

# **Equality of opportunity with categorical data**

---

**Carmen Herrero and Antonio Villar**

*Carmen Herrero and Antonio Villar*

### **Equality of opportunity with categorical data**

**Abstract:** *This paper refers to some social evaluation problems when equity matters. We propose here a way of assessing the equality of opportunity that is applicable to categorical data. It consists of dividing society into groups with similar characteristics and to measure the dispersion of outcomes within those groups, from the distribution of the population in the different categories. An empirical illustration on the equality of opportunity in education is provided, using the PISA 2012 data on mathematical knowledge regarding the Spanish regions.*

**Key words:** *Social evaluation, equality of opportunity, categorical data, PISA.*

**JEL Classification:** *I21, D63.*

### **Igualdad de oportunidades con datos categóricos**

**Resumen:** *este trabajo aborda el problema de la evaluación social cuando la equidad es un aspecto relevante. En él se propone una forma de valorar la igualdad de oportunidades aplicable en contextos donde los datos son categóricos. Este método consiste en dividir la sociedad en grupos de características similares y medir la dispersión de los resultados entre dichos grupos, a partir de la distribución de la población en diferentes categorías. El trabajo incluye una aplicación empírica sobre la igualdad de oportunidades educativas usando los datos del informe PISA 2012 sobre conocimientos matemáticos en las regiones españolas.*

**Palabras clave:** *evaluación social, igualdad de oportunidades, datos categóricos, PISA.*

**Clasificación JEL:** *I21, D63.*

### **L'égalité des chances avec des données catégoriques**

**Résumé:** *Cet article discute l'évaluation sociale du point de vue de l'équité. Pour ce faire, nous proposons une méthode pour valoriser l'égalité des chances avec des données catégoriques. Cette méthode consiste à diviser la société en groupes qui ont des caractéristiques similaires pour mesurer ensuite leur dispersion, à partir de la répartition de la population dans les différentes catégories. Cet article comprend également une application empirique concernant l'égalité des chances dans l'éducation en utilisant les données de PISA dans son rapport 2012, tout particulièrement sur les connaissances mathématiques dans les régions espagnoles.*

**Mots-clés:** *évaluation sociale, égalité des chances, données catégoriques, PISA.*

**Classification JEL:** *I21, D63.*

## Equality of opportunity with categorical data

Carmen Herrero and Antonio Villar\*

–Introduction. –I. The model. –II. An empirical illustration.  
–III. Final remarks. –References. –Appendix.

*Primera versión recibida el 28 de junio de 2013; versión final aceptada el 15 de noviembre de 2013*

### Introduction

The need to introduce distributive considerations in the analysis of social outcomes is already well established. Income distribution, health, education, or human development are instances in which this concern has been systematically incorporated from different perspectives (e.g. Roemer, 1996; Bleichrodt & van Doorslaer, 2006; Fleurbaey, 2008; Stiglitz, Sen & Fitoussi, 2009; OECD, 2010; United Nations, 2010). Equality of opportunity is one of the most relevant approaches to address distributive justice. There is a wide spectrum of views with respect to what is required for equality of opportunity,

---

\* *Carmen Herrero*: Professor of Economics, University of Alicante & Senior Researcher at the Ivie. Postal address: Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03080 Alicante, Spain. E-mail: cherreroblanco@gmail.com

*Antonio Villar*: Professor of Economics, Pablo de Olavide University & Visiting Professor at the European University Institute (Florence, Italy). Postal address: Departamento de Economía, Universidad Pablo de Olavide, Carretera Utrera km. 1, 41013 Sevilla. E-mail: avillar@upo.es

**Acknowledgements:** Thanks are due to Juan D. Moreno-Ternero for helpful discussions on an earlier version of this work and to Ildefonso Méndez for his assistance with the data. Financial support from Ministerio de Educación y Ciencia e Innovación, Junta de Andalucía, and FEDER, under projects ECO2012-34928, PROMETEO2013-037, ECO2010-21706, and SEJ-6882, is gratefully acknowledged.

from the non-discrimination viewpoint to the consideration that social provision should compensate for all forms of disadvantage. Common to those views is that individuals are accountable, to some extent, for the achievement of the advantage in question, whether this refers to health, education, income, utility or welfare. Indeed, the issue of responsibility has become prominent in some of the recent developments within the areas of political philosophy and welfare economics (see Arnesson, 1989; Cohen, 1989; Fleurbaey, 1995; Bossert, 1995; Bossert & Fleurbaey, 1996; Roemer, 1998; Fleurbaey, 2008; and the literature cited therein).

The bottom line behind the equality of opportunity principle is that people who are relatively disadvantaged due to external circumstances deserve some kind of compensation. And, complementarily, that we should not be concerned for those outcome differences that derive from other people's characteristics that could be deemed irrelevant for the problem under consideration.

We propose in this paper a way of applying the equality of opportunity approach to the evaluation of societies when individual outcomes are categorical (e.g. when the data that describe their achievements correspond to ordinal perceptions, positions in a ranking, or quality levels). To do so we start by grouping people according to their circumstances so that society is partitioned into a finite number of *types*, each of which gathers individuals with similar circumstances. In that way all people of the same type share the same opportunity and therefore, outcome differences within types will be deemed irrelevant from an ethical viewpoint. The differences in the distribution of individual achievements across types, on the contrary, can be regarded as differences in people's opportunities. This is the kind of inequality we are interested in.

