serie ad

WP-AD 2013-08

Analysis of group performance with categorical data when agents are heterogeneous: The case of compulsory education in the OECD

Carmen Herrero, Ildefonso Mendez and Antonio Villar

Ivie

Working papers

Working papers

Working papers

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

Ivie working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication. Ivie's decision to publish this working paper does not imply any responsibility for its content.

La Serie AD es continuadora de la labor iniciada por el Departamento de Fundamentos de Análisis Económico de la Universidad de Alicante en su colección "A DISCUSIÓN" y difunde trabajos de marcado contenido teórico. Esta serie es coordinada por Carmen Herrero.

The AD series, coordinated by Carmen Herrero, is a continuation of the work initiated by the Department of Economic Analysis of the Universidad de Alicante in its collection "A DISCUSIÓN", providing and distributing papers marked by their theoretical content.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie http://www.ivie.es, así como las instrucciones para los autores que desean publicar en nuestras series.

Working papers can be downloaded free of charge from the Ivie website http://www.ivie.es, as well as the instructions for authors who are interested in publishing in our series.

Versión: noviembre 2013 / Version: November 2013

Edita / Published by:
Instituto Valenciano de Investigaciones Económicas, S.A.
C/ Guardia Civil, 22 esc. 2 1º - 46020 Valencia (Spain)

WP-AD 2013-08

Analysis of group performance with categorical data when agents are heterogeneous: The case of compulsory education in the OECD

Carmen Herrero, Ildefonso Mendez and Antonio Villar*

Abstract

This paper analyses the evaluation of the relative performance of a set of groups when their outcomes are defined in terms of categorical data and the groups' members are heterogeneous. This type of problem has been dealt with in Herrero and Villar (2012) for the case of a homogenous population. Here we expand their model controlling for heterogeneity by means of inverse probability weighting techniques. We apply this extended model to the analysis of compulsory education in the OECD countries, using the data in the PISA. We evaluate the relative performance of the different countries out of the distribution of the students' achievements across the different levels of competence, controlling by the students' characteristics (explanatory variables regarding schooling and family environment). We find that differences in reading ability across OECD countries would lower by 35% if their endowment of students' characteristics would be that for the OECD average.

Keywords: Group performance, compulsory education, heterogeneity, categorical data, inverse probability weighting.

JEL classification numbers: I24, C14.

---

* C. Herrero: University of Alicante, Ivie and IAE. I. Mendez: University of Murcia. A. Villar: Pablo de Olavide University and Ivie. Corresponding author: I. Mendez, University of Murcia, Departamento de Economía Aplicada, Facultad de Economía y Empresa, 30100, Espinardo (Murcia), Spain. Phone: +34 868883732. Fax: +34 868883745. E-mail: ildefonso.mendez@um.es.

We thank the editor and an anonymous referee for their comments and suggestions that led to considerable improvements of the paper. Thanks are also due to Dirk Van de Gaer, X. Ramos, F. Goerlich and the participants in different seminars for helpful comments. This paper was written while the first author was visiting the IAE, under the Salvador de Madariaga project PR2010-0124. This paper is framed within the projects SEJ2007-62656, SEJ-6882/ECON, ECO2010-21706 and PROMETEO 2013/037.

1 Introduction

We consider here an evaluation problem in which we have to compare the relative performance of several groups, out of the distribution of the achievements of their members in a set of ordered categories. Think for instance of the comparison of the health situation of different countries out of the distribution of the population in four or five health statuses (e.g. from "excellent" to "very bad"). The key elements of the problem are, therefore, the presence of several groups, the qualitative nature of the outcome variable, which resolves into a given set of ordered categories, and the focus on relative performance.

This type of problem has been addressed recently by Herrero and Villar (2012). They start by considering pairwise comparisons between groups in terms of the probability that an agent picked at random in one group belongs to a higher category than an agent randomly chosen in the other. Then they extend those comparisons to all groups involved, by taking into account both direct and indirect relations. As a result they obtain an evaluation function that corresponds to the dominant eigenvector of a matrix that describes all those comparisons (see below). This evaluation function is characterized in terms of some ethical and operational properties.

An implicit assumption in their model is that groups are homogeneous so that the distribution of the outcome variable is the sole relevant information. Yet one might be interested in evaluating not only the observed outcomes but also the extent to which those outcomes reflect diverse structural characteristics of the population that affect the agents' performance. This may well be the case in the example mentioned above (comparing the health situation of different countries), regarding the influence of aspects such as age or wealth in the final outcomes. To deal with this type of evaluation we need a methodology that permits making comparisons in terms of a common set of characteristics. This is the key point of this paper. More specifically, we combine here the original model in Herrero and Villar (2012) with inverse probability weighting (IPW) techniques that permits one controlling for differences in the distribution of the determinants of the outcome variable.

Using this methodological approach we obtain a covariate-adjusted evaluation that allows isolating the impact of the selected explanatory variables, by comparing this evaluation with the unadjusted one. In that way we can separate the part of the observed differences that is explained by the covariates and the part which cannot be accounted for. The covariate-adjusted eigenvector tells us about the relative performance of the groups once their conditioning variables have been equalized. Comparing the covariate-adjusted and the unadjusted evaluations permits one to estimate the impact of the latent variables on the relative performance.

The interpretation of the differences between both evaluations depends on the problem at hand and, in particular, on the choice of the explanatory variables. In this respect our analysis is reminiscent of the "equality of opportunity" literature, as covariate-adjusted values might be interpreted as an expression of the differential "effort", whereas the unadjusted values would reflect the interplay of both effort and "opportunity".[1] Yet, this model does not provide a "measure" of equality of opportunity, as our comparison deals with relative performance both in the adjusted and unadjusted evaluations.

We apply this extended model to the evaluation of compulsory education in the OECD using the data provided by the Program for International Students Assessment (PISA). We evaluate the performance of schoolchildren regarding reading ability, out of the 2009 data set (the last one available). Our evaluation involves the estimation of the impact of the students' environment (parental and school characteristics) on the final scores. Comparing the adjusted and unadjusted evaluations allows concluding that the set of explanatory variables accounts for 35% of the differences in the relative performance. We also consider how those differences have evolved in the first decade of the 21st Century, by comparing the results in 2009 and those in 2000. We find that differences in students' reading ability across OECD countries, both adjusted and unadjusted, have substantially decreased during that period, particularly so for European OECD countries.[2]

The paper is organised as follows. Section 2 presents the formal model whereas Section 3 applies it to the results on reading competence out of the data in the PISA (2000 and 2009). Section 4 gathers a few final comments.

