in the negative, because the general baccalauréat has existed over the entire period and *the same increasing trend reappears when the analysis is restricted to French-born men and women who got general baccalauréat.*

THIRD ANALYSIS - ADDENDUM
ONLY INCUMBENTS OF GENERAL BACCALAUREAT ARE CONSIDERED HERE



—◇— FOURTH TRANSITION - TERTIARY DEGREE vs ONLY GENERAL BACCALAUREAT

CONCLUSION

Statistical analysis conducted on the *same data* therefore demonstrates that *two conclusions are simultaneously true*.

First, *when it is evaluated on the whole population, the general strength of the association between social origin and education obtained has rather monotonically decreased between birth cohorts that are about 50 years apart*.

Second, *when only the sub-population of incumbents of baccalauréat is considered, the general strength of the association between social origin and possession or non-possession of a tertiary degree has continuously increased between the same birth cohorts*.

This is what I propose to call '*the apparent paradox of the democratization of education*'. Temporal dynamics in inequality of educational opportunity therefore reflect two things:

- the general weakening of the association in the whole population;
- the fact that, with the expansion of education, the association strengthens in the last transition.

PLAUSIBLE EXPLANATION

To explain the latter result, Rob Mare proposed a tentative hypothesis in footnote 5 (page 82) of his 1981 ASR paper entitled “Change and Stability in Educational Stratification”:

“In brief, the trends in Table 3 result from the diminished selectivity of the educational process.

Within cohorts, socioeconomic effects on grade progression decline from the earliest to the latest school transitions. This reduction in socioeconomic effects results from differential dropout rates, which systematically reduce differences among children from different socioeconomic levels on unmeasured determinants of grade progression (for example, ability, motivation, etc.). Greater homogeneity on unmeasured factors at higher levels of schooling reduces the effects of observed socioeconomic variables.

[...] Between cohorts the impact of differential attrition changes because the degree of attrition at any schooling level is smaller in more recent than in earlier cohorts. That is, in more recent cohorts, larger fractions of each cohort remain in school at each schooling level. Thus, in recent cohorts there is greater heterogeneity on unmeasured variables at each level of schooling. This greater heterogeneity results in there being larger effects of observed socioeconomic factors in recent than in earlier cohorts.”

BUT THE SAME TENTATIVE ARGUMENT
CAN ALSO BE EXPRESSED IN A DIFFERENT WAY...

In the oldest cohorts, because of very strong differential elimination linked to social origin, children from the lower classes who were able to prepare and pass *baccalauréat* were probably particularly brilliant.

Being ‘over-selected’ so to speak, they probably did not encounter many difficulties in their subsequent paths within tertiary education. They were therefore able to compete with their schoolmates from the upper classes or to outclass them.

With the expansion of education in more recent cohorts, differential elimination linked to social origin is probably less strong and this might weaken the ‘over-selection’ mechanism.

THIS IS EXACTLY WHAT PIERRE BOURDIEU & JEAN-CLAUDE
PASSERON SUGGESTED ON PAGE 96 OF THEIR 1970 BOOK,
La reproduction. Éléments pour une théorie du système d'enseignement

(in a section entitled '*Inégalités devant la sélection et inégalités de sélection*')

« *Il suit aussi de ces analyses que si la part des étudiants des classes populaires accédant à l'Université venait à s'accroître de façon sensible, le degré de sélection relative de ces étudiants compenserait de moins en moins en s'affaiblissant les désavantages scolaires liés à l'inégalité de la répartition du capital linguistique et culturel entre les classes sociales. On verrait donc réapparaître la corrélation directe entre les résultats scolaires et la classe sociale d'origine qui, dans l'enseignement supérieur, ne s'observe pleinement que dans les domaines les moins directement contrôlés par l'École, tandis que, dans l'enseignement secondaire, elle se manifeste déjà jusque dans les résultats les plus scolaires.* »

IS THIS TENTATIVE EXPLANATION ACTUALLY TRUE?

Longitudinal version of the argument, i.e. within a given cohort

Let me suppose that we are able to observe the academic performance of pupils originating from different social classes across several transitions in the educational system...

To demonstrate the validity of the argument, we should observe that the performance distributions of the different social classes become more and more similar with the advancement in the educational system.

