Soc Indic Res
DOI 10.1007/s11205-007-9216-1

A Discrepancy Index for the Study of Participation with an Application to the Case of Higher Education in Italy

Valentino Larcinese

Accepted: 29 October 2007

Abstract This article proposes and applies a simple method to measure the distance from a situation of uniform participation. First, a discrepancy index based on the use of generalized Lorenz curves is presented. This index can be expressed in terms of means and Gini indices of relevant characteristics in the populations of participants and that of a control group. A multivariate extension is also illustrated. The discrepancy index is then used to analyse access into higher education in Italy during the nineties. Discrepancy between participants and non-participants is especially high in terms of parents' wealth and education, less in terms of income. Also, discrepancy is higher during recessions. Multivariate analysis reveals that "cultural constrains" are more important than financial constraints. Hence, efforts to equalize opportunity for access into higher education should be as much focussed on motivation and social background as they are on financial constraints.

Keywords Participation · Opportunity · Social mobility · Human capital · Inequality

1 Introduction

Participation can be defined as self selection to be part of a group and can be regarded as the outcome of the interaction between preferences and constraints. In the context of educational choices, this process is affected by individual skills and motivations as well as by financial constraints and other factors that may affect the expected returns of education.[1] An educational system that intends to be both fair and efficient, according to most definitions of these words, should aim at selecting the most skilled people, independently of their income and social background. Therefore, as social scientists and policy-makers are

[1] See, for example, Shavit and Blossfeld (1993) and Checchi (1999).

V. Larcinese (✉)
Department of Government and STICERD, London School of Economics, Houghton Street, London WC2A 2AE, UK
e-mail: V.Larcinese@lse.ac.uk

increasingly stressing the importance of promoting participation and of providing equal opportunities, it has become correspondently important to be able to assess the performance of education systems from this perspective, i.e. evaluating how much they promote equal opportunity and social mobility.

Aggregate measures of participation provide useful information but hide an important distributive dimension. Consider, for example, a country where 20% of the young population enters into the higher education system. This measure (20%) does not give any information on the social background of participants. On one extreme, participants could perfectly represent the composition of the country population across a number of relevant dimensions (income, residence, race etc). On the other extreme, participants could all have the same background (for example, it could be the richest 20% of the population). Clearly, each case would carry very different normative and policy implications. It seems, therefore, important, in studying access to education as well as any other form of participation, to be able to measure not only the aggregates but also the distributive dimension.

This paper proposes a new method to measure the distance from a situation of homogeneous participation and shows how this method can be applied to the case of access into higher education in Italy. The method consists in comparing the distribution of the relevant characteristics of the population of participants with that of a reference group.[2] This method delivers an index which measures the distance from an ideal situation of uniform participation, i.e. a situation in which the population of participants and the control population do not differ in terms of the chosen relevant variables. This index has a bounded support and can, therefore, be easily interpreted and used to compare different situations.

In a way, the discrepancy index reflects how much a given outcome is correlated with characteristics that are typically beyond individual control. If, therefore, we interpret the distributions involved as the conditional probabilities of participating, we can also regard the discrepancy index as a measure of the degree of equal opportunity, where uniform participation can be regarded as perfectly equal opportunity.[3]

I will now introduce the index with reference to the case of access into higher education. Its application to the study of other forms of participation is rather straightforward.

2 A Simple Discrepancy Index

Let us consider a reference population of households whose income[4] is distributed according to the distribution function $F(x)$, with density function $f(x)$. We will indicate with $f(x|H)$ the income distribution in the sub-population of households whose offspring are in

[2] An earlier work by Fry (1983) on Thailand uses the same idea of comparing the proportion of individuals in a given region or occupation with their relative access into various educational degrees. Fry (1983) uses the simple comparison of proportions to derive a "discrepancy ratio". This work represents an extension of that idea, in the sense of using well established tools in the theory of inequality measurement to derive a discrepancy index that embeds the same intuition of Fry's "discrepancy ratio".

[3] The fact that the relevant variables are beyond individual control is crucial for this interpretation to be valid. See for example Arneson (1989) and Le Grand (1991). This is the same idea expressed in Roemer (1998), who calls such variables "circumstances".

[4] I will refer to parents' income as the relevant dimension of heterogeneity. Other variables can be considered and, in the subsequent analysis, I will also consider parents' wealth and education.

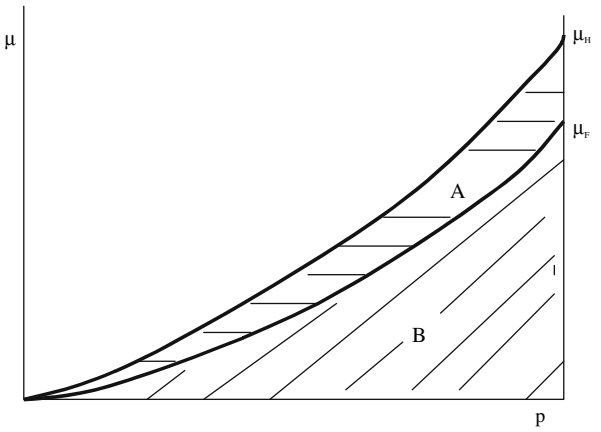

Fig. 1 Generalized Lorenz curves for the two populations

full time higher education. If income is not relevant for accessing education, then x and H are independent variables and we can write $f(x) = f(x|H)$. Consider then the corresponding generalized Lorenz curves,[5] defined as

$$\begin{aligned} p = F(y) &\Rightarrow GL_F(p) = \int_0^y xf(x)dx \\ p = F(y|H) &\Rightarrow GL_H(p) = \int_0^y xf(x|H)dx \end{aligned} \tag{1}$$

