Soc Choice Welfare (2002) 19: 795–810

Social Choice and Welfare

On the measurement of unfairness
An application to high school attendance in Argentina

Leonardo C. Gasparini

Universidad Nacional de La Plata, Departmento de Economía,
Facultad de Ciencias Económicas, Calle 6 e/47 y 48, Oficina 516, La Plata (1900), Argentina (e-mail: lgmm@netverk.com.ar)

Received: 25 August 2000/Accepted: 6 August 2001

Abstract. This paper presents a framework to measure unfairness in certain outcomes, like education attendance or basic health services consumption. The determinants of an individual outcome are divided into socially acceptable and unacceptable sources of differences in that outcome. To detect an unfair situation, comparisons are restricted to those individuals who share the same value of the vector of acceptable factors. The relevant argument to compare is the expectation of the outcome conditional on the vector of unacceptable variables. Unfairness is related to inequality in the distribution of those conditional expectations across individuals. An illustration of the framework is presented for the case of high school attendance in the Greater Buenos Aires area and other Argentine cities.

1 Introduction

Most of the studies in welfare economics aimed at measuring the fairness of social arrangements are focused on the distribution of individual utility, usually estimated by the distribution of income or total consumption. According to this utilitarian approach, the distribution of particular goods and services is not relevant since they are just arguments of the individual's utility, and only

---

I would like to thank Angus Deaton, Igal Hendel, Mariana Marchionni, Walter Sosa Escudero and seminar participants at Universidad Nacional de La Plata, Universidad de San Andrés, the 10th Annual Inter-American Seminar on Economics, NBER (Santiago), the 54th Congress of the International Institute of Public Finance (Córdoba), the LACEA meetings in Buenos Aires, and the meetings of the Asociación Argentina de Economía Política in Córdoba for helpful comments and suggestions. Of course, the usual disclaimer applies.

the latter should be of concern in a non-paternalistic society. However, in the real world, politicians, policy-makers, and people in general seem to care about the distribution of particular goods and services. Two prominent examples are education and basic health care.[1] Public programs aimed at reducing disparities in the consumption of education and basic health seem to be more popular than programs whose main goal is reducing income inequality. Rightists and leftists often agree upon the social desirability of a more fair distribution of education, but tend to disagree when discussing income distribution.

There are normative arguments behind this concern. It has long been sustained that in order to assess the fairness of a social arrangement, the emphasis should be placed on the distribution of the opportunities to attain certain outcomes, rather than on the distribution of those outcomes. Disparities in outcomes might be perfectly consistent with equal opportunities. Social scientists have championed different interpretations of the concept of equality of opportunity (see [10] and [12]). These ideas share the notion that the equalization of the "starting conditions" from where people shape their lives should be of primary social concern. It is relatively non-controversial to consider an individual's educational level and basic health status important factors in determining her set of opportunities. Therefore, the fairness in the distribution of at least certain basic levels of education and health should be of social concern.

Naturally, fairness in education does not necessarily mean equality in educational levels. Two individuals facing the same constraints may take different decisions about attending school. The inequality in educational levels that arises from those decisions may not be considered unfair. The same conclusion applies to the health case.

This paper deals with the measurement of fairness – not equality – in certain outcomes like education attendance and basic health consumption, since they are considered to be a crucial factor to attain fairness – not necessarily equality – in society. The approach calls for the partition of the variables that determine a given outcome into socially *acceptable* and *unacceptable* sources of differences in the outcome. Only those outcome differences that are due to differences in unacceptable variables are considered unfair. A particular problem is posed by the fact that variables are typically stochastic. If the intrinsic random component in the individual outcome is considered an acceptable source of inequality, then the expectation of the outcome conditional on the vector of unacceptable variables should be the object of comparison among individuals.

The rest of the paper is organized as follows. In Sect. 2 the basic framework is presented and some empirical implementation problems are discussed.

[1] Just to mention one of many examples, the Argentine Constitution establishes that it is the authority of the National Congress "to make laws regarding the organization of education which … ensure … the equality of opportunity and guarantee the principles of equity and free of charge provision of public education" (Constitución Argentina 1994, article 75, clause 19).

In Sect. 3 unfairness indices pertaining to secondary school in the Greater Buenos Aires area and other Argentine cities are calculated and interpreted. Finally, Sect. 4 concludes.

2 The framework

Although this article is mostly concerned with fairness in education attendance and basic health consumption this section presents a more general framework that can be applied to other outcomes, including income or total consumption.

