Robust comparisons of inequality of opportunity for skills acquisition under the veil of ignorance*

Francesco Andreoli[a,†], Edward Levavasseur[b], and Nicolas Gravel[c]

[a] *LISER - Luxembourg Institute for Socio-Economic Research*
[b] *Aix-Marseille University and AMSE*
[c] *Centre for Social Sciences and Humanities (CSH Delhi) and Aix-Marseille University*

January 2018

Preliminary draft. Do not cite or quote

Abstract

This paper examines a robust operational definition of "opportunity equalization" for skills acquisition among children of different family background. The criteria allow to compare societies are described as collections of distributions of outcome (lotteries), one such distribution for every family background type. Societies are confronted from the view point of a philosopher placed behind a veil of ignorance with respect to the type he/she sould have in this society. Under established axioms of choice under ambiguity, we show that the preferences of this philosopher involve comparison of the expectation of some concave function of the expectations of the consequences of skills lotteries on children well-being. We provide an empirical criterion for comparing societies that coincide with the unanimity of all rankings that would command agreement among these philosophers when they exhibit aversion to inequality of opportunity. This last result is used to assess the evolution of inequality of opportunity for skills acquisition across PISA countries, decomposing inequalities developed early in life from the contribution of the education systems.

Keywords: equality of opportunities, groups, zonoids.
JEL classification numbers: D63.

*We are grateful to Patrick Moyes for helpful discussion on ideas developed in this paper. This paper forms part of the research project "The Measurement of Ordinal and Multidimensional Inequalities" (Contract No. ANR-16-CE41-0005-01) of the French National Agency for Research whose financial support is gratefully acknowledged. Edward Levavasseur acknowledges hospitality from LISER.

†Corresponding author: Luxembourg Institute of Socio-Economic Research, LISER. MSH, 11 Porte des Sciences, L-4366 Esch-sur-Alzette/Belval Campus, Luxembourg. E-mails: francesco.andreoli@liser.lu, edward.levavasseur@univ-amu.fr and nicolas.gravel@csh-delhi.com .

1 Introduction

There is increasing evidence that inequalities in accumulation of skills early in life contribute to economic and health inequalities later in life. Evidence from recent PISA reports highlights growing disparities across countries, as well as across social groups, in the skills achievements of adolescents and young adults.

Part of the heterogeneity in skills can be traced down to innate abilities, to preferences and effort choices of individuals. The resulting inequality in skills distribution has to be respected as long as its determinants are regarded as fair sources of skills accumulation.

In this paper, we study criteria to quantify the effect of unfair sources of inequality in skills acquisition. Factors such as parental education may have implications for children skills development in early age, these effects cumulating throughout the life cycle into inequalities in human capital, in earnings and eventually in opportunity for well-being. We denote the contribution of these factors to overall skills heterogeneity in the population, defining inequalities in the distribution of skills across types in the population (a type gathers children with similar background), as *inequality of opportunity for skills acquisition*. This paper characterize preferences for a social planner who is averse to unequal distribution of skills opportunities and who values welfare distribution stemming from skills acquisition in the society. These preferences are used to construct and test robust ranking of societies according to the level of social welfare (and hence inequality of opportunity for skills acquisition).

More formally, let n identify the number of types in the society, a type being a group of people sharing the same circumstances. We assume a continuum of individuals populating the society, with p_i, $i = 1, \ldots, n$ the mass of individuals of type i. A society is represented by a matrix $\mathbf{p}$ whose element p_{ij} is the probability that an individual from type i achieves an outcome j. Hence, $\sum_{j=1}^{k} p_{ij} = 1$. We also denote $\overline{p} = (\overline{p}_1, \ldots, \overline{p}_k)$ the population probability of achieving an outcome j, with $\overline{p}_j = \sum_{i=1}^{n} p_i p_{ij}$. Taking an ex-ante perspective on society, i.e. before individual preferences and choices are realized, we expect $\overline{p}$ to be the skills opportunities distribution emerging if parental background circumstances had no effect on individual skills accumulation.

We measure the well-being of the type-specific skills opportunity distribution by $\sum_{j=1}^{k} u_j p_{ij}$, u_j being the specific level of well-being that one would expect to observe if skills level j is achieved. All individuals from the same type face the same distributions of opportunities for skills accumulation, hence the same well-being. If, for instance, the coefficient u_j measures the average wage attached to a given skill level j, then $\sum_{j=1}^{k} u_j p_{ij}$ would measure the expected wage of a given type in the population. If there were equal opportunity for skills acquisition, than every individual would ex-ante expect the same wage rate $\sum_{j=1}^{k} u_j \overline{p}_j$.[1]

Inequality of opportunity for skills acquisition emerges when the probabilities for skills acquisition differ across types and, generally, from the population distribution $\overline{p}$. It follow that the distribution of *consequences* of unequal opportunities for skills acquisition, measured by $\sum_{j=1}^{k} u_j p_{1j}, \sum_{j=1}^{k} u_j p_{2j}, \ldots, \sum_{j=1}^{k} u_j p_{nj}$, also display inequality. This last form of inequality can be well captured by a social welfare function that, under the veil of ignorance, attaches the same weight all profiles irrespectively of the group of origin (i.e. there is no intrinsic value attached the name of the group) while accounting for the demographic composition of the population. Let ϕ be an increasing (captures the idea that more advantageous distribution of consequences from opportunities in skills acquisition lead to larger welfare) and concave (captures the idea if inequality of opportunity aversion: increasing inequality of opportunity generated by family background characteristics are detrimental for societal welfare) function, then social welfare $W(\mathbf{p})$ is:

$$W(\mathbf{p}) \quad = \quad \sum_{i=1}^{n} p_i \phi \left(\sum_{j=1}^{k} u_j p_{ij} \right) \tag{1}$$

for any ϕ and u.

Our first contribution establishes that the preferences of a ethical observer who respond to meaningful properties can be represented by a social welfare function as in (1). We provide an axiomatic characterization.

Our second contribution establishes testable, empirical criteria that allow to compare two societies $\mathbf{p}$ and $\mathbf{p}'$ on the basis of the social welfare they display. For a given choice of

[1] Nevertheless, the actual average wage rate might differ across types because of differences in the ex-post effort choices, which might vary across individuals.

ϕ and u, evaluation is simple: $\mathbf{p}'$ is preferred to $\mathbf{p}$ if and only if $W(\mathbf{p}') \geq W(\mathbf{p})$. However, this evaluation is not robust to the choice of the individual well-being indicator, nor to the social evaluation function. Hence, we require $W(\mathbf{p}') \geq W(\mathbf{p})$ for all ϕ increasing concave and all u. We analyze a *statistical test* that, when applied to the data $\mathbf{p}$ and $\mathbf{p}'$, would allow to conclude on the preferred society. While the roust welfare criterion we identify is appealing, it lacks theoretical connections with inequality of opportunity normative setting. We analyze as well the *transformations* of the data that, if applied to $\mathbf{p}$ would produce a counterfactual society $\mathbf{p}'$ displaying larger social welfare and that unambiguously reduce inequality of opportunity for skills acquisition. We show that welfare dominance is supported by finite sequence of these transformations.

The main result of the paper consists in two theorems that show the equivalence between 1) transformations of the data, 2) welfare dominance 3) a statistical test. In the first theorem, the focus is on situations where opportunities for skills acquisition are defined over many (k) ordered outcomes and where the population distributions coincide.

In the second theorem, we focus is on social welfare functions as in (1) where well-being indicators u are *increasing* in the outcome class, i.e. $u_j \leq u_{j+1}$ for any skills level j. Increasing well-being indicators imply that societies can be compared according to the degree of inequality of opportunity as well as according to the *advantages in opportunities* they display. We stick to the notion of *stochastic dominance* to identify distributions that are more advantageous in terms of skills opportunities: if society $\mathbf{p}$ and society $\mathbf{p}'$ do not display inequality of opportunity but the distribution $\overline{p}$ stochastic dominates that $\overline{p}'$ (i.e., the distribution of skills in the population $\mathbf{p}$ is undisputed better than that in $\mathbf{p}'$), then society $\mathbf{p}$ has to be preferred to $\mathbf{p}'$ by all social evaluation functions. In this case, improvements in opportunities can be trade-off by increments in inequality of opportunity for skills acquisition.

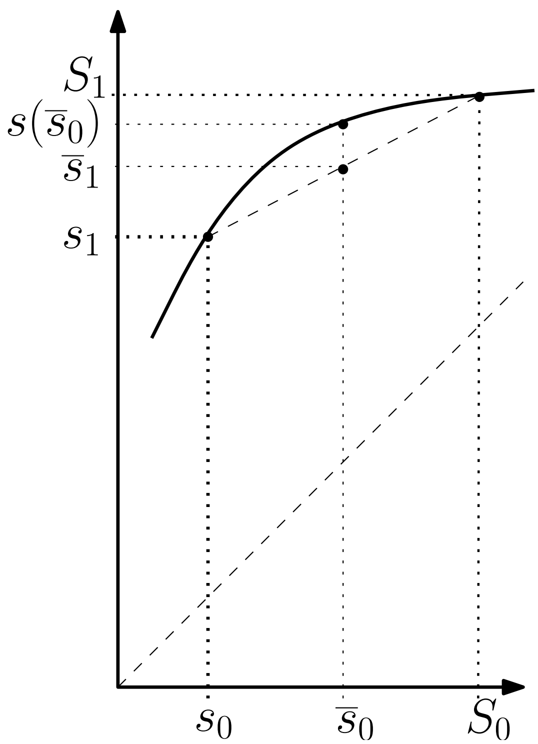
We apply the measurement model described above to assess the opportunity unequal distribution of basic skills around age 15, as measured in PISA data. PISA test outcomes are constructed in a way that they measure skills of the tested children coherently with the average degree of competencies expected at that age. Competencies are held fixed across countries in any give PISA year, making the PISA score comparable across countries. We are interested in the unequal opportunities for PISA-skills acquisition conditional on human capital, education and abilities of the parents of interviewed children. Our measurement model accommodates for the possibility that PISA test scores map unobservable real skills into PISA skills levels via a monotonic increasing transformation (not necessarily linear), thereby preserving only ordinal (but not cardinal) meaning to measured skills. Our model exploits exclusively ordinal information.

The scope of the analysis is to construct a robust, welfare consistent, ranking of OECD countries participating to PISA, based on distribution of skills across types defined by the quality of parental background experienced by the child. Even assuming that underling skills are cardinally comparable (but PISA test scores are not) we face two important measurement issues.

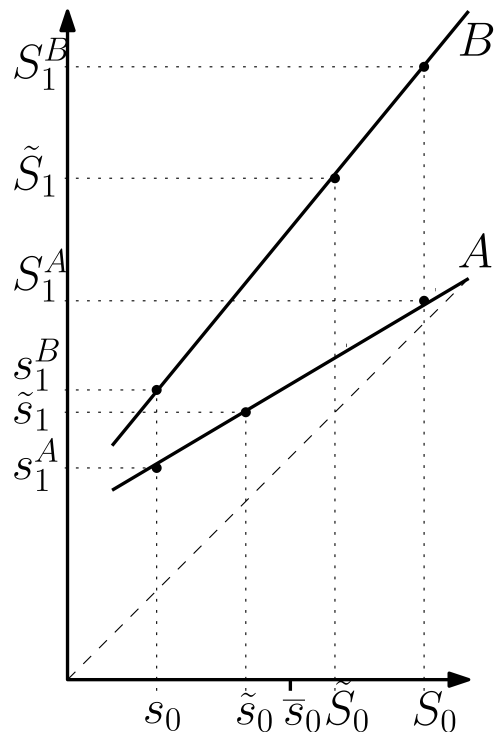
First, skills are malleable throughout the life cycle of the children. In particular, differences across parental background composition, parental resources and (time and monetary) investments may explain distribution of skills in early life. Differences in mechanisms of skills acquisition (for instance, via school segregation, or differences in public versus private schooling) as well as the existence of skills complementarities (even within the same system) and the degree of elitism in the education system, all operate during late childhood and early adolescence, thus modifying the patterns of initial inequalities in skills acquisition. Distinguishing the different cases, and understanding the empirical limitations, is important to infer differential in inequality of opportunity across countries.

Consider for instance the situation in Figure 2(a). There are two types in this society, one granting on average low initial skills $s_t = s_0$ to children, the other background granting on average high skill levels $s_t = S_0$. The concave function represents the skills cumulation process over the lifecycle, projeting skills at age 15 on the vertical axis for any level of skills at early age. Coherently with the literature (see Heckman), the process displays skills complementarities (high-skilled children tend to accumulate more skills on average) but returns to skills are decreasing (concavity). Overall, we register inequality of opportunity at age 15: $s_1 \neq S_1$ for children experiencing different parental background.

Figure 1: Inequality of opportunity and skills formation



(a)



(b)

Note: The process of skills s_t formation along children lifecycle: from early age ($t = 0$) to age 15 ($t = 1$). The example consider two types, one with low average skills (s) and one with high skills (S). The acquisition processes are represented by line segments A and B.

Nevertheless, the skill accumulation process is fair: The skill level at age 15 depends on the skill level at early age, but types are exposed to the same skills accumulation process. In this case, inequality of opportunity is generated by early life conditions. The fair distribution incorporating no inequality of opportunity, corresponding to $\overline{s}_1$ (the average of the distribution of skills at age 15 obtained by average skills distributions s_1 and S_1) is however underestimating the distribution of opportunities that would be observed if sources of unfair inequalities were removed (corresponding to $s(\overline{s}_0)$). This happens because the relevant source of inequality of opportunity takes place at early age, but there are non-linearities in the process of skills accumulation. The validity of the welfare model characterized in the paper, which is based on a linearity assumption, is hence informed on a test for the presence of linearities in the skills formation process. If rejected, the shape of the skills cumulation process plays a fundamental role in assessing inequality of opportunities. If the accumulation process is increasing and concave (as in the figure), then the welfare criterion we analyze would over-estimate inequality of opportunity.[2]

The second issue we address concerns the source of inequality of opportunity in skills acquisition. Evaluating and ranking opportunity distributions in two countries should single out the contribution of different sources. For policy purposes, it is for instance relevant to distinguish between the contribution of the skills cumulation process (where educational policies may play a role) from the contribution of differences in early conditions (that household-based policies may countervail). Countries may, in fact, differ not only on the process of accumulation of skills (where education system play a decisive role), but also on demographics and on the direct transmission process of parents skills into children skills.

[2]Consider the case where one has to compare two countries, both giving the same distribution of skills s_1 and S_1 with average $\overline{s}_1$ and thereby the same inequality of opportunity (Figure 2(a)). In both cases, the source of inequality of opportunity comes from differences in parental investments in early age, while all people are exposed to the same accumulation mechanism. Nevertheless, the country characterized by a concave accumulation process would display a larger average distribution of skills if sources of unequal distribution of opportunities were removed ($s(\overline{s}_0) > \overline{s}_1$), implying that actual inequality of opportunity in this country not only has produce a skill gap across children from different family backgrounds, but it has also deteriorated the overall average performance. All social evaluation functions would likely rank this country as less-opportunity egalitarian, implying that evaluations based on data (s_1, S_1) would over-estimate inequality of opportunity.

Consider for instance the situation in Figure 2(b). Let A and B represent two skills accumulation processes in two countries. Each country displays only one accumulation process, implying that inequalities in opportunities for skills acquisition has to be traced down to differences in the initial conditions, and how these differences interact with the skill accumulation process. In this example, comparing inequality of opportunity in the distribution with average skills levels (s_1^A, S_1^A) with the distribution with skills levels (s_1^B, S_1^B) would capture the implications that the skills accumulation process has on the initial inequality of opportunity, as measured by (s_0, S_0) in both cases. The skills accumulation processes in both countries are fair but different, but the initial distribution of skills is not. The interaction among the two components generates differences in inequality of opportunity.