2 The Model

Consider a set of g groups or societies, $G = \{1, 2, ..., g\}$, each of which consists of n_i agents, $i \in G$. We want to compare the relative performance of those groups with respect to a given aspect, when their achievements are given in categorical terms. More precisely, we assume that there is a set of categorical positions, $H = \{h_1, ..., h_s\}$, ordered from best to worst, $h_1 \succ \cdots \succ h_s$ (health statuses, educational levels, age intervals, professional positions, etc.). Each group presents a given distribution of achievements across those categories. Our goal is comparing their relative performance, taking into account the role of the differences in the structural characteristics that may influence the outcome variable. To do so we divide the evaluation problem into two parts. First,

[1] Equality of Opportunity (EOp) is one of the most prominent concepts of distributive justice. The key idea behind this concept is that the concern about inequality should not focus on the equality of outcomes but rather on the existence of a common *playing field* for all people. From this perspective agents' outcomes can be regarded as deriving from two different sources: *effort and opportunity*. Effort refers to people's decisions whereas opportunity refers to the agents' external circumstances. A fair society is one in which final outcomes do not depend much on the agents' external circumstances, that is, a society in which all people share similar opportunities. In that society outcome differences are basically determined by the agents' preferences and effort and not by aspects that are beyond their control and responsibility (see Arneson, 1989; Cohen, 1989; Roemer, 1993, 1998; Fleurbaey, 2008).

[2] Here again our analysis is very close to that of equality of opportunity in education. See on this respect Peragine and Serlenga (2008), Lefranc et al. (2008), Chechi and Peragine (2010), Calo-Blanco and Villar (2010), OECD (2010 b), Villar (2012), Calo-Blanco and García-Pérez (2013).

we assume that all groups are homogeneous regarding those characteristics, so that the evaluation only takes into account their relative achievements. The key point here is how to make systematic comparisons out of qualitative data. Second, we consider that groups are heterogeneous and provide a method to control for such heterogeneity.

2.1 The evaluation formula when groups are homogeneous

Let $G = \{1, 2, ..., g\}$ stand for a set of g groups under the assumption that they are homogeneous with respect to the aspect under evaluation. Let a_{ir}, for $i = 1, ..., g$, $r = 1, ...s$, be the proportion of people of group i in position r, and let A stand for the matrix that collects all those values (that is, the ith row of matrix A describes the distribution of achievements of group i across the different categories, in terms of relative frequencies).

An evaluation problem, or simply a problem, can be summarized by that matrix A, under the assumption that the set of groups, G, and the set of categorical positions, H, are given, and that we focus on the relative frequencies of the agents across the categorical positions, independently on the size of the groups (a property known as *group replication invariance*). Our target is to define a suitable evaluation function that enables comparing the relative performance of the different groups.

Given an evaluation problem A we say that group i dominates group j when it is more likely that an agent chosen at random from i occupies a higher position than an agent chosen at random from j. Let p_{ij} be the probability that an agent from group i occupies a higher position than an agent from group j, and let $e_{ij} = 1 - p_{ij} - p_{ji}$ (that is, the probability that an individual in i, chosen at random, belongs to the same category that an individual in j, randomly chosen). Note that, those probabilities can be easily computed as follows:

$$p_{ij} = a_{i1}(a_{j2} + \cdots a_{js}) + a_{i2}(a_{j3} + \cdots a_{js}) + \cdots + a_{i,s-1}a_{js}$$
$$e_{ij} = a_{i1}a_{j1} + \cdots + a_{is}a_{js}$$

In pairwise comparisons, the quotient p_{ij}/p_{ji} tells us the relative advantage of group i with respect to group j. That is, $p_{ij}/p_{ji} > 1$ implies that people in i have advantage over people in j, and viceversa.[3]

Remark 1 *When the distribution of the population in group i stochastically dominates that of group j, we have that $p_{ij} > p_{ji}$.*[4]

[3] Lieberson (1976) in a similar vein introduces the Index of Net Difference, $ND(i, j) = \mid p_{ij} - p_{ji} \mid$ to inform about inequalities between two groups. If $ND(ij) = 0$, then $p_{ij} = p_{ji}$ that is, it is equally likely, given an individual chosen at random in any of the two groups that the individual in i is at a better position than the individual in j than the other way around. The other extreme case is when $ND(i, j) = 1$, which happens whenever all individuals in one of the groups are at better positions than those in the other group. Intermediate positions provide with values of $ND(i, j)$ between 0 and 1.

[4] Lefranc, Pistolesi and Trannoy (2008, 2009) make use of stochastic dominance for income distributions to compare equality of opportunity among two different groups of people.

Note that, when there are more than two groups, pairwise comparisons only cover part of the relevant domination relationships. This is so because in that case one has to take into account not only direct dominance relations but also the indirect ones. That is, the relative position of group i with respect to group j also depends on how those groups relate to third parties.

Herrero and Villar (2012) introduce a summary measure of relative achievements that can take into account all those relations. They define the *relative advantage* of group i with respect to group j, RA_{ij}, as the ratio between the probability that i dominates j and the sum of the probabilities that group i be dominated by other groups (a number between 0 and $(g-1)$). Formally:

$$RA_{ij} = \frac{p_{ij}}{\sum_{k \neq i} p_{ki}}$$

That is, RA_{ij} is directly proportional to the probability of i dominating j and inversely proportional to the sum of the probabilities of i being dominated by some other group. Observe that when there are only two groups, $RA_{ij} = p_{ij}/p_{ji}$, whereas this value changes in the presence of more groups. A summary measure of the overall advantage of a given group can thus be obtained as a weighted average of the relative advantages with respect to all groups. That is,

$$RA_i = \sum_{j \neq i} \lambda_j RA_{ij}$$

where λ_j is the weight attached to group j.

It is natural to look for a weighting system consistent with the evaluation of relative advantages. That is, a weighting system such that $\lambda_j = RA_j$. That means that one has to find a vector $\mathbf{v} = (v_1,\ v_2,\ ...,\ v_g)$ such that:

$$v_i = \sum_{j \neq i} v_j RA_{ij} = \sum_{j \neq i} \frac{v_j p_{ij}}{\sum_{k \neq i} p_{ki}} \tag{1}$$

Herrero and Villar (2012) show that such a vector always exists, it is strictly positive and unique (up to normalization), and has an interesting number of additional properties. Moreover, that vector $\mathbf{v}$ is easy to compute, since it corresponds to the Perron eigenvector of the following matrix:

$$Q = \begin{pmatrix} g-1-\sum_{i \neq 1} p_{i1} & p_{12} & \cdots & p_{1g} \\ p_{21} & g-1-\sum_{i \neq 2} p_{i2} & \cdots & p_{2g} \\ \cdots & \cdots & \cdots & \cdots \\ p_{g-1,1} & p_{g-1,2} & \cdots & p_{g-1,g} \\ p_{g1} & p_{g2} & \cdots & g-1-\sum_{i \neq g} p_{ig} \end{pmatrix}$$

The interpretation of the components of matrix Q is the following. Off-diagonal elements (elements in place ij with $i \neq j$) are simply p_{ij}, that is, the probability that an individual chosen at random in group i is at a higher position than an individual chosen at random in group j. Thus all off-diagonal elements capture the relative dominance between pairs of groups. As for the elements in the diagonal, the element jj provides the probability of someone chosen at random in group j to be at least in a position as good as (or better than) anyone in any other group.