A concern for the distribution of a given outcome can take two different forms depending on whether the causes of that outcome are given relevance in assessing the fairness in the outcome distribution. If only the outcomes and not their causes are considered relevant, a situation will be regarded as unfair whenever two individual outcomes differ, regardless of the causes of that difference.[2] As argued above, people tend to go beyond outcomes and look at their determinants. An unequal distribution of an outcome may be labeled as fair if the process by which it is generated is considered fair.

But how should we assess the fairness of that process? The dominant approach in the field of economics is that of equality of choice sets (see [2], [5], [8] and [13]). Factors that determine an outcome are divided into those that are given to an individual, and those that she freely chooses. For a difference in outcomes to be considered unfair, it should be the result of differences in factors in the former group.

The problem with this approach is that in most practical situations the distinction between *constraint* and *choice* is not clear. One can argue that most, and probably all factors that determine an outcome are in a sense beyond individual control: a person does not choose her preferences, her talent, her cost of exerting effort, or her rationality. Therefore, all of these variables should be included in the constraint set. But as soon as we do so, the notion of choice becomes trivial.

I prefer to avoid this philosophical discussion and focus on the social *acceptability* of the sources of differences. Inequality in a given outcome across individuals can be thought of as the result of individual differences in its explanatory variables. People tend to consider inequality as fair or unfair depending on the sources of that inequality. Differences in school attendance among youths may be considered fair if they are the result of differences in talent, effort or luck. But the same attendance differences might be labeled as unfair if their sources are differences in parental income, race or gender. Notice that talent, the cost of exerting effort, luck, parental income, race and gender are all beyond individual control. However, for some reason, people tend to consider differences in some of them *acceptable* sources of inequality

[2] A typical income distribution analysis fits into this framework. The factors that determine incomes are not scrutinized. All that matters are the actual income values, and not the process by which they are generated.

in attendance, and differences in some others *unacceptable* sources.[3] Of course different people have different views about how to partition the set of explanatory variables. Some people, for instance, would regard ability as an acceptable source of differences in outcomes; while for some others that would be unacceptable.[4] Rightists surely have a larger set of *acceptable* variables than leftists do. Societies also differ in the sources of inequities that, on average, are prepared to accept.[5]

By changing the focus of the analysis from "variables in the constraint set" to "socially unacceptable variables" we make clear that the partition of the set of explanatory variables needed to assess the fairness of an outcome depends on value judgments and cannot be performed using any seemingly objective rule. Any unfairness analysis that goes beyond outcomes must face this subjectivity. It is the user of that analysis who should provide the criterion to split the explanatory variables. This is not a simple task. However, it seems that people do have opinions about what they consider acceptable or not, although perhaps they are not ready to offer a strong and coherent philosophical framework to back those opinions.

Suppose the set of explanatory factors of a stochastic outcome x is already divided into a vector of acceptable factors (labeled as A) and a vector of unacceptable ones (labeled as U). The following definition states the concept of unfairness used in this paper.

Definition. *The distribution of a stochastic outcome x is considered to be unfair if and only if there exists a vector A and two different vectors U_i, U_j s.t $E(x/A, U_i) \neq E(x/A, U_j)$*

where $E(x/A, U)$ is the expectation of x conditional on vectors A and U. The definition implies that for a situation not to be regarded as unfair, for every given vector A, the expected value of the outcome should be the same regardless of the value of vector U.[6] Notice that for a given A, differences in out-

[3] Roemer ([10] and [11]) pioneered this approach by proposing a distinction between factors for which society believes an individual should (*effort*) and should not (*circumstances*) be held accountable.
[4] See [1] for a discussion on the acceptability of ability in determining college admissions.
[5] "Americans commonly perceive differences of wealth and income as earned and regard the differential earnings of effort, skill, foresight, and enterprise as deserved. Even the prizes of sheer luck cause very little resentment" [14]. This statement would probably not be completely true in some, for example, European and Latin American countries.
[6] Notice that this is a weak condition since (i) it does not require equal expected outcomes for different values of each unacceptable variable, but for different values of the whole vector U; (ii) it does not compare two outcomes with different values of A despite the fact that the difference between those two outcomes might be mainly driven by differences in vector U; and (iii) it does not consider fairness *across* acceptable variables (e.g., if ability is considered an acceptable source of differences in education consumption, people might not only require equality of education within each ability group, but also that the expected education consumption for talented youngsters be not lower than for non-talented ones).

comes are not considered unfair if their conditional expectations are the same. Hence, the definition implicitly assumes that the "basic and unpredictable element of randomness in human responses" [7] that remains after including all explanatory variables into the analysis is regarded as an acceptable source of differences in outcomes.[7]

The main interest of this paper is to measure the degree of unfairness and not just the presence of it.[8] Ideally the choice of an unfairness measure should be guided by the social welfare cost of inequality in the distribution of the conditional expectations $E(x/A, U_i)$. Hence, our unfairness index would be some measure of dispersion in the distribution of those conditional expectations.