A second example is that in which country B, displaying a fair process but starting with a very unequal distribution of opportunities, displays skills distribution (s_1^B, S_1^B) and is compared with a country displaying $(\tilde{s}_1, \tilde{S}_1)$. In the latter society, which we label AB, there is less inequality in early skills, $(\tilde{s}_0, \tilde{S}_0)$, but the process of accumulation is not fair: children from the type with skills $\tilde{s}_0$ would face accumulation process A while children with skills $\tilde{S}_0$ would later face accumulation process B, more advantageous. This can be, for instance, an outcome consistent with early tracking, which imposes career assignment on the basis of early skills development. Evaluating the inequality of opportunity of society $(\tilde{s}_1, \tilde{S}_1)$ against society (s_1^B, S_1^B) would likely mask the consequences of different channels. On the one hand, the first society magnifies small initial gaps in opportunities though selective and unfair education systems, which calls for policy intervention. On the other hand, the system in the second society is fair, but the initial gap in skills endowment is large and it is then magnified by the system. This calls for different forms of intervention.

It is key for policy purposes to distinguish the role of different channels. We do so using counterfactual methods, which allow to keep constant some sources of inequality of opportunities across societies, and ranking the resulting distributions once these channels have been switched off. For instance, we use counterfactual methods to simulate the counterfactual distribution of skills that would have prevailed in society AB if the initial skills distribution was that of society B, namely (s_0, S_0). As figure 2(b) shows, one would have to compare the outcome distribution (s_1^B, S_1^B) with the counterfactual distribution (s_1^A, S_1^B).

To do so, we exploit variation across age of comparable children to identify and decompose the role of initial skills distribution and later accumulation process on the inequality of opportunity measured by raw PISA data. Since there are no large scale assessment surveys that follow a cohort of children, as they move up through the different grade years of their education system, we have merged data from TIMSS 2003, TIMSS 2007 and PISA 2009. Indeed, since the children from all three databases were mostly born in 1993/1994, we are able to assess the same birth-cohort in grade 4 with TIMSS 2003, then 4 years later in grade 8 with TIMSS 2007, and finally 2 years later when they are aged 15 to 16 using PISA 2009. Furthermore, since all three databases provide similar background information on the children (Parent's education, number of books at home, various material possessions...), one can create a new database by merging the 3 previously mentioned databases. Finally, since multiple countries are assessed in at least 2 of the 3 databases, we will be able to assess the evolution of 23 countries at 2 of the 3 grade years, and 11 countries - including the USA - at all 3 grade years.

2 Characterizing social welfare with equality of opportunity concerns

2.1 Notation

The sets of integers, non-negative integers, strictly positive integers, real numbers, non-negative real numbers and strictly positive real numbers are denoted by $\mathbb{N}$, $\mathbb{N}_+$, $\mathbb{N}_{++}$, $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ respectively. The cardinality of any set A is denoted by $\#A$ and the k-fold Cartesian product of a set A with itself is denoted by A^k. The inner product of an $n \times m$ matrix a by an $m \times r$ matrix b is denoted by $a.b$. The k-dimensional unit vector is denoted by 1^k. For any $k \in \mathbb{N}_{++}$ with $k \geq 2$, we denote by S^{k-1} the $k-1$ dimensional simplex defined by $S^{k-1} = \{s \in [0,1]^k : s_1 + ... + s_k = 1\}$. A *permutation matrix* π is a square matrix whose entries are either 0 or 1 and sum also to 1 in every line and every column. Our notation for vectors inequalities is $\geqq$, $\geq$ and $>$. A *binary relation* $\succsim$ on a set Ω is a subset of $\Omega \times \Omega$. Following the convention in economics, we write $x \succsim y$ instead of $(x, y) \in \succsim$. Given a binary relation $\succsim$, we define its *symmetric factor* $\sim$ by $x \sim y \iff x \succsim y$ and $y \succsim x$ and its *asymmetric* factor $\succ$ by $x \succ y \iff x \succsim y$ and not $(y \succsim x)$. A binary relation $\succsim$ on Ω is *reflexive* if the statement $x \succsim x$ holds for every x in Ω, is *transitive* if $x \succsim z$ always follows $x \succsim y$ and $y \succsim z$ for any x, y, $z \in \Omega$ and is *complete* if $x \succsim y$ or $y \succsim x$ holds for every distinct x and y in Ω. A symmetric, reflexive and transitive binary relation is called an *equivalence relation* and a reflexive, transitive and complete binary relation is called an *ordering*. Given an equivalence relation $\sim$ on Ω, and some $\omega \in \Omega$, we denote by $E_\sim(\omega)$ the equivalence class of ω under $\sim$ defined by $E_\sim(\omega) = \{\omega' \in \Omega \mid \omega' \sim \omega\}$. It can be seen immediately that if $\sim$ is an equivalence relation, one has $E_\sim(\omega) \neq \emptyset$ for every ω, $E_\sim(\omega) = E_\sim(\omega')$ or $E_\sim(\omega) \cap E_\sim(\omega') = \emptyset$ for every elements ω and ω' in Ω and $\bigcup\limits_{\omega \in \Omega} E_\sim(\omega) = \Omega$ so that the equivalence class of all elements of Ω under $\sim$ form a partition of Ω. Such a partition is called the *quotient of* Ω *under* $\sim$.

2.2 Basic Framework and Definitions

We are interested in comparing societies on the basis of their performances in *equalizing opportunities* among some exogenously given groups of individuals. These groups can be based on religion, race, gender, family backgrounds, etc. Our approach to appraising equality of opportunities does not enquire about the origin of these groups. We neither assume that the number of such groups is the same across societies. For instance, we may consider societies formed by one group only. Our approach would then view such one-group societies as achieving (trivially) perfect equality of opportunities. The opportunities offered to a group in a society are described by the *fraction of individuals* in this group who achieve any relevant outcome. We assume specifically that there are k such outcomes. We could view these outcomes as anything that individuals have reason to value and that are observable somehow. Examples would include income categories, or education levels.

It is somewhat important for the analysis conducted here that these various outcomes be in finite number.

We hence depict a society $\mathbf{p}$ as an $n(p) \times k$ row stochastic matrix:

$$\mathbf{p} = \begin{bmatrix} p_{11} & ... & p_{1k} \\ ... & ... & ... \\ p_{n(p)1} & ... & p_{n(p)k} \end{bmatrix}$$

where p_{ij}, for $i = 1, ..., n(s)$ and $j = 1, ..., k$ denotes the probability that an individual coming from group i achieves outcome j in society $\mathbf{p}$. We denote by p_i the lottery (probability distribution over $\mathbb{R}^k$) associated to group i in society $\mathbf{p}$. A society $\mathbf{p}$ with $n(\mathbf{p})$ groups is therefore an element of $(S^{k-1})^{n(s)}$ and the set of all logically conceivable such societies is $\bigcup\limits_{l \in \mathbb{N}_{++}} (S^{k-1})^l$. If $\mathbf{p}$ is a society in $(S^{k-1})^m$ and $\mathbf{p}'$ is a society in $(S^{k-1})^n$, we denote by $(\mathbf{p}, \mathbf{p}')$ the society in $(S^{k-1})^{m+n}$ where the m first groups face the lotteries associated to $\mathbf{p}$ and the n last groups get face the lotteries, in the same order, in $\mathbf{p}'$. For any lottery p in S^{k-1}, we denote by p the one group society in which all members face the lottery p. We notice that depicting societies as lists of probability distributions, with one such list for every group, makes sense only if one adopts an *anonymous* postulate that "the name of the group does not matter". The fact that these groups are cast, races or religions has no importance for appraising opportunities. The only relevant feature of the society is the distribution of the ethically relevant outcomes within each group.

Alternative societies are to be compared by an *ethical observer* or a philosopher who is placed behind a "veil of ignorance" as to the group to which he (she) would belong if he (she) was to live in the various societies. We assume that such ethical observer uses the ordering $\succsim$, with asymmetric and symmetric factors $\succ$ and $\sim$ respectively to compare these societies. We interpret the statement $\mathbf{p} \succsim \mathbf{p}'$ as meaning "I would weakly prefer starting my life in society $\mathbf{p}$ than in society $\mathbf{p}'$ ". A similar interpretation is given to the statements $\mathbf{p} \succ \mathbf{p}'$ (strict preference) and $\mathbf{p} \sim \mathbf{p}'$ (indifference). Since the ordering $\succsim$ is defined on the whole set $\bigcup\limits_{l \in \mathbb{N}_{++}} (S^{k-1})^l$, it is in particular defined on the set S^{k-1} of all conceivable one-group societies and, therefore, of all lotteries.

We start by identifying the properties (axioms) of $\succsim$ that are necessary and sufficient for the existence of a function $\Psi : S^{k-1} \to \mathbb{R}$ such that, for all societies $\mathbf{p}$ and $\mathbf{p}'$ in $\bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$, one has:

$$\mathbf{p} \succsim \mathbf{p}' \Longleftrightarrow \sum_{i=1}^{n(p)} \frac{\Psi(p_i)}{n(p)} \geq \sum_{i=1}^{n(p')} \frac{\Psi(p'_i)}{n(p')} \tag{2}$$

An ordering satisfying this property could therefore be thought of as resulting from the comparisons of the *average evaluation* of the lotteries offered by two compared societies for some evaluation function, under the assumption that the ethical observer is equally likely to fall in any group. Notice that formula (2) defines in fact a *family* of social criteria, with as many members as there are logically conceivable functions Ψ. We shall discuss below how one could restrict this family a bit by imposing some additional property on the ranking of single groups societies or, equivalently, lotteries. The key assumption that we shall use on this matter is that the ranking of lotteries obey the standard VNM properties. If this is the case, then one could write, for any $p \in S^{k-1}$, the function Ψ by:

$$\Psi(p) = \Phi(\sum_{h=1}^{k} p_h \alpha_h) \tag{3}$$

for some real numbers $\alpha_1, ..., \alpha_k$ and some function $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$. In this specification, the real numbers $\alpha_1, ..., \alpha_k$ are interpreted as the utility evaluation, made by the philosopher, of the various outcomes. Hence the expression $\sum_{h=1}^{k} p_h \alpha_h$ constructed with these numbers can be seen as the *expected utility* associated to the lottery p, and the function Φ can be seen as a transformation of this expected utility into some magnitude, which reflects the attitude of the ethical observer with respect to ambiguity.

Following (Gravel, Marchant, and Sen 2012), we refer to any ranking that satisfies (2) for some function Ψ as to a *Uniform Expected Utility* (UEU) ranking of societies. This name comes from the decision under ignorance context in which this family was studied. Indeed, any ranking of societies that is numerically represented by (2) for some function Ψ can be though of as resulting from the comparison of the expected utility of the various lotteries offered by the societies, under the (uniform) assumption that the ethical observer assigns an equal probability to belonging to every group.

We now introduce the four axioms that characterize the UEU family of social rankings.

In line with the discussion above, the first axiom requires the social ranking to be *anonymous*. That is, the names or the nature of the groups do not matter for appraising the disparity of opportunities between groups in a society. Hence all societies who offer the same distributions of outcomes among the same number of groups are normatively equivalent. We state formally this axiom as follows.

Axiom 1 *(Anonymity) For every society* $\mathbf{p} \in \bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$*, and all* $n(\mathbf{p}) \times n(\mathbf{p})$ *permutation matrix* π*, one has* $\pi.\mathbf{p} \sim \mathbf{p}$*.*

The second axiom imposed on $\succsim$ is a continuity condition, imposed on the comparison of a one-group society vis-à-vis any other society. It says, roughly, that the strict ranking of a single lottery associated to a one-group society vis-à-vis any other society should be robust to "small" changes in the probabilities of achieving any given outcome. Its formal statement is as follows.

Axiom 2 *(Continuity) For every society* $\mathbf{p}$*, the sets* $B(\mathbf{p}) = \{\rho \in S^{k-1} : \rho \succsim \mathbf{p}\}$*,* $W(p) = \{\rho \in S^{k-1} : \mathbf{p} \succeq \rho\}$ *are both closed in* $\mathbb{R}^k$*.*

The next axiom is called *averaging* in (Gravel, Marchant, and Sen 2012). In the current context, the axiom evaluates what happens to the disparity of opportunities in a given society when the number of groups is enlarged. It says that if the disparity of opportunities in the added groups is better (worse) than what they are in the initial society, then the addition of those groups improves (deteriorates) the disparity of opportunities. It says also, conversely, that if a society loses (gains) from identifying new groups with specific distributions of outcome among their members, then this can only be because the distribution of outcomes within those groups is worse (better) than that already present in the original society. This axiom is formally stated as follows.

Axiom 3 *(Averaging) For all societies* $\mathbf{p}$ *and* $\mathbf{p}' \in \bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$*,* $\mathbf{p} \succsim \mathbf{p}' \Leftrightarrow \boldsymbol{p} \succsim (\mathbf{p}, \mathbf{p}') \Leftrightarrow (\mathbf{p}, \mathbf{p}') \succsim \mathbf{p}'$*.*

When applied to an ordering, the Averaging axiom implies several other properties. One of them is the axiom called "replication equivalence" by (Blackorby, Bossert, and Donaldson 2005) (p. 197) in the somewhat different context of population ethics. This axiom states that, for societies where every group faces the same opportunities, the *number* of those groups does not matter. This property is rather natural in the context of equalizing opportunities. If all groups in a society were offering the same opportunities, then the number of those groups would be irrelevant. We state formally this property as follows.

Condition 4 *(Irrelevance of the number of groups in case of equal opportunities) For every lottery* $\rho \in \mathbb{R}^k$ *and every society* $\mathbf{p} \in \bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$ *such that* $p_i = \rho$ *for all* $i = 1, ..., n(p)$*, one has* $\mathbf{p} \sim \rho$*.*

This condition is implied by averaging if $\succsim$ is reflexive. The proof of this claim is left to the reader.

The next, and last, axiom that requires the ranking of any two societies with the same number of groups to be robust to the addition, to both societies, of a common distribution of opportunities. That is to say, the ranking of any two societies with the same number of group should be independent from any group that they have in common. Formally, this axiom is stated as follows.

Axiom 5 *(Same number group independence) For all societies* $\mathbf{p}$, $\mathbf{p}'$ *and* $\mathbf{p}'' \in \bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$ *such that* $n(\mathbf{p}) = n(\mathbf{p}')$, $(\mathbf{p}, \mathbf{p}'') \succsim (\mathbf{p}', \mathbf{p}'')$ *if and only if* $\mathbf{p} \succsim \mathbf{p}'$*.*

It can be checked that any UEU ranking satisfies anonymity, continuity, averaging and Same Number Group Independence. (Gravel, Marchant, and Sen 2012) (see also (Gravel, Marchant, and Sen 2011)) have established the converse implication. Hence, one has:

Theorem 6 *Let* $\succsim$ *be an ordering on* $\bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$ *satisfying anonymity, continuity, averaging and same number expansion consistency. Then* $\succsim$ *is a UEU social ordering. Furthermore, the function* Ψ *of Expression (2) is unique up to a positive affine transformation and is continuous.*

Theorem 6 does not restrict in any way the function Ψ. Yet, one may wish to do so. A somewhat natural restriction would be to require the ranking of one-group societies - for which the issue of disparities of opportunities among groups vanishes - to obey the well-known VNM axiom. This would amount to add the requirement that $\succsim$ satisfies the following axiom.

Axiom 7 *(VNM for One-Group societies) For every lotteries p, p' and p" $\in S^{k-1}$ and every number $\lambda \in [0,1]$, $p \succsim p'$ if and only if $\lambda p + (1-\lambda)p'' \succsim \lambda p' + (1-\lambda)p''$.*

It is then immediate to obtain the following result (see e.g. Proposition 6 in (Gravel, Marchant, and Sen 2012)).