The novelty of our contribution is that we provide a complete cardinal measure of equality of opportunity that is not based on decomposability principles (additive decomposability of inequality indices, as in Ruiz-Castillo, 2003; Villar, 2005; Lefranc, Pistolesi & Trannoy, 2008 and 2009 or Checchi & Peragine, 2010, or the decomposition of outcomes between characteristics and returns, as in Bourguignon, Ferreira & Menéndez, 2007 or Calo-Blanco

& García-Pérez, 2012).<sup>1</sup> Moreover, it is applicable to categorical data, a case in which only ordinal incomplete rankings are found, mostly based on the application of Lorenz dominance criteria (e.g. Allison & Foster, 2004; Abul-Naga & Yalcin, 2008; Zheng, 2011).

The paper is structured as follows. Section 1 presents the reference model. Section 2 presents an empirical application out of the last PISA data on fifteen-year old students' performance regarding mathematics in the Spanish regions. A few final words in Section 3 close the paper.

## I. The model

Consider the problem of providing an assessment of the equality of opportunity for a society with  $m$  agents,  $M = \{1, 2, \dots, m\}$ , when individual outcomes are categorical. That is, there is a partition of all possible individual outcomes into a set  $C = \{1, 2, \dots, \gamma\}$  of *categories* so that each individual outcome belongs to one of those categories. Such an assessment will be represented by an evaluation function,  $I_{opp}(\cdot)$ , a mapping from the set of individual realizations to the real numbers, which is to be interpreted as an index of *inequality of opportunity*.

Our approach is based on the idea that individual outcomes depend on two sets of variables that will be referred to as *choice variables* and *environment variables*. Choice variables are those that link outcomes with autonomous decisions and can be regarded as an expression of the agents' responsibility. They are also known in the literature as *effort variables*. Environment variables refer to the agents' external circumstances and may include aspects such as gender, inherited wealth, parental background, etc. They are also known as *opportunity variables*.

Note that the division between those types of variables involves compromises that affect the nature and the extent of the evaluation, as it implies

---

1 See, however, Almas et al. 2011, for a different approach based in the notion of the Unfair Lorenz Curve.

deciding on those aspects for which inequality is relevant and those for which it is not. Two remarks are worth considering in this respect.

First. External circumstances, in the sense used here, will typically depend on the problem under consideration. Take the case of wealth, for instance. It may well be that wealth derives from people's effort in many cases. Yet, when evaluating education or health, one usually assumes that wealth is part of the external circumstances, in the sense that it is a variable that should not affect people's outcomes in those respects (or, put differently, that education or health differences due to wealth are socially unfair).

Second. When applying the equity of opportunity principle to specific cases one may find that not all variables that affect agents' outcomes can be classified either as decision variables or environment variables. That may happen with some natural causes (e.g. age) or structural traits (e.g. social conventions). When comparing societies with different patterns for those types of variables, one has to find a way of neutralizing those differences. Age is a case particularly relevant in the evaluation of income, education or health because the outcome distribution is sensitive with respect to the age profile.

Suppose that, in spite of all those difficulties, we have already solved the question of which variables are decision variables and which ones are environment variables. Let  $T = \{1, 2, \dots, \tau\}$  denote the resulting set of *types* in which the population is partitioned, according to the agents' external circumstances. That is, all agents within a type have similar circumstances. There is a finite set of relevant categories,  $C = \{1, 2, \dots, \gamma\}$ , that summarize all possible individual outcomes.

### ***A. The evaluation formula***

We want to assess the equality of opportunity, regarding a given aspect, in a society made of  $m$  individuals of  $\tau$  different types that may produce outcomes belonging to one of  $\gamma$  different categories. The main idea behind our approach is that observed differences in the distribution of individual outcomes across types reflect the different opportunities that people enjoy.

Let  $m(t,c)$  be the number of agents of type  $t$  with outcome  $c$ ,  $m(c)=\sum_{t=1}^{\tau} m(t,c)$  the total number of agents with outcome  $c$  in society, and  $m(t)=\sum_{c=1}^{\gamma} m(t,c)$  the total number of type- $t$  agents. Call  $\alpha_{tc}$  the probability that an individual of type  $t = 1, 2, \dots, \tau$  will have outcome  $c = 1, 2, \dots, \gamma$ . That is,  $\alpha_{tc} = \frac{m(t,c)}{m(t)}$ . Let  $\pi = (\pi_1, \pi_2, \dots, \pi_{\tau})$  be the vector of population shares of the different types; that is,  $\pi_t = \frac{m(t)}{m}$ , with  $\pi_t > 0, \forall t$  and  $\sum_{t=1}^{\tau} \pi_t = 1$ .

Consider now the following matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1\gamma} \\ a_{21} & a_{22} & \dots & a_{2\gamma} \\ \dots & \dots & \dots & \dots \\ a_{\tau 1} & a_{\tau 2} & \dots & a_{\tau \gamma} \end{bmatrix} \quad (1)$$

Row  $t$  of matrix  $\mathbf{A}$ , denoted by  $\alpha(t)$ , represents the outcome distribution of type  $t$  and is to be interpreted as an expression of the differential effort of the agents of that type. Column  $c$  of matrix  $\mathbf{A}$ , denoted by  $\alpha(c)$ , describes the distribution of the categorical outcome  $c$  across types. The differences in the probabilities within a column correspond in our framework to differences among types due to their differential circumstances.<sup>2</sup>

The product  $\pi\mathbf{A} = \mathbf{b}$  yields a row vector of  $\gamma$  terms,  $\mathbf{b} = (b_1, b_2, \dots, b_{\gamma})$ , each of which describes the share of agents with outcome  $c$  in the population. That is,

$$b = \pi\alpha(c) = \frac{m(c)}{m}, \quad c = 1, 2, \dots, \gamma$$

Our target here is to define a measure that captures the inequality of opportunity associated with a matrix  $\mathbf{A}$  of relative frequencies of individual outcomes among a population consisting of  $\tau$  different types. Ideally, the dis-

---

2 Think, for the sake of illustration, of the case of two types, rich and poor, and two health states, good or bad. The rows of the corresponding matrix describe the distribution of health states for rich and poor people, respectively. The columns describe how good and bad health states are distributed between rich and poor. This type of difference is the one that we associate with the inequality of opportunity in society (here due to income factors).

tribution of every outcome across types should be uniform. That is, in a fair society the differences in people's external circumstances should not affect their realizations. As a consequence, the observed differences in the distribution of realizations across types derive from characteristics that involve diverse opportunities for the members of this society.