Historical version of the argument, i.e. between two or more cohorts

Let me suppose that we can do the same for not only one cohort, but two cohorts that are several decades apart, separated by educational expansion...

To demonstrate the validity of the argument, we should observe that, for the same transitions, the performance distributions of the different social classes are more differentiated in the recent cohort than in the ancient cohort.

THIS IS ACTUALLY WHAT WE CAN OBSERVE ON FRENCH DATA!

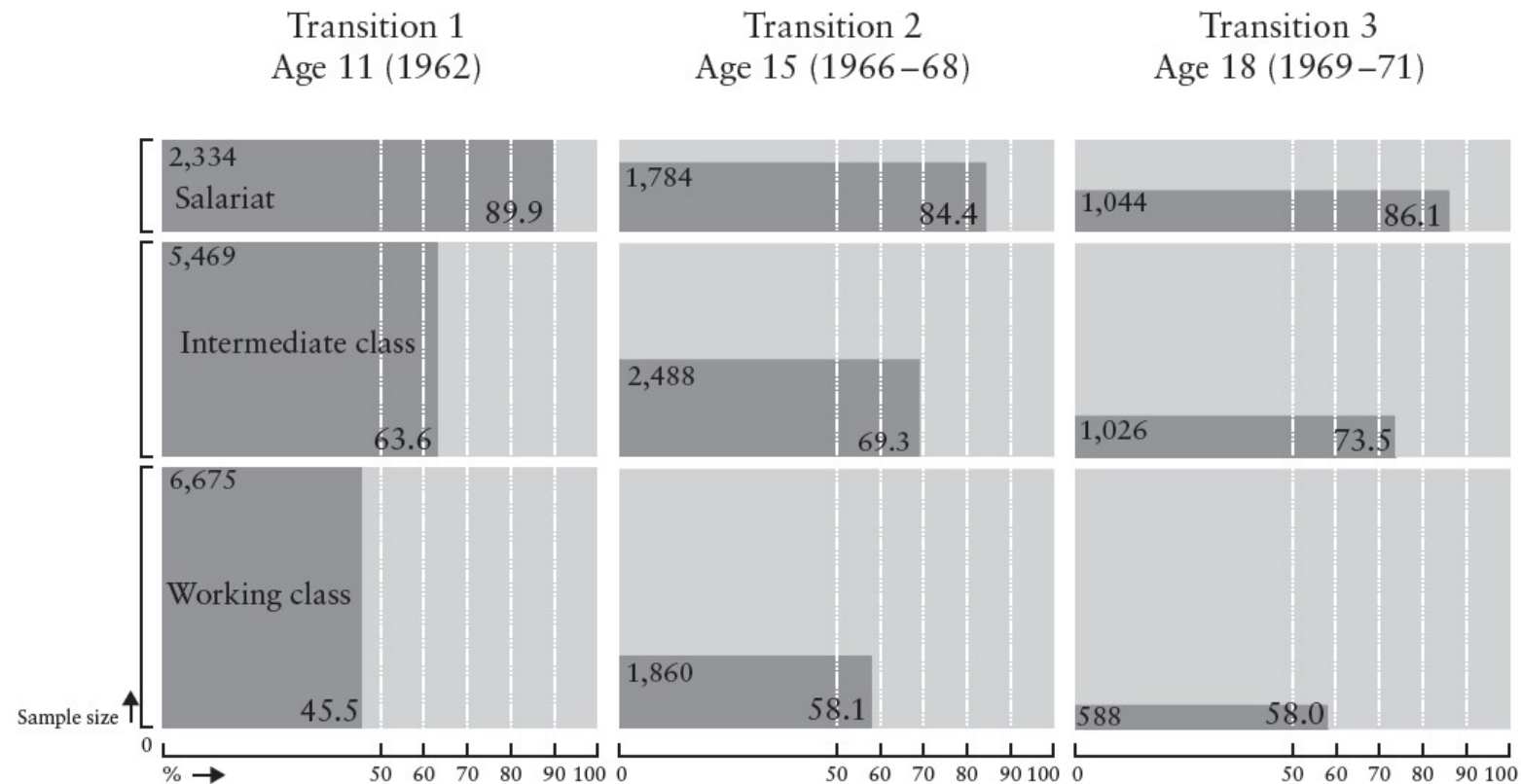


Figure 5.3. Transition rates showing class inequality at each transition (1951 cohort)

SOURCE: 1962–1972 INED survey; our calculations.

