On the Comparison of Group Performance with Categorical Data

Carmen Herrero
University of Alicante & Ivie

Antonio Villar
University Pablo de Olavide & Ivie

March 29, 2013

Abstract

This paper provides a criterion to evaluate the overall performance of a given number of groups whose members belong to different categories, out of the analysis of their relative frequencies. Assuming that categories can be ordered from best to worse, the starting point is that of dominance relations in pair-wise comparisons. We say that group i dominates group j when the expected category of a member of i is higher than the expected category of a member of j. We extend this principle to multi-group comparisons in an endogeneous way. The evaluation function associates, to each evaluation problem, the (unique) dominant eigenvector of a matrix whose entries describe the dominance relations between groups in pairwise comparisons. An application to the structure of human capital in Europe illustrates the working of this model.

1 Introduction

There are many situations in which one is interested in making group comparisons in terms of some ordered characteristic. Think for instance of the evaluation of health in different countries, out of the data of Self Assessed Health Status Surveys; or the comparison of human capital in terms of the educational levels achieved by the working age population. In the first case people usually declare that their health situation is one among four or five categories, ranging from “excellent” to “very bad”. In the second case we find a distribution of the population into different categories, from no education to university studies. Comparing societies out of those data requires either finding a way of scaling the different categories (health statuses or education levels in our examples)

*Acknowledgements: Thanks are due to M. Ballester, G. Yanoletzky,
and summarizing those values by some aggregation procedure, or else devising a criterion that is capable of dealing with qualitative data.

We shall follow here the latter approach. More precisely, this paper presents a procedure to make group comparisons when the traits or achievements of their members are described by ordered categorical data such as age intervals, income brackets, health statuses, education levels, prestige positions, etc. The setting consists, therefore, of a finite set of groups whose members are classified into a given number of categories, which describe their characteristics or their realizations. Categories are assumed to be ordered, so that one can unambiguously say that a category “precedes”, is “higher than”, or is “better than” another. Our goal is finding a suitable summary measure that synthesizes the key features of the different groups taking into account the distribution of its members along those categories. The evaluation will focus on the relative frequencies of the agents in the different cells that arise from the double partition into groups and categories. Comparing groups will amount, therefore, to comparing the distribution patterns of their categories.

These ideas are related to the statistical analysis dealing with the similarity between rank distributions and the sociological and economic literature about segregation (see for instance Reardon & Firebaugh (2002), Grannis (2002), Reardon & O’Sullivan (2004), del Río & Alonso (2010), or Yalonztky (2012)). An early contribution worth stressing is that of Lieberson (1976). He introduces the notion of Net Difference for the evaluation of pairs of groups, as the absolute difference between the probability that an agent from group $i$ be in a better position than an agent from group $j$ and the probability that an agent from group $j$ be in a better position than an agent from group $i$.

Here we extend this idea to a more general setting in order to compare any finite number of groups. Such an extension is not trivial because it implies taking into account the direct and indirect relationships between all the groups. The evaluation function is presented in Section 2 in a constructive way. We start by considering uniform problems, i.e., problems in which all groups have identical population size. Here, we give two interpretations of our evaluation function: on the one hand, as a way of balancing the aggregate value of gains and losses for each group, then, as the average time each group is selected in a "competition" procedure that provides naturally with a Markov process. Then, we extend these ideas to the general case, under an implicit assumption: group replication invariance, that amounts to consider immaterial the sizes of the groups. In either case, we select an invariant weighting system that properly evaluates the relative performance of the different groups. We call the worth of a group the value so obtained.

The resulting evaluation function corresponds to the eigenvector of a suitably constructed matrix that incorporates the information about the distribution of the population in the groups across the different categories. The solution is thus a fixed point of a mapping that can be interpreted as the limit process of an evaluation procedure in which the weights with which we calculate the overall advantage of a group are progressively adjusted according to the outcome of the evaluation process. This type of evaluation has a similar flavour to some
of the ways of evaluating the impact of scientific journals (see Pinski & Narin (1976), Laband & Piette (1994), Palacios Huerta & Volij (2004), Serrano (2004), Waltman & Jan van Eck (2010) or the construction of the Eigenfactor), as well as with the assignment of scores in tournaments (see Moon (1968), Laslier (1997), Slutzki & Volij (2006), Gonzalez-Diaz, Hendrichx & Lohmann (2013)]. Indeed, our evaluation function is similar to the fair-bets scoring in tournaments (see Moon & Pullman (1970), Slutzky & Volij (2006), Chevotarev & Shamis (1998)].