It is easy to check that matrix Q is a Perron matrix all whose columns add up to $(g-1)$. Therefore, assuming that matrix Q is irreducible, there is a unique eigenvector, $\mathbf{v} >> \mathbf{0}$, absorbent, and such that its components add up to g. The components of such an eigenvector satisfy equation [1], and thus, provide the evaluation we are looking for. Vector $\mathbf{v}$ can thus be regarded as a summary measure of the relative performance of the different groups.

2.2 Controlling for heterogeneity

We now show how to combine this model with inverse probability weighting (IPW) techniques that permit one controlling for differences in the distribution of the determinants of the outcomes in the different groups. We obtain in this way a covariate-adjusted eigenvector that provides an evaluation of the relative performance once the impact of the differences in the distribution of the covariates has been cancelled. Comparing covariate-adjusted and unadjusted evaluations tells us about the influence of the explanatory variables in the observed performance.

The IPW estimators are easy to implement, allow for an undetermined amount of heterogeneity in the estimates, and make no assumption on the distribution of the outcome variable H. Additionally, it has been shown in the treatment effects literature that the IPW estimators provide consistent and in some cases asymptotically efficient estimates of the parameter of interest under fairly standard regularity conditions. Furthermore, Busso et al. (2009) showed that the IPW estimators exhibit the best overall finite sample performance among the broad class of treatment effect estimators. This is particularly relevant in the current context since estimation samples are of modest size in many empirical applications.

The goal is to ensure the same distribution of the covariates (X) in each category (h) of each group (r) used to calculate the eigenvector. We first choose one of the $C = g \cdot s$ subsamples in which the total sample is partitioned, labeled c_r, as the reference sample, i.e. that whose distribution of covariates is to be used in the remaining $C-1$ subsamples. Alternatively, we could use the total sample for group g or the overall sample for all the groups as the reference sample. Later in this section we analyze how the reference sample affects the outcome of the evaluation tool and we also provide some insights on how to select the reference sample.

Next, we generate a set of dummy indicator variables Z_c that equal one if an observation belongs to the reference sample and zero if it belongs to subsample c, for $c \in C, c \neq c_r$. We then estimate the conditional probability of being in the reference sample given X, i.e. $p_c(x) = P(Z_c = 1/X = x)$ for each observation in subsample c, for $c \in C, c \neq c_r$. This variable is known as the *propensity score* in the treatment effects literature. The research value of the propensity score rests on its power to solve the dimensionality problem, since adjusting for between-groups differences on a high dimensional vector of covariates can be either difficult or impossible. Rosenbaum and Rubin (1983) show that the propensity score captures all of the variance on the covariates relevant for adjusting between-group comparisons, that is, treated ($Z_c = 1$) and control ($Z_c = 0$) units with the same value of the propensity score have the same distribution of the elements in X.

The propensity score can be estimated by means of a simple binary choice model like a logit or a probit model or by nonparametric methods like power series regression. The distribution of the covariates in a particular subsample is changed for that in the reference sample by simply introducing the appropriate weighting function λ_c. Formally, let $g(X)$, $g(X/Z_c = 1)$ and $g(X/Z_c = 0)$ be the joint density of X in the estimation sample, in the reference subsample and in subsample c, respectively, and observe that by definition,

$$g(X) = \frac{g(X/Z_c = j)\, P(Z_c = j)}{P(Z_c = j/X)}, \text{ for } j = \{0, 1\}.$$

Then, it follows that:

$$\underbrace{\frac{P(Z_c = 1/X)\,(1 - P(Z_c = 1))}{(1 - P(Z_c = 1/X))\, P(Z_c = 1)}}_{\lambda_c} \cdot g(X/Z_c = 0) = g(X/Z_c = 1), \tag{2}$$

where $P(Z_c = 1)$ is the proportion of observations from the reference sample in the estimation sample. This equation suggests a simple three-step method to change the distribution of X in each subsample for that in the reference sample. First, get an estimate of the propensity score for each observation in the sample. Second, plug the estimated propensity score and the proportion of observations from the reference sample in the estimation sample into the sample analog of λ_c to obtain an estimate of the weighting function.[5] Next, use the estimated λ_c to weight observations of subsample c and calculate, for each group, the proportion of observations in each category. The covariate-adjusted eigenvector is obtained by simply applying the evaluation tool due to Herrero and Villar (2012) to the new percentages.

This weighting scheme works by weighting-down (-up) the distribution of 1s of the dummy indicator variable $(1 - Z_c)$ for observations in subsample c for those values of the elements of X that are (over-) under-represented among observations in the reference sample. The following overlap assumption on the joint distribution of Z and X is necessary for the estimation problem to be well defined: $0 < P(Z_c = 1/X) < 1$, for $c \in C$. This common support condition states that for a given value of X there is some fraction of the estimation sample in the reference sample and in subsample c to be compared. Lack of overlap as well as estimated propensity scores close to one can lead to imprecise estimates of λ_c. To overcome this limitation we follow Crump et al. (2009). They propose

[5] The weights are normalized so that they add up to one within each subsample.

a systematic approach to addressing lack of overlap in estimation of average treatment effects by characterizing optimal estimation subsamples for which the average treatment effect can be estimated most precisely. Remarkably, the optimal rule depends solely on the propensity score in most cases. We apply the procedure in Crump et al. (2009) for the treatment effect on the treated, the parameter whose weighting function is that in [2], to each of the comparisons that we perform.

That could be argued that only the observations of the reference sample that belong to all the optimal estimation subsamples defined according to Crump et al. (2009) should be used to calculate the covariate-adjusted eigenvector. However, that requirement can only be satisfied if C is not very large and the subsamples being compared do not differ to a great extent in the distribution of the variables in X. Otherwise, the cost in terms of sacrificed external validity would be higher than the improvement in internal validity.

The choice of the reference sample affects the relative importance of each category within each group (a_{ij}), the probability that an agent from group i occupies a higher position than an agent from group j (p_{ij}), the ratio between any two p's and, thus, the outcome of the evaluation process (v_i).[6] The more selected the reference sample, the higher the cost in terms of sacrificed external validity. That is the case for the lowest and highest proficiency levels, particularly so in the latter case for the less developed OECD countries. We use the overall sample of OECD countries as the reference sample in our application. This way we analyze how the relative scholastic performance of the OECD countries would change if their endowment of environmental variables would be that for the OCDE average.[7]

The covariate adjusted matrix corresponds to the homogenous groups case. That is, the information has been transformed so that it describes the outcomes that would have been obtained if all groups were similar regarding the explanatory variables. The covariate-adjusted eigenvector, $\mathbf{v}^*$, tells us the relative differences in the groups that are not accounted for those variables. And, consequently, the difference between the eigenvector obtained with the original data and the covariate-adjusted one tells us about the impact of leveling the environmental variables in the evaluation of the relative performance of the different groups.