Clearly,

$$f(x) = f(x|H) \Rightarrow GL_F(p) = GL_H(p).$$

If instead, as it is reasonable to expect, income matters for participation, then households with offspring into higher education should be richer than average. Our index can be built starting from this simple and intuitive observation. Imagine therefore that households with offspring in higher education are richer than average at *all levels of* p. This implies that $GL_F(p) < GL_H(p) \forall p$. Thus, the distance $GL_H(p) - GL_F(p)$ measures the advantage (in terms of income) of the poorest fraction p of the population H over the corresponding poorest fraction p of the population F (see Fig. 1).

The area A in Fig. 1 depends on the metric chosen, which in this case is income. To have an index that is easily interpretable and permits comparisons across different situations we can consider the areas below GL_F and GL_H and take their ratio:

$$D(F,H) = \frac{\int GL_F(p)dp}{\int GL_H(p)dp} \tag{2}$$

We will typically have $\int GL_H(p)dp \geq \int GL_F(p)dp$ and therefore $0 \leq D(F,H) \leq 1$.

Now note that $\int GL(p)dp = \mu \int L(p)dp$, where μ and $L(p)$ are respectively the mean income and the Lorenz curve corresponding to a generic income distribution $F(x)$. The Gini index G referred to the same distribution F can be calculated as $G = 1 - 2\int L(p)dp$.

This means that

[5] For comprehensive surveys of the literature on inequality measurement see Champernowne and Cowell (1998) or Lambert (1993).

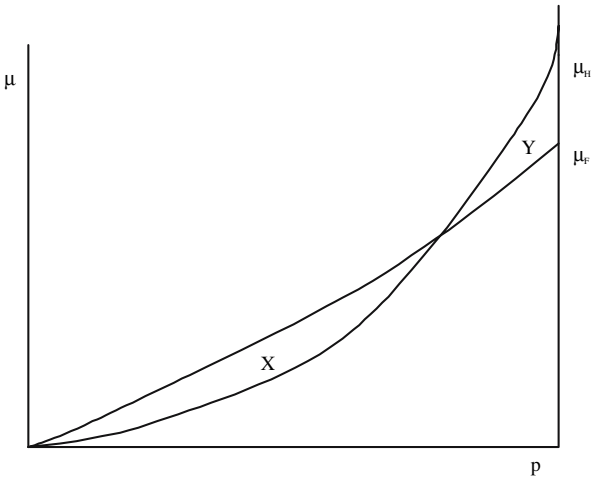

Fig. 2 Generalized Lorenz curves crossing once (X = Y⇒$D = 1$ with x and H not independent)

$$\int L(p)dp = \frac{(1-G)}{2} \Rightarrow \int GL(p)dp = \frac{\mu(1-G)}{2} \tag{3}$$

and

$$D(F,H) = \frac{\mu_F(1-G_F)}{\mu_H(1-G_H)} \tag{4}$$

This index is easy to calculate and has an immediate interpretation grounded in the theory of inequality measurement. If $GL_F(p) \leq GL_H(p) \forall p$ then the D index goes from 0 to 1, where a value of 1 implies equal participation across different income classes and a lower value implies a larger distance from equal participation.

One caveat is necessary in order to correctly interpret the D index. Although independence between the relevant characteristic and access to education implies $D = 1$, the reverse is not true. It is possible to have $D = 1$ when generalized Lorenz curves intersect. In Fig. 2 we have a case where they cross once. If the area below the intersection (X) and that above the intersection (Y) are equivalent then D is equal to 1 even though x and H are clearly not independent. In such case comparing the two generalized Lorenz curves can still deliver a number of insights but relying only on the D index would be misleading. However, our intuition suggests (and the empirical analysis of this paper will confirm this intuition) that the generalized Lorenz curves for the two populations are always quite far apart.

It should also be stressed that this analysis does not imply any direct causality between the variables involved. The purpose of the discrepancy index is only to measure the distance from a situation of uniform participation and not to establish causal relationships, for which a regression analysis would clearly be more appropriate.

3 Access to Higher Education in Italy

The Italian higher education system has often been criticized in the public debate for its poor performance (for example because the ratio of graduates over enrolled students is about 0.3), but far less attention has been placed on its limited ability to increase social

mobility. This is probably due to the fact that the formal right for everyone to access higher education has always been guaranteed and there are no screening mechanisms in place in most universities.[6] It seems, therefore, important to ask if this "open" system is actually effective in promoting opportunity and social mobility. This is especially the case as formal openness has been coupled with very limited availability of grants and loans for students.

Using data from the Household Survey of Wealth and Income produced by the Bank of Italy, I calculate the D indices (for the years 1989, 1991, 1993, 1995 and 1998) referred to head of households' income, wealth and education, by comparing the population of households with students[7] (H) with the population of all households with offspring aged 19–26 (F). The choice of the control group is only dictated by the purposes of the analysis.[8] In our case we want to compare the population of students only with the population that could potentially be in higher education: this makes some difference especially for income and wealth, as households with offspring aged 19–26 are normally at the peak of their income path.