Section 3 considers the evaluation problem from a different perspective, by looking at the properties that a generic evaluation function may satisfy and providing a characterization of the one presented in Section 2. There are three axioms that yield the characterization result. The first one, reciprocity, establishes that for those problems involving only two groups the relative value of the groups should correspond to their relative domination probabilities. The second axiom, agreement, refers to the evaluation of a group in a given problem and its evaluation in the two-group problem that results from merging all other groups into a single one. This axiom says that the ratio between the evaluation in both cases should coincide with the corresponding ratio of the evaluation of their beats, measured as a weighted average of its domination probabilities. The third axiom, group replication invariance, says that the size of the groups does not affect their evaluation, which only depends on the relative frequencies of their types within each group. We also show that our evaluation function satisfies some other properties that are common in this context (anonymity, symmetry, monotonicity, and stochastic dominance).

We present in Section 4 an application of this evaluation procedure to the comparative analysis of human capital in Europe. Here the society is the set of European citizens aged between 25 and 64, groups are the European countries, and categories are the levels of educational achievements (grouped into three categories out of the ISCED classification). The data correspond to the year 2010 and are obtained from Eurostat.

A few final comments in Section 5 close the work.

2 The Model

Consider a society \( N = \{1, 2, ..., n\} \) made of \( g \) nonempty groups, \( G = \{1, 2, ..., g\} \), with \( n > g \geq 2 \), and let \( n_i \) denote the number of members in group \( i \in G \). We assume that the individual characteristics of the groups’ members induce a partition in terms of \( s \) categorical positions, \( C = \{c_1, c_2, ..., c_s\} \), ordered from best to worst, \( c_1 \succ c_2 \succ ... \succ c_g \). We denote by \( n_{ir} \) the number of members of group \( i \) in category \( r \). By definition, \( n_i = \sum_{r=1}^{s} n_{ir}, \ n = \sum_{i=1}^{g} n_i \).

An evaluation problem is a tuple \((N, G, C)\). Our goal is to compare the relative performance of the different groups by confronting agents from them. Let \( B_{ij} \) stand for the number of times that an agent in group \( i \) is at a higher position than an agent in group \( j \) (how many times \( i \) beats \( j \) in pair-wise comparisons,
so to speak). As categories are ordered, this number is given by:

\[ B_{ij} = n_{i1}(n_{j2} + \cdots + n_{js}) + n_{i2}(n_{j3} + \cdots + n_{js}) + \cdots + n_{is}(s-1)n_{js} \]  

[1]

If we normalize this number by taking into account all possible comparisons, we simply get the probability that a member of group \( i \) chosen at random is in a better position than a member of group \( j \). That is:

\[ p_{ij} = \frac{B_{ij}}{n_i n_j}. \]

Similarly, \( p_{ji} \) denotes the probability that a representative member of group \( j \) be at a higher position than a representative member of group \( i \). And, consequently, \( e_{ij} = 1 - p_{ij} - p_{ji} \) stands for the probability that a member from \( i \), picked at random, be at the same position than a member from \( j \).

Notice that \( e_{ij} = \frac{E_{ij}}{n_i n_j} \), where \( E_{ij} = n_{i1}n_{j1} + n_{i2}n_{j2} + \cdots + n_{is}n_{js} \) is the number of times that an agent in group \( i \) is at the same position than an agent in \( j \).

**Remark 1** We shall assume, for the sake of simplicity in exposition, that all \( B_{ij} \) are strictly positive. This point will be discussed later on.

Consider now the following:

**Definition 2** We say that group \( i \) **dominates group \( j \)** in a pair-wise comparison whenever the number of times an individual in \( i \) is in a better position than an individual in \( j \) is larger than the other way around. That is, 

\[ i \succ j \iff B_{ij} > B_{ji}. \]

Note that previous definition is equivalent to say that the probability of an individual chosen at random in \( i \), to be at a higher position than an individual randomly chosen in \( j \), than the other way around, i.e., \( i \succ j \iff B_{ij} > B_{ji} \iff p_{ij} > p_{ji} \). When there are only two groups involved this is a sound criterion that permits one to evaluate their relative performance in an unambiguous way. This type of pair-wise comparison is reminiscent of Lieberson’s (1976) Index of Net Difference.\(^1\) Extending this principle to a more general setting, involving any finite number of groups, is not trivial and requires some additional elaboration. We have to devise a way of comparing the relative position of members in each group with respect to all other groups, taking into account both direct and indirect relationships.