This model is applied in the next Section to evaluate schoolchildren scholastic performance across 29 OECD countries. We do so by considering the distribution of the students in five categories of reading competence (the six determined by PISA, after merging levels 5 and 6). This leaves us with 145 subsamples to compare. Moreover, there exists a large heterogeneity in the distribution of the variables in X both between and within OECD countries. For these reasons,

[6] Conversely, the ratio between any two λ's is invariant to the selection of the reference sample.

[7] Our results remain qualitatively largely unchanged when we use particular subsamples as the reference sample. In general, we find that the more selected is the reference subsample, the higher is the difference that accounting for covariates makes with respect to the unadjusted eigenvector.

we cannot restrict the reference subsample to those observations satisfying the optimal selection rule in all the comparisons. The same holds when we restrict the analysis to the 21 OECD European countries.

3 The evaluation of compulsory education in the OECD through PISA

Compulsory education is probably the most basic instance of social insurance as it guarantees minimal levels of knowledge to all citizens in a given society, which in turn conditions their opportunities regarding access to the labour market, further education, and the extent of social interactions. Most OECD countries have established a ten- to eleven-year period of compulsory education (from 6 to 15 or 16 years). It is important to know how effective this education is, how different are the educational outcomes both between and within countries, and how much those differences depend on the countries' observable characteristics.

Here we apply the model presented above to evaluate compulsory education in the OECD countries out of the data provided by the Programme for International Student Assessment (PISA). This Programme provides the broadest dataset for the evaluation of schoolchildren performance and the characteristics of their schooling and family environments. It is a triennial worldwide test of 15-year-old schoolchildren's scholastic performance, the implementation of which is coordinated by the OECD. PISA surveys started in 2000 with the aim of evaluating the students' ability, about the end of compulsory education, in three different domains: reading, mathematics and science.

Every period of assessment specialises in one particular category, but it also tests the other two main areas studied. The 2009 report has focused on reading ability, which is a key determinant of individuals' learning capacity and conditions their participation in social life. "Levels of reading literacy are more reliable predictors of economic and social wellbeing than is the quantity of education as measured by years at school or in post-school education… It is the quality of learning outcomes, not the length of schooling, that makes the difference." (OECD (2010a), p. 32).

We focus on the data in PISA 2009 to analyze the different performance of educational systems regarding schoolchildren's reading ability across OECD countries.[8] The focus of our analysis is threefold. First, we investigate the relative performance of OECD countries in 2009 comparing the distribution of mean score tests and their relative performance, as measured by the corresponding (unadjusted) eigenvector. Second, we estimate the students' relative achievements once their external circumstances have been equalized, using two different sets of explanatory variables. And third, we provide a rapid overview of the changes experienced in 2009 with respect to 2000. Given this purpose, we

[8] One of the assets of the PISA report is that it provides a unified scoring system to evaluate the performance of 15-year-old students in very different countries. The units of those scores are set with respect to the values obtained in the 2000 wave of the report, by taking a value of 500 for the average of the OECD Member States with a standard deviation of 100.

can only consider 29 out of the 34 OECD countries, since Estonia, the Slovak Republic, Slovenia and Turkey did not participate in the 2000 wave and Japan participated but did not inform on the educational level of the students' parents.

3.1 Comparing the relative performance of OECD educational systems in 2009

The PISA report provides a classification of the students in six different levels of reading competence, from Level 1 (the lowest level) to Level 6 (the highest one). Those levels are defined in terms of the capacity of the students to master certain cognitive processes and operationalized in terms of ranges of the scores obtained by the students (see Figure 1.2.12 in OECD (2010a) for details). Table 1 summarizes the scoring intervals that parameterize those levels of competence.

We combine levels 5 and 6 of reading competence into a unique upper level since the share of students in the sixth level is quite low in most countries and it does not allow us to obtain accurate estimates of the propensity score for the students in that level.[9] The reduction in the number of levels of reading competence leaves almost unchanged the unadjusted eigenvectors but it substantially increases the stability of the covariate-adjusted eigenvectors as we increase the number of covariates used to estimate the propensity score. The distribution of students within the five levels of reading competence in 2009 is presented in Table 2.

In order to compare the performance of educational systems out of the distribution of outcomes in Table 2 we rely on the model of Section 2. Now *groups* are OECD countries, *members* are the 15-year old students within each country, and *categories* correspond to the five levels of reading competence. Table 3 below presents the eigenvector for the year 2009 and compares those values with the mean scores of the PISA tests.[10] We present both unadjusted and covariate-adjusted eigenvectors (indeed we present two different covariate adjusted vectors, to be explained below). The eigenvectors have been normalized so that the sum of its components equals the number of components. That is, values above one indicate that the country performs over the mean, while components below one are to be interpreted as performing below the mean.

[9] Less than 1% of the students have a reading competence of level 6 in 26 out of the 29 OECD countries.

[10] Test results are not presented in PISA as point estimates. Rather, PISA reports student performance through five plausible values that can be defined as random values from the posterior distribution of an student's performance (see OECD (2009) for details). As indicated in OECD (2009), we perform our statistical analysis independently on each of the five plausible values. The final eigenvector is the average of the eigenvectors obtained for each of the five plausible values. Anyway, the resulting eigenvectors are almost identical to those obtained using the average of the five plausible values as the summary measure of the students' performance.

Table 1. Levels of reading competence. PISA 2009.

Level of competence		Score range
Level 6		>698
Level 5		626-698
Level 4		553-626
Level 3		480-553
Level 2		407-480
Level 1	Level 1a	335-407
	Level 1b	262-335

Table 2. Share of students at each proficiency level on the reading scale. OECD countries. PISA 2009.