Figures 3–5 report the generalized Lorenz curves referred respectively to head of households' wealth, income and education. A direct comparison of the generalized Lorenz curves offer a number of insights. First of all, the generalized Lorenz curves referred to households with offspring in higher education are always (with very limited exceptions that will be discussed below) above the corresponding curves for the control population. Thus, in accessing higher education there are hurdles that are either represented by any of the variables chosen, or correlated with them. It is also evident that the bigger discrepancy is in wealth and the smallest in income. It is not surprising that, when undertaking human capital investment, families do not look only at their current income but rather to their ability to finance education over a period of at least 4–5 years. Both wealth and education are better indicators of this capability. At the same time, parents' education might be capturing "cultural constraints". We will return on this in the next section.

From Figs. 3–5, the income discrepancy appears increasing over time, while the education discrepancy remains fairly stable. Also, the education discrepancy can probably be attributed in large part to the different education levels in the lowest deciles, a sign that parents with low education levels are rather rare in the population of households with offspring into higher education. On the contrary, most of the income discrepancy seems to

[6] To understand some of the present limits of this system, it is useful to recall that universities were essentially conceived for an *èlite*, and not for mass education. The increase in participation in the last 50 years has been dramatic; at the same time, investments and supply have not increased at the same rate, resulting in a fall in a number of indicators of performance. In spite of having mainly a public higher education system, public expenditure in higher education in Italy has been increasing quite slowly if compared with other European countries. Also, total expenditure (both private and public funding) in higher education, shows a remarkable difference with other developed countries.

[7] Given the high number of dropouts, it would be interesting also to use the population of those who actually manage to complete tertiary education rather than the population of students. Unfortunately we cannot recover this information from the available data. However, if family characteristics have an impact also on the probability of completion (which appears extremely likely), participation and opportunity should probably be regarded as more skewed of what can appear by looking at the students' population only.

[8] Obviously not all students are aged 19–26: however, the number of those who are not is very limited. Moreover, individuals within 19–26 (both students and not) are more likely to be dependent on parents, and this is particularly the case in Italy. An alternative possibility is to focus on households and compare those with at least one student with those without students; in this case it is possible to exclude those households that could not possibly have students among their members. Alternative cases have been considered and the results do not differ from those reported here. One important advantage of focusing on individuals is that this takes into account the possibility that some households have more than one student.

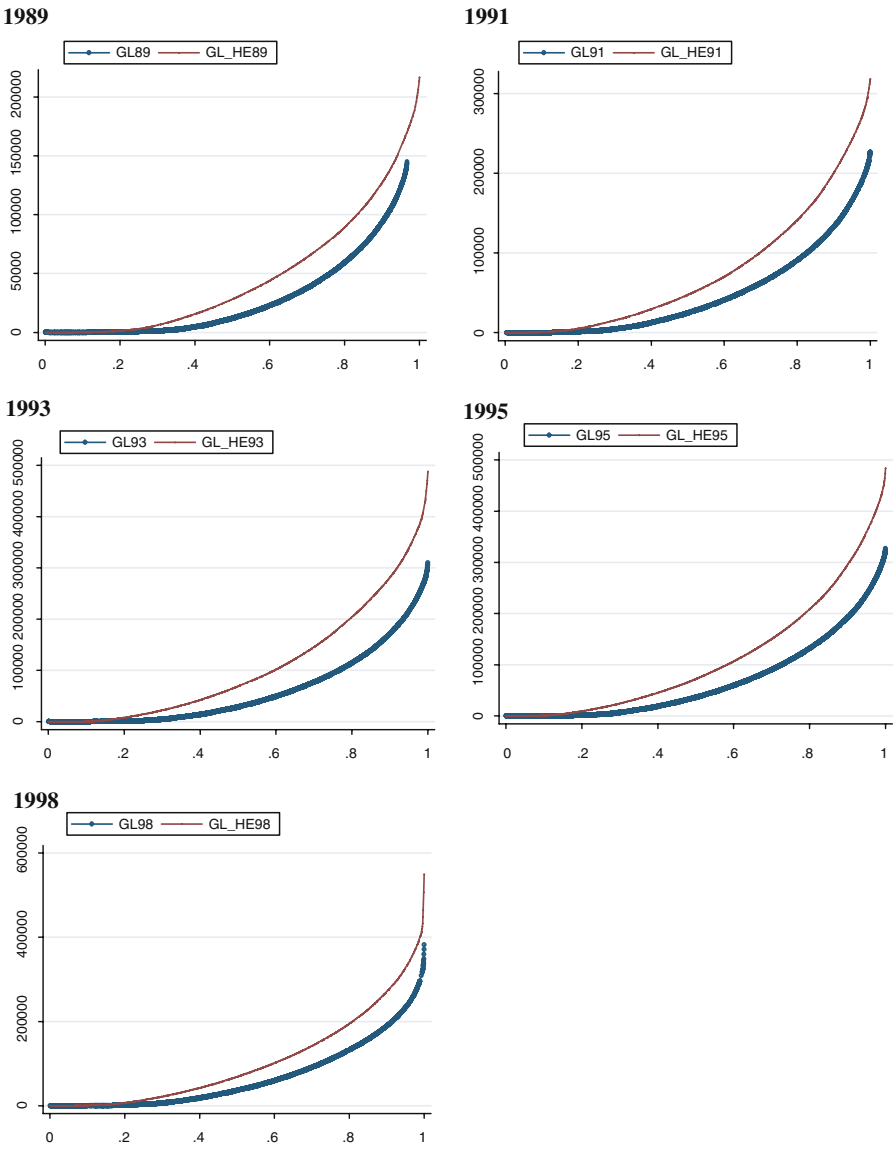

Fig. 3 Generalized Lorenz curves of wealth (in thousands of Italian Lire) for all households with off-springs aged 19–26 (GL) and for households with off-springs in higher education (GL_HE)

appear at high income levels, which indicates a high concentration of offspring from very high earners among university students. These last two pieces of information are lost when the curves are translated into the synthetic discrepancy index: this suggests that looking directly at the generalized Lorenz curves can always be useful in its own.