Proposition 1 *Let $\succsim$ be an ordering on $\bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$ satisfying anonymity, Continuity, Averaging, Same Number Group Independence and VNM for One-Group Societies. Then $\succsim$ is a UEU social ordering and the function Ψ of Expression (2) can be written as per Expression (3) for some function $\Phi : \mathbb{R} \to \mathbb{R}$ and some real numbers $\alpha_1, ..., \alpha_k$.*

Ethical observers who rank societies behind a veil of ignorance may be distinguished according to what could be called "aversion to inequality of opportunities". Intuitively, aversion to inequality of opportunities would correspond to a preference for societies who exhibit no disparity of opportunities - say because they are made of one single group - over societies who exhibit some disparity of opportunities among their different groups. This suggests the following notion of comparative aversion to inequality of opportunities among ethical observers.

Definition 8 *Given two rankings $\succsim_1$ and $\succsim_2$ of all societies in $\bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$, we say that $\succsim_1$exhibits at least as much aversion to inequality of opportunity as $\succsim_2$ if, for every lottery $\rho \in S^{k-1}$ and society $\mathbf{p} \in \bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$, $\rho \succsim_2 \mathbf{p} \Longrightarrow \rho \succsim_1 \mathbf{p}$.*

In words, an ethical observer who compares societies by means of the binary relation $\succsim_1$ exhibits at least as much aversion to inequality of opportunities as another who bases his/her comparisons on $\succsim_2$ if any preference that the later will have for a society with no inequality of opportunities (as compared to any reference society) would also be endorsed by the former. It is not difficult to see that this notion of "comparative aversion to opportunity inequality" can translate, when expressed for UEU rankings, into a statement of "comparative concavity" applied to the function Ψ of Expression (2). Specifically, the following proposition can be established (see (Gravel, Marchant, and Sen 2012) for a proof).

Proposition 2 *Let $\succsim_1$ and $\succsim_2$ be two orderings on $\bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$ which can be represented as per (2) for some functions Ψ^1 and Ψ^2 respectively (both having S^{k-1} as a domain and $\mathbb{R}$ as a range). Then $\succsim_1$ exhibits at least as much aversion to inequality of opportunity as $\succsim_2$ as per Definition 8 if and only if there exists some increasing and concave function g having the image of $\mathbb{R}$ under Ψ^2 as domain and $\mathbb{R}$ as range such that, for every $p \in S^{k-1}$, one has $\Psi^1(p) = g(\Psi^2(p))$.*

Hence, for comparisons of societies made by a UEU criterion, the statement "has more aversion to opportunity inequality as" can be translated into "has a more concave evaluation function Ψ as". While this is reminiscent of standard definition in the context of standard inequality measurement, there is an important difference. In the usual income inequality setting, there is a (natural ?) benchmark to define "neutrality to income equality". An ethical observer concerned about distributions of incomes is usually considered as being neutral vis-à-vis income equality if it considers as equivalent all income distributions that have the same per capita income. Given this benchmark, it is standard to define someone has exhibiting aversion to inequality "in the absolute" if this person exhibits more aversion to income inequality than a person who is neutral to inequality. In the current setting, we are not aware of the existence of a well-accepted standard of neutrality toward equality of *opportunities*. One such benchmark could be to consider all societies with the (symmetric) average lottery (with the symmetric average calculated across groups) to be equivalent. Formally, this would amount to define neutrality with respect to equality of opportunities as follows.

Definition 9 *An ordering $\succsim$ on $\bigcup_{l \in \mathbb{N}_{++}} (S^{k-1})^l$ is said to exhibit neutrality with respect* *to equality of opportunities if for any two societies* $\mathbf{p}$ *and* $\mathbf{p}'$ *in* $\bigcup_{l\in\mathbb{N}_{++}}(S^{k-1})^l$ *such that* $\frac{\sum_{i=1}^{n(\mathbf{p})}p_i}{n(\mathbf{p})} = \frac{\sum_{i=1}^{n(\mathbf{p}')}p'_i}{n(\mathbf{p}')}$ *, one has* $\mathbf{p} \sim \mathbf{p}'$.

When applied to a UEU ranking of societies, this definition of neutrality with respect to equality of opportunities implies that the function Ψ that represents such a ranking as per Expression (2) is linear. We state this formally as follows.

Proposition 3 *Let* $\succsim$ *be an ordering on* $\bigcup_{l\in\mathbb{N}_{++}}(S^{k-1})^l$ *that can be represented as per (2) for some functions* Ψ*. Then* $\succsim$ *exhibits neutrality with respect to equality of opportunities if and only if, for every lottery* $p \in S^{k-1}$*, one has* $\Psi(p) = \sum_{j=1}^{k}\beta_j p_j$ *for some real numbers* $\beta_1, ..., \beta_k$*.*

Proof. *TO BE PROVIDED (EASY).* ■

If one agrees with this standard of neutrality with respect to equality of opportunities, then would could define an ethical observer as exhibiting aversion to inequality of opportunities whenever the observer has more aversion to inequality of opportunities than an observer who exhibits neutrality with respect to equality of opportunities.

2.3 Extensions: the empirical case

Now, a society can be seen as a population of N individuals partitioned into n types, each type $i = 1, \ldots, n$ gathers N_i individuals each facing the same distribution of opportunities for skills acquisition.

With ideal data, we would like to infer opportunity distribution for each of the N individuals, thus capturing the full set of implications related to each individual own background. In this case, the welfare model in (**??**) perfectly suits the purpose of assessing inequality of opportunities for skills acquisition, where each parental background would represent a type of family background with its direct influences on the individual distribution of opportunities for skills acquisition.

With less then ideal data, *estimation* of opportunity profile has to be limited to the n types, the N_i individuals in type i serve to estimate the type i opportunity profile. Hence, some heterogeneity is lost in the process and inequality of opportunity comparisons concentrate on fewer types in the population ($n < N$). With replication invariance properties of the welfare ordering, one can always represent a society by a matrix with N rows, the distribution of opportunities attached to each individual in type i being replicated N_i times. Welfare with this matrix evaluated with (**??**) should coincide with welfare evaluated with (1).

I will focus in what follows on the welfare model (1), and show which types of implications we face when instead we want to represent data with model (**??**).

3 Comparing situation on the basis of their social welfare: general case

3.1 Inequality of opportunity reducing transformations

We investigate transformations that, when applied to the data $\mathbf{p}$, give $\mathbf{p}'$. We study transformations that have a clear interpretation in terms of their consequences on inequality of opportunity and that do not produce effects on the population distribution of skills. We consider distribution matrices with $n(\mathbf{p})$ types, where types have variable weights $p_1, p_2, \ldots, p_{n(\mathbf{p})}$. We consider a *population model*: we assume a uniform population which is distributed on $[0, 1]$, implying $p_i \in \mathbb{R}_+$. This assumption is valid and meaningful for a population model, where one can assume an infinity of agents populating the economy (but a finite number of types $n(\mathbf{p})$, which might differ across distributions).[3]

Transformation T1 A *permutation of types* transformation (T1) occurs whenever $\mathbf{p}' = \mathbf{\Pi} \cdot \mathbf{p}$ and $(p'_1, \ldots, p'_{n(\mathbf{p}')}) = (p_1, \ldots, p_{n(\mathbf{p})}) \cdot \mathbf{\Pi}^t$, where Π is a $n(\mathbf{p}) \times \mathbf{p}$ permutation matrix.

[3]We will discuss how the operations can be eventually redrafted when one considers an *empirical model*, that is one is interested in the welfare generated in the underlying sample. Under the assumption that all $N(\mathbf{p})$ individuals receive a uniform sample weight $1/N(\mathbf{p})$, the type i size is $p_i = N_i(\mathbf{p})/N(\mathbf{p})$ which is a rational number. The empirical model sets restrictions on the admissible operations that we can explicitly consider.

Any permutation of types transformation preserves welfare and inequality of opportunities as long as evaluations are independent on the name of the types (but depends on their skills opportunities distribution). In practice, a distribution of skills opportunities associated to a type i is regarded to as "detrimental" to social welfare because it implies high chances that low realizations occurs (for instance, because p_{i1} is very high) rather that because type i is perceived as an unfavorable type (for instance, it represent the most economically deprived parental background).

Transformation T2 An *empty type* transformation (T2) involves adding or deleting an additional type "0" with distribution of opportunities $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ and population weight $p_0 = 0$ so that the matrix $\mathbf{p}'$ coincides with matrix $\mathbf{p}$ augmented by one additional row π and $n(\mathbf{p}') = n(\mathbf{p}) + 1$. Type 0 is a phantom. Hence it should not affect measured welfare or mobility. One implication of the operation is that irrelevant types, i.e. types that are not likely to be observed in a society, do not contribute to welfare and inequality of opportunity.

Transformation T3 A *replication of a types* transformation (T3) is implemented by splitting the population of a type i, with $1, \ldots, i, \ldots, n(\mathbf{p})$ of size p_i into two new types i_1 and i_2 such that all individuals in types i_1 and i_2 face the same distribution of opportunities of group i, that is $p_{ij} = p'_{i_1 j} = p'_{i_2 j}$ and $p_{hj} = p'_{hj}$ for every outcome class j and every individual $h \neq i$, and $n(\mathbf{p}') = n(\mathbf{p}) + 1$. After the transformation, the society $\mathbf{p}'$ displays one additional type gathering individuals that face the same skills opportunities distribution as the individuals in group i (compared to $\mathbf{p}$, the matrix $\mathbf{p}'$ has one additional row which is a replica of row i), and have weights $p'_{i_1} + p'_{i_2} = p_i$, where $p'_{i_1} = \alpha p_i$ with $\alpha \in [0, 1]$. This operation does preserve welfare and inequality of opportunities: it says that if a given type is split according to an irrelevant characteristics (such as splitting the group of people with low parental background into those having parents with light-colored eyes and those having parents with dark-colored eyes), then one expect both new types to face the same opportunity distributions (it is reasonable to assume that parents eye color does not affect one opportunity profile per se). If this is the case, then welfare and inequality of opportunity do not change.[4]

Transformation T4 A *uniform mixture of distributions of skills opportunities* transformation (T4) occurs whenever the distributions of skills opportunities of any two types i and i' are compounded into a new distribution according to the demographic weight of each type, and individuals from both types experience this new distribution of skills opportunities after the transformation. Formally, $p'_{ij} = p'_{i'j} = w_{ii'}p_{ij} + (1 - w_{ii'})p_{i'j}$ with $w_{ii'} = p_i/(p_i + p_{i'})$ and $p_{hj} = p'_{hj}$ for any outcome class j and every types $h \neq i, i'$. After the transformation, both groups i and i' share the same compounded distribution. The intuition behind this transformation is opposite to that of the replication of a type. Consider two types of low economic parental background, one gathers people grown in urban areas (type i), supposedly facing larger opportunities of acquiring skills, and other raised as child in rural areas (type i'). Since being raised in urban or rural areas has implications on skills opportunities, we expect the distributions associated to types i and i' differ. Coherently with the overall model, which assumes that circumstances have a distributional effect on skills but are irrelevant from an efficiency perspective (urban and rural areas differ because opportunities are not equally available in both locations because agglomeration plays an important role in directing public spending and investment on skill-enhancing public goods for a given budget), then if differences across two groups were neutralized (this could be the case if skills-enhancing technologies are made available at the same intensity in urban and rural areas), we expect that differences across groups induced by the additional irrelevant circumstance (living in urban or rural areas) are smoothed and distributions converge to the average distribution.[5] We regard the situation in which distributions converge uniformly to the expected distribution as reducing inequality of opportunity and improving welfare.

This last operation has many attractive features. First, it preserve the population average distribution. Second, it preserve the number and size of the groups. Third, it is consistent with the replication of types operation: if a uniform mixture transformation

[4]The transformation has an appealing interpretation within the population model, where population is uniform distributed within each type.

[5]Example kindergarten

involves two types that differ exclusively for an irrelevant attribute (such as the color of the eyes before, we hence expect both types to face the same skills opportunity distribution), then after the operation they will still share the same distribution of skills.

3.2 Testing the welfare criterion

A Zonotope set associated to any given matrix represents all the vectors that can be reached from the Minkowsky sum of the matrix rows. Consider a *column-stochastic matrix* $\mathbf{A}$ of size $n \times k$ such that each of the k column vectors of $\mathbf{A}$ belongs to the k-variate simplex. Let use $\mathbf{a}_i$ with $i = 1, \ldots, n$ indicate a row vector of the matrix. We define the zonotope set $Z(\mathbf{A}) \in [0,1]^k$ as

$$Z(\mathbf{A}) = \left\{ \mathbf{z} = (z_1, \ldots, z_k) : \ \mathbf{z} = \sum_{i=1}^{n} \theta_i \mathbf{a}_i, \ \theta_i \in [0,1] \ \forall i = 1, \ldots, n \right\}$$

For a given matrix $\mathbf{p}$ representing distributions of opportunities defined over k outcomes for $n(\mathbf{p})$ types (the matrix is hence row stochastic), we denote the $n(\mathbf{p}) \times k$ matrix $\widetilde{\mathbf{p}}$ where $\widetilde{p}_{ij} = p_i p_{ij}$ for any i, j the *joint distribution density* of types and outcomes in the society. Formally:

$$\widetilde{\mathbf{p}} = \begin{pmatrix} \widetilde{p}_{11} & \ldots & \ldots & \ldots & \widetilde{p}_{1k} \\ \ldots & \ldots & \widetilde{p}_{ij} & \ldots & \ldots \\ \widetilde{p}_{n(p)1} & \ldots & \ldots & \ldots & \widetilde{p}_{n(p)k} \end{pmatrix}$$

Here, $\widetilde{p}_{ij}$ is the probability that a person is of type i *and* achieves outcome j. Aggregating the densities across the classes of realizations, one obtains the population marginal weights $\sum_{j=1}^{k} \widetilde{p}_{ij} = p_i$ for any i. The column vector of population weights is simply denoted $p = (p_1, \ldots, p_{n(\mathbf{p})})^t$. Aggregating the densities across types one obtains the marginal distribution of skills opportunities across the population, which coincide with the average distribution of skills across the population: $\sum_{i=1}^{n(\mathbf{p})} \widetilde{p}_{ij} = \overline{p}_j$ for all $j = 1, \ldots, k$.

The focus of this section being on comparisons of distribution matrices that have the same skills margins (but not necessarily the same types margins), we standardize the joint density by $\overline{p}_1, \ldots, \overline{p}_k$, which gives the $n(\mathbf{p}) \times k$ matrix $\widetilde{\mathbf{p}} \cdot diag(\overline{p})^{-1}$. Any element of this matrix writes as $\frac{\widetilde{p}_{ij}}{\overline{p}_j} = \frac{p_i p_{ij}}{\overline{p}_j}$ for any type i and outcome class j, where p_i is the demographic size of group i. Formally:

$$\widetilde{\mathbf{p}}.diag(\overline{p})^{-1} = \begin{pmatrix} \frac{\tilde{p}_{11}}{\overline{p}_1} & \dots & \dots & \dots & \frac{\tilde{p}_{1k}}{\overline{p}_k} \\ \dots & \dots & \frac{\tilde{p}_{ij}}{\overline{p}_j} & \dots & \dots \\ \frac{\tilde{p}_{n(p)1}}{\overline{p}_1} & \dots & \dots & \dots & \frac{\tilde{p}_{n(p)k}}{\overline{p}_k} \end{pmatrix}$$

We interpret $\frac{\widetilde{p}_{ij}}{\overline{p}_j}$ as the conditional probability that a person achieving outcome j is of types i (the construction exploits the Bayes rule). Hence $\widetilde{\mathbf{p}} \cdot diag(\overline{p})^{-1}$ is *column stochastic.*

In the spirit of the Lorenz curve, we study the *Lorenz Zonotope* $LZ \in [0,1]^{k+1}$ of a matrix $\mathbf{p}$, which is a $k+1$ dimensional Zonotopes reporting the conditional groups distributions and the population margin, and is defined as:

$$LZ(\mathbf{p}) := Z\left((p, \widetilde{\mathbf{p}} \cdot diag(\overline{p})^{-1})\right),$$

where

$$\left(p, \widetilde{\mathbf{p}}.diag(\overline{p})^{-1}\right) = \begin{pmatrix} p_1 & \frac{\tilde{p}_{11}}{\overline{p}_1} & \dots & \dots & \dots & \frac{\tilde{p}_{1k}}{\overline{p}_k} \\ p_i & \dots & \dots & \frac{\tilde{p}_{ij}}{\overline{p}_j} & \dots & \dots \\ p_{n(p)} & \frac{\tilde{p}_{n(p)1}}{\overline{p}_1} & \dots & \dots & \dots & \frac{\tilde{p}_{n(p)k}}{\overline{p}_k} \end{pmatrix}$$

The LZ satisfies interesting properties. If opportunities for skills acquisition are independently distributed with respect to individual circumstances (implying that the rows of $\mathbf{p}$ coincide), then the Lorenz Zonotope coincides with the diagonal of cube that contains it.