Country	Level 1 (or below)	Level 2	Level 3	Level 4	Level 5	Sample size
Australia	0.117	0.204	0.296	0.260	0.123	9,975
Austria	0.215	0.249	0.288	0.203	0.045	4,571
Belgium	0.098	0.183	0.280	0.316	0.123	5,731
Canada	0.107	0.216	0.305	0.264	0.107	16,782
Chile	0.248	0.352	0.284	0.105	0.011	3,803
Czech Republic	0.145	0.235	0.277	0.252	0.091	4,283
Denmark	0.155	0.272	0.325	0.213	0.034	3,991
Finland	0.073	0.165	0.302	0.329	0.131	4,372
France	0.134	0.207	0.302	0.255	0.102	2,787
Germany	0.118	0.210	0.305	0.280	0.088	2,853
Greece	0.165	0.266	0.316	0.205	0.048	3,676
Hungary	0.142	0.254	0.346	0.217	0.040	3,203
Iceland	0.143	0.215	0.332	0.231	0.079	2,581
Ireland	0.132	0.228	0.329	0.244	0.067	2,596
Israel	0.202	0.230	0.281	0.207	0.080	3,639
Italy	0.163	0.241	0.315	0.229	0.051	22,617
Korea	0.036	0.135	0.323	0.376	0.130	3,554
Luxembourg	0.198	0.239	0.293	0.210	0.059	3,053
Mexico	0.337	0.373	0.236	0.052	0.002	21,134
Netherlands	0.082	0.221	0.311	0.284	0.102	3,304
New Zealand	0.093	0.168	0.271	0.286	0.182	2,885
Norway	0.126	0.227	0.326	0.243	0.079	3,299
Poland	0.104	0.234	0.328	0.260	0.075	3,370
Portugal	0.163	0.257	0.332	0.210	0.037	4,748
Spain	0.159	0.250	0.352	0.206	0.033	18,947
Sweden	0.134	0.221	0.327	0.225	0.094	3,132
Switzerland	0.138	0.238	0.325	0.234	0.065	8,502
United Kingdom	0.137	0.241	0.310	0.229	0.084	8,288
United States	0.148	0.247	0.287	0.229	0.089	3,569
OECD total	0.161	0.247	0.306	0.219	0.066	185,245
OECD average	0.145	0.234	0.307	0.236	0.078	6,388

Equivalently, the mean scores of the PISA tests have been normalized by setting its average equal to one.

The unadjusted eigenvectors and the mean scores of the PISA tests yield similar results in terms of ranking but rather different pictures in terms of cross-country differences, with the distribution of mean score values being far more concentrated around the mean than that of eigenvector components. Indeed, while the Kendall's rank order correlation coefficient between the two evaluation measures is 0.92, the coefficient of variation of the eigenvector components is eight times that of the mean scores.

The normalized mean score underestimates the relative advantage of the countries that perform best at reading according to the eigenvector and overestimates the relative position of the countries that perform worst. For example, the normalized mean score of Korea, New Zeland, Finland and Belgium, the four countries at the top of the distribution according to the eigenvector, are 46%, 37%, 31% and 26% lower than the corresponding eigenvector components, respectively. Conversely, the mean scores of Greece, Austria, Chile and Mexico, the four countries at the bottom of the distribution according to the eigenvector, are 3, 2.2, 1.4 and 1.3 times larger than the corresponding eigenvector components, respectively. The difference between the two evaluation methods is of modest size for the countries that perform close to the mean.

3.2 Accounting for differences in characteristics

We now move to the covariate-adjusted eigenvectors in column (1). These were obtained by using the set of students' external factors that is common to both the 2000 and the 2009 reports in the estimation of the propensity score.[11] In this specification we control for children's sex and for the employment status and the educational level of their parents by means of a set of dummy indicator variables that inform on whether the father or the mother are employed and whether their highest educational level is the secondary or the tertiary level as defined in the International Standard Classification of Education (ISCED). We also control for whether at least one of the parents was born in a country different from their current country of residence and for whether there are at least one hundred books at home or not.

Regarding their schooling environment, we control for the share of the school total funding for a typical school year that comes from the government, whether the assessments of students are used to make judgments about teachers' effectiveness or not and whether the principal, the department head or the teachers have the main responsibility for hiring teachers or not.[12] All these familiar and

[11] The propensity scores are estimated using binary logit models. The reference subsample is the average of the OECD, and for the first proficiency level.

[12] The relevance of government funding is preferred to the indicator of whether the school is public or private because of the large number of missing values in the latter variable in the 2000 report.

Table 3. Eigenvectors and mean score. OECD, 2009. Reading competence.

Country	Unadjusted eigenvector	Mean score[a]	Covariate-adj. eigenvector (1)	Covariate-adj. eigenvector (2)
Australia	1.213***	1.030	0.986	1.014
Austria	0.676***	0.961	0.772**	0.768
Belgium	1.416***	1.042	1.121	1.137
Canada	1.187***	1.031	1.036	1.034
Chile	0.409***	0.910	0.632***	0.585
Czech Republic	0.976***	1.005	1.179***	1.146
Denmark	0.756***	0.979	0.753	0.763
Finland	1.541***	1.059	1.212**	1.235
France	1.082***	1.019	1.081	1.067
Germany	1.115***	1.024	0.932**	0.963
Greece	0.744***	0.968	0.764	0.740
Hungary	0.937***	0.995	0.994***	1.028
Iceland	0.982	0.998	0.939	0.935
Ireland	0.958***	1.003	0.953**	0.942
Israel	0.867***	0.979	0.786***	0.789
Italy	0.823***	0.983	1.013***	0.996
Korea	1.999***	1.082	1.420**	1.359
Luxembourg	0.779***	0.981	1.075***	0.999
Mexico	0.286	0.868	0.578	0.582
Netherlands	1.226***	1.043	1.114**	1.146
New Zealand	1.686***	1.063	1.493	1.506
Norway	0.989	1.009	0.963	0.970
Poland	1.023***	1.016	1.568***	1.626
Portugal	0.747***	0.975	0.991	0.953
Spain	0.763	0.973	0.859	0.853
Sweden	1.007***	1.008	0.922	0.955
Switzerland	0.936**	0.993	0.913**	0.920
United Kingdom	0.955***	1.002	0.961**	0.973
United States	0.921**	1.003	0.989	1.015
Coeff. of variation	0.343	0.042	0.223	0.228

Notes: [a] The mean is set equal to one. The symbols ** and *** indicate that the difference between the component and that for the same country and eigenvector (adjusted or unadjusted) in the year 2000 is significant at the 5% and 1% significance level, respectively. We use bootstrap hypothesis testing. The covariate-adjusted eigenvector in (1) controls for students' sex, their parents' employment status and educational level, whether at least one of the parents was born in a country different from their current country of residence, whether there are at least one hundred books at home, the percentage of the school total funding for a typical school year that comes from the government, whether the assessments of students are used to make judgements about teachers' effectiveness and whether the principal, the department head or the teachers have the main responsibility for hiring teachers or not. The eigenvector in (2) further controls for whether the school is located in a city with over one million people, the share of teachers fully certified by the appropriate authority in the shool, whether there is at least one other school in the same area, whether students are grouped by ability, whether achievement data are used to evaluate the principal's or the teachers' performance, whether the school monitors the practice of teachers and whether external examination boards exert a direct influence on decision making or not. The additional controls in (2) are present only in the 2009 report.

schooling factors are relevant determinants of international differences in students' educational achievement according to Hanushek and Woessmann (2011). Table 4 summarizes the distribution of these variables for the 29 OECD countries.