The inequality that is relevant for our purposes is, therefore, that *within* the columns of matrix  $\mathbf{A}$ , which describes the agents' chances of having a given categorical outcome depending upon their type. Clearly, there is no point in fostering an egalitarian distribution *within* a type, as those internal differences are, by construction, ethically irrelevant.

Given the distribution of outcome  $c$  across types,  $\alpha(c)=(\alpha_{1c}, \alpha_{2c}, \dots, \alpha_{\tau c})$ , we denote by  $i(c)$  the associated inequality measure, where  $i(\cdot)$  is a *relative inequality index* applied to such a distribution. The inequality of opportunity associated with matrix  $\mathbf{A}$  can be obtained as the weighted sum of the inequality across categories, with weights equal to the corresponding population shares. That is,

$$I_{opp}(\mathbf{A}, \mathbf{i}) = \pi \mathbf{A} \mathbf{i} = \sum_{c=1}^{\gamma} \frac{m(c)}{m} i(c), \quad (2)$$

where  $\mathbf{i} = [i(1), i(2), \dots, i(\gamma)]$  is a vector of dispersion measures among the types by categorical outcomes. This inequality of opportunity index can also be interpreted as a summary measure of the differences between the rows of matrix  $\mathbf{A}$ , which describe the distribution of outcomes across types.

Regarding the choice of the appropriate inequality index, we propose using the index of Atkinson (1970) for the value  $\varepsilon=1$ , which can be expressed as:

$$i_{Ar(1)}(\cdot) = 1 - \frac{\tilde{\mu}(\cdot)}{\mu(\cdot)}, \quad (3)$$

where  $\mu(\cdot)$  is the arithmetic mean and  $\tilde{\mu}(\cdot)$  the geometric mean. This index has several advantages: it is scale independent, has a range between 0 and 1, has a clear intuitive meaning and has been recently adopted by the United Nations as the standard to incorporate inequality concerns into the human development index (see United Nations, 2010 and Goerlich & Villar, 2009 for a discussion on inequality measures).



### ***B. Overall social evaluation and inequality-adjusted values***

In some cases, it could be interesting to get an evaluation of the society's overall performance and not only of inequality of opportunity. This can be done by means of an evaluation function  $\eta(\mathbf{A}, \cdot)$ , which associates real values to social outcomes. A particular case is that in which we have a vector  $\mathbf{q} = (q_1, q_2, \dots, q_\gamma)$  of cardinal values associated with the different categories of outcomes. Those values correspond to the weights attached to the different achievements in our evaluation. That is, each term  $q_c$  can be interpreted as the unitary contribution of an agent with outcome  $c = 1, 2, \dots, \gamma$  to the total.<sup>3</sup> Then, column vector  $\mathbf{d}$ , given by  $\mathbf{A}\mathbf{q} = \mathbf{d}$ , provides an evaluation of the average outcome of the different types, conditional on the cardinalization scheme  $\mathbf{q}$ . That is,

$$d_t = \alpha(t)\mathbf{q} = \sum_{c=1}^{\gamma} \frac{m(t,c)}{m} q_c, \quad t = 1, 2, \dots, \tau$$

The overall social evaluation of society, derived from a matrix  $\mathbf{A}$  and an evaluation criterion  $\mathbf{q}$ , will be given by

$$\eta(\mathbf{A}, \mathbf{q}) = \pi \mathbf{A} \mathbf{q} \tag{4}$$

That is,

$$\eta(\mathbf{A}, \mathbf{q}) = \sum_{c=1}^{\gamma} \frac{m(c)}{m} q_c = \sum_{t=1}^{\tau} d_t \pi_t$$

The last expression states that our evaluation criterion is embedded in an outcome evaluation function, a mapping  $\eta: \mathbf{R}_+^{\tau \times \gamma} \times \mathbf{R}_+^{\gamma} \rightarrow \mathbf{R}_+$  which associates real numbers to categories taking into account their distribution across the different types. That criterion consists of a weighted average of the values of the outcomes, with weights given by the corresponding population shares. Or, alternatively, it consists of the weighted sum of the evaluation of the types' average outcomes, with weights equal to the shares of the population types.

---

3 In some cases, particularly when dealing with categorical data, those values correspond to a particular cardinalization introduced by the analyst out of some external source. When categories correspond to intervals, those weights may reflect a centrality measure of the interval (mean or median).

We can therefore identify  $q_c \frac{m(c)}{m}$  as the implicit evaluation of outcome  $c$ , and  $d_t \pi_t$  as the contribution of type  $t$  to aggregate value.

We can combine the evaluation of social outcomes and the assessment of the equality of opportunity in a single measure,  $H(\cdot)$ , that describes the *inequality-adjusted evaluation function*. This measure can be defined as follows:

$$H(\mathbf{A}, \mathbf{q}, \mathbf{i}) = \eta(\mathbf{A}, \mathbf{q}) [1 - I_{opp}(\mathbf{A}, \mathbf{i})] \quad (5)$$

That is, function  $H(\cdot)$  applies a discount to the overall evaluation of social outcomes,  $\eta(\cdot)$ , equal to the weighted inequality across types (the inequality of opportunity index). The term  $\eta(\mathbf{A}, \mathbf{q}) / I_{opp}(\mathbf{A}, \mathbf{i})$  describes the total amount of such a reduction.