A very important reason to look at the generalized Lorenz curves is also to make sure that they do not intersect. From Figs. 3–5, there appears to be no obvious intersections, but the situation is not completely clear for very low levels of p. For this reason, Figs. 6–8

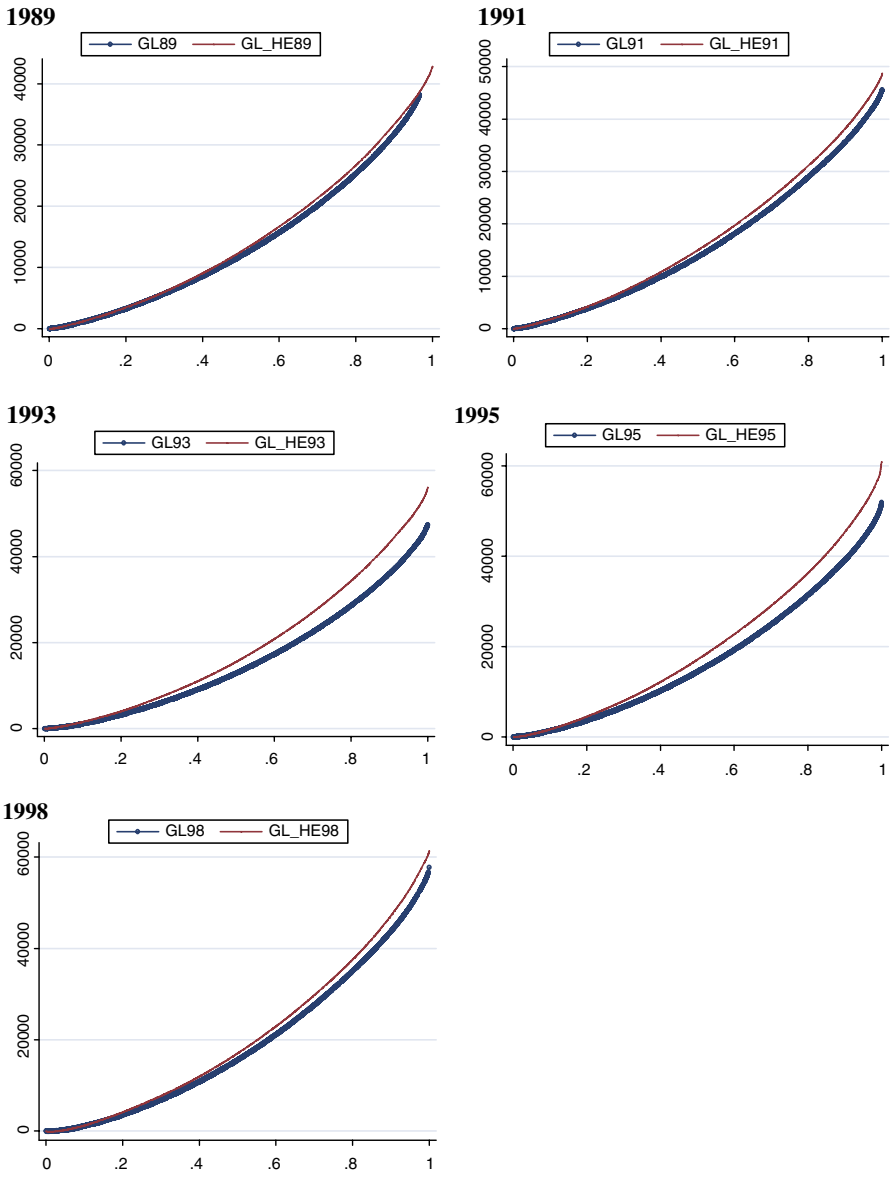

Fig. 4 Generalized Lorenz curves of income (in thousands of Italian Lire) for all households with off-springs aged 19–26 (GL) and for households with off-springs in higher education (GL_HE)

report the generalized Lorenz curves for the first decile only, thus magnifying the differences between the curves.