There are at least three reasons why the task of measuring unfairness becomes much harder than measuring outcome inequality. First, we have to find the factors that determine an outcome. Second, we need to split the set of explanatory variables into acceptable and unacceptable sources of differences in outcomes. Finally, while in an outcome inequality analysis the target variable is usually observable, in an unfairness analysis the conditional expected value of an outcome needs to be estimated. The rest of this section briefly discusses two estimation problems.

If we consider only the set of observations that share a given value of the acceptable vector A, we can write $x_i = E(x/U_i) + e_i$. Consider that the error term e_i is just acceptable uncorrelated randomness. The typical way to tell $E(x/U_i)$ and e_i apart from each observation x_i is to express the conditional expectation as a function of U_i and to assume some structure for that function. But that structure, which is crucial to determine the division between $E(x/U_i)$ and e_i, is essentially arbitrary. To illustrate this point, suppose that non-parametric estimation is chosen. To apply this method first we have to solve the *smoothing parameter selection problem* (see [6]). In this context, that problem has both a statistical and a conceptual dimension. On the one hand, the choice of a bandwidth (or any other smoothing parameter) is a sample size issue: as the number of observations tends to infinity, the bandwidth should tend to zero. However, given a small sample size, the choice of the bandwidth becomes also a conceptual issue. The selection of the bandwidth implicitly determines the partition between expected value and error. If conceptual considerations and/or additional information lead us to believe that differences in outcomes x are mostly attributable to differences in $E(x/U_i)$, we would choose a small bandwidth that does not smooth the data very much. On the

[7] If, for instance, people decide the value of x by rolling a dice, it is relatively non-controversial to consider the outcome distribution as acceptable. But even situations where people are forced to accept the allocation of x generated by chance are also likely to be considered fair by many people. One example is the draft for the military service. Differences in outcomes are large (especially in war times). Yet, outcome differences are not seen as unfair if they are entirely due to chance.

[8] Notice that according to the definition given above most real-world situations in services like education and health would be considered unfair.

other hand, if the acceptable error term is thought to be responsible for most of the differences across individuals in the data, a larger bandwidth should be selected to be sure to eliminate the stochastic component. The same kind of considerations determines the choice between non-parametric and parametric estimation and the selection among different parametric specifications. When we are uncertain about the relative relevance of the error term the natural recommendation is to try with several smoothing parameters and parametric specifications and check for robustness.

The typical omitted variables problem is also troublesome in this context. In practice e_i may include variables we are unable to measure or detect as relevant explanatory factors. The main problem arises when some of the unobservable variables are acceptable and correlated with variables in U.[9] In that case we may incorrectly label a situation as unfair if differences in expected values across individuals with different values of U are caused by unobservable acceptable variables correlated with the unacceptable factors. In [3] it is shown that the distortion in assessing unfairness caused by this problem essentially depends on the degree of correlation between unobservable and observable explanatory factors. In addition, and since we are mainly interested in the *comparison* of unfairness measures between two outcomes (e.g., unfairness in the high school attendance decision in two different years), the key element in that bias turns out to be the difference between those outcomes in the degree of correlation between their unobservable and observable explanatory factors.

3 An application to high school attendance in Argentina

The approach outlined in the last section is applied to study unfairness in the access to secondary education in Argentina. The analysis of the primary level does not seem to be relevant since attendance rates were always close to 100% in the last decade. Secondary school is a 5-years educational level usually attended by youths from 13 to 17 years old.

Youths, or their parents, take many decisions regarding high school. They choose whether to attend or not, they select a school, and they decide the allocation of time and effort between studying and other activities. Naturally, all these decisions determine their educational outcomes and the set of opportunities they will face in the future. In this section the focus is only on the most basic decision: whether to attend high school or not. I take inequality in the probabilities of attending secondary school for groups that share the same value of the acceptable variables as sign of unfairness in the access to that educational level.

Probabilities are estimated using conventional non-parametric and parametric techniques. All the non-parametric estimations are locally weighted re-

[9] Ability in the education choice and need in the health case are typical acceptable unobservable variables, correlated with unacceptable factors (e.g., income).

gressions (*lowess*). The smoothed value of the dependent variable x_i is obtained by running a regression of x on the vector of unacceptable variables U using only the observation i and some observations close to i. The number of observations used in a regression is determined by the bandwidth.[10] The regression is weighted using a tricube function that assigns the highest weight to i. The estimated regression is used to predict the smoothed value for x_i. The procedure is repeated for each observation. The resulting curve of smoothed values is adjusted so that the mean coincides with the mean of the unsmoothed values. The smoothed value for x_i is interpreted as the estimated probability of attending high school for individual i and is used to compute the unfairness indices. The same procedure is applied using different bandwidths to check for robustness in the order of the indices. The parametric estimations are standard logit regressions. The predicted values of these regressions are used as inputs of the unfairness measures.