From that expression, we can estimate the *relative social loss* due to the inequality of opportunity,  $RSL_{opp}(\mathbf{A}, \mathbf{i})$ , as the ratio between the social loss,  $\eta(\mathbf{A}, \mathbf{q}) / I_{opp}(\mathbf{A}, \mathbf{i})$ , and the inequality-adjusted social evaluation,  $H(\mathbf{A}, \mathbf{q}, \mathbf{i})$ ; that is,

$$RSL_{opp}(\mathbf{A}, \mathbf{i}) = \frac{I_{opp}(\mathbf{A}, \mathbf{i})}{1 - I_{opp}(\mathbf{A}, \mathbf{i})} \quad (6)$$

This expression gives us the share of the social value that is lost due to the inequality of opportunity. Interestingly enough, equation 6 turns out to be independent of the cardinalization assumed. That is, we can have a measure of the relative social loss without actually knowing the social evaluation function. Even though the information provided by equations 2 and 5 is equivalent, the figures in expression 5 are usually easier to interpret.

## II. An empirical illustration: Inequality of opportunity in scholastic performance in the Spanish regions according to PISA 2012

To illustrate the workings of this evaluation method, we analyse the inequality of opportunity in scholastic performance of fifteen-year old Spanish students using the latest wave of the Programme for International Student Assessment (PISA), focussing on mathematical knowledge. PISA is a triennial worldwide test of 15-year-old schoolchildren's scholastic performance, the implementation of which is coordinated by the OECD. The survey provides

information about the students' test scores and their personal background, learning habits, attitudes, engagement and motivation, and characteristics of their schools. It provides information on the students' ability in three different domains: reading comprehension, mathematics, and science. Every period of assessment specialises in one particular category, but it also tests the other two. The subject specialisation is rotated through each PISA wave. The 2012 report focuses on mathematics. See OECD (2013) for details.

Most of the Spanish regions have produced an enhanced sample of PISA so that they have statistically significant data at the regional level, which allows making comparisons among them.<sup>4</sup> In order to get an estimate of the inequality of opportunity in the educational systems of the Spanish regions, we define the types and categories correspondent to the model above as follows. Types are given by the deciles of the distribution of the *PISA Index of Economic, Social and Cultural Status* (IESCS). This index combines information on parents' education and occupations and home possessions. Regarding categories, we take the distribution of the students' outcomes in the five *levels of competence* defined by PISA after aggregating levels 5 and 6 into a single one.<sup>5</sup>

Our data are thus given by thirteen matrices, as there are thirteen Spanish regions with relevant data, each consisting of ten rows and five columns. Rows correspond to the deciles of the distribution of the IESCS and columns to the level of competence. That is, the term  $\alpha'_{ic}$  corresponds to the share of agents in the decile  $t$  with competence level  $c$  in region  $i$ . Those data are incorporated into the Appendix.

Table 1 contains information regarding inequality of opportunity by levels of competence in the different regions. It is interesting to observe a common pattern between the regions: higher inequality of opportunity occurs at both the highest and the lowest level of competence. Yet the regions behave rather differently, as the coefficient of variation shows. The higher diversity

---

4 The only regions without statistically significant data in 2012 are Canarias, Castilla La Mancha and Comunidad Valenciana. Those of Extremadura were incorporated later in the report.

5 Those levels of competence are qualitative in nature, even though they are parameterized by some thresholds of the test scores.

also occurs at the highest level of competence, whereas the smaller one appears at the lowest level of competence. Andalusia, the Balearic Islands and Murcia show the highest inequality values for level 5 but rather low inequality at level 1. Navarre and the Basque Country exhibit very low inequality in both extremes. Catalonia and Madrid show mild inequality on both levels.

**Table 1.** *Inequality of opportunity by deciles of the IESCS in Spanish regions, per level of competence*

| Regions             | Levels of competence |       |       |       |       |
|---------------------|----------------------|-------|-------|-------|-------|
|                     | 1                    | 2     | 3     | 4     | 5     |
| Andalusia           | 0.142                | 0.034 | 0.028 | 0.171 | 1.000 |
| Aragon              | 0.207                | 0.024 | 0.017 | 0.112 | 0.164 |
| Asturias            | 0.242                | 0.045 | 0.012 | 0.064 | 0.252 |
| Balearic Islands    | 0.144                | 0.016 | 0.028 | 0.090 | 1.000 |
| Cantabria           | 0.119                | 0.057 | 0.010 | 0.055 | 0.314 |
| Castille and León   | 0.188                | 0.059 | 0.012 | 0.076 | 0.212 |
| Catalonia           | 0.236                | 0.057 | 0.019 | 0.156 | 0.262 |
| Galicia             | 0.112                | 0.043 | 0.011 | 0.053 | 0.253 |
| La Rioja            | 0.191                | 0.051 | 0.019 | 0.052 | 0.204 |
| Madrid              | 0.200                | 0.050 | 0.008 | 0.086 | 0.227 |
| Murcia              | 0.143                | 0.016 | 0.032 | 0.138 | 1.000 |
| Navarre             | 0.128                | 0.081 | 0.015 | 0.039 | 0.260 |
| Basque Country      | 0.140                | 0.032 | 0.003 | 0.059 | 0.213 |
| Mean                | 0.169                | 0.043 | 0.016 | 0.089 | 0.412 |
| Coeff. of Variation | 0.262                | 0.431 | 0.525 | 0.483 | 0.817 |