Using this "magnifying lens" it is actually possible to notice few intersections. These can be found for the wealth curves of 1993 and for the income curves of 1989 and 1998. In the first two cases the distance between the curves before the intersection is infinitesimal. For the income curves referred to 1998 there is a neat intersection occurring at $p = 0.05$. This means that the poorest students are slightly poorer than the poorest population in the

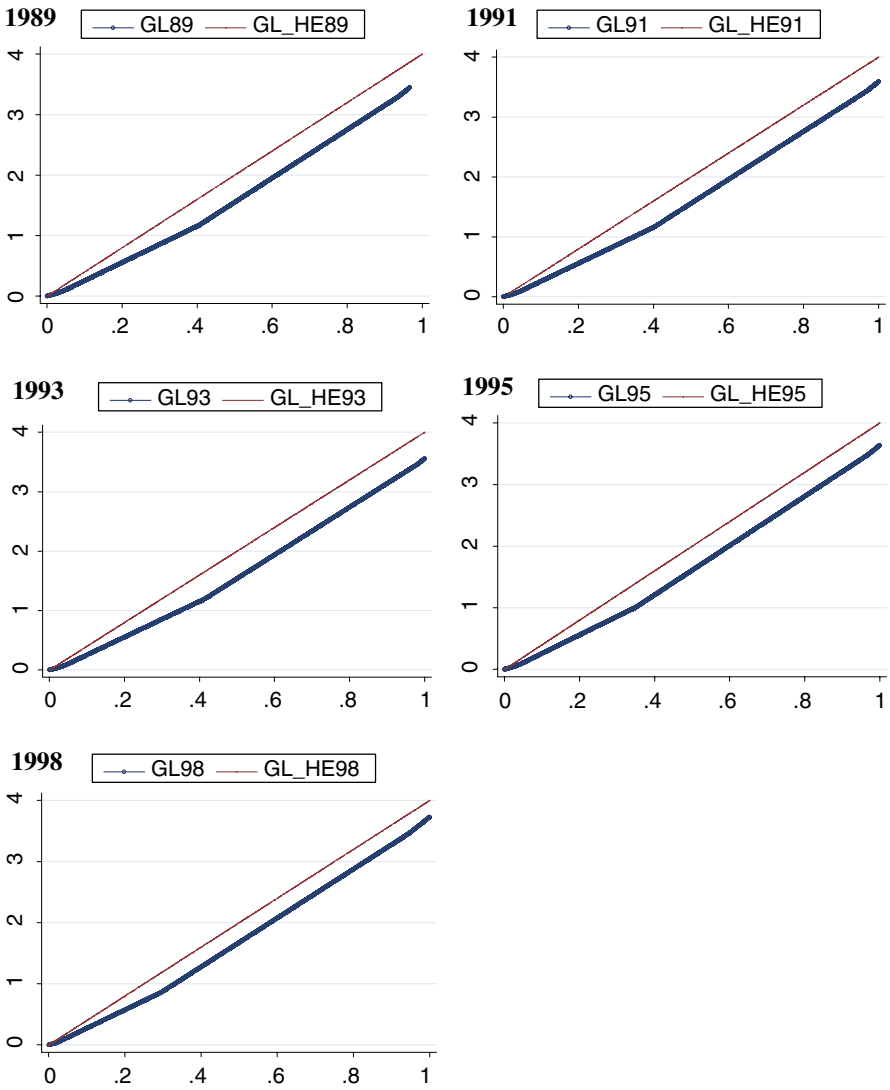

Fig. 5 Generalized Lorenz curves of parents' education (head of household) for all households with off-springs aged 19–26 (GL) and for households with off-springs in higher education (GL_HE). 1 = primary education (5 years); 2 = intermediate (8 years); 3 = secondary (13 years); 4 = higher education

relevant age.[9] We should bear this in mind when discussing the discrepancy indices below. However, by comparing Fig. 7 with Fig. 4, i.e. by putting in perspective the size of the distance before the intersection, it is fairly obvious that the impact of the intersection on the size of the index should be virtually negligible. Apart from these cases, the Generalized

[9] This could be due to the functioning of some limited support mechanism in favour of very poor students: this creates a "kink" in the incentives to enter into higher education. At the same time, appearing only in a single year, this feature could just be due to sampling variability.

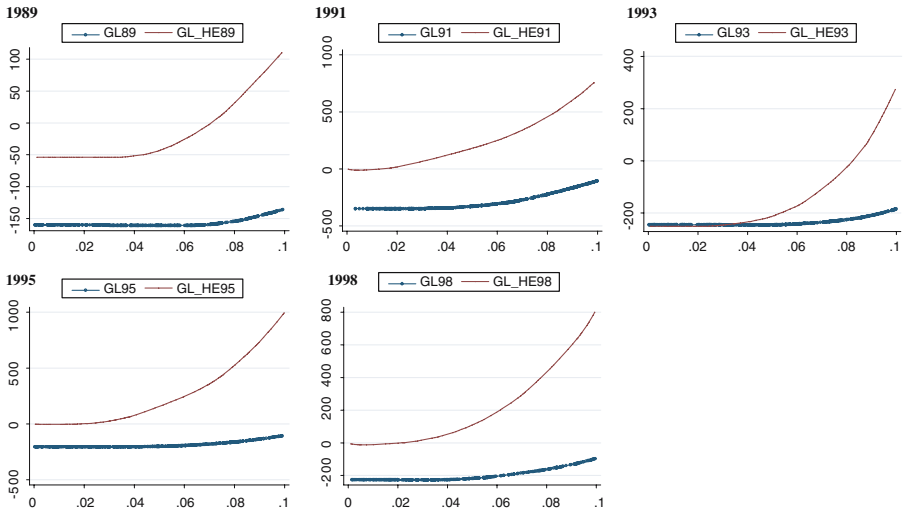

Fig. 6 Generalized Lorenz curves of wealth (in millions of Italian Lire) for all households with off-springs aged 19–26 (GL) and for households with off-springs in higher education (GL_HE). First decile