\(^1\)The Index of Net Difference, \( ND(i, j) \), is defined as the absolute value of the difference \( p_{ij} - p_{ji} \). This index provides an estimate of the relative advantage of any two groups. When \( ND(i, j) = 0 \) a member chosen at random from group \( i \) has the same probability of being at a better position than a member picked at random from group \( j \) than the other way around. On the opposite extreme, we find the case \( ND(i, j) = 1 \), which happens whenever all members of one group occupy better positions than those in the other. Intermediate cases generate values in the interior of the interval \([0, 1]\).
2.1 Uniform problems

Suppose that all groups have the same population, \( n \). We want to evaluate the importance of the different beatings in an endogeneous way. Take group \( i \). The beatings of group \( i \) over any other group, \( j \neq i \) are given by the numbers \( B_{ij} \), \( j \neq i \). But the importance of those beats change with the group \( j \) considered. Assume that we attach values \((v_1, \ldots, v_g) \gg 0\) to the different beats. Namely, each beat on group 1 is evaluated by \( v_1 \), and so for. Then, the aggregate value of the dominations of group \( i \) over the different groups is given by \( \sum_{j \neq i} B_{ij} v_j \).

On the other hand, the aggregate value of the dominations of the different groups over group \( i \) is given by \( \sum_{j \neq i} B_{ji} v_i \). We ask whether there is a vector of sensible valuations, in the sense that the value of the dominations of any group equilibrates the value of the dominations of the rest of the groups over this particular one, namely whether there is a system of values \((v_1, \ldots, v_g) \gg 0\) so that for all \( i = 1, \ldots, g \)

\[
\sum_{j \neq i} B_{ji} v_i = \sum_{j \neq i} B_{ij} v_j \quad [1]
\]

Notice that if such a system of values exist, the information they provide is relative, namely, there is a degree of freedom in choosing the scale (normalization).

Equation (1) can be rewritten as

\[
v_i = \frac{\sum_{j \neq i} B_{ij} v_j}{\sum_{j \neq i} B_{ji}} \quad \text{for all } i, j = 1 \ldots g \quad [2]
\]

By using the right hand side of Equation (2), we can define a transformation of "tentative valuations", \((\lambda_1, \ldots, \lambda_g)\), into new ones, so that \( \Phi_i(\lambda_1, \ldots, \lambda_g) = \frac{\sum_{j \neq i} B_{ij} \lambda_j}{\sum_{j \neq i} B_{ji}} \). Then, the existence of the values \((v_1, \ldots, v_g) \gg 0\) fulfilling Equation (1) is equivalent to the existence of a fixed point of function \( \Phi \).

Let us now formalize previous ideas. A uniform evaluation problem is an evaluation problem \((N, G, C)\), such that all groups have the same size (i.e., \( N = gn \), where \( n \) stands for the common groups population size). For uniform evaluation problems, the value \( v_i \) in Equation (1) is called the worth of group \( i \) and define the following:

Definition 3 A consistent evaluation function is a mapping \( F \) such that, for each uniform evaluation problem \( P = (N, G, C) \), it associates to each group its worth. That is, for each problem \( P \) we have: \( F(P) = v \), with:

\[
v_i = \sum_{j \neq i} B_{ij}(P) v_j \quad \text{for all } i, j = 1 \ldots g \quad [3]
\]

Notice that the worth of a group is higher, other things equal, the higher the value of the groups it dominates.

In the case in which there are only two groups involved the ratio between their valuations coincides with the ratio of their relative beatings, and thus with
the ratio of the probability of one dominating the other. That is,

\[
\frac{p_{ij}}{p_{ji}} = \frac{v_i}{v_j} = \frac{B_{ij}(P)}{B_{ji}(P)}
\]

As pointed out before, finding a consistent evaluation vector amounts to finding a fixed point of function \(F\), for each possible problem. We show next that such a fixed point always exists and it is unique, once the scale has been chosen and bearing in mind Remark 1.