Our estimates attest that accounting for heterogeneity makes a relevant difference. On the one hand, the coefficient of variation of the eigenvector lowers by 35%, once we control for differences in characteristics. That is, the explanatory variables in (1) account for more than a third of the variation in schoolchildren's scholastic relative performance in reading ability across OECD countries. On the other hand, the Kendall's correlation coefficient between the unadjusted and the covariate-adjusted eigenvectors is of approximately 0.56. This indicates that differences in characteristics across OECD countries account for approximately 45% of the sorting of OECD countries that results from the unadjusted eigenvector.

The relative advantage of children coming from the countries that do best at reading according to the unadjusted eigenvector lowers once we control for heterogeneity. That is the case for Australia, Belgium, Finland and Korea, whose relative advantage lowers by at least 20%. Conversely, the relative position of the countries that do worst according to the unadjusted eigenvector improves once we control for X. Chile, Italy, Luxembourg, Mexico, Poland and Portugal improve their relative performance by approximately 55%, 23%, 38%, 102%, 53% and 33%, respectively. Australia, Germany and Sweden move from performing over the mean to performing slightly below the mean once we control for students' external circumstances. The opposite holds for Czech Republic, Italy and Luxembourg.[13]

One of our main findings is that at least half of the inequality in students' reading ability across OECD countries is explained by country differences in students' family and schooling characteristics. How much of the remaining differences can be accounted for by introducing additional explanatory variables? To investigate this issue we have included in Table 3 another covariate-adjusted eigenvector in column (2), estimated by expanding the set of controls in specification (1) with some determinants of students' achievements.[14] This allows us to analyze how the coefficient of variation of the eigenvector components changes as we expand the set of covariates. In particular, in (2) we additionally control for whether the school is located in a large city, i.e. a city with over one million people, or not, for the percentage of teachers fully certified by the appropriate authority in the school, whether there is at least one other school in the same area or not, whether students are grouped by ability or not, whether achievement data are used to evaluate the principal's or the teachers' performance or not, whether the school monitors the practice of teachers or not

[13] The relative position of Spain also improves once we control for X. This result is in line with that in Ciccone and Garcia-Fontes (2009). They found that there is a sizeable increase in Spain's PISA scores relative to the rest of Europe when parental schooling is accounted for.

[14] Those are present only in the 2009 report and cannot be considered when comparing the results in 2009 and in 2000.

Table 4. Students' personal and familiar characteristics. Descriptive statistics. PISA 2009.

Country	Women	Mother works	Father works	Foreign parent[a]	Mother's educ. Secondary	Mother's educ. Tertiary	Fathers' educ. Secondary	Fathers' educ. Tertiary	100 books[b]
Australia	51.4	76.8	90.5	39.1	56.2	42.9	59.1	39.7	52.3
Austria	51.7	76.2	93.2	20.5	73.8	24.6	55.2	43.8	42.9
Belgium	49.7	77.1	90.9	25.4	41.7	54.4	46.4	49.1	42.6
Canada	51.3	81.4	90.9	24.8	38.6	60.3	46.8	51.1	48.6
Chile	50.8	52.0	89.5	2.0	63.7	26.1	61.8	29.1	20.4
Czech Republic	48.0	84.0	94.3	8.3	73.9	25.9	72.5	27.2	49.3
Denmark	51.6	80.1	87.1	29.2	43.1	52.1	56.0	40.7	37.7
Finland	51.7	87.4	89.0	7.1	24.1	72.6	28.3	66.6	52.3
France	53.0	77.3	90.7	23.6	55.4	42.1	58.8	38.9	41.3
Germany	50.4	75.9	91.1	19.7	69.7	27.4	55.9	42.0	50.0
Greece	51.8	63.1	89.4	15.6	54.4	38.7	51.1	40.3	39.1
Hungary	50.0	75.6	86.0	5.1	69.2	30.3	74.1	25.4	52.0
Iceland	52.3	85.5	92.4	9.1	55.1	44.0	57.0	42.0	58.5
Ireland	49.8	67.8	86.6	23.7	57.1	39.5	56.8	36.2	43.1
Israel	54.9	69.0	86.0	31.6	46.4	48.4	49.4	46.7	41.6
Italy	50.1	65.3	93.1	12.0	72.4	24.4	71.9	23.6	41.2
Korea	45.9	52.3	88.1	0.2	63.2	34.3	48.6	48.6	57.2
Luxembourg	52.8	67.2	92.1	55.5	47.8	38.0	44.9	42.2	55.7
Mexico	52.8	39.5	86.7	3.0	40.6	28.3	38.5	33.0	11.2
Netherlands	51.4	78.3	91.8	18.9	53.5	40.7	47.3	46.1	38.6
New Zealand	51.4	79.0	91.2	40.8	56.4	41.1	60.6	36.7	50.6
Norway	49.1	87.6	92.8	14.4	38.9	60.3	44.7	54.1	54.4
Poland	50.3	69.3	85.7	0.6	79.2	20.6	84.9	14.8	37.9
Portugal	52.2	76.3	90.1	19.4	45.5	21.9	43.9	18.9	31.2
Spain	49.8	68.4	90.1	14.5	50.5	37.0	46.3	38.7	49.5
Sweden	50.4	87.4	92.0	21.3	35.4	62.9	46.3	50.6	54.7
Switzerland	49.4	72.4	93.9	40.4	59.5	36.3	48.1	48.6	40.8
United Kingdom	51.1	77.3	88.9	13.1	51.3	47.6	55.4	43.1	42.3
United States	49.4	73.7	85.1	26.4	48.0	47.8	54.0	41.1	35.0

Notes: [a] Informs on the percentage of the school total funding for a typical school year that comes from the government. [b] Indicates whether there are at least one hundred books at home or not.

and whether external examination boards exert a direct influence on decision making or not. Table 5 summarizes the distribution of these variables across the 29 OECD countries. As before, these variables are relevant determinants of students' achievement tests according to Hanushek and Woessmann (2011).

The difference between the two covariate-adjusted eigenvectors is of modest size, below 4% for all but three out of the 29 components. Moreover, the coefficient of variation remains largely unchanged when we expand the set of controls. This finding suggests that there is a part of the difference in the evaluation of students' reading ability across OECD countries that cannot be explained by the differences in the characteristics that the theory suggests. That is to say, there is a set of unobservable factors that induce the existence of international differences, beyond the action of the standard determinants. Those unobservable variables may refer to cultural and organizational factors, such as the design of the educational system, the commitment of families and society with education, the implementation of educational policies, the teachers' attitudes and involvement, and, of course, the average effort of the students. The low sensitivity of the coefficient of variation of the covariate-adjusted eigenvector, with respect to the inclusion of seven relevant additional covariates, points out the relevance of those factors. This is supported by the fact that the same type of results appears between the regions of some countries.[15]

3.3 The evolution of performance: changes between 2000 and 2009

Let us now comment on how those values have evolved between 2000 and 2009, so that we can have an idea of the dynamics of the educational systems.

