Sur la dissemblance et l'égalisation des chances

Francesco Andreoli

Université de Cergy-Pontoise

Année 2012

N° attribué par la bibliothèque

|_|_|_|_|_|_|_|_|_|_|

THESE

pour l'obtention du grade de

docteur de l'Université de Cergy-Pontoise

en cotutelle avec l'Université de Vérone

Discipline : Sciences économiques

Présentée et soutenue publiquement

par

Francesco ANDREOLI

le 6 décembre 2012

Titre :

SUR LA DISSEMBLANCE ET L'EGALISATION DES CHANCES

Codirecteurs de recherche :

M. Arnaud LEFRANC
professeur à l'Université de Cergy-Pontoise

M. Eugenio PELUSO
professeur à l'Université de Vérone

JURY

M. Gorkem CELIK,	professeur à l'ESSEC Business School *Examinateur*
M. Daniele CHECCHI,	professeur à l'Université de Milan *Examinateur*
M. Michel LE BRETON,	professeur à l'Université de Toulouse I *Président*
M. Arnaud LEFRANC,	professeur à l'Université de Cergy-Pontoise *Codirecteur*
M. Eugenio PELUSO,	professeur à l'Université de Vérone *Codirecteur*
M. Alain TRANNOY,	directeur d'études à l'EHESS *Rapporteur*
M. Dirk VAN DE GAER,	professeur à l'Université de Gand *Rapporteur*

Remerciements

Je tien tout d'abord à remercier le professeur Arnaud Lefranc et le professeur Eugenio Peluso. Ils ont accepté de codiriger cette thèse et ont fait un remarquable travail d'encadrement, en démontrant toujours leur disponibilité et leur ouverture intellectuelle. J'ai eu la chance d'être tout d'abord étudiant et ensuite coauteur du professeur Arnaud Lefranc. Sa rigueur, tant sur le fond que sur la forme, a contribué à élever mes exigences quant à mes travaux de recherche. Le professeur Eugenio Peluso a éveille mon intérêt pour l'économie des inégalités et l'analyse, dans une perspective scientifique et rigoureuse, des problèmes sociaux. Je suis très honoré d'avoir bénéficié des conseils, des efforts et du temps que mes deux codirecteurs de recherche m'ont consacrés. Cela m'a toujours encouragé à me montrer à la hauteur de leur investissement.

Je remercie également les membres du jury : le professeur Michel Le Breton, le professeur Gorkem Celik et le professeur Daniele Checchi, pour l'honneur qu'ils m'ont témoigné en faisant partie de cette commission. Je suis très reconnaissant aux professeurs Alain Trannoy et Dirk Van de gaer d'avoir accepte d'être les rapporteurs de cette thèse.

Toute ma gratitude va au professeur Claudio Zoli, non seulement pour ses enseignements, sa disponibilité et son encadrement mais, au delà, pour son amitié et pour l'exemple qu'il a constitué au niveau tant professionnel que personnel. Nos longues séances de travail ont été agréables, et de véritables occasions de croissance intellectuelle.

La rédaction de cette thèse a bénéficié de l'accord de cotutelle entre l'Université de Vérone et l'Université de Cergy-Pontoise. J'ai commencé mon travaux de recherche au sein du Département de Sciences Economiques de l'Université de Vérone, où j'ai trouvé un environnement très dynamique, ouvert à la recherche et à l'internationalisation. Je remercie toute l'équipe et aussi son directeur, le professeur Federico Perali.

C'est au THEMA que j'ai été chaleureusement accueilli pour continuer mes recherches. J'y ai bénéficié de conditions de travail exceptionnelles et de moyens très importants. Je tiens à remercier tous mes collègues au sein du laboratoire et son directeur, le professeur Olivier Donni, ainsi que le professeur Régis Renault. L'ESSEC Business School a été un excellent partenaire pour l'échange et le développement de mes idées. Que le professeur Raymond Thietart, doyen du programme doctoral, en soit remercié.

Enfin, tous les échanges entre Vérone et Cergy n'auraient pas été possibles sans la contribution accordée par l'Università Italo Francese/Université Franco Italienne au sein du Bando Vinci 2010.

Je remercie tous les amis et collègues thésards à Vérone et à Cergy. Ma reconnaissance va à Céline Lecavelier, Romain Legrand, Anca Matei et Davide Romelli pour leur précieux et généreux travaille de lecture de cette thèse. Je me sens redevable envers la famille Magnani, Nicolas Pelet et Marjorie Trocq pour m'avoir aidé et accueilli quand j'en ai eu besoin.