The decision to attend school presumably depend on many factors. Unfortunately, given the relative small number of observations available in a typical study, the analysis should keep the dimensionality low and ignore many of those factors. Also, from a practical point of view it is likely that the decision-maker's fairness concerns be posed in low dimensional terms (e.g., being worried about the relation between high school attendance and income). Four explanatory variables are used in this analysis: age, sex, income and family education. Income refers to household income adjusted by demographics.[11] When analyzing education choices for youths, their earnings are subtracted from family income to get *parental income*. Family education is approximated by the maximum of the educational levels attained by the household heads.

Parental income is considered an unacceptable source of differences in high school attendance. That will be also the case for gender in most of the exercises. Although it is sensible to consider age as an acceptable factor, lack of observations and simplicity of presentation led me in most cases to treat it as unacceptable (within the 13–17 years-old group). Parental education might be considered unacceptable in some cases, and acceptable in others, depending on the interpretation of what it is proxy for, and on value judgments.[12] Both cases are treated in this paper.

---

[10] A bandwidth of b means that $b.N$ observations are used to smooth each point in the data. The exceptions are the end points, where smaller subsets are used.

[11] Household income is divided by the number of equivalent adults in the family raised to the power of 0.8 to capture some degree of household consumption economies of scale. The equivalence scale is taken from the agency that calculates official poverty statistics in Argentina (INDEC), while 0.8 is taken arbitrarily from a sample of parameters estimated in other studies.

[12] For instance, differences in family education might be thought of as been caused by differences in wealth, and therefore considered unacceptable. If the user of the unfairness analysis is paternalistic, differences in family education might be considered unacceptable, even if those differences are driven mainly by preferences. On the other hand, family education will be regarded as an acceptable variable if preferences are fully respected.

Table 1. Attendance rates. Secondary school. Greater Buenos Aires, 1980–2000

	Averages				1996	1997	1998	1999	2000
	1980–1983	1984–1987	1988–1991	1992–1995					
All	0.66	0.71	0.73	0.74	0.74	0.80	0.86	0.87	0.91
Gender									
Male	0.62	0.67	0.71	0.70	0.71	0.78	0.83	0.85	0.90
Female	0.69	0.75	0.76	0.78	0.77	0.83	0.89	0.89	0.91
Family education									
Low	0.58	0.63	0.63	0.62	0.63	0.70	0.81	0.81	0.87
High	0.90	0.93	0.94	0.94	0.94	0.97	0.95	0.96	0.96
Age									
13–14	0.80	0.84	0.84	0.84	0.85	0.94	0.96	0.98	0.98
15–17	0.59	0.65	0.68	0.69	0.68	0.74	0.82	0.83	0.86
Income quintil									
1st	0.50	0.53	0.58	0.56	0.55	0.59	0.75	0.78	0.82
2nd	0.53	0.60	0.62	0.64	0.63	0.77	0.80	0.81	0.85
3rd	0.59	0.71	0.72	0.73	0.76	0.78	0.88	0.88	0.91
4th	0.74	0.76	0.77	0.78	0.81	0.89	0.89	0.91	0.96
5th	0.84	0.90	0.91	0.91	0.92	0.92	0.96	0.94	0.98

Source: Own calculations based on the EPH, GBA, October. *All* refers to all youths between 13 and 17 who finished primary school. Group L comprises those youngsters from families where none of the household heads has a high school degree. Income quintiles are constructed sorting individuals by parental income.

To deal with the problem of separating out the expected value and the error term I check for robustness using several bandwidths for the non-parametric estimation and several specifications for the parametric model. The omitted variables problem is present here, since we do not observe some potentially correlated explanatory variables, typically natural ability. This problem would not be very harmful for the analysis if the degree of correlation between ability and the unacceptable explanatory variables included in the regressions (typically, household income) did not significantly change in the period being analyzed. I implicitly make that assumption in what follows.

The first part of this section is devoted to measure changes in unfairness in secondary school attendance from 1980 to 2000 in the Greater Buenos Aires (GBA) area. Then, the analysis is extended to other Argentine cities. The basic information is taken from the Encuesta Permanente de Hogares (EPH), the main household survey in Argentina.

The Greater Buenos Aires area has around 12 million inhabitants, around a third of Argentina's total population. The EPH covers around 4,500 households (more than 11,000 people) in GBA. Table 1 shows attendance rates for youths in high school age (between 13 and 17) who finished primary school.[13]

[13] The surveys usually capture around 900 youths in that condition.