The first point to be noted is that the unadjusted eigenvector has substantially changed during the first decade of the 21st Century. Indeed, the difference between the eigenvector components for the years 2000 and 2009 is significantly different from zero at the 1% significance level for 23 out of the 29 OECD countries and at the 5% significance level for other two developed countries.[16] Most

[15] This is the case of Italy and Spain, countries that show an internal diversity of educational outcomes (differences between the regions within each country), similar to that of the OECD. This is so in spite of formally having a common educational system.

[16] We use bootstrap to test whether the difference between a country's eigenvector component in 2000 and in 2009 is statistically significant or not. We implement bootstrap hypothesis testing as follows. For each country, let n_t^c indicate the number of students with level of reading competence c in year t, for $t = \{2000, 2009\}$. For each country and for each c, we merge the samples of the two years into one sample of $\left(n_c^{2000} + n_c^{2009}\right)$ observations. We draw a bootstrap sample of $\left(n_c^{2000} + n_c^{2009}\right)$ observations with replacement from the merged sample and we assign the first n_c^{2000} observations to the first year. We then calculate the eigenvectors of the two years and we compute the difference between them. We repeat these steps 1000 times. The p-value is then estimated as the number of times the difference between the eigenvectors coming from bootstrap samples exceeds that observed in the original sample, divided over the number of repetitions.

Table 5. Schooling characteristics. Descriptive statistics. PISA 2009.

Country	Large city[c]	Fund govern.[d]	Evaluate teachers[e]	Control[f]	Certified teachers[g]	Track[h]	Compete[i]	Perform[j]	Monitor[k]	External exams[l]
Australia	34.1	57.7	38.6	77.8	63.9	69.0	75.7	40.9	86.6	54.1
Austria	33.3	60.6	40.5	81.8	59.9	64.9	71.8	40.2	86.1	51.6
Belgium	34.8	60.5	43.8	78.3	61.1	64.6	72.9	41.3	85.8	53.8
Canada	33.4	45.1	38.6	62.4	74.7	68.2	75.8	40.5	84.1	43.4
Chile	35.3	61.0	50.3	86.4	59.1	62.0	79.2	48.5	87.6	53.1
Czech Republic	32.9	60.6	40.7	83.9	61.6	63.5	76.5	43.9	88.4	50.4
Denmark	35.2	61.8	39.9	82.2	58.7	66.4	75.7	44.1	86.5	51.4
Finland	33.9	62.3	45.9	87.3	61.9	64.0	76.1	48.2	86.6	50.9
France	36.7	56.0	53.5	87.4	65.5	65.0	78.2	49.5	90.7	52.0
Germany	34.0	60.4	42.7	83.6	62.2	63.2	75.9	42.7	85.4	49.8
Greece	34.6	59.0	46.6	85.5	58.1	59.7	77.6	47.6	85.6	50.3
Hungary	38.0	59.1	47.6	87.1	65.9	61.6	77.9	44.6	88.9	54.3
Iceland	44.9	50.6	43.2	85.4	60.8	60.8	81.2	44.0	89.0	49.1
Ireland	37.7	58.9	51.7	86.4	65.7	64.0	80.2	50.5	88.3	51.7
Israel	35.3	59.4	51.5	87.7	63.0	64.9	77.8	48.4	87.7	55.9
Italy	35.3	45.7	40.5	59.0	78.7	66.1	78.6	43.9	84.6	44.7
Korea	37.8	41.9	51.0	87.2	61.8	64.1	79.6	50.5	88.7	51.7
Luxembourg	45.3	43.1	50.2	80.7	73.6	77.7	92.4	42.4	95.1	58.4
Mexico	37.1	43.1	45.3	60.5	80.1	65.6	80.3	48.2	86.4	46.7
Netherlands	34.3	42.2	48.6	86.4	61.5	63.6	76.8	47.1	87.2	52.2
New Zealand	36.1	42.6	51.8	86.6	61.0	59.3	77.6	48.1	89.0	52.0
Norway	35.1	42.1	46.8	85.9	61.0	64.7	77.2	46.2	86.4	50.8
Poland	36.0	42.2	49.1	86.7	60.8	65.6	78.1	46.2	87.5	54.6
Portugal	33.2	42.1	42.2	85.6	60.9	62.1	75.3	44.1	85.1	51.6
Spain	34.3	43.5	41.1	61.6	77.2	67.0	78.4	42.1	85.0	45.7
Sweden	34.0	41.8	49.1	87.7	58.2	62.7	75.2	45.9	87.3	54.5
Switzerland	30.6	43.5	41.2	73.3	67.9	67.6	75.4	43.7	86.2	50.3
United Kingdom	32.6	43.8	40.3	72.7	67.8	66.7	77.1	41.6	87.3	48.9
United States	37.4	42.6	52.3	88.3	65.9	64.9	79.1	47.6	88.3	54.2

Notes: [c] Indicates whether the school is located in a city with over one million people or not. [d] Informs on the percentage of the school total funding for a typical school year that comes from the government. [e] Indicates whether the assessments of students are used to make judgements about teachers' effectiveness or not. [f] Indicates whether the principal, the department head or the teachers have the main responsibility for hiring teachers or not. [g] Informs on the share of teachers fully certified by the appropriate authority in the school. [h] Indicates whether students are grouped by ability or not. [i] Indicates whether there is at least one other school in the same area or not. [j] Indicates whether achievement data are used to evaluate the principal's or the teachers' performance. [k] Indicates whether the school monitors the practice of teachers. [l] Indicates whether external examination boards exert a direct influence on decision making or not.

of the countries that performed over the mean in the year 2000 lowered their relative advantage in students' reading ability during the period of analysis, with the reduction being largest for the Netherlands (36%), Ireland (34%), the United Kingdom (31%) and Finland (22%). Belgium, Korea and New Zealand are the exception to the latter rule. They all performed over the mean in 2000 but they improved their relative advantage from 2000 to 2009. The improvement is particularly relevant for Korea (59%), that moves from the 9th position in the 2000 ranking to the first position in 2009. We also find that 13 out of the 16 countries that performed below the mean in 2000 lowered their relative disadvantage during the first decade of the 21th Century. The improvement is particularly relevant for Luxembourg (83%), Chile (77%) and Poland (72%). Indeed, the latter country moves to performing over the mean in the year 2009.