By construction, the Lorenz Zonoid is always a k-dimensional hyperplan lying on a $k+1$ dimensional space. However, one dimension here is redundant. From $\widetilde{\mathbf{p}}$ one can derive both margins, implying that the population weights distribution p is redundant. Nevertheless, LZ inclusion is equivalent to inclusion of the LZ projections. We produce results based on the LZ and resort to its projection only for graphical/expositional purposes.

3.3 A preliminary result

The first result concerns welfare comparisons of two societies $\mathbf{p}$ and $\mathbf{p}$' that differ in the number and population density of the types (p and $n(\mathbf{p})$ possibly different from p' and $n(\mathbf{p'})$). For each society we study the inequality of opportunity for skills acquisition by focusing on how dissimilar are the distributions of skills opportunities across types. We compare distributions with the same margin, i.e. with fixed population distributions ($\overline{p} = \overline{p}'$). Hence, the welfare measure that we consider are capturing exclusively differences across societies in terms of inequality of opportunity.

Theorem 10 *For any pair of societies represented by opportunity distribution matrices* $\mathbf{p}$ *and* $\mathbf{p}'$ *with* $\overline{p} = \overline{p}'$*, the following statements are equivalent:*

1. *The distribution* $\mathbf{p}'$ *is obtained from* $\mathbf{p}$ *through a finite sequence of transformations T1, T2, T3, T4.*
2. $\sum_{i=1}^{n(\mathbf{p}')} p_i' \phi \left(\sum_{j=1}^{k} u_j p_{ij}' \right) \geq \sum_{i=1}^{n(\mathbf{p})} p_i \phi \left(\sum_{j=1}^{k} u_j p_{ij} \right)$ *for all* ϕ *increasing concave and for any* $u_1, \ldots, u_k$.
3. $LZ(\mathbf{p}') \subseteq LZ(\mathbf{p})$.

Proof. We prove that 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 1). See appendix A.1 ■

Discussion evolves on two separate lines. First, the theorem allows to produce welfare comparisons when the opportunity marginals coincide. When they do not, LZ inclusion can still be tested. However, equivalences with 1) breaks down: T1-T4 preserve the opportunity distribution margin. Also condition 2) has to be refined: as in Koshevoy Mosler, the welfare criterion produce relative evaluations of inequality of opportunity based on the function $\sum_{i=1}^{n(\mathbf{p})} p_i \phi \left(\sum_{j=1}^{k} u_j \frac{p_{ij}}{\overline{p}_j} \right)$, in a similar way the welfare function in KM is based on the distribution of *goods shares* rather than the distribution of goods (given that two allocations may differ in terms of goods endowments). The ratio $\frac{p_{ij}}{\overline{p}_j} = \frac{\widetilde{p}_{ij}}{p_i \overline{p}_j}$ captures the discrepancy between the actual joint density of parental background and opportunity distribution, and a counterfactual distribution obtained under the assumption that parental background circumstances and skills opportunities are independently distributed. The ration is then a measure of associated between skills opportunities and parental background.

By considering relative skills distributions, we can find an upper bound for welfare $\phi(\sum_{j=1}^{k}(u_j))$, which coincides with the case $\frac{\widetilde{p}_{ij}}{p_i \overline{p}_j} = 1$ for all i and j. This number can be used to construct a *relative inequality of opportunity measure* $IOP_{\phi,u} \in [0,1]$ defined over distributions p (population weights), $\widetilde{\mathbf{p}}$ (the joint density) and $\overline{p}$ (the opportunity distribution in the population). The index $IOP_{\phi,u}$ measures inequality of opportunity as a form of distance from a welfare-maximizing allocation of opportunities across types. When welfare is maximal, i.e. $W(\mathbf{p}) = \phi(\sum_{j=1}^{k}(u_j)) = \max W$, inequality of opportunity should be zero. Otherwise the measure should increase with the degree of association of opportunities and parental background. Hence, we can construct a ratio-scale measure of inequality of opportunity:

$$IOP_{\phi,u} = 1 - \frac{W(\mathbf{p})}{\max W} = 1 - \frac{1}{\phi(\sum_{1}^{n(\mathbf{p})} u_j)} \sum_{i=1}^{n(\mathbf{p})} p_i \phi \left(\sum_{j=1}^{k} u_j \frac{\widetilde{p}_{ij}}{p_i \overline{p}_j} \right).$$

Different indices of relative IOP can be crafted upon suitable choices of ϕ and u.[6] Nevertheless, one might be interested in welfare evaluations even when averages differ. We cover this case in section 4.

A second line of discussion concerns the interpretation in cases where the population is finite and countable (of size N) and types weights can be expressed as $p_i = \frac{N_i}{N}$. In these situations, axioms have to be refined slightly, as we motivate in the following section.

[6] If, for instance, $u_j = 2(1 - \frac{j}{k})$ (the Gini social welfare function weights) and ϕ linear, implying $\max W = \phi(\sum_{j=1}^{k} u_j) = k - 1$, then $IOP = \sum_{i=1}^{n(\mathbf{p})} p_i \frac{1}{k-1} \sum_{j}^{k} 2(1 - j/k)\rho_{i(j)} = \sum_{i=1}^{n(\mathbf{p})} p_i (\mu_i(1 - G_i))$, where $\rho_{ij} = \frac{\widetilde{p}_{ij}}{p_i \overline{p}_j}$ and $\rho_{i(j)}$ are arranged in increasing order (implying $\rho_{i(j)} \leq \rho_{i(j+1)}$ for every j), $\mu_i = \sum_{j=1}^{k} \frac{1}{k} \rho_{ij}$ and $G_i = 1 - \frac{2}{\mu_i} \sum_{j=1}^{k} (1 - j/k)\rho_{i(j)}$ is the Gini index of the distribution $\rho_{i1}, \ldots, \rho_{ik}$. According to this parametric choice of ϕ and u, inequality of opportunity for skills acquisition can be measured as an average of type-specific Gini indices, assessing the intensity of association in parental background and skills opportunities.

3.4 The case with discrete (empirical) populations

The ”population model” (continuous population) can be meaningfully adapted to measure social welfare from a sample of data of size N where type $i = 1, \ldots, n(\mathbf{p})$ population is N_i. Individuals are uniformly weighted $\frac{1}{N}$, implying that $p_i = \sum_{h=1}^{N_i} \frac{1}{N} = \frac{N_i}{N}$. In this context, transformations T2, T3 and T4 are unappropriate. We replace them with the equivalent counterparts.

We assume from the outset that data are reported at the individual level: welfare in a sample of $N(\mathbf{p})$ individuals is represented by a matrix $\mathbf{p}$ of size $N \times k$. Each row of the matrix reports a distribution of opportunities faced by an individual i in the society. This distribution is replicated across all individuals belonging to the same type. Each individual weight is $p_i = \frac{1}{N(\mathbf{p})}$.

Transformation T5 A *population replication* transformation (T5) allows to replicate each of the individuals in the distribution α-times, with α an integer positive number larger than one. This operation has implications on the way data are represented. Each replica of individual i shares the same distribution of opportunity as i. After the transformation T5 generating situation $\mathbf{p}'$ from $\mathbf{p}$, the population weights are adjusted so that $p'_i = \frac{1}{\alpha N(\mathbf{p})}$ for every $i = 1, \ldots, \alpha N(\mathbf{p}')$. The matrix $\mathbf{p}$ is extended accordingly: each row is replicated α-times, to obtain a new distribution matrix $\mathbf{p}'$ of size $\alpha N(\mathbf{p}) \times k$. The operation is nevertheless regarded as neutral with respect to social welfare and inequality of opportunity.

Transformation T6 A *uniform transfer of opportunities* transformation (T6) consists of a uniform Pigou-Dalton transfer of probability masses involving the opportunity profiles of two individuals. The intensity of the transfer of probabilities is regulated by a parameter $\alpha \in [0, 1]$, such that T6 can be equivalently described as a compounding of opportunity distributions across two types. When transformation T6 is applied to matrix $\mathbf{p}$ and concerns two individuals i and i', it produces matrix $\mathbf{p}'$ such that $N(\mathbf{p}') = N(\mathbf{p})$, $p'_{ij} = (1-\alpha)p_{ij} + \alpha p_{i'j}$ and $p'_{i'j} = \alpha p_{ij} + (1-\alpha)p_{i'j}$ while $p'_{hj} = p_{hj}$ for all $h \neq i, i'$ and for all $j = 1, \ldots, k$. The operation is such that $\overline{p}'_j = \overline{p}_j$ for any realization class. This transformation corresponds to compounding opportunity profiles across at least two individuals, thus reducing dissimilarities and the intrinsic association between background and opportunity distributions. In the spirit of income inequality measurement, we retain this operation to be welfare improving.[7]

Corollary 11 *For any pair of societies represented by opportunity distribution matrices* $\mathbf{p}$ *and* $\mathbf{p}'$ *with* $\overline{p} = \overline{p}'$*, the following statements are equivalent:*

1. *The distribution* $\mathbf{p}'$ *is obtained from* $\mathbf{p}$ *through a finite sequence of transformations T1, T5, T6.*

2. $\sum_{i=1}^{N(\mathbf{p}')} \frac{1}{N(\mathbf{p}')}\phi\left(\sum_{j=1}^{k} u_j p'_{ij}\right) \geq \sum_{i=1}^{N(\mathbf{p})} \frac{1}{N(\mathbf{p})}\phi\left(\sum_{j=1}^{k} u_j p_{ij}\right)$ *for all* ϕ *increasing concave and for any* $u_1, \ldots, u_k$.

3. $LZ(\mathbf{p}') \subseteq LZ(\mathbf{p})$.

Proof. We prove that 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 1). See appendix A.2. ∎

In what follows, we discuss and demonstrate extensions of the model to the case of variable means within the framework of the corollary.

4 Comparing situation on the basis of their social welfare: increasing utilities

Theorem 10 offers equivalent criteria to test a very restrictive notion of social welfare. Some meaningful restrictions however apply. For instance, outcomes categories $1, \ldots, k$ are arranged in increasing order, implying that well-being comparisons should be limited to all utilities $u_1, \ldots, u_k$ such that $u_j \leq u_{j+1}$. This additional requirement has three implications for the results in Theorem 10.

[7] Any T6 operation can be equivalently reformulated through a T-transform matrix, justifying the reference to the Pigou-Dalton transfers in the multidimensional settng. Following Andreoli Zoli, it can be shown that every transformation T6 can be seen as the outcome of a specific sequence of transformations involving T2, T3 and T4. This result is consistent with the informational loss that applies to the current setting, where individual weights are fixed to $1/N$.

First, it provides an additional restriction to be satisfied by the welfare criterion in statement 2) of the theorem, implying less demanding welfare criterion for ranking societies. It can in fact be the case that there exist pairs of societies that fail to be ranked according to all vectors $u_1, \ldots, u_k$, but can still be ranked by those vectors displaying increasing well-being coefficients.

The second implication is that the underlying concept of welfare is now sensitive to *improvements in opportunities* transformation (T7). This consists in obtaining matrix $\mathbf{p}'$ from $\mathbf{p}$ by transporting a small probability mass ε assigned to outcome j by the opportunity distribution of individual i towards outcome $j+1$, such that $p'_{ij} = p_{ij} - \varepsilon$, $p'_{ij+1} = p_{ij+1} + \varepsilon$ and $p'_{ij} = p_{ij}$ in all other cases. Any such operation induces a *first order stochastic dominance* improvement in individual i opportunities in matrix $\mathbf{p}'$ compared to $\mathbf{p}$, implying an improvement in social welfare. The implications of the improvement in individual is opportunities are reflected in the improvement of the average opportunity distribution $\overline{p}'$ which stochastic dominates $\overline{p}$.[8]

Consistently, the third implication is that LZ inclusion cannot be used as a valid empirical test for the restricted welfare model. LZ inclusion criterion has to be weakened, we propose to do so by studying the inclusion of what we call the *Generalized Lorenz Zonotoep*. We discuss the implications of weakening the welfare dominance criterion in the simple case with just two outcomes, which allows to pin down the different steps behind the construction of a generalized LZ. We then provide the result within the general setting.

4.1 Example: The case with two outcomes

We consider distributions of opportunities for bad (1) and good (2) outcomes. In the examples, we consider distributions of opportunities for N individuals represented by $N \times 2$ matrices. For some of these individuals, distributions of opportunities can coincide. We are interesting in ranking matrices $\mathbf{p}$ and $\mathbf{p}'$ when the population distributions of

[8] An improvement in opportunities produce efficiency gains and in the population distribution even when they increase inequality of opportunity. On the contrary, changes in the degree of association of parents background and opportunities produce welfare consequences only when they reduce underlying inequality of opportunities.

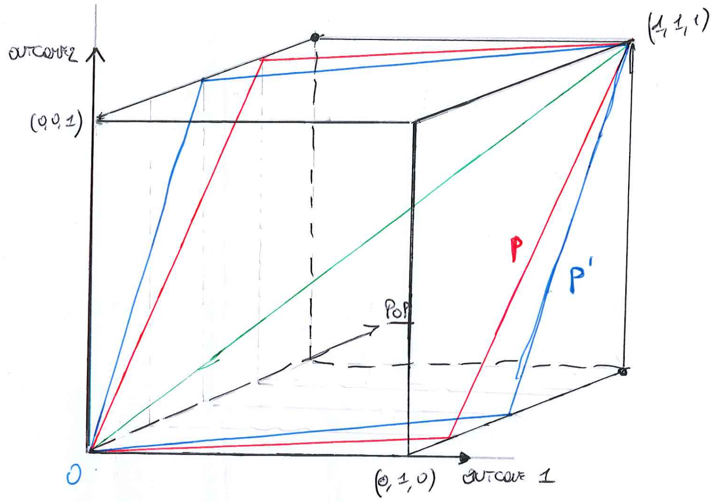


Figure 2: LZ for matrices $\mathbf{p}$ (blue) and $\mathbf{p}'$ (red).

opportunities, $\overline{p}$ and $\overline{p}'$ may differ.

Within this setting, let assume first that distribution $\mathbf{p}$ and $\mathbf{p}'$ are such that $N(\mathbf{p}) = N(\mathbf{p}') = 4$, and

$$\mathbf{p} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \qquad \mathbf{p}' = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix},$$

such that $\overline{p} = (\frac{1}{2}, \frac{1}{2})$ and $\overline{p}' = (\frac{1}{4}, \frac{3}{4})$. The Lorenz Zonotopes of these matrices are as in figure **??**. Both Lorenz Zonotopes lie on a separate hyperplane, for the reasons explained in the examples reported in the appendix. However, inclusion of Lorenz Zonotopes cannot be claimed as the two intersect along the hypercube diagonal.[9]

Projections of the LZ on the $k = 2$ dimensional square generated by distributions of outcome 1 and outcome 2 would coincide, thus failing to register this overlapping. This happens because the two distribution display different population distributions of opportunities, with $\overline{p}'$ stochastic dominating $\overline{p}$. To incorporate these differences in evaluation,

[9]In fact, the population weights (here constant and equal to $\frac{1}{4}$) can be obtained by a linear combination of the columns of the matrices $\mathbf{p}$ and $\mathbf{p}'$ respectively. In the respective cases, the weights should coincide with population distribution of opportunities $\overline{p}$ and $\overline{p}'$. The fact that $\overline{p} \neq \overline{p}'$ implies the differences in slopes of the hyperplane supporting each LZ.