NOTE: The horizontal width of the rectangles is proportional to the transition rates. The vertical height of the rectangles corresponds to the numerical proportion of each class.

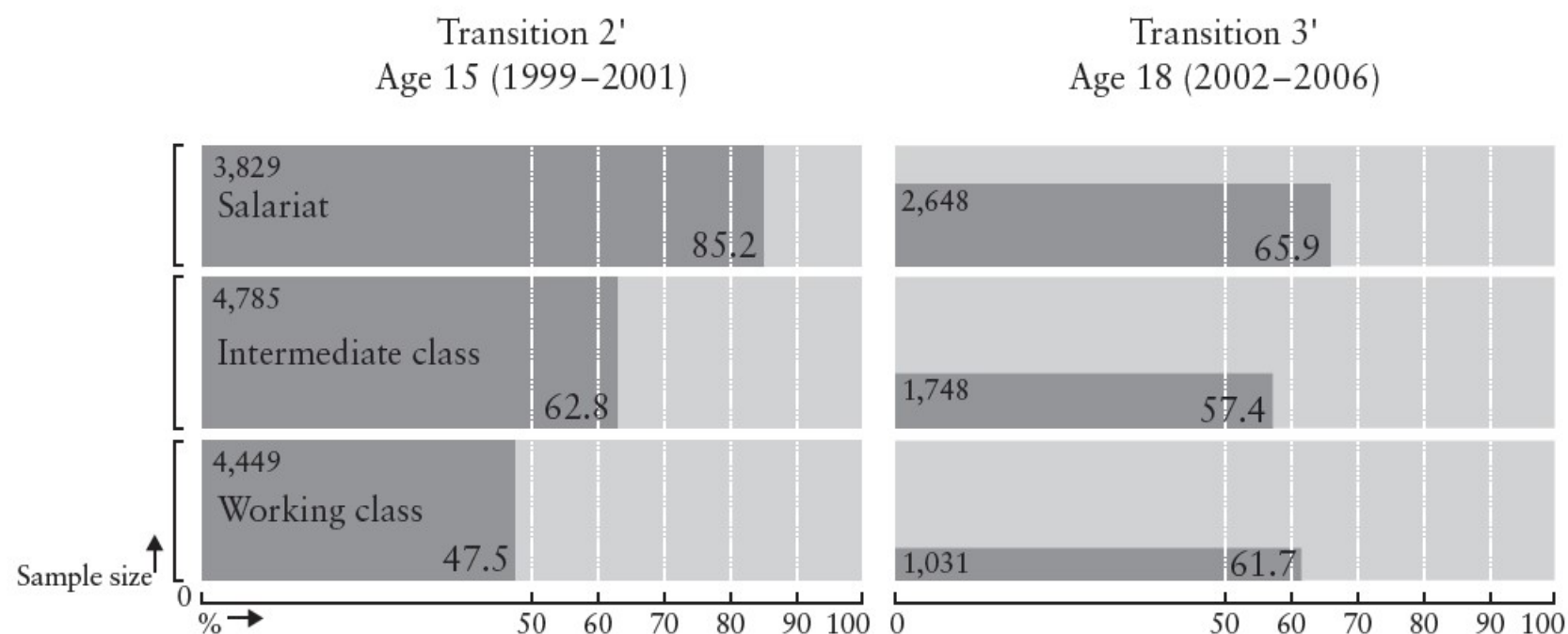


Figure 5.4. Transition rates showing class inequality at each transition (1984 cohort)

SOURCE: 1995–2006 Ministry of Education Panel study; our calculations.

NOTE: The horizontal width of the rectangles is proportional to the transition rates. The vertical height of the rectangles corresponds to the numerical proportion of each class.