As a result of these changes, the coefficient of variation of the unadjusted eigenvector lowered by 17.7% from 2000 to 2009, attesting that inequality in students' reading ability across OECD countries lowered during the first decade of the 21th Century. Additionally, the Kendall's correlation between the unadjusted eigenvectors in 2000 and in 2009 is 0.58, indicating that the ordering of countries in terms of relative performance in reading competence has substantially changed from 2000 to 2009. The relative position of 15 out of the 29 OECD countries analyzed is no more than three positions further in 2009 than it was in 2000. Conversely, that differential amounts to at least eight positions for seven countries. While Denmark, the United Kingdom, Austria and Ireland lose 8, 10, 10 and 11 positions, respectively, Germany, Korea and Poland improve their relative position in 7, 8 and 16 positions, respectively.

The relationship between the covariate-adjusted eigenvector and the unadjusted one in 2000 is pretty much the same as in 2009, so that we shall not repeat the analysis here. More interesting is to analyze the difference between the covariate-adjusted eigenvector components for the years 2000 and 2009. We find that it is significantly different from zero at conventional significance levels for 15 out of the 29 OECD countries. In particular, Poland, Luxembourg and Chile stand out among the countries that significantly improve their relative position since they increase their component by 117%, 72% and 61%, respectively. The former two countries move from performing below the mean in the base year to a component larger than one in 2009. Conversely, Finland, one of the best-performing countries both in the unadjusted and in the covariate-adjusted measures of inequality in both years, lowers its relative advantage once accounting for differences in characteristics by almost 35% from 2000 to 2009. As a result, the Kendall's correlation between the 2000 and the 2009 covariate-adjusted eigenvectors is of only 0.24. Additionally, the coefficient of variation of the covariate-adjusted eigenvector (1) lowers by approximately 35% between the years 2000 and 2009.

Our results remain almost unchanged when we restrict the analysis to the 21 European OECD countries in Table 2. In particular, we find that students' external factors account for slightly more than one half of the differences in relative performance in 2009 and that differences across European OECD countries in students' reading ability, both adjusted and unadjusted, have decreased from

2000 to 2009 more than they have done in the OECD.[17]

4 Final Remarks

We have presented here a model that combines the one developed in Herrero and Villar (2012), that permits evaluating group performance with categorical data, with inverse probability weighting (IPW) techniques that control for differences in the distribution of the determinants of the outcome variable. We obtain in this way a covariate-adjusted eigenvector that, when compared with the unadjusted one, allows us to estimate the impact of the difference in characteristics over the relative performance.

We have applied this methodology to the evaluation of compulsory education in the OECD through PISA 2000 and 2009. We find that differences in reading ability across OECD countries would lower by more than one third if their endowment of schooling and family characteristics would be that for the OECD average. We have also found that the differences in students' reading ability across OECD countries substantially lowered during the first decade of the 21th Century.

There are two related questions that come to mind when considering this particular application. First, why making an evaluation out of categorical data (the distribution of students across the different levels of competence) when we have all cardinal information that might be required? The reason is that rather than relying on summary statistics (e.g. means and inequality measures) as it is the case in most of the cardinal approaches, we are able to deal with discretized versions of the whole distributions in a relatively simple way and so to extract more information. Second, why to use just five levels of competence rather than richer distribution profiles (e.g. using percentiles)? The answer here is twofold. On the one hand, a small number of levels permits a richer set of covariates. On the other hand, in this particular case, those levels are given externally so that there is less arbitrariness in deciding the clusters by the analysts.

[17] These estimates are available upon request to the authors.

5 References

Arneson, R.J., 1989. Equality and equality of opportunity for welfare. Philosophical Studies 56, 77-93.

Busso, M., DiNardo, J., McCrary, J., 2009. New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators. IZA Discussion Paper 3998.

Calo-Blanco, A., Villar, A., 2010. Quality of Education and Equality of Opportunity in Spain. Lessons from PISA, Fundación BBVA, working paper 6-2010.

Calo-Blanco, A. García-Pérez, J.I., 2013. On the Welfare Loss Caused by Inequality of Opportunity. Journal of Economic Inequality. DOI: 10.1007/s10888-013-9247-6.

Checchi, D., Peragine, V., 2010. Inequality of opportunity in Italy. Journal of Economic Inequality 8, 429-450.

Ciccone, A., Garcia-Fontes, W., 2009. The quality of the Catalan and Spanish education systems: A perspective from PISA. IESE Research Papers D/810, IESE Business School.

Cohen, G.A., 1989. On the currency of egalitarian justice. Ethics, 99(4), 906-944.

Crump, R.K., Hotz, V.J., Imbens, G.W., Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. Biometrika 96(1), 187-199.

Echenique, F., Fryer, R.G., 2005. On the measurement of segregation. Working Paper 11258, NBER.

Fleurbaey, M., 2008. Fairness, responsibility and welfare. Oxford University Press, Oxford.

Hanushek, E.A., Woessmann, L., 2011. The Economics of International Differences in Educational Achievement. In E.A. Hanushek, S. Machin, L. Woessmann (eds.), Handbook of the Economics of Education, Vol. 3, pp. 89-200, Amsterdam: North Holland.

Herrero, C., Villar, A., 2012. On the comparison of group performance with categorical data. Working Paper AD 2012-17, Ivie.

Lefranc, A., Pistolesi, N., Trannoy, A., 2008. Inequality of opportunities vs. inequality of outcomes: Are western societies all alike?. Review of Income and Wealth 54, 513-546.

Lefranc, A., Pistolesi, N., Trannoy, A., 2009. Equality of opportunity and luck: definitions and testable conditions, with an application to income in France. Journal of Public Economics 93, 1189-1207.

Lieberson, S. 1976. Rank-Sum comparisons between Groups. Sociological Methodology 7, 276-291.

OECD, 2009. PISA Data Analysis Manual: SPSS. Second Edition. OECD, Paris.

OECD, 2010a. PISA 2009 Results: What Students Know and Can Do. Student Performance in Reading, Mathematics and Science (Volume I). OECD, Paris.

OECD, 2010b. PISA 2009 Results: Overcoming Social Background. Equity in learning opportunities and outcomes (Volume II). OECD, Paris.

Peragine, V., Serlenga, L., 2008. Higher education and equality of opportunity in Italy. Research on Economic Inequality 16, 67-97.

Roemer, J.E., 1993. A pragmatic theory of responsibility for the egalitarian planner. Philosophy and Public Affairs 22(2), 146-166.

Roemer, J.E., 1998. Equality of opportunity. Harvard Univeristy Press, New York.

Rosenbaum, P., Rubin, D.B.,1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70, 41-55.

Villar, A., 2012. The Educational Development Index. Working Paper Econ n^o 1206, Pablo de Olavide University.

serie ad

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Website: www.ivie.es
E-mail: publicaciones@ivie.es