Mes pensées vont aussi à ma famille, elle a été toujours proche et prête à me soutenir, surtout avant le début de cette thèse. Le mots me manquent pour exprimer ma gratitude à Silvia Salisburgo, ma source d'inspiration et de motivation. À elle, qui a supporté héroïquement le fardeau des longues périodes de pèlerinage transalpin, je veux dédier cette thèse.

Table of Contents

List of Figures

List of Tables

Introduction

Equality of opportunity has gained popularity, in scholarly debates as well as among policymakers, for defining the relevant equalitarian objective for the distribution and redistribution, among individuals, of a broad range of social and economic outcomes. Public policy may "level the playing field" through different types of interventions that may directly affect the process of opportunity formation. Alternatively, the same objective can be reached by redistributing resources and, eventually, income so that many important dimensions of the individual well-being, such as education, health and life chances, are equalized across groups of the population. Equality of opportunity concerns in policy intervention are, for instance, the triggering motivations of the *Europe 2020* policy agenda.[1] This agenda is focused on the promotion of social inclusion across the European Union labor market, the public education system and in the social life, and it guides the policymaker decisions towards equality of opportunity objectives. This is the view of Atkinson and Marlier, who assert that "an inclusive society is one that rises above differences of race, gender, class, generation and geography to ensure equality of opportunity regardless of origin" (Atkinson and Marlier 2010b, p. 3).

In their report, Atkinson and Marlier stress the importance of providing the policymaker with tools that allow to incorporate equality of opportunity concerns in the evaluation of the policymaker's activity. The Europe 2020 report prescribes the guidelines for the policymaker, and suggests some indicators (known as the *Laeken social inclusion indicators*, see Atkinson and Marlier 2010a) that are supposed to quantify the impact of policy intervention. European policymakers are required to shape policies towards these objectives, while

[1] For a detailed description of the Europe 2020 agenda, see Atkinson and Marlier (2010a).

maintaining the freedom to choose the type of, and the budget allocated to, policy intervention. This long list of indicators puts the emphasis on the individual outcomes realizations, and on the individual position in the social and economic hierarchy, while overlooking the differences between outcomes distributions across groups of the population. Although the Europe 2020 indicators may call for intervention on the causes of social exclusion, they cannot be used alone to judge the effect of the policy itself, as "social exclusion is not only a matter of ex post trajectories, but also of ex ante expectations" (Atkinson 1998, p. 14).

These concerns on the evaluation of policy intervention in the light of the European Union agenda undoubtedly motivate the definition of an appropriate methodology for measuring and comparing changes in equality of opportunity. In fact, equality of opportunities appears to be the relevant ethical foundation for social inclusion arguments. This is particularly so because of the relative nature of the concept. In fact, the condition of being socially excluded involves comparisons between individuals, and often this is a manifestation that affects groups and not just individuals. Therefore, assessing the degree of social inclusion between different ethnic groups requires, in this respect, to verify whether the distributions of outcomes across these groups are sufficiently similar.

This thesis approaches the normative evaluation of equality of opportunity objectives by building on the equality of opportunity theoretical and empirical literature (discussed in Fleurbaey 2008, Ramos and Van de gaer 2012). In line with these approaches, normative evaluations should involve comparisons of conditional distributions of outcomes between groups of individuals. In particular, *equalization of opportunity* criteria are formalized, analyzed and implemented by measuring the differences between conditional distributions of groups. These methods can be used, for instance, to asses the welfare implication of the Europe 2020 agenda implementation.

Scholars in economics have reached agreement on a very ideal notion of what is equality of opportunity, that has been introduced in the philosophical literature (Dworkin 1981a, Dworkin 1981b) and formalized by Roemer (1998). Given a well defined set of morally irrelevant circumstances, partitioning a population into social groups, equality of opportunity holds whenever the income distributions associated to these groups coincide. This definition rules out concerns on overall inequality of outcomes, and focuses the attention on between groups comparisons. This demanding notion of equality of opportunity may well define the final objective guiding the policymaker intervention, although it can be hardly exploited to assess empirically the impact of the intervention. In fact, it is very likely that the equality of opportunity criterion is not satisfied before *and* after policy intervention.

This observations, along with the necessity to find tools to evaluate the policymaker intervention, leave space for debate on one important question: How to evaluate the distributional impact of a policy concerned with equalization of opportunities?