Each performance variable is standardized: Mean=0 Standard Deviation=1

TABLE 5.1
*Means of standardized test scores at each transition by class in the 1951
and 1984 cohorts*

<i>Class</i>	<i>Mean</i>	<i>N</i>	<i>Mean</i>	<i>N</i>	<i>Mean</i>	<i>N</i>
1951 COHORT						
	TRANSITION 1: AGE 11 (1962)		TRANSITION 2: AGE 15 (1966–1968)		TRANSITION 3: AGE 18 (1969–1971)	
Salariat	0.426	2,334	0.075	1,784	0.028	1,044
Intermediate	0.047	5,469	–0.012	2,488	–0.009	1,026
Working	–0.188	6,675	–0.056	1,860	–0.034	588
Total	0	14,478	0	6,132	0	2,658
Salariat/ working gap	0.614		0.131		0.062	
1984 COHORT						
			TRANSITION 2': AGE 15 (1999–2001)		TRANSITION 3': AGE 18 (2002–2006)	
Salariat			0.383	3,829	0.163	2,648
Intermediate			–0.034	4,785	–0.120	1,748
Working			–0.292	4,449	–0.214	1,031
Total			0	13,063	0	5,427
Salariat/ working gap			0.675		0.377	

SOURCE: 1962–1972 INED survey and 1995–2006 Ministry of Education Panel Study; our calculations.

Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males

Stephen V. Cameron

Columbia University

James J. Heckman

University of Chicago and American Bar Foundation

This paper examines an empirical regularity found in many societies: that family influences on the probability of transiting from one grade level to the next diminish at higher levels of education. We examine the statistical model used to establish the empirical regularity and the intuitive behavioral interpretation often used to rationalize it. We show that the implicit economic model assumes myopia. The intuitive interpretive model is identified only by imposing arbitrary distributional assumptions onto the data. We produce an alternative choice-theoretic model with fewer parameters that rationalizes the same data and is not based on arbitrary distributional assumptions.

What is an economist to make of this empirical regularity? To answer this question, this paper investigates the following five questions. (1) Is there any reason to be interested in the coefficients of logistic regression transition probabilities? (2) Can the logit model of grade transitions be linked to an economic model with dynamic choice by rational agents? (3) Can the intuitive explanation for the regularity provided by sociologists—that it is a consequence of dynamic selection bias or uncontrolled unobservables—be formally justified? (4) How sensitive is this explanation to the choice of functional forms for transition equations and the distributions of unobservables? (5) Is it possible to devise a robust econometric scheme that corrects for dynamic selection bias but does not rely on arbitrary distributional assumptions?

Briefly, our answers are as follows. (1) Logit regression coefficients, while widely used to interpret data, are of little interest in their own right. As a purely statistical matter, a variety of alternative measures of the effects of variables on the transition probabilities of educational attainment might be used. When they are used, many show a *reversal* of the empirical regularity displayed by logit regression coefficients. The “empirical regularity” depends on the choice of statistic used to summarize the evidence. The issue of which statistic to use can be settled only by an appeal to behavioral theory and cannot be resolved on purely statistical grounds. (2) The logit model of grade transitions implicitly assumes myopia on the part of agents. It is not an attractive interpretive framework for economists because it is difficult to justify on choice-theoretic grounds. (3) It is possible to formally define dynamic selection bias, but it does not universally operate in the way hypothesized in the literature on educational attainment. Selective survival of the “fittest” students can lead to *overstatement* of the estimated effects of family background on higher-grade transitions. (4) A pattern of declining logit coefficients for higher-grade transitions is critically dependent on choices of functional forms for the distribution of unobservables. (5) Under certain

ADDITIONAL CONCLUSION

Should we conclude that IEO has *decreased* (because of results of unconditional analyses)?

Or should we conclude that IEO has *increased* (because of results of conditional analysis in the last transition)?

My own answer to that question is that we must “prefer” the former conclusion, i.e. *decreasing inequality*. Why? Not for ideological reasons of course! But for serious methodological reasons...

As sociologists, our responsibility is to express diagnosis about inequality in society that is primarily based on consideration of the *whole* population, i.e. in *unconditional* analyses ... rather than diagnosis based on consideration of only a *fraction* of the whole population, i.e. in conditional analyses.

Finally, if, for good reasons, we are interested in conditional analyses, we should take account of the possibility of ‘dynamic selection bias’ or ‘changing unobserved heterogeneity over birth cohorts’ as a consequence of the educational expansion.

I would not have been able to deliver this presentation without the remarkable work that has been undertaken for long by French Public Statistics ...and without the alliance between French Public Statistics and the French Social Science Data Archive, *Réseau Quetelet*.

Thank you for your attention!