**Theorem 4** Let \(P\) be a uniform evaluation problem regarding \(g \geq 2\) groups whose members are classified into \(s\) ordered categories. There exists a unique consistent evaluation function \(F\), with \(F(P) = v > 0, v_i = \sum_{j \neq i} B_{ij}(P)v_j,\) and \(\sum_{i=1}^{g} v_i = 1.\)

**Proof.** Let us arrange the domination information in problem \(P\) in a matrix so that the off-diagonal elements are \(B_{ij}\), and the elements in the diagonal (ii) are the total comparisons of group \(i\) with any other group, in which the elements in \(i\) are not beaten, i.e.,

\[
M(P) = \begin{pmatrix}
\sum_{j \neq 1}(B_{1j} + E_{1j}) & B_{12} & \ldots & B_{1g} \\
B_{21} & \sum_{j \neq 2}(B_{2j} + E_{2j}) & \ldots & B_{2g} \\
\vdots & \vdots & \ddots & \vdots \\
B_{g1} & B_{g2} & \ldots & \sum_{j \neq g}(B_{gj} + E_{gj})
\end{pmatrix}
\]

Notice that \(M(P) > 0\), and the sum of its columns is constant: for all \(k = 1, \ldots, g,\)

\[
\sum_{j \neq k}(B_{kj} + E_{kj}) + \sum_{j \neq k} B_{jk} = \sum_{j \neq k}(B_{kj} + B_{kj} + E_{kj}) = \sum_{j \neq k} n^2 = (g-1)n^2
\]

Consequently, and due to the Perron-Frobenius Theorem, \((g-1)n^2\) is the dominant eigenvalue of matrix \(M(P)\), and such an eigenvalue has associated an eigenvector \(v = (v_1, \ldots, v_g) > 0,\) such that \(M(P)v = (g-1)n^2v.\)

This eigenvector, \(v = (v_1, v_2, \ldots, v_g)\), is unique up to a scalar multiplication, so that we can assume, without loss of generality, that \(v_1 + v_2 + \cdots + v_g = 1.\) Observe that the ith entry of that eigenvector can be written as:

\[
v_i = \frac{\sum_{j \neq 1} B_{ij}v_j}{\sum_{j \neq i} B_{ji}} \quad [4]
\]

and thus we obtained the desired result. \(\square\)
2.1.1 A probabilistic interpretation

Let us consider now the following protocol. Again we are in a uniform problem, i.e., all groups have the same population, \( n \). Imagine that a certain group \((i)\) is initially selected, and another group \((j)\) is chosen at random, and two individuals (one in group \(i\), and another one in group \(j\)) are randomly selected to be compared. If the individual in \(j\) beats the individual selected in \(i\), then \(j\) is now selected and another group comes randomly to be compared with \(j\); otherwise, \(i\) keeps being the selected group, and another group comes randomly to be compared with it. The process is repeated infinitely many times. Then, the probability that group \(j\) becomes the new group selected is given by

\[
\frac{1}{g} B_{ji} = \frac{1}{g} \sum_{j \neq i} (p_{ij} + e_{ij}),
\]

while the probability that \(i\) keeps being selected is given by

\[
\frac{1}{g-1} (B_{ij} + E_{ij}) = \frac{1}{g-1} \sum_{j \neq i} (p_{ij} + e_{ij}).
\]

Previous protocol define a Markov process whose stochastic matrix is the following

\[
S(\mathbf{P}) = \frac{1}{g-1} \begin{pmatrix}
\sum_{j \neq i} (p_{1j} + e_{1j}) & p_{12} & \ldots & p_{1g} \\
p_{21} & \sum_{j \neq 2} (p_{2j} + e_{2j}) & \ldots & p_{2g} \\
\ldots & \ldots & \ldots & \ldots \\
p_{g1} & p_{g2} & \ldots & \sum_{j \neq g} (p_{gj} + e_{gj})
\end{pmatrix}
\]

The stationary distribution of this protocol provides a vector that indicates the proportion of time (in the long run) in which each group is selected.

Now, notice that \( S(\mathbf{P}) = \frac{1}{(g-1)n^2} M(\mathbf{P}) \), and thus, the ergodic distribution of \( S(\mathbf{P}) \) is exactly the worth vector \( \mathbf{v} = (v_1, ..., v_2) \).

2.2 The general case

The results for the case of uniform problems can be straightforward extended whenever we assume that the size of the groups is immaterial for the comparison exercise. This amounts to assume a property of group replication invariance that reduces any problem to a uniform problem, by simply replicating each group a certain number of times, so that we end up with an associated problem in which all groups have the same population.