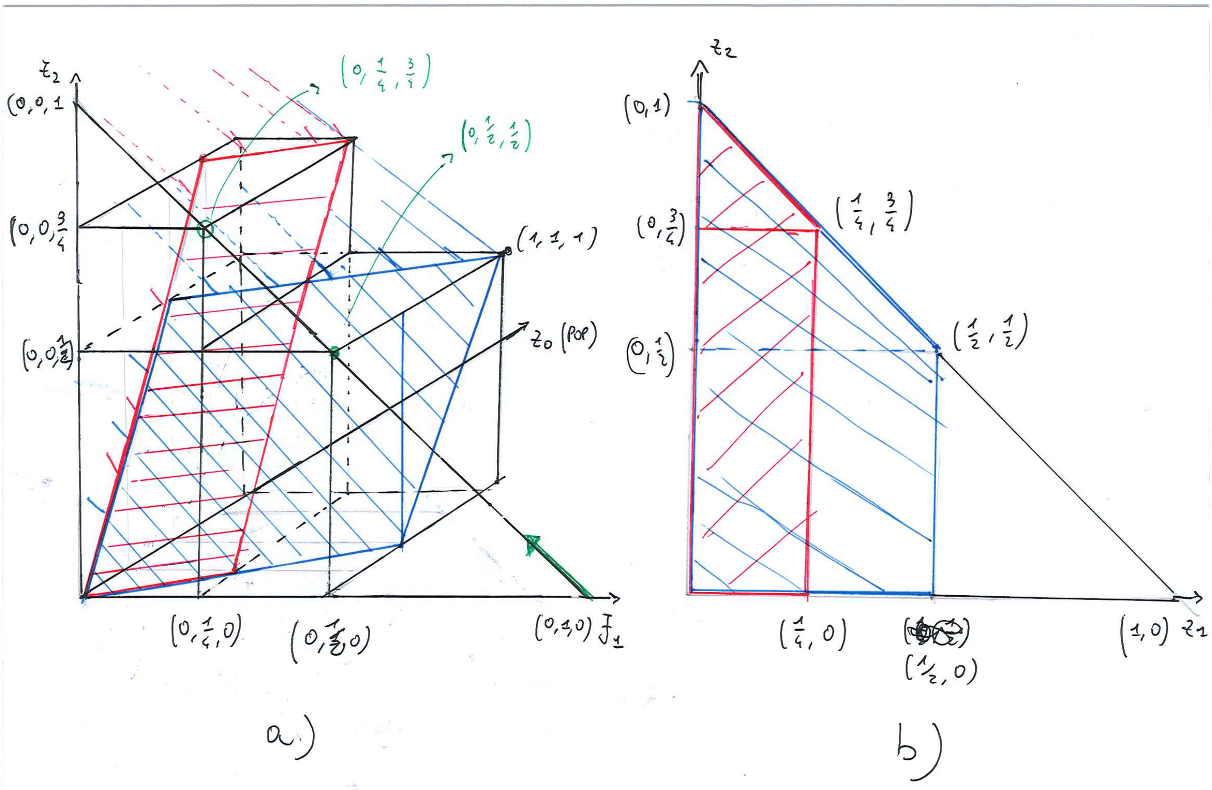


Figure 3: The unscaled LZ (z_0 is the population ordinate, z_1, z_2 are for bad and good outcomes) of $\mathbf{p}$ (blue) and $\mathbf{p}'$ (red) along with the extensions towards all distributions that are dominating in the sense of SD1, panel a), and their projection on the outcome axis.

we represent the data by the *Unscaled Lorenz Zonotopes* $Z((p, \widetilde{\mathbf{p}}'))$ and $Z((p, \widetilde{\mathbf{p}}))$, where $p = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. The original hypercube containing the LZ is now stretched to reach size $(1, \overline{p}_1, \overline{p}_2) = (1, \frac{1}{2}, \frac{1}{2})$ for $\mathbf{p}$ and $(1, \overline{p}'_1, \overline{p}'_2) = (1, \frac{1}{4}, \frac{3}{4})$ for $\mathbf{p}'$. Panel a) in figure 3 reports such zonotopes.

The diagonal line connecting the extremes of the axis reporting the zonotope coordinates z_1 and z_2 represent all admissible combinations of population distribution probabilities $\overline{p}_1$ and $\overline{p}_2$ such that $\overline{p}_1 + \overline{p}_2 = 1$. What the picture intuitively shows is that moving north-west welfare increases due to (first order SD) improvements in the population distribution. Welfare deteriorates moving south-east, unless the utility weights are such that $u_1 \geq 0$ and $u_2 \leq 0$, a case that clearly violates the underlying assumption of increasing utility weights ($u_1 \leq u_2$).

We then consider the possibility of constructing an *Generalized Lorenz Zonotope* based on the knowledge of $(p, \widetilde{\mathbf{p}})$ distribution and on the assumption of increasing weights. The generalized LZ, $GLZ \in [0, 1]^3$, extends in additive form the zonotope $Z((p, \widetilde{\mathbf{p}}))$ towards the set of points where there is clear sign of welfare dominance (i.e. to all points on the $2 + 1$ space that can be reached from any point $\mathbf{z} \in Z((p, \widetilde{\mathbf{p}})))$. This set is identified in different colors on the graph and is identified as follows:

$$\begin{aligned} GLZ(\mathbf{p}) &= Z((p, \widetilde{\mathbf{p}})) + (\mathbb{R}_+ \times \mathbb{G}_2), \\ \mathbb{G}_2 &= \{(-1, 1)z : \ z \in \mathbb{R}_+\}. \end{aligned}$$

As the graph shows, we now conclude that $GLZ(\mathbf{p}') \subseteq GLZ(\mathbf{p})$ as somehow expected: in fact, $\mathbf{p}'$ is obtained from $\mathbf{p}$ by an improvement in the situation of individual $i = 3$, which leads to an unambiguous increase in social welfare, despite both matrices display cases where inequality of opportunity is maximal.

The two zonotopes $Z((p, \widetilde{\mathbf{p}}))$ and $Z((p, \widetilde{\mathbf{p}}'))$ lie again on hyperplanes (in two dimensions). Contrary to the LZ case, however, the two zonotopes now lie on the *same* hyperplane. They can be hence projected on the $k = 2$ dimensional space (z_1, z_2) without loss of information. Their projections are marked with solid bold lines of different colors on the picture in panel $b)$ of figure 3. On a similar vein, projections of $GLZ(\mathbf{p}')$ and $GLZ(\mathbf{p})$

can also be represented on the $k = 2$ dimension space. These correspond to the perimeters in the same figure as above. As expected, the projection of $GLZ(\mathbf{p}')$ is included in that of $GLZ(\mathbf{p})$, and it turns out to be a sufficient test for Generalized Lorenz Zonotope inclusion.

Using projections on $k = 2$ dimension space allows to represent the same information as GLZ while bringing evident representational advantages. In figure 4 we report graphs corresponding to different situations where dominance is and is not satisfied. In panel a), we compare two situations displaying same population distributions of opportunities (they lie in the same square) but different degrees of inequality of opportunities. The projection of GLZ of the case displaying equal opportunities (the lower bound coinciding with the line segment connecting the origin to the point $(\overline{p}_1, \overline{p}_2)$) welfare dominates the other case.

In panel b), we compare projections of GLZ for two situations displaying perfect equality of opportunity but different population distributions of opportunities. The line segment connecting the origin with $(\overline{p}'_1, \overline{p}'_2)$ lies above segment connecting the origin with $(\overline{p}_1, \overline{p}_2)$, consistently with Generalized Lorenz Zonotope inclusion. In fact, $(\overline{p}'_1, \overline{p}'_2)$ first order stochastic dominates $(\overline{p}_1, \overline{p}_2)$ implying higher welfare.

In panel c) it is shown that a situation displaying inequality of opportunity can welfare-dominate another distribution displaying equality of opportunity, provided that the improvement in the distribution of opportunities of the dominating distribution is large enough. This implies that the opportunity distribution of the worse-off individual in the dominant situation should be considered an improvement over $(\overline{p}_1, \overline{p}_2)$.

Finally, panel d) shows that GLZ inclusion may be violated even when two situations have very similar distributions of population opportunities but display significantly different levels of inequality of opportunity.

The logic of GLZ inclusion test, as shown by the projections, is that of the GL curve. Parallel to this result, we can show that GLZ inclusion is a necessary and sufficient test for welfare dominance with increasing utility prizes associated to outcome lotteries. We produce a unified result for the k-dimensional case. Intermediate cases with $k = 3$ and

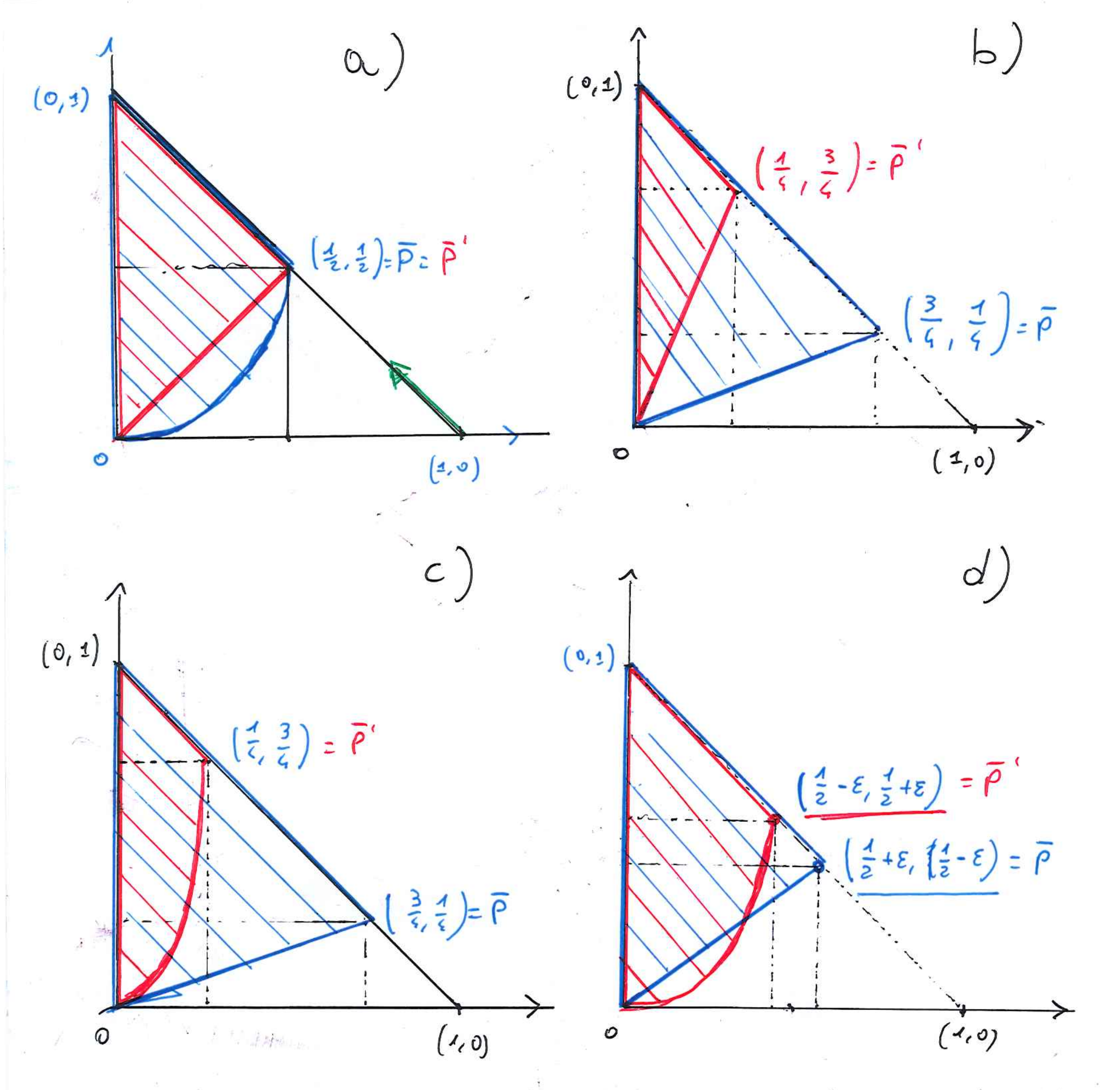


Figure 4: Comparisons of GLZ projections for situations that differ in average opportunities distribution and in inequality of opportunity.

$k = 4$ are discussed in appendix B.

4.2 Main result

We regard an improvement in opportunity distributions in term of first order stochastic dominance as a natural candidate for producing increments in social welfare for all increasing well-being functions associated to the classes of realizations. Furthermore, stochastic dominance is well defined even when outcomes have only an intrinsically ordinal meaning.

Lorenz Zonotopes can be extended in a way consistent with the notion of stochastic dominance. The Generalized Lorenz Zonotope provide representations of distributions of opportunities within and across types of the population that reflect both the extent of inequality of opportunity (i.e., dissimilarity between distributions across types) as well as (first order SD) improvements in population opportunities (i.e., account for differences in the welfare attribute to having opportunities equally distributed as the average population distributions $\overline{p}$).

As intuitively motivated in the case $k = 2, 3, 4$, the GLZ can be fully characterized on the basis of the underlying improvements in opportunities that induce stochastic dominance. The row vector $\boldsymbol{\gamma}^j = (\gamma_1^j, \ldots, \gamma_j^j, \ldots, \gamma_{k-1}^j)$ of size $k-1$ such that $\gamma_j^j = -1$, $\gamma_{j+1}^j = 1$ and $\gamma_h^j = 0$ for all $h \neq j, j+1$, identifies the direction of changes in the distribution of opportunities due to an improvement in probabilities from class j to $j+1$. Compounding all these vectors, one obtains all possible improvements in the distribution $\overline{p}$ which give distribution $\overline{p}'$ that stochastic dominate $\overline{p}$. The GLZ at k dimensions is extended accordingly, which gives:[10]

$$\begin{aligned} GLZ(\mathbf{p}) &= Z((p, \widetilde{\mathbf{p}})) + (\mathbb{R}_+ \times \mathbb{G}_k), \\ \mathbb{G}_k &= \left\{ \sum_{j=1}^{k-1} z_j \boldsymbol{\gamma}^j : \ (z_1, \ldots, z_{k-1}) \in \mathbb{R}_+^{k-1} \right\}. \end{aligned}$$

The main result of the paper is obtained for empirical distributions of welfare where

[10]The GLZ can be compared to the definition of Extended LZ in Koshevoy and Mosler, where the set of directions of the extensions is extended to $\mathbb{G}_k = \mathbb{R}_+^k$.

individual have uniform weight and there are as much opportunity distribution in the empirical population as individuals, although some individuals belonging to the same type may face the same opportunity distribution. We refer to the empirical population distribution weights by the vector $p = (\frac{1}{N}, \dots, \frac{1}{N})^t$.[11]

We show that the inclusion test based on the empirical GLZ provides a necessary and sufficient statistics for robust welfare evaluations of empirical opportunity distributions even when comparing distributions $\mathbf{p}$ and $\mathbf{p}'$ that differ in the marginal distributions of empirical populations ($N(\mathbf{p}') \neq N(\mathbf{p})$) and in the distributions of opportunities in the population ($\overline{p}' \neq \overline{p}$). We also show that welfare dominance is supported by the existence of simple transformations of the data that have a compelling interpretation: either they preserve of decrease inequality of opportunity, or they have an unambiguous implications for opportunities improvements.