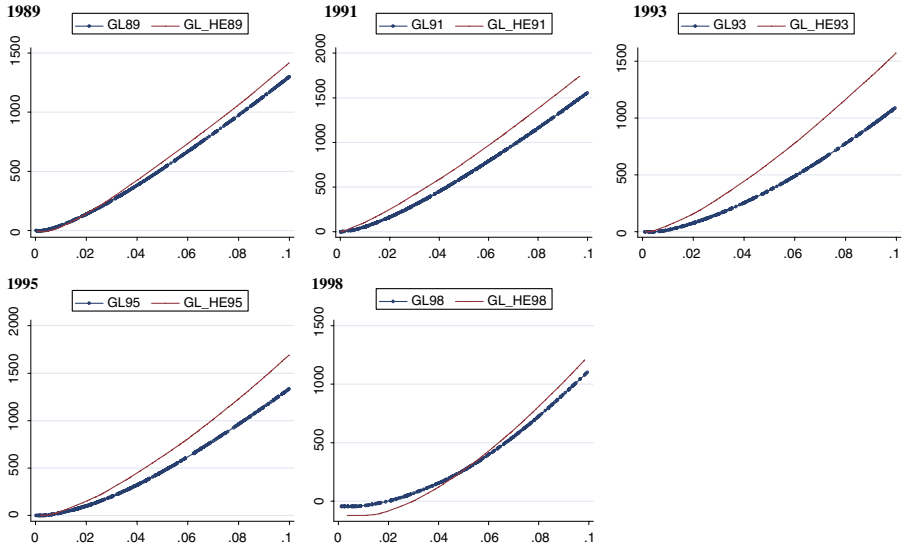

Fig. 7 Generalized Lorenz curves of income (in millions of Italian Lire) for all households with off-springs aged 19–26 (GL) and for households with off-springs in higher education (GL_HE). First decile

Lorenz curves for the student population are always substantially above those of the reference population. Hence, we can safely interpret the case $D = 1$ as independence.

Moving to the D-index, we can now quantify the distance between the generalized Lorenz curves for the various cases. As expected, the D indices are below 1 in all cases and for all years (Fig. 9). Confirming what could be guessed from the Fig. 3–5, the biggest gap between students and non-students is in wealth, the lowest in income, while parents' education is always in between.

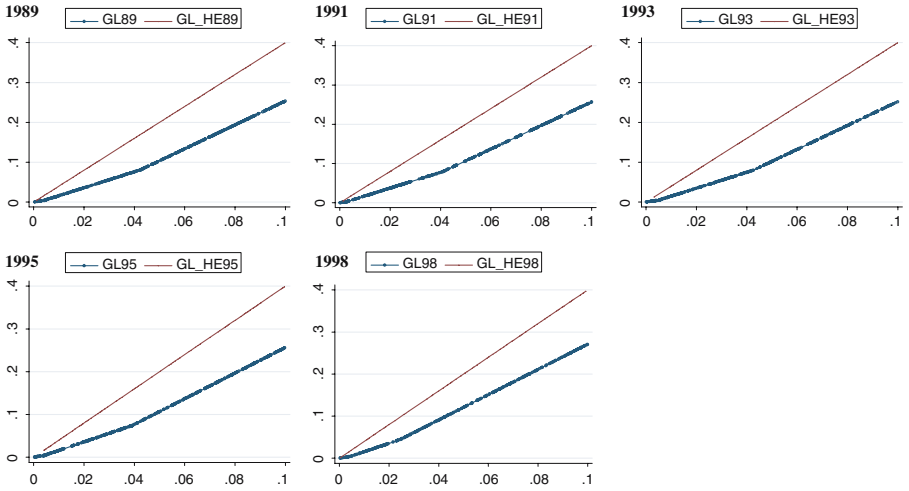

Fig. 8 Generalized Lorenz curves of parents' education (head of household) for all households with offsprings aged 19–26 (GL) and for households with off-springs in higher education (GL_HE). First decile. .1 = primary education (5 years); .2 = intermediate (8 years); .3 = secondary (13 years); .4 = higher education

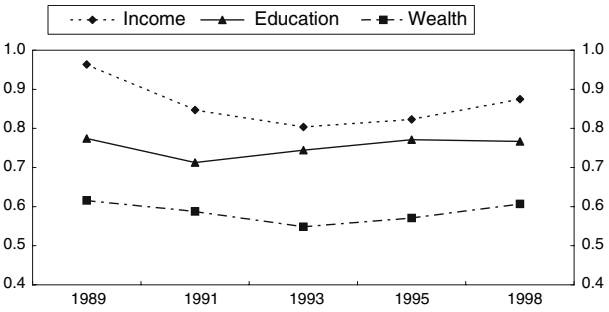

Fig. 9 The *D* index during the nineties

The indices are remarkably stable over time, although some variations can be noted.[10] For what concerns differences in income and wealth, our index decreases along the first half of the nineties, with a minimum in 1993, and a recovery in the last years. When we consider parent's education the pattern is similar, with the minimum in 1991 and a slight fall in 1998 as compared to 1995. In all cases the maximum is attained at the start of the period we analyse, in 1989: thus, discrepancy during the nineties has always been above its late-eighties level (Table 1).