Attendance rates rapidly increased in the first half of the eighties and remained more or less stable until 1996. The rate dramatically increased in the late nineties due to the extension of compulsory schooling to the first two years of secondary school and the launching of an extensive system of scholarships.

High school attendance rates were always higher for women than for men. It seems that there have not been systematic changes in this gap. Naturally, the group of youths from families with low education has lower attendance rates. It is interesting to notice the substantial increase in attendance in this group in 1997/1998. Attendance grew in those years for the 13–14 years-old group, for whom secondary school was made compulsory. However, attendance rates also sharply rose for youngsters in the 15–17 age range. High school attendance appears to be associated to parental income. The dispersion of attendance rates across income quintiles has shrunk in the last two decades.

The individual probabilities of attending secondary school needed to construct the unfairness measures are estimated by parametric and non-parametric techniques. Due to the relative small number of observations, the non-parametric analysis is limited to two explanatory variables: log parental income and parental education.[14] Also, individuals are divided into only two groups (L and H) according to their family education.

For the case where the latter is taken as an unacceptable factor, all youngsters are considered together into a single unfairness index (I). On the other hand if parental education is considered an acceptable factor, two indices should be calculated, one for each family education group (I_l and I_h). The arguments of these unfairness measures should be the probabilities of high school attendance conditional on parental income for all youngsters who qualify to attend high school and who belong to a given family education group. Results for two selected years, 1992 and 1998, using a bandwidth of 0.8 are shown in Figs. 1 and 2. Observations marked with a circle (plus sign) are the estimated probabilities of youths from more-educated (less-educated) families. Only the estimated values marked with a circle are used to obtain I_h, plus signs are used to get I_l and both circles and plus signs are used to calculate I. From Figs. 1 and 2 it is clear that parental income affects the schooling attendance decision, even when controlling for parental education. That effect is more dramatic in group L. From the inspection of both figures inequality in the probabilities of high school attendance seems to be lower in 1998: the curve of predicted probabilities for group L seems flatter, and in addition the distance between curves L and H seems smaller. Of course, these presumptions should be given precise meaning: that is the purpose of the unfairness indices.

Parametric estimation allows for a richer specification. A logit regression of the attendance decision is run on log parental income, family education, a gender dummy and age.[15] Table 2 shows the results for 1992 and 1998.

---

[14] This implies that gender and age (within the group of individuals in high school age) are considered unacceptable.
[15] The parametric specification also allows working with more groups in the family education variable.

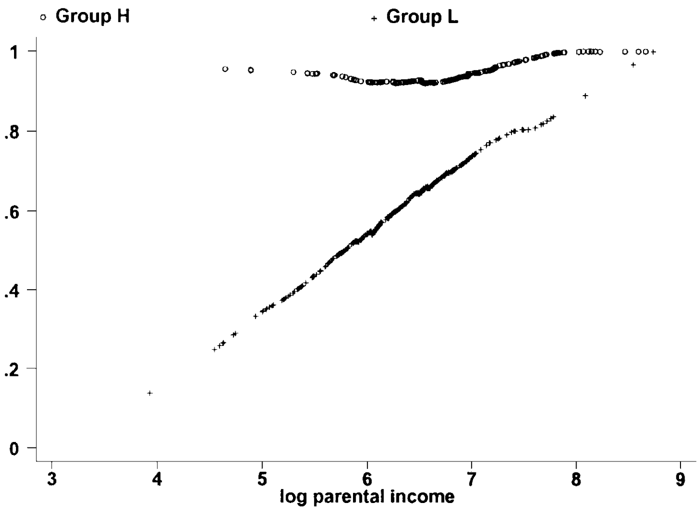

Fig. 1. Probability of attending high school. Lowess estimates GBA, 1992

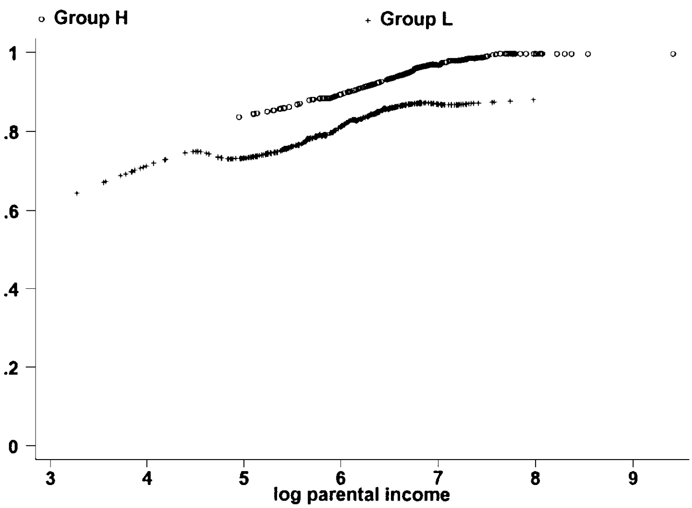

Fig. 2. Probability of attending high school. Lowess estimates GBA, 1998

Figures 3 and 4 show the predicted values from each regression for each family education group, setting age equal to 15. From the figures, predicted probabilities seem to be more concentrated around its mean in 1998, implying lower unfairness.