This procedure is implicit in the following formulation of the problem. Let \((N, G, C)\) be a general problem. Our goal is to compare the relative performance of the different groups by confronting representative agents from the different groups. In such a case, we move from the actual number of beatings, i.e., the numbers \(B_{ij}, i, j = 1, ..., g\), to the probability that an agent chosen at random from group \(i\) is at a better position than an agent randomly chosen from group \(j\), \(p_{ij}, i, j = 1, ..., g\). In so doing, we now look for a system of weights \(v = (v_1, ..., v_g) >> 0\), so that, for all \(i, j = 1, ..., g\),

\[
v_i = \frac{\sum_{j \neq i} p_{ij} v_j}{\sum_{j \neq i} p_{ji}} \quad [5]
\]
Irrespective of the size of the groups, we can also now construct the following matrix:

$$\Pi(P) = \begin{pmatrix}
    \sum_{j \neq 1} (p_{1j} + e_{1j}) & p_{12} & \ldots & p_{1g} \\
p_{21} & \sum_{j \neq 2} (p_{2j} + e_{2j}) & \ldots & p_{2g} \\
    \vdots & \vdots & \ddots & \vdots \\
p_{g1} & p_{g2} & \ldots & \sum_{j \neq g} (p_{gj} + e_{gj})
\end{pmatrix}$$

whose interpretation is straightforward: off-diagonal elements are $p_{ij}$, i.e., the probability that an agent chosen at random from group $i$ is at a higher category than an agent in group $j$, also randomly chosen. Diagonal elements are the aggregate probability of a representative agent in $i$ not being strictly dominated by representative agents in any of the other groups.

Notice that $\Pi(P)$ is a matrix with positive entries so that its columns add up to a constant number, $g - 1$. Thus, a vector with positive components exists, $v = (v_1, \ldots, v_n)$ so that $\Pi(P)v = (g - 1)v$, and thus, $v >> 0$ exists and its $i$-th components satisfy Equation (5). As before, vector $v$, also in this case, can be interpreted as the stationary value of a function $F$, defined by Equations (5).

**Remark 5** The uniqueness and strict positiveness of the weighting system is ensured by the implicit assumption that $B_{ij} > 0$, (and thus, $p_{ij} > 0$), for all $i, j$. Indeed to guarantee this result it is enough to assume a much weaker condition, namely, the irreducibility of matrix $P^*$. That is, that there is no partition of the set of groups $G$ into two parts, $G_1, G_2$, so that $p_{ji} = 0$, for all $i \in G_1, j \in G_2$. The interpretation of previous condition is straightforward. It simply means that any group in $G_1$ fully dominates any group in $G_2$. If that is the case, the valuation vector associated to the dominated groups (those in $G_2$) could be zero, and we only can guarantee the strict positiveness for the valuation of groups in $G_1$. The reducibility of matrix $\Pi(P)$ indicates that groups $G_1$ and $G_2$ belong somehow to "different" worlds, and it only makes sense to compare the relative performance within each class.²

## 3 Properties of the evaluation function

In this Section we will provide additional support to our evaluation function by showing that it satisfies a number of interesting properties that reinforce its operational and normative appeal.

Let $P = (N, G, C)$ be an evaluation problem involving $n$ agents, $g$ non-empty groups, and $s$ ordered categories, and let $P(g)$ stand for the set of problems $(N, G, C)$ with $g$ groups and any given (finite) number of agents. An evaluation function over the domain of all problems $(N, G, C)$ is denoted by $\Omega$, whereas $\Omega^g$ describes the restriction of $\Omega$ over $P(g)$. More formally:

²In a similar vein, Slutzki & Volij (2006) consider only tournament matrices that are irreducible, interpreting that the reducibility of matrices implies a division in different equivalence classes, and it is only meaningful to compare elements within the same class.
Definition 6 An evaluation function $\Omega$ is a mapping defined over all admissible problems $P = (N, G, C)$, such that, that for any $g \in \mathbb{N}$, $g < n$, $\Omega^g : M(g) \rightarrow \mathbb{R}^2_+$ is a vector such that $\sum_i \Omega_i^g(P) = g$.

The interpretation of $\Omega_i^g(P)$ is simply the value we attach to group $i$ in problem $P$. We adopt the convention of setting the mean value of the evaluation equal to one, in order to facilitate the interpretation, even though this is immaterial for the analysis.

We present now different properties that the evaluation function described in the former section enjoys. The first one worth mentioning is that of anonymity, which says that the evaluation only depends on the characteristics of the groups and not on other aspects such as labels or names. Formally:

- **Anonymity**: For any problem $P = (N, G, C)$, for any permutation $\sigma : G \rightarrow G$, and for all $i \in G$, it happens that $\Omega^g_{\sigma(i)} = \sigma(\Omega^g_i)$.