The period considered is probably too short to provide any long-term interpretation. Nevertheless, if we look for a possible explanation of this pattern, the first possibility is to relate the index to an increase in university fees. If credit constraints are important then an

[10] This is especially surprising for the education *D*-index. Although the time span is limited, the education levels of parents increase over time, hence moving upward the generalized Lorenz curves of the general population. This should induce an upward trend in the *D*-index, even in the absence of any change in the generalized Lorenz curve referred to students. Hence, the stability of the *D*-index probably indicates a deteriorating degree of equality of opportunity.

Table 1 Average income, wealth and education per population group

	1989	1991	1993	1995	1998
Income (F)	32731.58	37184	39722.59	42906.99	48594.23
Income (F, 19–26)	40923.34	47259.07	49373.55	53535.35	60901.1
Income (H)	45775.45	51585.85	59780.99	63716.3	64804.32
Wealth (F)	123864.4	186537	238078.6	259894.1	303482.2
Wealth (F, 19–26)	149878.9	226628.6	310189.7	327020	382287.6
Wealth (H)	216749.2	318336.7	488368.7	484401.9	549501.9
Education (F)	7.783093	8.038197	7.942632	8.043659	8.451096
Education (F, 19–26)	8.048636	8.227871	8.144739	8.38958	8.813812
Education (H)	10.18597	10.3325	10.50772	10.48396	10.67269

Note. All variables refer to the head of household. Income and wealth are expressed in thousands of Italian Lire, education in number of years. (F) refers to the whole population; (F, 19–26) to the subpopulation of households with offspring aged 19–26, H to the subpopulation of households with offspring in higher education. Income is net of taxes and includes earnings, transfers and capital income (no equivalence scales have been used). Wealth includes financial assets, businesses, real estate and is net of debt

Table 2 *D* index and GDP growth rate

	Income	Wealth	Education	GDP growth %	GDP growth % (3 years average)
1987				2.3	
1988				4.5	
1989	0.963661	0.615532	0.774377	4.3	3.7
1990				2.5	
1991	0.847199	0.587603	0.712538	0.8	2.5
1992				1.2	
1993	0.803667	0.548681	0.744481	−1.3	0.2
1994				2.8	
1995	0.823088	0.570793	0.771119	2.1	1.2
1996				0.9	
1997				1.5	
1998	0.874816	0.606788	0.766626	1.3	1.2

increase in fees might have had an impact on the poorest. This, however, does not explain the recovery in the second half of the period considered.[11]

Another possibility is to link the index with the economic cycle. Financial constraints tend to be binding more frequently during a recession. This is especially true for students at the border between entering and not, as income tends to fall for some households and the probability to loose jobs increases. Although I do not perform a formal test of this hypothesis, the correlation between the *D* index and the GDP growth rate is quite striking. Table 2 reports the *D* indices, real GDP growth rates for each year of the survey, and real

[11] Although figures are different across universities, the increase in fees has been quite generalized and constant during the nineties. This calls for further analysis on both the policies that have been implemented (i.e. beyond the increase in fees) and the other elements that might have affected participation. On the impact of fees on enrolment in higher education see Heller (1997).

GDP growth rates averaged over the year of the survey and the two previous years. The *D* indices for income and wealth clearly show patterns similar to those of the GDP growth rates, especially of the averaged one. The *D* index referred to education shows instead no clear link with the GDP growth rates. Thus, the hypothesis that the degree of equal opportunity falls during recessions seems to deserve better attention in future research, where using data that cover a longer time span could allow a more precise identification of the impact of the economic cycle on equal opportunity.

4 Multivariate Analysis

The population characteristics that we considered in the previous section are likely to be positively correlated. Hence, further insights can be derived by focussing on appropriate sub-groups of the population, in the same way as one would use control variables in regressions: this allows us to extend the analysis to consider the role of one characteristic while "controlling" for another. In other terms, we can compare $F(x|Z_i)$ with $F(x|H,Z_i)$, where Z_i represents the variable we want to keep constant and $i = 1,..., n$ are the subgroups derived according to the characteristic Z.

We start by isolating the role of income and wealth within sub-populations divided according to the level of education of the head of the household: (1) primary school or below; (2) intermediate (8 years full time education); (3) high school; (4) further studies. We can then calculate the *D* indices referred to income and wealth within each of the four groups: these measure how far we are from independence between income (wealth) and participation within groups that are homogeneous with respect to the level of parents' education.

Figure 10 shows the index calculated for the year 1998. Again, wealth seems more relevant than income in discriminating between participants and not. However, it is now possible to notice that the role of wealth is decreasing (in the sense that the index increases towards 1) as the level of education of the household's head increases. Thus, among better educated parents, differences in wealth are less important. This could be regarded as evidence of the "cultural constraint effect". On the other hand, this result could be the consequence of the better educated having expectations of higher streams of income in the future. However, we are considering parents who are at the peak of their career (having offspring aged 19–26) and, therefore, the impact of such expectations should be rather limited. This consideration seems to point towards the presence of cultural constraints as an important limit to participation. Moreover, the role of income within groups with the same level of education is far from clear. The *D* indices in this

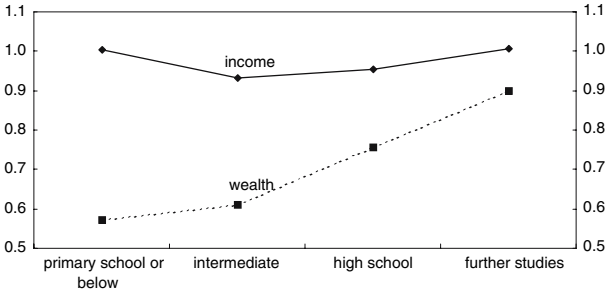

Fig. 10 *D* index within education subgroups (1998)