Table 3 shows the Gini coefficient for the distribution of conditional probabilities of attending high school, which is interpreted as an unfairness mea-

Table 2. Logit regressions of the high school attendance decision. Greater Buenos Aires, 1992 and 1998

	1992	1998
Log parental income	0.783	0.535
	(0.170)	(0.144)
Primary school degree	0.912	0.597
	(0.282)	(0.293)
High school degree	2.701	1.353
	(0.414)	(0.416)
College degree	4.159	2.730
	(1.053)	(1.068)
Male	−0.772	−0.459
	(0.198)	(0.223)
Age	−0.330	−0.590
	(0.075)	(0.098)
Constant	0.199	7.332
	(1.493)	(1.710)
No observations	598	792
Log likelihood	−308.1	−275.7
Pseudo $R2$	0.184	0.160

Note: log parental income is the log of household income (net of the youth's income) adjusted by demographics. *primary school degree* = 1 if the maximum educational degree attained by the household heads is primary school. The rest of the family educational groups are defined in a similar way. Standard errors in parenthesis.

sure for secondary school attendance.[16] The first three rows presents results obtained by estimating the probabilities using the lowess model with bandwidth 0.8 while the rest are calculated from the logit estimates. The lines labeled *all* present measures which take age, gender, income and parental education as unacceptable sources of differences in high school attendance, so all the individual probabilities are dumped together in one index. The rest of the rows are obtained under the assumption that parental education is an acceptable factor. In lines 7 and 8 gender and age are also acceptable. Results do not significantly vary across most rows. Unfairness in secondary school attendance drop in mid-eighties and slowly grew until 1996. There was a dramatic fall in all indices in the late nineties, especially during 1997/1998.[17] Unfairness seems to be closely related to attendance rates. Given that most youths from

[16] Other inequality indices, including some that are based on an absolute concept of inequality that might be more appealing to some people in this context, yield similar results so they are omitted to save space. Also results do not substantially vary by changing the bandwidth in a sensible range. Results for other indices, other bandwidths and other parametric specifications are shown in [3] and can be provided by the author upon request.

[17] The only exception to this behavior is when restricting the analysis to the group of youths from well-educated parents: there is not a clear pattern in the Gini coefficients, which in any case are extremely low.

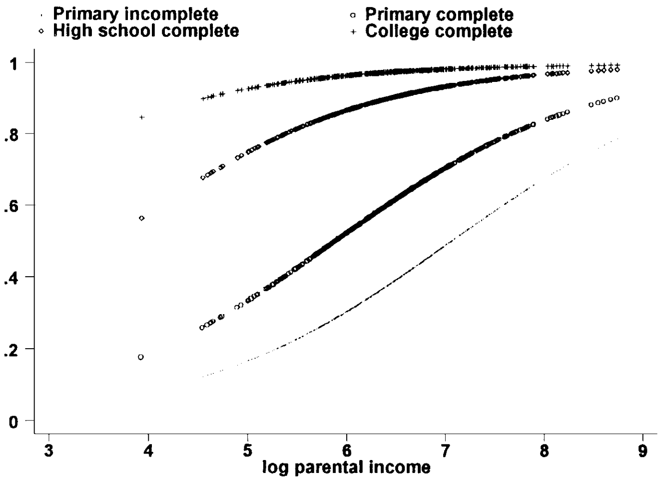

Fig. 3. Probability of attending high school. Logit estimates GBA, 1992

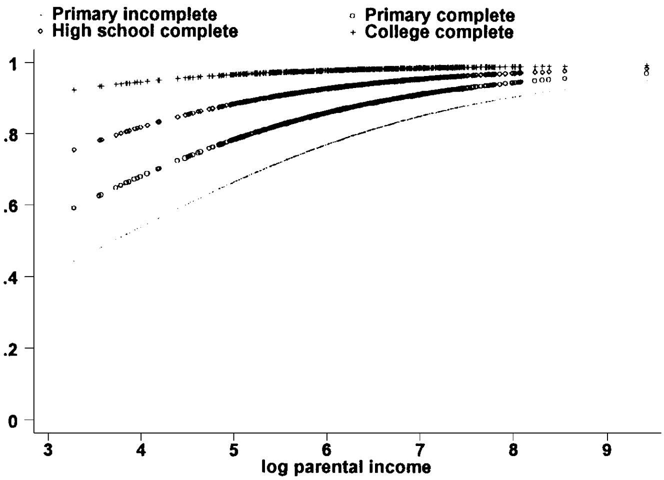

Fig. 4. Probability of attending high school. Logit estimates GBA, 1998

rich and well-educated families do already attend high school, an increase in attendance rates basically means that a higher proportion of socially disadvantaged youth make it to high school, thus lowering unfairness.