A second standard property has to do with the situation in which for all pairs of groups, $i, j$, it happens that $B_{ij} = B_{ji}$, that is, in all pairwise comparisons there is no advantage of any group over the other. If this is the case, then all groups should be given identical value.

- **Symmetry**: For any problem $P = (N, G, C)$, such that for all $i, j \in G$, $B_{ij} = B_{ji}$, we have that $\Omega_i^g(P) = \Omega_j^g(P)$ for all $i, j \in G$.

In the above situation, matrix $\Pi(P)$ is symmetric, and $\frac{1}{g-1}\Pi(P)$ is a bistochastic matrix, and thus, the dominant eigenvalue is $(1, ..., 1)$.

Consider now the particular case of evaluation problems involving only two groups. The next three properties refer to that case, and somehow are, in turn, more demanding. We start by introducing a monotonicity property that describes the behavior of the evaluation function when there is an improvement in one of the groups. Let $P = (N, \{i, j\}, C)$ be a two-group problem and suppose that some members of group $i$ improve their situation by moving to a higher position (from $s$ to $r$, say, with $r < s, s, r \in C$). Let $(N, \{i', j\}, C)$ denote the new situation and call the change from $i$ to $i'$ an unambiguous improving. Then, the relative valuation of group $i'$ with respect to group $j$ should be higher in problem $(N, \{i', j\}, C)$ than it was in the original problem. Formally:

- **Monotonicity**: Let $P = (N, \{i, j\}, C)$ be a two-group problem, and let $P' = (N, \{i', j\}, C)$ be a problem that results from an unambiguous improving of group $i$. Then, $\Omega_i^g(P)/\Omega_j^g(P) < \Omega_i^g(P)/\Omega_j^g(P')$.

The next property, stochastic dominance, can be regarded as an implication of the monotonicity and anonymity properties presented above. It relates the valuation of two groups of the same size, when one of them stochastically dominates the other one.

\footnote{With this normalization we immediately differentiate the groups above the mean (those with a worth larger than one) from the groups below the mean (those with a worth smaller than one).}
• **Stochastic Dominance:** Let $P = (N, \{i, j\}, C)$ be a two-group problem, s.t. $n_i = n_j$, and suppose that $n_{i1} \geq n_{j1}, n_{i1} + n_{i2} \geq n_{j1} + n_{j2}; \cdots; n_{i1} + n_{i2} + \cdots + n_{i,s-1} \geq n_{j1} + n_{j2} + \cdots + n_{j,s-1}$, with some strict inequality. Then $\Omega_i^2(P)/\Omega_j^2(P) > 1$.

The dominance relation between groups $i$ and $j$ before can be interpreted as $i$ being an unambiguous improving of $j$ (either in one or several steps). Thus, combining monotonicity with anonymity, we get the stochastic dominance property. Also note that, in this context, stochastic dominance implies dominance in pair-wise comparisons and dominance according to the relative advantage criterion (see the comment after equation [4]).

Consider finally, the idea of providing a specific form of the relative evaluation of pair of groups, whenever we face a two-groups problem. **Reciprocity** establishes that, in those problems involving only two groups of the same size, the relative valuation of the groups should coincide with the ratio between the number of times a group beats the other.\(^4\) Formally:

• **Reciprocity:** For any problem $P = (N, \{i, j\}, C) \in P(2)$, we have:

$$\frac{\Omega_i^2(P)}{\Omega_j^2(P)} = \frac{B_{ij}}{B_{ji}}$$

Previous property implies both monotonicity and stochastic dominance.

Now we introduce the idea that what matters is the relative frequencies (the shares of the agents of a group in the different categories) and not the size of the groups. That is,

• **Group Replication Invariance:** Let $P = (N, G, C)$ be a problem in $P(g)$, and let $\alpha = (\alpha_1, \ldots, \alpha_g) \in \mathbb{N}^g$ be a vector of natural numbers. Call $P \otimes \alpha \in P(g)$ a problem in which the population of group $i$ in $P$ has been replicated $\alpha_i$ times, for $i = 1, 2, \ldots, g$. Then, $\Omega(g)(P) = \Omega(g)(P \otimes \alpha)$.