*Source:* OECD (2013) and own elaboration

Table 2 provides the overall measure of Inequality of Opportunity proposed and relates it to some other variables. The first two columns of the table give information about the Inequality of Opportunity index (Iopp), both in absolute terms and relative to the mean. This second column permits one to both visualize easily the rapport with the mean and to compare the Iopp with another inequality index that appears in columns 3 and 4. This index is the coefficient of determination (the square of the linear coefficient

of correlation) relating test scores and values of the IESCS. This is one of the inequality measures used by the OECD in order to assess the equity of educational systems. Even though the correlation between both measures is positive and relatively important (0.68), the data show that our Iopp measure discriminates more than the coefficient of determination and that there are some regions changing from above to below the corresponding average (Balearic Islands, La Rioja and Madrid).

**Table 2.** *Inequality of opportunity and relative social loss in the Spanish educational systems, regarding mathematical knowledge*

|                  | (1)<br>Iopp | (2)<br>Iopp /<br>mean | (3)<br>R <sup>2</sup> | (4)<br>R <sup>2</sup> /<br>mean | (5)<br>Test<br>Scores | (6)<br>(%) Relative<br>social loss |
|------------------|-------------|-----------------------|-----------------------|---------------------------------|-----------------------|------------------------------------|
| Andalucía        | 0.107       | 1.328                 | 0.16                  | 1.159                           | 472                   | 12                                 |
| Aragón           | 0.082       | 1.026                 | 0.157                 | 1.138                           | 496                   | 9                                  |
| Asturias         | 0.081       | 1.008                 | 0.156                 | 1.130                           | 500                   | 9                                  |
| Baleares         | 0.085       | 1.054                 | 0.125                 | 0.906                           | 475                   | 9                                  |
| Cantabria        | 0.067       | 0.839                 | 0.105                 | 0.761                           | 491                   | 7                                  |
| Castilla y León  | 0.071       | 0.880                 | 0.12                  | 0.870                           | 509                   | 8                                  |
| Cataluña         | 0.102       | 1.267                 | 0.175                 | 1.268                           | 493                   | 11                                 |
| Galicia          | 0.078       | 0.975                 | 0.1                   | 0.725                           | 489                   | 8                                  |
| La Rioja         | 0.078       | 0.975                 | 0.154                 | 1.116                           | 503                   | 8                                  |
| Madrid           | 0.074       | 0.928                 | 0.16                  | 1.159                           | 504                   | 8                                  |
| Murcia           | 0.097       | 1.214                 | 0.146                 | 1.058                           | 462                   | 11                                 |
| Navarra          | 0.070       | 0.867                 | 0.128                 | 0.928                           | 517                   | 7                                  |
| País Vasco       | 0.051       | 0.638                 | 0.105                 | 0.761                           | 505                   | 5                                  |
| Coeff. Variation | 0.188       |                       | 0.182                 |                                 | 0.032                 | 0.204                              |

Source: OECD (2013) and own elaboration

Column 5 provides the average test scores in order to compare performance and equity. The good news is that there is no trade-off between both aspects. Indeed, the coefficient of correlation between Inequality of Opportunity and test scores is -0.69 (even higher, in absolute terms, than the correlation between Iopp and R<sup>2</sup>). The correlation between test scores and the coefficient of determination is also negative, but much smaller (-0.13). It

is worth noting that the coefficient of variation of the inequality of opportunity is almost six times higher than that of the test scores (see the values of the coefficient of variation in the last row).

The last column of the table provides the estimate of the relative social loss due to inequality of opportunity. Andalusia, Catalonia and Murcia, which are the regions with higher inequality of opportunity, show social losses above 10 %. In the other extreme we find Cantabria, Navarre and the Basque Country with values below 7 %.

### **III. Final remarks**

Equality of opportunity is a powerful evaluation principle of the highest pertinence when dealing with such essential aspects of life as income, health or education. Applying this principle usually requires a number of compromises and case-specific adaptations that determine the extent and relevance of the analysis. Those difficulties increase when the original data are categorical.

We have presented in this paper a model for the evaluation of equality of opportunity applicable to the case in which outcomes can be categorical. The key idea is that of partitioning society in a finite set of types that collect agents with the same relevant circumstances and then compare the frequency distribution of the realizations across types.

When categories can be given a cardinal evaluation, we can also provide an overall evaluation of society's performance taking into account both average levels and inequality of opportunity. One may reasonably argue that such an evaluation becomes fully dependent on the weighting system of categories, which may derive from different sources or admit a number of alternatives. Yet it is worth noting that the analysis of equality of opportunity is independent on any cardinal valuation of the categories, which makes this analysis very robust.

In a complementary line of research, the authors have developed a model for the analysis of categorical data with an endogenous weighting system

(see Herrero & Villar, 2013) permitting one to address this problem from a slightly different viewpoint (see Herrero, Méndez & Villar, 2014). The key difference is that the model presented here provides a summary measure of inequality of opportunity for the whole society, as an aggregate of inequality of opportunity in the different categories, whereas in those papers equality of opportunity only allows comparing the relative opportunity values of the different types. Both lines of research are, therefore, complementary.