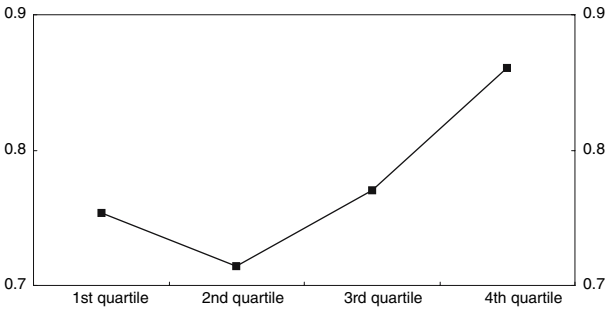

Fig. 11 *D* index of education within wealth subgroups (1998)

case do not show any obvious pattern. The index is approximately 1 both for the least and the most educated groups, thus revealing that, for the least and the most educated parents, income plays essentially no role: access into higher education is basically constant across different income classes within these groups. Instead, the index is well below 1 for groups 2 and 3.

Some further insights may come from reversing the order of the analysis and assessing the role of parents' education within given income or wealth groups. As wealth differentials seem more important than income ones, in Fig. 11 we divide the population into four groups according to the household's wealth, from the poorest 25% to the richest 25%. We then calculate the *D* indices using the head of the household's level of education. In this case the index does not show a monotone pattern, being at its minimum for the second wealth group (from $p = 0.25$ to $p = 0.5$). The distribution of the parents' level of education is clearly less important for the richest group but, quite remarkably, the index is anyway always well below 1, for all wealth groups. This means that the distribution of the parents' education matters even within groups that are homogeneous with respect to wealth. This result, once again, points to the crucial importance of considering cultural constraints when analysing access to higher education.[12]

5 Conclusion

This paper presents a simple method to assess the distance from a situation of uniform participation and equal opportunity. The method consists in comparing generalized Lorenz curves for variables beyond individual control referred to the population of participants and to a relevant control population. This method delivers an index which has then been used to provide an assessment of the ability of the Italian higher education system to select students independently of their parents' endowments.

Our results show that the distribution of participants differs in important and systematic ways from the general population in the same cohort, particularly for what concerns parents' level of education and wealth. It also shows that, during the decade 1989–1998, the situation has not improved: on the contrary, the discrepancy between the two distributions has increased, although not constantly, over time. Parents' income turns out to be the least important of the three variables considered, suggesting that efforts to equalize opportunity by providing financial support to students on the basis of their parents' income (as currently happens) could be quite misplaced. Finally, "cultural constraints" seem to

[12] These can only partially be captured by variables like parents' education. In ethnically diverse societies, for example, traditional values can play an important role: see for example Below (2007).

play a substantial role on the ability and willingness to enter into higher education, independently of possible financial constraints.

To conclude, it is worth highlighting that this methodology is not specific to participation to higher education. It can be applied to analyse a variety of other situations when it is possible to identify both a desirable outcome and the corresponding relevant population variables.

Acknowledgements I wish to thank Frank Cowell, Cristiana Vitale and participants in seminars in Banca d'Italia, Bocconi University (Workshop on Income Distribution and Welfare) and University of Chieti-Pescara for useful comments and suggestions. The usual caveat applies.

References

Arneson, R. (1989). Equality of opportunity for welfare. *Philosophical Studies, 56*, 77–93.
Banca d'Italia (1989). I Bilanci delle Famiglie Italiane nell'Anno 1989. *Supplemento al Bollettino Statistico*.
Banca d'Italia (1991). I Bilanci delle Famiglie Italiane nell'Anno 1991. *Supplemento al Bollettino Statistico*.
Banca d'Italia (1993). I Bilanci delle Famiglie Italiane nell'Anno 1993. *Supplemento al Bollettino Statistico*.
Banca d'Italia (1995). I Bilanci delle Famiglie Italiane nell'Anno 1995. *Supplemento al Bollettino Statistico*.
Banca d'Italia (1998). I Bilanci delle Famiglie Italiane nell'Anno 1998. *Supplemento al Bollettino Statistico*.
von Below, S. (2007). What are the chances of young turks and Italians for equal education and employment in Germany? The role of objective and subjective indicators. *Social Indicators Research, 82*, 209–231.
Checchi, D. (1999). Inequality in incomes and access to education. A cross-country analysis (1960–1995). *Centro Studi Luca d'Agliano, Development Studies* W.P. n. 125.
Champernowne, D. G., & Cowell, F. A. (1998). *Economic inequality and income distribution*. Cambridge University Press.
Fry, G. W. (1983). Empirical indicators of educational equity and equality: a Thai case study. *Social Indicators Research, 12*, 199–215.
Heller, D. E. (1997). Student price response in higher education: An update to Leslie and Brinkman. *Journal of Higher Education, 68*, 624–659.
Lambert, P. J. (1993). *The distribution and redistribution of income*. Manchester University Press.
Le Grand, J. (1991). *Equity and choice*. London: Harper Collins.
Roemer, J. E. (1998). *Equality of opportunity*. Cambridge: Harvard University Press.
Shavit Y., & Blossfeld H. (Eds.) (1993). *Persistent inequality: Changing educational stratification in thirteen countries*. Boulder: Westview Press.