Theorem 12 *For any pair of societies represented by opportunity distribution matrices* $\mathbf{p}$ *and* $\mathbf{p}'$, *the following statements are equivalent:*

1. *The distribution* $\mathbf{p}'$ *is obtained from* $\mathbf{p}$ *through a finite sequence of transformations T1, T5, T6, T7.*

2. $\sum_{i=1}^{N(\mathbf{p}')} \frac{1}{N(\mathbf{p}')} \phi \left(\sum_{j=1}^{k} u_j p'_{ij} \right) \geq \sum_{i=1}^{N(\mathbf{p})} \frac{1}{N(\mathbf{p})} \phi \left(\sum_{j=1}^{k} u_j p_{ij} \right)$ *for all* ϕ *increasing concave and for any* $u_1, \dots, u_k$ *increasing (i.e.,* $u_j \leq u_{j+1}\ \forall j$*).*

3. $GLZ(\mathbf{p}') \subseteq GLZ(\mathbf{p})$.

Proof. We prove that 1) ⇒ 2) ⇒ 3) ⇒ 1).

1) ⇒ 2) is mechanical. 2) ⇒ 3) follows if the GLZ is specified in the right way. We think the actual representation serves the scope. 3) ⇒ 1) is by construction. You need to show that GLZ can be written in a convenient mathematic form, and that show that inclusion has implications that are compatible only with the transformations in 1). ■

[11] Referring to this framework simplifies exposition of results. Moreover, the setting is consistent with the idea of using GLZ inclusion tests on the data.

5 Empirical Application

The aim of this paper is to assess the effect of educational systems on inequality of oppotunity, where oppotunity is assumed to be determined - or at least affected - by parental background. This requires that we follow a pupil cohort over time, so that we may evaluate how inequality of opportunity widens or is reduced by the education system, as the cohort advances along the grade years.

In this section, we will first explain how the database was constructed by merging 3 surveys: PISA 2009, TIMSS 2007 and TIMSS 2003. Then we will proceed to explain how the empirical results were obtained by applying our criteria to the data. And finally we will show and comment the results.

5.1 The Data

In this paper we have chosen to study the pace at which different education systems widen inequality of opportunity along grade years. This requires that one should compare the same birth cohort in different countries, and at different points in the education system. Thus, one needs to have a database with not only multiple countries, but also multiples grade years.

Unfortunately, although educational data which follow a cohort of students over time do exist, these tend to focus on a specific country. The British Household Panel Survey (BHPS) and the Panel Study of Income Dynamics (PSID) are two examples of panel data which provide information on education, but only for the UK and the US. Using a variety of country specific databses - each one constructed autonomously - would arguably produce questionable results, and would raise cross-country comparability issues.

Conversely, many large scale assessment surveys have been developed on education since the mid-1990s which include multiple countries, but they focus exclusively on a specific grade year or age group. The Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA) are the most prominent examples of such large scale assessment surveys.

In order to construct a database with a birth cohort assessed at different grade years, and in multiple countries, we have chosen to merge 3 databases: TIMSS 2003, TIMSS 2007 and PISA 2009. The approach we use here is consistent with **AIGUL PAPER + CHECCHI PAPER**

In this section we will present each one of the 3 databases, and then we will explain how they were merged together into a cohort database.

5.1.1 Large Scale Assessment Surveys

The construction of any large scale assessment survey on education poses a number of challenges: some of which concern multi-country surveys in general, others of which are specific to educational surveys. The former kind - applicable to all international surveys - concern the ability of TIMSS or PISA to produce unbiased and comparable results across countries: First, how can one make sure that a sampled population is representative of the actual country it accounts for? Second, how does one ensure that the test results are comparable across all coutries? And third, how could one insure that the questions in the tests do not artificially favour some countries over others?

To meet any issues relating to sample-representativeness, TIMSS and PISA both resort to a two-stage sampling process. This implies first randomly selecting a number of schools within a country, and then randomly choosing pupils within the selected schools. In particular, this adresses any concerns of possible selections biases, which might have occured in the absence of random selection. Then, to correct for any possible shortcomings of the process' ability to produce a sample which exactly reflects the country's population, both TIMSS and PISA proceed to re-weight the pupils in such a way as to fit the actual population of their respective countries.

The issues relating to the comparability of the results across all countries in the TIMSS and in the PISA surveys, is solved by producing test sheets which are identical for all countries. This ultimately leads to the third point, concerning the exact nature of the test, which should be designed to measure educational skills in a meaningful way, and without favouring some countries over other countries.

Contrary to the PISA surveys, TIMSS studies have developed surveys which are based on what is considered to be fundamental concepts in Maths and Science, and which are taught as part of all national curricula. In the maths tests for instance, typical questions included solving simple problems or computing basic arithmetics. In no case were children tested on specific skills, excluded from their national curriculum. This means that no nation's curriculum could prove more suited for maximizing the TIMSS tests scores. In contrast, the PISA surveys have developed tests which are not curricula-based, and which instead measure the ability of children from different countries to apply Science and Maths concepts to reel life situations. Typical questions from the PISA Maths tests require the use of basic arithmetic to solve reel life problems. Thus the main difference between the PISA tests and the TIMSS tests, is that the former supposedly requires slightly more abstract thinking, whereas TIMSS requires more formal mathematical knowledge. We argue further in the paper that the actual skills needed to complete either of the tests are not substantially different.

Finally, TIMSS and PISA - like any other educational survey - have to deal with the problem of grading the tests. Whereas in many education systems grading is equated with computing the percentage of correct answers over the total number of questions, TIMSS and PISA rely instead on what is called *Item Response Theory (IRT)*. Commonly used in Psychology research, IRT has been invented to facilitate the statistical analysis of multiple answer sheets, by including what can effectively be assimilated to an error terms. Thus IRT does not merely grade an answer sheet, but instead can provide the probability distribution of an individual's grade. The need for IRT stems from the fact that an answer sheet helps to make an estimation of an individual's skills, but it cannot be used as an absolute evaluation of his skills. Indeed, the factors affecting an individual's answers include his skills, but also his motivation and Luck. Extremely poor answer sheets for instance are likely to have been plagued by low motivation, thereby providing an underestimate of the individual's actual skills. Conversely, extremely accurate answer sheets are more likely to have been affected by positive luck or cheating, thereby providing a likely overestimation of the individual's skills. Bearing that in mind, it follows that the error terms increase as one moves towards the tails of the distribution.

One of the most widespread mathematical models of IRT - and which is used in both TIMSS and PISA - is the Rasch Model. In both surveys, one is thus provided with 5 plausible values computed using the Rasch Model, and which corresponds to 5 likely evaluations of the individual's skills, and based on the aggregation of his correct and incorrect answers to the questions. As one might expect, the 5 plausible values (each one varying between 0 and 1,000) are increasing with the number of correct answers, and decreasing with the number of wrong answers. One might also note that the variance between the 5 plausible values of an individual, is increasing as he moves towards to tails of the distribution, since tail estimations have higher errors terms.

Having explained how the TIMSS and PISA surveys manage to produce internationally comparable results, we will now proceed to detail the specific contents of the 2003 and 2007 waves of the TIMSS surveys, and then we will detail the contents of the PISA 2009 survey.

5.1.2 TIMSS 2003 and TIMSS 2007

The first TIMSS study - which was conducted in 1995 - was at the time the largest international survey on education ever conducted. It measured the Science and Maths abilities of children from 40 different countries at 5 separate grade years. Ever since, TIMSS has been releasing new surveys every 4 years (1999, 2003, 2007, 2009, 2015), and has come to focus on the measurement of children's skills in grades 4 and 8, thus choosing to drop grades 3, 7 and 12. The geographical scope of the surveys has been broadening over the years, with the number of countries tested in the surveys growing from 40 in 1995 to 57 countries in 2015.

Of particular interest to this paper, is also the increasing quality and quantity of information on the background of the pupils. Variables such as the immigration status of each parent and the child himself, can prove useful at the very least as control variables, and will help cancel out the effects of having a large share of immigrants on national test results. Other variables on parental background - such as education, social status and

Country	Mean Math Score	Mean Parental Education	Mean number of books	Obervations
Armenia	456	NA	86	4828
Australia	499	NA	116	4321
Cyprus	510	NA	72	4191
Hong Kong	575	NA	55	4557
Hungary	529	NA	95	3200
Iran	389	NA	35	4080
Italy	503	NA	65	4229
Japan	565	NA	67	4507
Latvia	536	NA	99	3580
Lithuania	534	NA	61	4165
Moldova	504	NA	49	3830
Netherlands	540	NA	89	2878
New Zealand	493	NA	99	4200
Norway	451	NA	100	4167
Philippines	358	NA	36	4419
Russia	532	NA	80	3929
Singapour	594	NA	79	6587
Slovenia	479	NA	83	2988
Tunisia	339	NA	40	3502
USA	518	NA	87	9628
Taiwan	564	NA	82	4642
Yemen	278	NA	37	3193

Table 1: TIMSS 2003 - 4^{th} Graders

number of books at home - are indispensable to this paper since they will be used as the basis on which to measure and compare equality of opportunity.

The specific databases of interest in this paper are the 4^{th} graders in TIMSS 2003, and the 8^{th} graders in TIMSS 2007 - both of which account for the same cohort observed at a 4-year interval. As one can see in tables 1 and 2, there are respectively 95.621 fourth graders across 22 countries in TIMSS 2003, and 205.230 eighth graders across 47 countries in TIMSS 2007. Crucial to this paper, there are 16 countries which are observed in both TIMSS 2003 and TIMSS 2007, namely Armenia, Australia, Cyprus, Hong Kong, Hungary, Iran, Italy, Japan, Norway, Russi, Singapore, Slovenia, Tunisia and the USA. This will enable us to compare how these 16 country-based cohorts have evolved over the 4 grade years that separates them.

Country	Mean Math Score	Mean Parental Education	Mean number of books	Obervations
Algeria	387	NA	31	5339
Armenia	499	NA	88	4581
Australia	496	NA	110	3995
Bahrain	398	NA	71	4194
Bulgaria	464	NA	97	3966
Bosnia	456	NA	36	4191
Botswana	364	NA	38	4136
Columbia	380	NA	34	4863
Cyprus	465	NA	83	4377
Czech Republic	504	NA	89	4834
Dubai	461	NA	73	2744
Egypt	391	NA	40	6438
El Salvador	340	NA	32	4044
England	513	NA	92	3964
Georgia	410	NA	92	4112
Ghana	309	NA	38	5176
Hong Kong	572	NA	60	3449
Hungary	517	NA	117	4103
Indonesia	397	NA	28	4119
Iran	403	NA	39	3959
Israel	463	NA	103	3216
Italy	480	NA	101	4408
Japan	570	NA	88	4278
Jordan	427	NA	60	5161
Malaysia	474	NA	51	4452
Malta	488	NA	994	4650
Morocco	380	NA	48	2896
Mongolia	432	NA	28	4237
Norway	469	NA	114	4561
Oman	372	NA	62	4654
Palestine	367	NA	48	4292
Qatar	307	NA	82	7082
Romania	461	NA	61	4177
Russia	512	NA	95	4461
Saudi Arabia	329	NA	54	4118
Singapore	593	NA	80	4589
Slovenia	501	NA	78	4019
South Korea	597	NA	122	4236
Sweden	491	NA	117	5065
Syria	395	NA	44	4562
Scotland	487	NA	78	4022
Thailand	441	NA	35	5397
Tunisia	420	NA	36	3971
Turkey	432	NA	47	4469
Ukraine	462	NA	80	4412
USA	508	NA	91	7261

Table 2: TIMSS 2007 - 8^{th} Graders

5.1.3 PISA 2009

In 2000 the OECD mandated the first competing version of a large scale assessment survey on education, which would come to be named PISA. It looked to assess the Math, Science and Reading abilities of children aged 15 across 32 countries - including 28 OECD coutries and 4 non-OECD countries. Contrary to TIMSS, its primary focus was to develop an international large scale assessment survey which would not be curriculum-based, and which would instead measure the abilities of children to utilize their Maths, Science and Reading abilities in real-life situations.

Since the first survey in 2000, PISA has been releasing new waves every 3 years (2003, 2006, 2009, 2012 and 2015). The latter waves have been including greater numbers of countries and increasing the amount of detailed information on the pupils (immigration status, language spoken at home, material possessions at home,...). This - as with the TIMSS databases - will prove helpful to single out the effects of the education system, by controling for the factors out of the education system's responsability. PISA also provides increasingly detailed information on the parental background of the pupils (Parental Education, Parental Social Status, and number of books at home, ...), which is crucial to construct our inequality of opportunity criteria.

The PISA 2009 database - on which will rely our empirical application - surveys 515.958 pupils aged 15 accross 73 countries. Crucial to this paper, one can see that there are 15 countries which are surveyed in both TIMSS 2003 and PISA 2009: Australia, Hong Kong, Hungary, Italy, Japan, Latvia, Lithuania, Netherlands, New Zealand, Norway, Russia, Singapore, Slovenia, Tunisia and the United States. 24 countries out of the 73 countries in PISA 2009, are also observed in TIMSS 2007: Australia, Bulgaria, Colombia, Czech Republic, Georgia, Hong Konk, Hungary, Indonesia, Israel, Italy, Japan, Jordan, Malaysia, Norway, Qatar, Romania, Russia, Singapore, Slovenia, South Korea, Sweden, tunisia, Turkey and Sweden. This implies that we will be able to study how equality of opportunity evolves along the grade years of the different education systems.

Country	Mean Math Score	Mean Parental Education	Mean number of books	Obervation
Albania	377	12	52	4596
Azerbaijan	431	13,6	63	4691
Argentina	388	11,9	74	4774
Australia	514	12,8	178	14251
Austria	496	13,1	149	6590
Belgium	515	13,5	137	8501
Brazil	386	10,1	41	20127
Bulgaria	428	13,5	124	4507
Canada	527	14,5	163	23207
Chile	421	12,1	74	5669
Shanghai-China	600	12,6	116	5115
Chinese Taipei	543	12,6	138	5831
Colombia	381	10,7	45	7921
Costa Rica	409	11,2	47	4578
Croatia	460	13,4	84	4994
Czech Republic	493	13,2	145	6064
Denmark	503	13,8	139	5924
Estonia	512	13,9	173	4727
Finland	541	14,7	164	5810
France	497	12,2	135	4298
Georgia	379	13,4	160	4646
Germany	513	12,4	162	4979
Greece	466	13,5	135	4969
Hong Kong-China	555	10,7	93	4837
Hungary	490	12,7	191	4605
Iceland	507	15,5	191	3646
India	349	9,2	39	4826
Indonesia	371	9,8	56	5136
Ireland	487	12,8	144	3937
Israel	447	12,7	154	5761
Italy	483	12,7	139	30905
Japan	529	13,6	153	6088
Kazakhstan	405	14,1	94	5412
Jordan	387	12,7	69	6486
Korea	546	13,6	184	4989
Kyrgyzstan	331	13,1	57	4986

Table 3: PISA 2009 - Pupils aged 15 (1/2)

Country	Mean Math Score	Mean Parental Education	Mean number of books	Oberva
Latvia	482	13,5	146	450
Liechtenstein	536	13,1	173	32
Lithuania	477	13,3	115	45
Luxembourg	489	13,3	194	46
Macao-China	525	10,2	64	59
Malaysia	404	13	80	49
Malta	463	11,7	171	34
Mauritius	420	12,2	93	46
Mexico	419	10,7	50	382
Republic of Moldova	397	13,5	66	51
Montenegro	403	12,7	119	48
Netherlands	526	13,5	136	47
New Zealand	519	11,9	165	46
Norway	498	13,6	183	46
Panama	360	10,8	53	39
Peru	365	11,3	50	59
Poland	495	11,9	128	49
Portugal	487	10,7	113	62
Qatar	368	13,9	133	90
Romania	427	13,1	103	47
Russia	468	13,2	143	53
Serbia	442	13,2	90	55
Singapore	562	10,9	124	52
Slovak Republic	497	13,2	118	45
Slovenia	501	12,8	120	61
Spain	483	11,9	162	258
Sweden	494	13,2	192	45
Switzerland	534	13,6	144	118
Thailand	419	9,6	66	62
Trinidad and Tobago	414	11,9	145	47
United Arab Emirates	421	13,2	107	108
Tunisia	371	10,5	41	49
Turkey	445	8,6	81	49
United Kingdom	492	13	144	121
United States	487	13,4	123	52
Uruguay	427	10,9	76	59
Miranda-Venezuela	397	11,8	98	290

Table 4: PISA 2009 - Pupils aged 15 (2/2)

5.2 Creating a birth Cohort

In order to assess how education systems widen inequality of opportunity along grade years, one must evaluate inequality of opportunity at different points in the education system. Fortunately, PISA and TIMSS measure children's skills at 3 different points in the education system, albeit using different methodologies. In this subsection, we will first explain why TIMSS and PISA skills are sufficiently comparable for the analysis of Inequality of opportunity. Then we will explain why we chose the particular waves of PISA 2009, TIMSS 2007 and TIMSS 2003. And lastly, we will justify why we chose to construct an age-based cohort rather than a grade-based cohort.