### References

- ABUL-NAGA, Ramses & YALCIN, Tarik (2008). “Inequality measurement for ordered response health data”, *Journal of Health Economics*, Vol. 27, pp. 1614-1625.
- ALLISON, R. Andrew & FOSTER, James (2004). “Measuring Health Inequality Using Qualitative Data”, *Journal of Health Economics*, Vol. 23, Issue 3, pp. 505-524.
- ARNESON, Richard (1989). “Equality and equal opportunity for welfare”, *Philosophical Studies*, Vol. 56, pp. 159–194.
- ALMAS, Ingvild; CAPPELEN, Alexander; LIND, Jo Thori; SORENSEN, Erik & TUNGODDEN, Bertil (2011). “Measuring unfair (in)equality”, *Journal of Public Economics*, Vol. 95, Issue 7-8, pp. 488-499.
- BLEICHRODT, Han & VAN DOORSLAER, Eddy (2006). “A welfare economics foundation for health inequality measurement”, *Journal of Health Economics*, Vol. 25, Issue 5, pp. 945-957.
- BOSSERT, Walter (1995). “Redistribution mechanisms based on individual characteristics”, *Mathematical Social Sciences*, Vol. 29, pp. 1–17.
- BOSSERT, Walter & FLEURBAEY, Marc (1996). “Redistribution and compensation”, *Social Choice and Welfare*, Vol. 13, Issue 3, pp. 343–355.

- BOURGUIGNON, François; FERREIRA, Francisco H. & MENÉNDEZ, Marta (2007). “Inequality of opportunity in Brazil”, *Review of Income and Wealth*, Vol. 53, Issue 4, pp. 585–618.
- CALO-BLANCO, Aitor & GARCÍA-PÉREZ, J. Ignacio (2012). “On the welfare loss caused by inequality of opportunity”, *The Journal of Economic Inequality*, DOI 10.1007/s10888-013-9247-6.
- CHECCHI, Daniele; PERAGINE, Vito (2010). “Inequality of opportunity in Italy”, *Journal of Economic Inequality*, Vol. 8, Issue 4, pp. 429–450.
- COHEN, G.A. (1989). “On the currency of egalitarian justice”, *Ethics*, Vol. 99, No. 4, pp. 906–944.
- FLEURBAEY, Marc (1995). “Equality and responsibility”, *European Economic Review*, Vol. 39, Issue 3-4, pp. 683–689.
- FLEURBAEY, Marc (2008). *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press.
- GOERLICH, Francisco & Villar, Antonio (2009). *Desigualdad y Bienestar Social: de la Teoría a la Práctica*. España: Fundación BBVA.
- HERRERO, Carmen & VILLAR, Antonio (2013). “On the Comparison of Group Performance with Categorical Data”, *PLoS ONE*, Vol. 8, No. 12.
- HERRERO, Carmen; MÉNDEZ, Ildefonso & VILLAR, Antonio (2014). “Analysis of Group Performance with Categorical Data when Agents are Heterogeneous: The Evaluation of Scholastic Performance in the OECD through PISA”, *Economics of Education Review*, Vol. 40, pp. 140–151.
- LEFRANC, Arnaud; PISTOLESI, Nicolas & TRANNOY, Alain (2008). “Inequality of opportunities vs. inequality of outcomes: Are western societies all alike?”, *Review of Income and Wealth*, Vol. 54, pp. 513–546.
- LEFRANC, Arnaud; PISTOLESI, Nicolas & TRANNOY, Alain (2009). “Equality of opportunity and luck: definitions and testable conditions, with an



- application to income in France”, *Journal of Public Economics*, Vol. 93, Issue 11-12, pp. 1189-1207.
- OECD (2010). *PISA 2009 Results: Overcoming Social Background* (vol. II). París: OECD Publishing.
- OECD (2013). *PISA 2012 Results: Excellence through Equity* (vol. II). París: OECD Publishing.
- OECD (2014). *PISA 2012 Results: What Students Know and Can do: Student Performance in Mathematics, Reading and Science* (vol. I). París: OECD Publishing.
- ROEMER, John. E. (1996). *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.
- ROEMER, John E. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- RUIZ-CASTILLO, Javiel. (2003). “The Measurement of Inequality of Opportunities”, *Research in Economic Inequality*, Vol. 9, pp. 1-34.
- STIGLITZ, Joseph; SEN, Amartya & FITOUSSI, Jean Paul (2009). “Report by the Commission on the Measurement of Economic Performance and Social Progress”. Retrieved from: <http://www.stiglitz-sen-fitoussi.fr/en/index.htm>
- UNITED NATIONS (2010). *Human Development Report. The Real Wealth of Nations: Pathways to Human Development*. New York: United Nations Development Programme.
- VILLAR, Antonio (2005). “On the Welfare Evaluation of Income and Opportunity”, *Contributions to Theoretical Economics*, Vol. 5, Issue 1.
- ZHENG, Buhong (2011). “A new approach to measure socioeconomic inequality in health”, *Journal of Economic Inequality*, Vol. 9, Issue 4, pp. 555-577.

**APPENDIX:** Distribution of the population of students in the Spanish regions by deciles of IESCS and levels of competence<sup>6</sup>

| <b>Andalusia</b> |          |          |          |          |
|------------------|----------|----------|----------|----------|
| <b>1</b>         | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.408            | 0.354    | 0.197    | 0.041    | 0.000    |
| 0.305            | 0.440    | 0.170    | 0.078    | 0.007    |
| 0.299            | 0.361    | 0.272    | 0.054    | 0.014    |
| 0.248            | 0.355    | 0.305    | 0.092    | 0.000    |
| 0.189            | 0.399    | 0.273    | 0.119    | 0.021    |
| 0.162            | 0.317    | 0.359    | 0.141    | 0.021    |
| 0.156            | 0.333    | 0.362    | 0.113    | 0.035    |
| 0.152            | 0.290    | 0.345    | 0.179    | 0.034    |
| 0.056            | 0.231    | 0.343    | 0.287    | 0.084    |
| 0.076            | 0.160    | 0.333    | 0.306    | 0.125    |