Indices can also be compared across cities. Table 4 shows a substantial change in the ordering of cities according to their degree of inequality in the probabilities of attending high school between 1996 and 1998. While Greater Buenos Aires was an area of relative high inequality in 1996, two years later its Gini was among the lowest. Again, the dramatic increase in attendance in

Table 3. Gini coefficient of the distribution of probabilities of attending high school. Greater Buenos Aires, 1980–2000

		1980–1983	1984–1987	1988–1991	1992–1995	1996	1997	1998	1999	2000
1. All	Lowess	15.4	12.5	13.2	13.7	14.1	10.2	5.5	5.4	4.3
2. Group L	Lowess	10.2	9.8	9.8	10.3	11.1	6.7	3.7	3.4	3.1
3. Group H	Lowess	2.7	2.5	1.5	1.9	1.3	1.9	2.5	0.7	1.9
4. All	Logit	19.3	15.3	15.3	15.8	16.5	14.5	8.0	8.1	6.8
5. Group L	Logit	18.0	15.7	14.5	15.0	16.6	15.7	8.8	9.7	8.4
6. Group H	Logit	3.2	3.0	2.8	3.6	2.7	2.8	2.8	2.1	2.2
7. Male L > 14	Logit	20.0	18.0	15.5	16.2	18.2	16.9	10.6	11.3	8.9
8. Female L > 14	Logit	15.7	13.8	12.8	11.9	15.0	13.4	6.8	8.6	8.6

Note: All refers to all youths between 13 and 17 who finished primary school. *Group L* comprises those youngsters from families where none of the household heads has a high school degree. *Lowess* and *logit* are the models used to estimate the probabilities. *Male L* > 14 refers to males in group L in the 15–17 age range.

Table 4. Gini coefficient of the distribution of probabilities of attending high school. Several Argentine cities 1996–1998

	All			Group L		
	1996	1997	1998	1996	1997	1998
Greater Buenos Aires	16.5	14.5	8.0	16.6	15.7	8.8
Mendoza	16.3	15.3	10.2	17.0	17.6	11.5
Rosario	13.5	17.5	15.2	12.1	17.4	15.8
Corrientes	12.1	12.1	8.4	13.5	14.9	8.1
La Plata	11.1	7.7	6.8	12.4	8.0	7.8
Santa Fe	9.7	11.1	12.5	11.8	11.2	16.0
Salta	5.0	10.0	7.0	5.0	9.4	6.6

Note: Probabilities are estimated using logit models.

that city accounts for this change. While the attendance rate grew 12 points between 1996 and 1998 in Greater Buenos Aires, the average increase for the sample of seven Argentine cities was only 4 points. The rate even decreased in some cities like Rosario, Santa Fé and Salta.

3.1 Decompositions of changes in unfairness

The change in an unfairness measure for school attendance between two years can be the consequence of changes in the characteristics of the population and/or changes in the way these characteristics are linked to the schooling decision. For instance, the dispersion in the probabilities of high school attendance among youths probably shrinks if the household income distribution becomes more equal and if the sensitivity of the attendance decision to household income decreases. Microeconometric decompositions techniques can be applied to separate out the effect of these two distinct phenomena (see [4] and [9]).

Let I be an inequality index computed over the distribution of predicted probabilities of attending school among individuals who share the same value of vector A of acceptable factors. Taking the case of parametric estimation, these probabilities are a function of the individual unacceptable characteristics U and of a vector of parameters α linking U with the attendance decision. Formally, for time t $I_t = H(\alpha_t, U_t)$.