One might argue that TIMSS and PISA measure different sets of skills which are not comparable to one another: the former measuring curriculum-based skills, the latter the ability for pupils to use their Maths in real-life situations. However, as shown in Rindermann (2007), the results obtained by countries at PISA tests are strongly correlated with their results at grade 8 TIMSS tests. At the country level, the results obtained in PISA 2003 and TIMSS 2003 for 8^{th} graders for instance, have a correlation of 0.88. In any case, should there be any scepticism as regards the comparability of TIMSS and PISA results, this would concern one's ability to comment on the evolution of scores between TIMSS and PISA. However, in this paper we will not study how scores evolve, but rather how the parental background of the children affects their scores throughout the education system. This implies that we will only be concerned with the stickiness between parental background and children's abilities, as opposed to the children's achievement *per se*.

Having put to rest the possible qualms relating to the comparability of PISA and TIMSS results, there still remains the issue of selecting the appropriate waves of PISA and TIMMS in order to build a cohort. Since TIMSS grade 8 and grade 4 are separated by a 4-year grade gap, if one chooses to rely on a TIMSS survey in year n for the fourth graders, then in order to assess the same cohort at a different point in time, one must additionally use the TIMSS wave of year $n + 4$ for the eighth graders. Luckily, as TIMSS surveys are released every 4 years, this simply implies taking one wave for the fourth graders, and the next wave for the eighth graders.

To compare the same cohort further along the education system, one can rely on PISA data. However, the children assessed in PISA being on average 2 years older than in TIMSS grade 8, one must use a wave which is two-year remote from the TIMSS survey. Unfortunately, PISA releasing one survey every 3 years and TIMSS every 4 years, a two year gap between the two only occurs every 12 years (2009, 2021, . . .). This means that – until PISA 2021 is released – the only possibility for constructing a cohort observed at 3 points along the education system using TIMSS and PISA data, is with PISA 2009, TIMSS 2007 and TIMSS 2003.

Furthermore, in order to be able to merge the 3 survey-waves into a cohort database, it is necessary that some countries participate in at least in 2 of the 3 surveys. Indeed, any country which is exclusive to either PISA 2009, TIMSS 2007 or TIMSS 2003 will be useless for a cohort analysis. Out of all the countries which are surveyed in either of the 3 surveys, table 5 shows that there are 34 countries that are surveyed at least twice - 11 of which appear in all 3 surveys. This implies that whereas 23 countries will be compared at 2 points in the education system, Australia, Hong Kong, Hungary, Italy, Japan, Norway, Russia, Singapore, Slovenia , Tunisia and the USA will be assessed in 2003, 2007 and 2009 for a given cohort.

Lastly, the basis on which to construct the cohort is possibly less straightforward than it first seems. Whereas TIMSS assesses children at specific grade years (i.e. 4 and 8), PISA focuses on the specific age group of pupils aged 15 years old. Both surveys also provide the grade year and age of each pupil. Thus, one of the first issues in order to create a cohort (which we will observe at different points along the education system), is on which *basis* to construct the cohort.

Considering the information provided by these databases, one could chose to construct either an age-based cohort or a grade-based cohort. The first would require keeping only the children born in 1992/1993 - thus excluding all children who were in grade 4 in 2003 or in grade 8 in 2007, but who were not born in 1992/1993 - and the second would require keeping only the children who were in grade 4 in 2003, in grade 8 in 2007 and in grade 10 in 2009 - thereby excluding all pupils in the PISA 2009 database who are not in grade 10.

Due to children getting held back one or several years - and in some cases children skipping years - along the education system, one finds that the share of children who are not in the expected grade year at age 15 is quite sizeable. Thus excluding the children born in the right year, but at the wrong grade year in PISA 2009, would significantly bias our results. Conversely, upon reaching grade 4 and grade 8, since there are fewer children having gotten held back, there fewer children whose age and grade year do not match. Thus, we have chosen to exclude the pupils in grade 4 and grade 8 who were not born in 1992/1993 from the TIMSS database. We have therefore constructed what is effectively an age-based cohort rather than a grade-based cohort.

6 Conclusions direction of future research

To do list:

- Proof of main theorem.
- verify IOP indices and their possible use in empirical analysis.
- If we feel the need of introducing the restriction u_j is increasing and concave, then I guess we should look at generalized zonotopes that are consistent with 1st and 2nd order stochastic dominance. I think we can work on the simplex to derive the set of admissible extensions based on the relevant notion of dominance.
- If we value concavity, we should also be willing to attribute an intrinsic meaning to the outcome classes. In the example of skills, outcomes are skills intervals (for instance, delimited by population quantiles?) and the outcome function (the u_j in out case) can be earning associated to these classes. For instance, one can associate to the x decile of skills the average earnings of people in that decile of skills.
- We should discuss relations with existing literature. Most of EOP literature focuses on opportunities for earnings acquisition. As above, we can use u_j to measure outcomes in a given class.

Country	TIMSS 2003 - Grade 4	TIMSS 2007 - Grade 8	PISA 2009
Armenia	✓	✓	✗
Australia	✓	✓	✓
Bulgaria	✗	✓	✓
Cyprus	✓	✓	✗
Colombia	✗	✓	✓
Czech Republic	✗	✓	✓
Georgia	✗	✓	✓
Hong Kong	✓	✓	✓
Hungary	✓	✓	✓
Indonesia	✗	✓	✓
Iran	✓	✓	✗
Israel	✗	✓	✓
Italy	✓	✓	✓
Japan	✓	✓	✓
Jordan	✗	✓	✓
Korea	✗	✓	✓
Latvia	✓	✗	✓
Lithuania	✓	✗	✓
Malaysia	✗	✓	✓
Malta	✗	✓	✓
Moldova	✓	✗	✓
Netherlands	✓	✗	✓
New Zealand	✓	✗	✓
Norway	✓	✓	✓
Qatar	✗	✓	✓
Romania	✗	✓	✓
Russia	✓	✓	✓
Singapore	✓	✓	✓
Slovenia	✓	✓	✓
Sweden	✗	✓	✓
Thailand	✗	✓	✓
Tunisia	✓	✓	✓
Turkey	✗	✓	✓
USA	✓	✓	✓

Table 5: Country presence in the 3 surveys

- When ϕ is linear and $u_1 = -1$ while $u_j = 0$ if $j \neq 1$, then welfare is the sample average of the intensity of occurrence of the minimum outcome in any given distribution. This principle is very much in line with the "mean of mins" principle by Roemer.
- When ϕ is extremely concave, reflecting a max-min welfare evaluation, the sum of evaluations ϕ across individuals can be formalized with the function $\min_i\{\sum_j u_j p_{ij}\}$. If $u_j = \mu_j$, the average earnings attached to the skills class j, then $\sum_j u_j p_{ij} = \sum_j \mu_j p_{ij} = \mu^i$, the average earnings of the type individual i belongs to. Hence, social welfare give prevalence to the average earnings of the more disadvantaged type, consistently with a "min of mean" criterion a la Van de gaer.
- Other criteria can be constructed accordingly.

- Statistical treatment of zonotope inclusions. Ideally, one would like to estimate zonotopes for the population distribution using information about the sample distribution.

- GLZ is developed building on a parallel with the generalized Lorenz curve. Can we develop further on this?

- GLZ inclusion can be a very demanding criterion. Indeed, welfare dominance for all ϕ increasing concave and any u increasing is a very strong requirement. It might make sense to use only specific functions ϕ, such as $\phi = \mu(1 - I)$ where μ is the average evaluation of each type opportunity profile and I the implied inequality. Alternatively, we can check dominance only when $u_j = \mu_j$, the average earnings in skills class j. Can we derive meaningful restriction of the class of welfare functions in the main theorem by looking at projections or integral (volumes) of the GLZ?

A Proofs

A.1 Proof of Theorem 10

Proof. We prove that 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 1).

1) $\Rightarrow$ 2). Linearity of the welfare function implies that transformations T1, T2 and T3 do not change welfare. Concavity of ϕ implies that any T4 transformation cannot increase welfare.

2) $\Rightarrow$ 3). Consider vectors $p = (p_1, \ldots, p_{n(\mathbf{p})})$ and $p' = (p'_1, \ldots, p'_{n(\mathbf{p}')})$ and a third vector p^* of size $n \geq n(\mathbf{p}'), n(\mathbf{p})$ with $p^* = (p^*_1, \ldots, p^*_n)$ such that $p^*_i \in [0,1]$ and $\sum_i p^*_i = 1$. Let the vector p^* constructed in such a way that its elements can be related to the elements of p and p' as follows. Denote $\mathcal{I} = \{h\}_{h=1}^n$ the set of natural numbers indicating elements of p^*. For all $i = 1, \ldots, n(\mathbf{p})$ exists $\mathcal{I}(i) \subseteq \mathcal{I}$ satisfying $1 \leq |I(i)| \leq n$, $\bigcup_{i=1}^{n(\mathbf{p})} \mathcal{I}(i) = \mathcal{I}$ and $\mathcal{I}(i) \cap \mathcal{I}(i') = \emptyset$ for any $i \neq i'$, such that $p_i = \sum_{h \in \mathcal{I}(i)} p^*_h$ and equivalently for p'. Since every individual in type $h \in \mathcal{I}(i)$ faces the same distribution of opportunities $(p_{i1}, \ldots, p_{ik})$, i.e. $p_{hj} = p_{ij}$ $\forall j$ for all $h \in \mathcal{I}(h)$, welfare in society $\mathbf{p}$ can be represented as $\sum_{i=1}^{n(\mathbf{p})} \sum_{h \in \mathcal{I}(i)} p^*_h \phi(\sum_{j=1}^k u_j p_{ij}) = \sum_{i=1}^{n^*} p^*_i \phi\left(\sum_{j=1}^k u_j p'_{ij}\right)$, and equivalently for $\mathbf{p}'$. Statement 2) hence implies $\exists p^*$ such that $\sum_{i=1}^{n^*)} p^*_i \phi\left(\sum_{j=1}^k u_j p_{ij}\right) \geq \sum_{i=1}^{n^*} p^*_i \phi\left(\sum_{j=1}^k u_j p_{ij}\right)$ for all ϕ increasing concave and for any $u_1, \ldots, u_k$. Following **?**, the condition is equivalent to $\mathbf{p}'$ *Generalize Lorenz dominates* $\mathbf{p}$ for any $u_1, \ldots, u_k$. GL dominance should be verified at any proportion $q \in [0,1]$ of the population distribution. Let $\theta_i(q) \in [0,1]$, $i = 1, \ldots, n$ be such that $\sum_{i=1}^n \theta_i(q) p^*_i = q$. For a given choice of $u_1, \ldots, u_k$, we define lorenz dominance as follows (**?**):

$$\min_{\theta'_1(q), \ldots, \theta'_n(q)} \sum_{i=1}^n \theta'_i(q) p^*_i \sum_{j=1}^k u_j p'_{ij} \geq \min_{\theta_1(q), \ldots, \theta_n(q)} \sum_{i=1}^n \theta_i(q) p^*_i \sum_{j=1}^k u_j p_{ij}, \quad p \in [0,1]. \tag{4}$$

We show that this condition induce zonotopes inclusion. Statement 2) in the theorem implies that for any vector $u = (u_1, \ldots, u_k) \in \mathbb{R}^k$ and for any $q \in [0,1]$ there exist $\theta_i(q), \theta'_i(q) \in [0,1]$, $i = 1, \ldots, n$ satisfying $\sum_{i=1}^n \theta_i(q) p^*_i = \sum_{i=1}^n \theta'_i(q) p^*_i = q$ such that (4) can be equivalently stated as:

$$u\cdot(\sum_{i=1}^{n}\theta_i'(q)p_i^*p_{i1}',\dots,\sum_{i=1}^{n}\theta_i'(q)p_i^*p_{ik}')^t\geq u\cdot(\sum_{i=1}^{n}\theta_i(q)p_i^*p_{i1},\dots,\sum_{i=1}^{n}\theta_i(q)p_i^*p_{ik})^t,\quad p\in[0,1],$$

where t stands for transpose. The inequality can be equivalently formulated as:

$$(u,1)\cdot(\sum_{i=1}^{n}\theta_i'(q)p_i^*p_{i1}',\dots,\sum_{i=1}^{n}\theta_i'(q)p_i^*p_{ik}',\sum_{i=1}^{n}\theta_i'(q)p_i^*)^t\geq \tag{5}$$

$$(u,1)\cdot(\sum_{i=1}^{n}\theta_i(q)p_i^*p_{i1},\dots,\sum_{i=1}^{n}\theta_i(q)p_i^*p_{ik},\sum_{i=1}^{n}\theta_i(q)p_i^*)^t,\quad p\in[0,1], \tag{6}$$

More generally, there is an infinity of vectors $v_q'=(\sum_{i=1}^n\theta_i'(q)p_i^*p_{i1}',\dots,\sum_{i=1}^n\theta_i'(q)p_i^*p_{ik}',\sum_{i=1}^n\theta_i'(q)p_i^*)$ and $v_q=(\sum_{i=1}^n\theta_i(q)p_i^*p_{i1},\dots,\sum_{i=1}^n\theta_i(q)p_i^*p_{ik},\sum_{i=1}^n\theta_i(q)p_i^*)$ corresponding to sequences of coefficients $\theta_i(q),\theta_i'(q)\in[0,1]$, $i=1,\dots,n$ satisfying $\sum_{i=1}^n\theta_i(q)p_i^*=\sum_{i=1}^n\theta_i'(q)p_i^*=q$. Let denote these vectors by sets $\mathcal{V}_q$ and $\mathcal{V}_q'$ for situations $\mathbf{p}$ and $\mathbf{p}'$ respectively. Building on the fact that a column vector $v\in\mathbb{R}^k$ belongs to the convex hull of a of a set of vectors $v_1,\dots,v_n$ if and only if $u\cdot v\geq\min_{1\leq i\leq n}u\cdot v_i$ for any row vector $u\in\mathbb{R}^k$, we conclude that condition (6) is satisfied if and only if

$$\mathcal{V}_q'\subseteq conv\mathcal{V}_q,\ \ q\in[0,1]. \tag{7}$$

For a given q, the set $\mathcal{V}_q$ identifies the intersection of an hyperplane in $\mathbb{R}^{k+1}$ with slopes $(0,1,\dots,1)$ and the zonotope $Z\left((p,\widetilde{\mathbf{p}})\right)$ associated to the situation $\mathbf{p}$ (and similarly for $\mathbf{p}'$). Hence (7) is equivalent to

$$Z\left((p',\widetilde{\mathbf{p}'})\right)\ \ \subseteq\ \ Z\left((p,\widetilde{\mathbf{p}})\right). \tag{8}$$

The inclusion implies that for any $\mathbf{z}'\in Z\left((p',\widetilde{\mathbf{p}'})\right)$ there exists $\theta_1,\dots,\theta_n$ with $\sum_{i=1}^n\theta_i=1$ such that $\mathbf{z}'=\sum_{i=1}^n\theta_i(p_i^*,\widetilde{p}_{i1},\dots,\widetilde{p}_{ik})$, which implies

$$\mathbf{z}'\cdot diag((1,\overline{p}))^{-1}=\sum_{i=1}^{n}\theta_i(p_i^*,\widetilde{p}_{i1},\dots,\widetilde{p}_{ik})\cdot diag((1,\overline{p}))^{-1}.$$

Hence (8) is equivalent to $Z\left((p^*, \widetilde{\mathbf{p}'^*} \cdot diag(\overline{p})^{-1})\right) \subseteq Z\left((p^*, \widetilde{\mathbf{p}^*} \cdot diag(\overline{p})^{-1})\right)$. The fact that $Z\left((p^*, \widetilde{\mathbf{p}'^*} \cdot diag(\overline{p})^{-1})\right) = Z\left((p, \widetilde{\mathbf{p}} \cdot diag(\overline{p})^{-1})\right)$ (and similarly for $\mathbf{p}'$) implies statement 3).