  

| <b>Aragon</b> |          |          |          |          |
|---------------|----------|----------|----------|----------|
| <b>1</b>      | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.417         | 0.309    | 0.194    | 0.058    | 0.022    |
| 0.299         | 0.255    | 0.277    | 0.146    | 0.022    |
| 0.246         | 0.289    | 0.289    | 0.141    | 0.035    |
| 0.156         | 0.193    | 0.356    | 0.237    | 0.059    |
| 0.147         | 0.213    | 0.382    | 0.221    | 0.037    |
| 0.080         | 0.248    | 0.350    | 0.263    | 0.058    |
| 0.078         | 0.220    | 0.262    | 0.326    | 0.113    |
| 0.044         | 0.228    | 0.265    | 0.353    | 0.110    |
| 0.084         | 0.161    | 0.287    | 0.350    | 0.119    |
| 0.082         | 0.150    | 0.340    | 0.340    | 0.088    |

  

| <b>Asturias</b> |          |          |          |          |
|-----------------|----------|----------|----------|----------|
| <b>1</b>        | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.270           | 0.374    | 0.227    | 0.104    | 0.025    |
| 0.258           | 0.321    | 0.270    | 0.132    | 0.019    |
| 0.162           | 0.329    | 0.341    | 0.138    | 0.030    |
| 0.158           | 0.272    | 0.304    | 0.190    | 0.076    |
| 0.156           | 0.269    | 0.338    | 0.206    | 0.031    |
| 0.082           | 0.314    | 0.352    | 0.195    | 0.057    |
| 0.108           | 0.190    | 0.411    | 0.222    | 0.070    |
| 0.113           | 0.225    | 0.338    | 0.238    | 0.088    |
| 0.024           | 0.154    | 0.320    | 0.331    | 0.172    |
| 0.019           | 0.146    | 0.285    | 0.348    | 0.203    |

<sup>6</sup> All the tables in this Appendix have been elaborated by the authors, from the data provided by the OECD (2013) vols. I and II.

**Balearic Islands**

| <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
|----------|----------|----------|----------|----------|
| 0.368    | 0.347    | 0.208    | 0.069    | 0.007    |
| 0.393    | 0.290    | 0.214    | 0.097    | 0.007    |
| 0.262    | 0.345    | 0.290    | 0.083    | 0.021    |
| 0.203    | 0.297    | 0.304    | 0.196    | 0.000    |
| 0.150    | 0.314    | 0.286    | 0.200    | 0.050    |
| 0.184    | 0.286    | 0.299    | 0.211    | 0.020    |
| 0.128    | 0.248    | 0.411    | 0.191    | 0.021    |
| 0.092    | 0.284    | 0.369    | 0.227    | 0.028    |
| 0.064    | 0.213    | 0.461    | 0.248    | 0.014    |
| 0.105    | 0.190    | 0.333    | 0.275    | 0.098    |

**Cantabria**

| <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
|----------|----------|----------|----------|----------|
| 0.320    | 0.314    | 0.229    | 0.118    | 0.020    |
| 0.219    | 0.311    | 0.318    | 0.132    | 0.020    |
| 0.192    | 0.377    | 0.245    | 0.166    | 0.020    |
| 0.199    | 0.371    | 0.258    | 0.146    | 0.026    |
| 0.183    | 0.320    | 0.327    | 0.150    | 0.020    |
| 0.159    | 0.258    | 0.351    | 0.192    | 0.040    |
| 0.098    | 0.366    | 0.327    | 0.170    | 0.039    |
| 0.107    | 0.228    | 0.336    | 0.268    | 0.060    |
| 0.046    | 0.151    | 0.336    | 0.309    | 0.158    |
| 0.088    | 0.126    | 0.296    | 0.308    | 0.182    |

**Castille and Leon**

| <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
|----------|----------|----------|----------|----------|
| 0.262    | 0.335    | 0.274    | 0.091    | 0.037    |
| 0.162    | 0.364    | 0.260    | 0.188    | 0.026    |
| 0.106    | 0.319    | 0.375    | 0.181    | 0.019    |
| 0.086    | 0.247    | 0.426    | 0.191    | 0.049    |
| 0.106    | 0.250    | 0.381    | 0.238    | 0.025    |
| 0.076    | 0.236    | 0.357    | 0.255    | 0.076    |
| 0.037    | 0.221    | 0.325    | 0.331    | 0.086    |
| 0.052    | 0.162    | 0.377    | 0.312    | 0.097    |
| 0.038    | 0.146    | 0.306    | 0.401    | 0.108    |
| 0.037    | 0.112    | 0.292    | 0.379    | 0.180    |

| <b>Catalonia</b> |          |          |          |          |
|------------------|----------|----------|----------|----------|
| <b>1</b>         | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.382            | 0.340    | 0.208    | 0.056    | 0.014    |
| 0.238            | 0.329    | 0.343    | 0.063    | 0.028    |
| 0.223            | 0.338    | 0.264    | 0.149    | 0.027    |
| 0.120            | 0.408    | 0.289    | 0.162    | 0.021    |
| 0.177            | 0.284    | 0.369    | 0.156    | 0.014    |
| 0.090            | 0.262    | 0.386    | 0.234    | 0.028    |
| 0.072            | 0.216    | 0.367    | 0.295    | 0.050    |
| 0.049            | 0.160    | 0.403    | 0.319    | 0.069    |
| 0.028            | 0.156    | 0.383    | 0.348    | 0.085    |
| 0.074            | 0.149    | 0.277    | 0.345    | 0.155    |