What would have been the change in I if only the characteristics of the population U had changed between t and t'? The following two equations answer this question keeping vector α alternatively fixed at t and t' values:

$$CE_t = H(\alpha_t, U_{t'}) - H(\alpha_t, U_t) \quad CE_{t'} = H(\alpha_{t'}, U_{t'}) - H(\alpha_{t'}, U_t)$$

where CE_s stands for the "characteristics effect", keeping parameters α fixed at time s values. The "parameters effect" (PE) captures the change in I in case only the parameters α changed. Again, we can compute this effect keeping U fixed at t or t' values.

$$PE_t = H(\alpha_{t'}, U_t) - H(\alpha_t, U_t) \quad PE_{t'} = H(\alpha_{t'}, U_{t'}) - H(\alpha_t, U_{t'})$$

It is straightforward to show that the actual change in I is the sum of the averages of both effects.

$$I_{t'} - I_t = \frac{(CE_t + CE_{t'})}{2} + \frac{(PE_t + PE_{t'})}{2}$$

The implementation of this decomposition requires estimating inequality indices I with the population of one year and the parameters α of a different year. Table 5 shows these estimates for the Gini coefficient of the distribution of conditional probabilities of attending high school among all youths between 13 and 17 who finished primary school. The actual Gini in Greater Buenos Aires fell from 17.7 to 8.0 between 1992 and 1998. That drop would have been

Table 5. Simulated and actual Gini coefficients. Distribution of probabilities of attending high school. Greater Buenos Aires, 1992 and 1998

Population	Parameters	
	1992	1998
1992	17.7	6.6
1998	21.7	8.0

Note: Distribution of estimated probabilities of attending high school for all youths between 13 and 17 who finished primary school. Parameters estimated using logit models.

greater (from 17.7 to 6.6) if the characteristics of the population had not changed between 1992 and 1998.

The results of the decomposition confirm the relevance of the parameters effect to "explain" the change in the unfairness index. The Gini decreased 9.7 percentage points between 1992 and 1998: the average parameters effect accounts for a drop of 12.4 points while the average characteristics effect actually implies an increase of 2.7 points. The positive sign of the latter effect means that changes in the characteristics of the population had an inequality increasing effect on the distribution of probabilities of attending high school. This result seems sensible since household income inequality substantially increased in Argentina during the nineties: larger income disparities surely translate into larger disparities in schooling choices. In contrast, the sensitivity of the attending decision to household income and parental education seems to have changed in a dramatic way so as to drive the dispersion in the probability of attendance to significantly lower levels. As it was mentioned before it is likely that the extension of the compulsory schooling two additional years and the launching of a large program of scholarships have been the main contributors to that phenomenon.

4 Final remarks

The need for empirical work on the measurement of unfairness in the distribution of some outcomes has been repeatedly stressed. This paper takes a step in that direction by presenting a framework based on the idea that only differences in outcomes caused by differences in some socially unacceptable variables are regarded as unfair. This leads to the necessity to identify the explanatory variables and classify them according to their acceptability as sources of outcome differences. Given the stochastic nature of the social phenomena, it also introduces the need to work with conditional expected values of the outcomes, a fact that generates various estimation problems. Traditional inequality indices can be applied to measure the degree of unfairness by using the estimated conditional expectations as arguments of those indices. The paper illustrates the approach with an application to secondary education in the Greater Buenos Aires area some other Argentine cities. It is believed that the framework presented in this paper can be readily extended to measure unfairness in other outcomes, like the access to health services and unemployment.

References

[1] Denison E (1970) An aspect of inequality of opportunity. J Polit Econ 78 (5): 1195–1202

[2] Fleurbaey M (1995) Three solutions to the compensation problem. J Econ Theory 65: 505–521

[3] Gasparini L (2000) Measuring unfairness in the distribution of education. Documento de trabajo, Departamento de Economía, Universidad Nacional de La Plata
[4] Gasparini L (2001) Microeconometric decompositions of aggregate variables. An application to labor informality in Argentina. Appl Econ (forthcoming)
[5] Gravel N (1994) Can a ranking of opportunity sets attach an intrinsic importance to freedom to choose? Am Econ Rev 84 (2): 454–458
[6] Härdle W (1990) Applied nonparametric regression. Cambridge University Press, Cambridge, Mass
[7] Johnston J (1984) Econometric methods, 3rd ed. McGraw-Hill
[8] Le Grand J (1991) Equity and choice: an essay in economics and applied philosophy. Harper Collins Academic, London
[9] Oaxaca R (1973) Male-female wage differentials in urban labor market. Int Econ Rev 14 (3): 693–709
[10] Roemer J (1996) Theories of distributive justice. Harvard University Press, Cambridge, Mass
[11] Roemer J (1998) Equality of opportunity. Harvard University Press, Cambridge, Mass
[12] Sen A (1992) Inequality reexamined. Cambridge University Press, Cambridge, Mass
[13] Thomson W (1994) Notions of equal, or equivalent opportunities. Soc Choice Welfare 11: 137–156
[14] Tobin J (1970) On limiting the domain of inequality. J Law Econ 13: 263–277