The properties analyzed so far involve a fixed number of groups. We now introduce a property that, instead, relates problems with different number of groups. This is a property in which we require a sort of bilateral consistency in the way the solution behaves when considering a series of two-group associated problems.\(^5\)

Given an evaluation function $\Omega$ and a problem $P \in P(g)$, recall that the aggregate valuation of the beats of group $i$, is given by

$$W_i(\Omega, P) = \sum_{j \neq i} B_{ij} \Omega_j^2(P) \quad [8]$$

$W_i(\Omega, P)$ increases both with the number of times an agent of this group beats an agent of other groups and with the value attached to those groups.

---

\(^4\)A similar property appears in Slutzki & Volij (2006)

\(^5\)On the relevance and ideas about bilateral consistency in different settings, see Hokari & Thomson (2008)
Now we relate the behavior of the evaluation function in a given problem with a series of associated problems involving two groups each. Let $P = (N, G, C)$ be a problem in $P(g)$ such that all groups have identical size, i.e., $n_1 = n_2 = \cdots = n_g$. For a group $i \in G$, let $P^i = [N, \{\{i\}, G_{-i}\}, C] \in P(2)$ be the two-group problem consisting of group $i$ and a group made out of the merging of all groups other than $i$. How should the evaluation function value group $i$ in the original problem and group $i$ in the associated two-group problem? The following property, agreement, establishes that the ratio between the evaluation of group $i$ in $P \in P(g)$ and in $P^i \in P(2)$ should coincide with the ratio between the aggregate valuation of the beats of $i$ in $P$ and in $P^i$. Formally:

- **Agreement:** Let $P = (N, G, C)$ be a problem in $P(g)$ with $n_1 = n_2 = \cdots = n_g$. Then, for all $i \in G$,

$$\frac{\Omega_i^g(P)}{\Omega_i^2(P^i)} = \frac{W_i(\Omega^g, P)}{W_i(\Omega^2, P^i)}$$

Among the mentioned properties, the last three, proportionality in the two-group case, group replication invariance, and agreement, characterize the valuation function that attaches to each group its worth. We omit the easy proof.

**Theorem 7** An evaluation function $\Omega$ satisfies reciprocity, group replication invariance, and agreement, if and only if, for any given problem $P = (N, G, C)$, it associates to each group its worth. That is, given a problem $P = (N, G, C) \in P(g)$, we have, for all $i \in G$:

$$\Omega_i^g(P) = \frac{\sum_{j \neq i} p_{ij} \Omega_j^g(P)}{\sum_{j \neq i} p_{ji}}$$ [9]

### 4 An illustration: Comparing the quality of human capital in Europe

Human capital is one of the key determinants of human development and long run economic growth (see for instance the recent contribution by Acemoglu & Robinson (2012) and the references provided there). Countries show a wide variety of educational structures, regarding the qualification of their labour forces, even in relatively homogenous environments. According to Eurostat, the data on the 2010 show that in Portugal 68.1% of the adult population had only completed primary studies, 16.5% secondary studies, and 15.4% tertiary studies. Those percentages in the case of Latvia were 11.5%, 61.6% and 26.9%, respectively, whereas in Belgium the distribution was 29.5% with primary studies, 35.5 with secondary studies, and 35 with tertiary education. Such a diversity of realizations points out the difficulty of getting an overall comparative measure. Our model can be fruitfully applied to this context, in order to obtain a sensible
evaluation of the relative quality of the human capital under the assumption that higher levels of education are preferable.\textsuperscript{6}

We compare here the composition of human capital in Europe in 2010, in terms of the distribution of the adult population across the different educational levels. By adult population we mean the population between 25 and 64 years old. Educational levels are defined in terms of the International Standard Classification of Education (ISCED), using three different levels. We compare the relative educational achievements of 30 European countries out of the data provided by Eurostat. Table 1 below describes, in the first three columns of its body, the distribution of the population of the different countries in three educational levels: primary studies (ISCED 0-2), secondary studies (ISCED 3-4), and tertiary studies (ISCED 5-6). The table shows the wide diversity in the human capital structure mentioned above. Roughly speaking, Europe presents a distribution in which one half of the adult population has secondary studies, whereas one fourth has primary studies and the remaining fourth tertiary studies. The extreme values for tertiary studies correspond to Finland (37.6 \%) and, somehow unexpectedly, to Italy (14.8). The extreme values for primary studies correspond to Lithuania (8 \%) and Malta (70.3 \%). Secondary studies range between 75.2 \% (Czech Republic) and 16.5 \% (Portugal). The corresponding coefficients of variation are 0.3 for tertiary studies, 0.3 for secondary studies, and 0.6 for primary studies.