  

| <b>Galicia</b> |          |          |          |          |
|----------------|----------|----------|----------|----------|
| <b>1</b>       | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.299          | 0.305    | 0.286    | 0.097    | 0.013    |
| 0.258          | 0.290    | 0.271    | 0.168    | 0.013    |
| 0.208          | 0.318    | 0.318    | 0.143    | 0.013    |
| 0.192          | 0.385    | 0.269    | 0.135    | 0.019    |
| 0.221          | 0.266    | 0.305    | 0.175    | 0.032    |
| 0.137          | 0.268    | 0.353    | 0.176    | 0.065    |
| 0.096          | 0.223    | 0.420    | 0.229    | 0.032    |
| 0.088          | 0.238    | 0.347    | 0.259    | 0.068    |
| 0.071          | 0.200    | 0.374    | 0.265    | 0.090    |
| 0.083          | 0.115    | 0.395    | 0.306    | 0.102    |

  

| <b>La Rioja</b> |          |          |          |          |
|-----------------|----------|----------|----------|----------|
| <b>1</b>        | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.354           | 0.291    | 0.196    | 0.127    | 0.032    |
| 0.275           | 0.336    | 0.208    | 0.141    | 0.040    |
| 0.222           | 0.288    | 0.275    | 0.163    | 0.052    |
| 0.142           | 0.271    | 0.258    | 0.226    | 0.103    |
| 0.133           | 0.253    | 0.313    | 0.233    | 0.067    |
| 0.135           | 0.219    | 0.342    | 0.271    | 0.032    |
| 0.094           | 0.188    | 0.336    | 0.282    | 0.101    |
| 0.077           | 0.160    | 0.372    | 0.288    | 0.103    |
| 0.053           | 0.127    | 0.267    | 0.347    | 0.207    |
| 0.038           | 0.127    | 0.268    | 0.338    | 0.229    |

| <b>Madrid</b> |          |          |          |          |
|---------------|----------|----------|----------|----------|
| <b>1</b>      | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.299         | 0.351    | 0.234    | 0.104    | 0.013    |
| 0.247         | 0.292    | 0.305    | 0.136    | 0.019    |
| 0.181         | 0.331    | 0.306    | 0.156    | 0.025    |
| 0.121         | 0.275    | 0.389    | 0.188    | 0.027    |
| 0.113         | 0.294    | 0.344    | 0.213    | 0.038    |
| 0.074         | 0.196    | 0.338    | 0.304    | 0.088    |
| 0.047         | 0.180    | 0.373    | 0.353    | 0.047    |
| 0.052         | 0.155    | 0.316    | 0.387    | 0.090    |
| 0.064         | 0.128    | 0.333    | 0.359    | 0.115    |
| 0.045         | 0.192    | 0.321    | 0.327    | 0.115    |

  

| <b>Murcia</b> |          |          |          |          |
|---------------|----------|----------|----------|----------|
| <b>1</b>      | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.479         | 0.300    | 0.164    | 0.050    | 0.007    |
| 0.314         | 0.343    | 0.241    | 0.095    | 0.007    |
| 0.324         | 0.388    | 0.173    | 0.094    | 0.022    |
| 0.286         | 0.353    | 0.301    | 0.060    | 0.000    |
| 0.236         | 0.354    | 0.313    | 0.083    | 0.014    |
| 0.167         | 0.409    | 0.303    | 0.098    | 0.023    |
| 0.219         | 0.277    | 0.285    | 0.161    | 0.058    |
| 0.117         | 0.314    | 0.321    | 0.204    | 0.044    |
| 0.085         | 0.298    | 0.397    | 0.163    | 0.057    |
| 0.075         | 0.209    | 0.313    | 0.313    | 0.090    |

  

| <b>Navarre</b> |          |          |          |          |
|----------------|----------|----------|----------|----------|
| <b>1</b>       | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.242          | 0.340    | 0.248    | 0.150    | 0.020    |
| 0.115          | 0.293    | 0.344    | 0.197    | 0.051    |
| 0.155          | 0.270    | 0.358    | 0.182    | 0.034    |
| 0.137          | 0.203    | 0.379    | 0.248    | 0.033    |
| 0.071          | 0.221    | 0.409    | 0.253    | 0.045    |
| 0.072          | 0.243    | 0.289    | 0.303    | 0.092    |
| 0.061          | 0.196    | 0.318    | 0.324    | 0.101    |
| 0.065          | 0.110    | 0.368    | 0.348    | 0.110    |
| 0.045          | 0.117    | 0.305    | 0.364    | 0.169    |
| 0.058          | 0.090    | 0.231    | 0.346    | 0.276    |

| <b>Basque Country</b> |          |          |          |          |
|-----------------------|----------|----------|----------|----------|
| <b>1</b>              | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> |
| 0.244                 | 0.341    | 0.305    | 0.097    | 0.013    |
| 0.149                 | 0.295    | 0.362    | 0.169    | 0.025    |
| 0.143                 | 0.285    | 0.360    | 0.179    | 0.033    |
| 0.121                 | 0.281    | 0.366    | 0.206    | 0.026    |
| 0.087                 | 0.267    | 0.350    | 0.241    | 0.055    |
| 0.096                 | 0.234    | 0.376    | 0.256    | 0.037    |
| 0.061                 | 0.213    | 0.369    | 0.285    | 0.072    |
| 0.056                 | 0.177    | 0.321    | 0.364    | 0.082    |
| 0.028                 | 0.147    | 0.360    | 0.326    | 0.140    |
| 0.109                 | 0.175    | 0.297    | 0.306    | 0.113    |