3) $\Rightarrow$ 1). Condition 3) implies that $Z(\widetilde{\mathbf{p}'} \cdot diag(\overline{p}')^{-1}) \subseteq Z(\widetilde{\mathbf{p}} \cdot diag(\overline{p})^{-1})$. We can follow Andreoli Zoli to show that Zonotope inclusion implies the existence of a finite sequence of transformations T1-T4. ■

A.2 Proof of Corollary 11

Proof. We prove that 1) $\Rightarrow$ 2) $\Rightarrow$ 3) $\Rightarrow$ 1)..

1) $\Rightarrow$ 2). Linearity of the welfare function implies that transformations T1 and T2 do not modify welfare. Any operation T6 modifies the initial distribution of individual expected utilities $\sum_j u_j p_{ij}$ for every $i = 1, \ldots, N(\mathbf{p})$ into $\sum_j u_j p'_{ij} = \sum_j u_j((1-\alpha)p_{ij} + \alpha p_{i'j}) = \sum_j u_j p_{ij} + \alpha \sum_j u_j(p_{i'j} - p_{ij})$ and $\sum_j u_j p'_{i'j} = \sum_j u_j((1-\alpha)p_{i'j} + \alpha p_{ij}) = \sum_j u_j p_{i'j} - \alpha \sum_j u_j(p_{i'j} - p_{ij})$, while $\sum_j u_j p'_{hj} = \sum_j u_j p_{hj}$ for every $h \neq i, i'$. It holds that $\sum_j u_j p'_{ij} \geq (<) \sum_j u_j p_{ij}$ and $\sum_j u_j p'_{i'j} < (\geq) \sum_j u_j p_{i'j}$ if and only if $\sum_j u_j(p_{i'j} - p_{ij}) \geq (<) 0$, implying a transfer of implicit well-being from the individual with larger $\sum_j u_j p_{ij}$ to that with smaller level. Concavity of the function ϕ in statement 2) picks up this effects and register higher welfare in $\mathbf{p}'$ compared to $\mathbf{p}$.

2) $\Rightarrow$ 3). Follows from Koshevoy 1995

3) $\Rightarrow$ 1). It can be demonstrated (lengthy, non diffuclt) from Theorem 1. ■

B GLZ inclusion: Cases with 3 and 4 outcomes

Consider now a case with $k = 3$ outcomes: a bad ($j = 1$), an average ($j = 2$) and a good ($j = 3$) outcome. GLZ are now defined on the $k + 1 = 4$ dimensional space. Even in this case, we can assess inclusion by studying the GLZ projections on the $k = 3$ space. The GLZ of a distribution $\mathbf{p}$ can be constructed in this framework by first defining the set of population distributions $\overline{p}' = (\overline{p}'_1, \overline{p}'_2, \overline{p}'_)$ that first order stochastic dominate the reference distribution $\overline{p} = (\overline{p}_1, \overline{p}_2, \overline{p}_3)$.

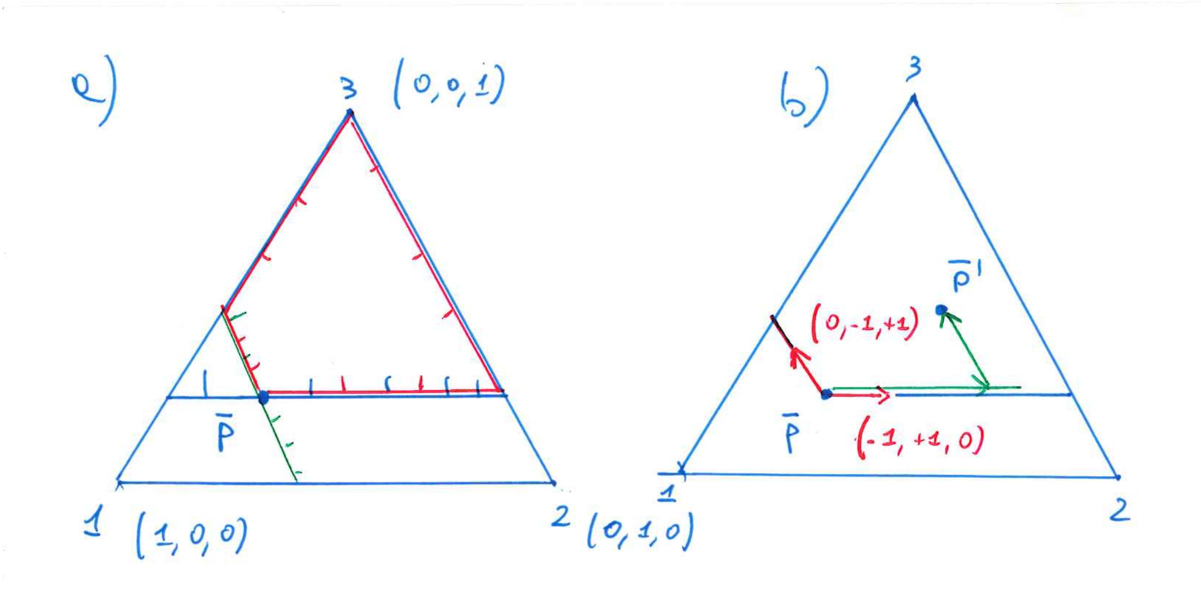


Figure 5: Identification of the set of distributions that first order stochastic dominates $\overline{p}$ on the $k = 3$ simplex.

To verify stochastic dominance, one has to check that $\overline{p}'_1 - \overline{p}_1 \leq 0$ and that $\overline{p}'_1 + \overline{p}'_2 - (\overline{p}_1 + \overline{p}_2) \leq 0$. We identify the set of distributions satisfying these constraints on the simplex represented in figure 5, reporting probability masses of groups 1, 2 and 3 at the extremes. Panel a) of the figure reports one of such cases. In the figure, the green line parallel to the simplex line segment connecting $(0, 1, 0)$ and $(0, 0, 1)$ displays the lower bound of the set of all configurations where the constraint $\overline{p}'_1 - \overline{p}_1 \leq 0$ is satisfied. On a similar vein, the blue line parallel to $(1, 0, 0)$ and $(0, 1, 0)$ connecting classes 1 and 3 displays all vectors $\overline{p}'$ such that $\overline{p}'_1 + \overline{p}'_2 - (\overline{p}_1 + \overline{p}_2) \leq 0$. The intersection of the two sets, represented in red, gathers all situations satisfying both conditions at once.

Elements in this set provide possible directions toward which the LZ can be extended, representing all situations in which a clear improvement in opportunities occurs and distributions are the ranked accordingly. Hence, the LZ extension follows any direction implied by points lying on the set of dominant distributions on the simplex with respect to $\overline{p}$. As shown in panel b), the set of points in the set of distributions that first order stochastic dominate $\overline{p}$ is spanned by two vectors with directions $(-1, +1, 0)$ and $(0, -1, +1)$. For instance, the sum of the two vectors represented in blue on the figure, when properly scaled gives vector $\overline{p}'$.

The directions implied by each of the two vectors $(-1,+1,0)$ and $(0,-1,+1)$ is that of a sequence of improvements. The improvement from outcme level $j=1$ to $j=3$, for instance, is $(-1,0,+1)$ and can be simply obtained as $(-1,+1,0)+(0,-1,+1)$. This offers an intuitive way of representing the Generalized Lorenz Zonotope of the matrix $\mathbf{p}$ (with $n(\mathbf{p})$ individuals and $k=3$ outcomes) on the $k+1=4$ dimensions:

$$\begin{aligned} GLZ(\mathbf{p}) &= Z((p,\widetilde{\mathbf{p}}))+(\mathbb{R}_+\times\mathbb{G}_3)\,, \\ \mathbb{G}_3 &= \{z_1(-1,+1,0)+z_2(0,-1,+1):\ (z_1,z_2)\in\mathbb{R}^2_+\}. \end{aligned}$$

Similarly to the case of the GLZ with $k=2$, even in this case the extension of the GLZ obtain by extending each point of the zonotope $Z((p,\widetilde{\mathbf{p}}))$ in any possible direction consistently with the limitation imposed by the set $\mathbb{G}_3$, which identifies all situations that displays stochastic dominance in the respective population distributions.

Building on these arguments, we are able to identify the set supporting the GLZ even when $k=4$. In this case, increments imply moving mass from class $j=1$ to $j=2$, or from $j=2$ to $j=3$ or from $j=3$ to $j=4$. Any of these transfers, or linear combination of them, guarantee stochastic dominance. The Generalized Lorenz Zonotope hence writes:

$$\begin{aligned} GLZ(\mathbf{p}) &= Z((p,\widetilde{\mathbf{p}}))+(\mathbb{R}_+\times\mathbb{G}_4)\,, \\ \mathbb{G}_4 &= \{z_1(-1,+1,0,0)+z_2(0,-1,+1,0)+z_3(0,0,-1,+1):\ (z_1,z_2,z_3)\in\mathbb{R}^3_+\}. \end{aligned}$$

The case $k=4$ is interesting because the simplex for distributions defined over four outcomes can be represented by a tetrahedron with unitary edges. We use the figure to motivate that the three vectors identifying the set $\mathbb{G}_4$ are indeed sufficient to identify any population distribution $\overline{p}$ that first order stochastic dominate any given distribution $\overline{p}$. One such simplex is reported in panel a) of figure 6. We use $1,\dots,4$ to identify vertices, representing lotteries that put the mass into one specific outcome. We consider the point $\overline{p}$, which falls within the simplex.

The first condition guaranteeing that $\overline{p}'$ stochastic dominates $\overline{p}$ is that $\overline{p}'_1-\overline{p}_1\leq 0$.

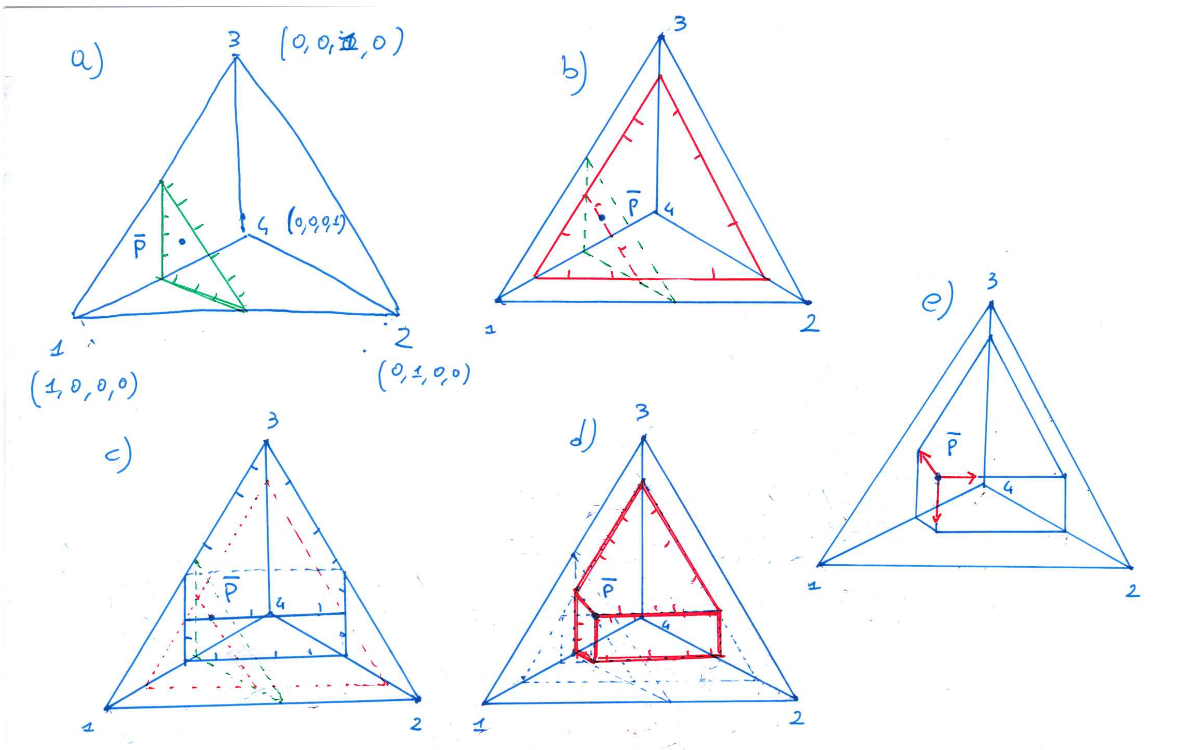


Figure 6: Identification of the set of distributions that first order stochastic dominates $\overline{p}$ on the $k = 4$ simplex.

The set of distributions satisfying this condition is identified by the green hyperplane containing $\overline{p}$ in the figure. Similarly, we identify conditions that guarantee $\overline{p}'_1 + \overline{p}'_2 - (\overline{p}_1 + \overline{p}_2) \leq 0$ by the blue hyperplane in panel c) of the figure, and of $\overline{p}'_1+\overline{p}'_2+\overline{p}'_3-(\overline{p}_1+\overline{p}_2+\overline{p}_3) \leq 0$ by the red hyperplane in panel b) of the same figure. The three conditions need to hold simultaneously to guarantee stochastic dominance, hence the set of admissible population distribution is identified by the intersection of the simplex subsets identified in panels a), b) and c). The intersection generates the set of admissible directions that guarantee to extend any point of the LZ towards stochastic dominance (panel d)). Even in this case, the vectors with directions $(-1, +1, 0, 0)$, $(0, -1, +1, 0)$ and $(0, 0, -1, +1)$, marked in red in panel e), allow to identify any point belonging to $\mathbb{G}_4$ in the figure, when properly scaled.

References

Blackorby, C., W. Bossert, and D. Donaldson (2005): *Population Issues in Social Choice Theory*. Cambridge University Press, Cambridge UK.

Gravel, N., T. Marchant, and A. Sen (2011): "Comparing Societies with Different Numbers of Individuals on the Basis of their Average Advantage," in *Social Ethics and Normative Econonomics: Essays in Honour of Serge-Christophe Kolm*, ed. by M. Fleurbaey, M. Salles, and J. A. Weymark, pp. 261–277. Springer Verlag.

——— (2012): "Uniform Utility Criteria for Decision Making under Ignorance or Objective Ambiguity," *Journal of Mathematical Psychology*, 56, 297–315.