The last column of Table 1 provides the evaluation of the relative quality of human capital in those countries according to our evaluation formula (the normalized eigenvector of the associated $P^*$ matrix). The worth vector shows that Lithuania, Estonia, Switzerland, Finland, Sweden, and United Kingdom are the countries with relatively better human capital structure, whereas Malta, Portugal, Italy, Romania, Spain and Greece are at the other end of the quality distribution. The coefficient of variation of the worth vector takes on the value 0.35, which is rather large. Figure 1 plots the distribution of this vector.

Table 1 and Figure 1 around here

5 Final Comments

There are many different evaluation problems that involve several groups whose members are classified into ordered categories. The solution proposed here exploits the information on the distribution of agents across categories in order to provide an estimate of their relative situation. Our contribution consists of framing the evaluation problem so that we can rely on a conventional solution (the fixed point of a linear mapping) to provide the type of evaluation we are looking for. Moreover, as this fixed point corresponds to the dominant eigenvector of a Perron matrix, the proposed solution exhibits simple, useful, and well-known properties (e.g. existence, uniqueness, positiveness, stability, and

\textsuperscript{6}Notice that this is a safe assumption from a microeconomic point of view but not necessarily so from a macroeconomic perspective (at least in the short run). This is so because the economic structure of a country may call for a particular distribution of qualifications.
regular behaviour regarding changes in the parameters). The way of constructing such an evaluation function is, furthermore, justified on the basis of the properties it fulfills and it is axiomatically characterized.

To guarantee that all groups have a positive worth we have to assume that matrix $\Pi(P)$ is irreducible, namely, that there is no way of splitting the set of groups $G$ into two parts, $G_1$ and $G_2$ so that any group in $G_1$ fully dominates any group in $G_2$. The simplest case in which irreducibility is violated appears when there is a group fully dominated by all the remaining groups. In such a case, we may think of the fully dominated group as a dummy, and it happens that its valuation is zero. Moreover, the relative valuation of the remaining groups does not change whether we take the dummy group into consideration or not. Indeed, if in a particular group all the population is concentrated in the worst category, then such a group becomes a dummy. In some types of tournaments, as in sport tournaments, the irreducibility of the tournament matrix is related to the idea of all teams belonging to the same league. If a problem $P$ presents an associated reducible matrix $\Pi(P)$, then an equivalence relation can be defined by considering direct and indirect dominations in the following way: "group $i$ belongs to the same class as group $j$ if there exists a finite sequence of groups: $i_1 = i, \ldots, i_k = j$, so that $p_{i_1} p_{i_2} \cdots p_{i_k} > 0$. Then $\Pi(P)$ can be decomposed into equivalence classes, and the relative valuation only makes sense within each class (see Slutzky & Volij (2006), or Chebotarev & Shamis (1998)).

Another special case is that in which there is no pairwise advantage between any two groups. In such a case, the valuation of all groups is identical. A special case of this situation is that in where there is perfect equality in the distribution of the types within the groups.

Allison & Foster (2004) analyze inequality relations for distributions involving categorical data, looking for comparison criteria independent of any particular cardinalization on the categories. They show that stochastic dominance provides a partial ordering robust to changes in the cardinalization. Their framework is similar to that considered here regarding the design of a measure that is independent on cardinalizations, even though the focus is different (we are concerned with relative valuations rather than with inequality relations). The fact that our measure unambiguously ranks group $i$ over group $j$ when $i$ stochastically dominates $j$ could be interpreted as a criterion consistent with Allison & Foster’s result. Nonetheless, there are two significant differences between these results. On the one hand, our measure provides a complete ordering rather than a partial ordering. On the other hand, our model goes beyond bilateral comparisons as it can be applied to any number of groups.

We have provided an illustration regarding the evaluation of human capital in 30 European countries. This allowed us to obtain insights on the comparative educational structure of those countries. Our instrument is flexible enough to be applied not only to similar problems (e.g. the evaluation of health achievements), but also to rather different scenarios. Think for instance of the case in which we have to compare customer satisfaction in different firms or between several branches of a given firm, when customers declare their satisfaction level by choosing one between four or five categories (e.g. from "fully satisfied" to
"not satisfied at all"). Our procedure provides a consistent and endogenous way of making those comparisons that would otherwise be dependent on some arbitrary weighting of categories.

References