This thesis shows that the evaluation of policy intervention motivated by equality of opportunity concerns always relies on *dissimilarity* comparisons between social and economic outcomes distributions. Two or more distributions are dissimilar whenever they can be distinguished one from the others. This very general definition entails two key aspect of dissimilarity that are exploited in this thesis: first, the dissimilarity is a multidimensional concept; second, the notion of dissimilarity does not require that comparisons across distributions are made with respect to some well defined reference distribution. Equalization of opportunity is grounded on a multidimensional comparison of (the dissimilarity between) outcome distributions made conditional upon the background of origin of individuals, and only these distributions should matter for assessing the changes in equality of opportunity.

The issue of *quantification* of the degree of dissimilarity has an old tradition in statistics, and it dates back to the earliest works of Gini (1914, 1965), who defines dissimilarity as lack of similarity. In a discrete setting, and for a given (ordered or non ordered) categorical variable, the notion of similarity between two distributions made conditional on different population groups can be formulated as in Gini (1914, p. 189): Similarity is achieved when "the overall populations of the two groups take the same values [of the categorical variable] with the same [relative] frequency."[2]

Building on this simple notion of similarity in a discrete framework, chapter 1 tackles the issues of *comparisons* of the degree of dissimilarity between sets of distributions. The objective of the chapter consists in constructing a *ranking* of sets of distributions according

[2] The text is translated from the original in Italian.

to the degree of dissimilarity between the distributions in these sets. The ranking of each set is determined by how far one given configuration is from the case of perfect similarity. There are, however, many possible configurations that satisfy the definition of perfect similarity, and are all indifferent among them (i.e. equally placed as the less dissimilar cases). It is therefore possible to compare configurations that have potentially different references for what is the perfect similarity configuration. Moreover, the dissimilarity ranking does not stem from the comparison of the distribution within a given configuration with a well defined reference distribution,[3] but only on the information provided by the distributions under comparison.

This property has two important implications, studied in chapter 1. The first implication is that virtually all the multidimensional configurations of frequencies distributions can be compared and, possibly, ordered according to a dissimilarity partial order. This is done by exploiting transformations of the data that allow to move from one configuration to the other towards perfect similarity. Chapter 1 treats only a partial order of configurations, based on a limited amount of possible transformations. Adding more structure would give more discriminatory orderings or even a dissimilarity indicator (like the well known dissimilarity index by Duncan and Duncan 1955).

The second implication is that dissimilarity is a primitive notion for many other relevant dimensions such as the inequality or the discrimination and, eventually, for equality of opportunity. Both these implications are extensively discussed in chapter 1, because they provide the building blocks for applying the notion of dissimilarity to the assessment of opportunity equalization.

The equalization of opportunity criterion introduced in chapter 3 invokes a dissimilarity comparison in a different context. The background of analysis is different, since opportunity equalization gathers together concerns of the degree of dissimilarity between continuous outcome distributions (thus taking into account also cardinal information) and relevant normative principles for the analysis of equality of opportunity (Roemer 1998, Fleurbaey 2008).

[3] As it is usually done in inequality comparisons, where the reference distribution is the one assigning everybody with the average income.

However, the notion of dissimilarity studied in chapter 1 is reflected on the equalization of opportunity criterion: equalization of opportunity defines a partial order of configurations, such that the score in this order depends on how far a given configuration is from granting equality of opportunity.[4] Equalization of opportunity is satisfied if, after implementation of a policy, the set of conditional distributions is "closer" to satisfy equality of opportunity than it was before policy implementation. In this context, the notion of "closer" rests on a generalization of the equality of opportunity analysis in Lefranc, Pistolesi and Trannoy (2009).

The organization of the thesis follows the same order of exposition of the arguments developed in the previous paragraphs. The thesis is divided into three chapters and two supporting appendices. The first chapter studies a very general notion of dissimilarity. It relies on a very simple but flexible structure, that can be exploited in a variety of contexts such as equality of opportunity, segregation, discrimination, intergenerational mobility and multidimensional inequality. All these phenomena are rationalized by dissimilarity comparisons between distributions of groups across well defined ordered, or non-ordered, realizations. For instance, the problem of segregation and social exposure at the individual level is studied in the second chapter by mean of the dissimilarity order. Finally, an opportunity equalization criterion is described in the third chapter. This criterion, suitable for assessing changes in achievements, looks at equality of opportunity as a form of lack of consensus among preferences in establishing the distribution of economic advantage among social groups. The two appendices provide supporting material for the empirical application in chapter 3. The first appendix covers new inferential results for inverse stochastic dominance at order higher than the second.[5] The second appendix is a review of the models, identifying assumptions and estimators for testing quantile treatment effects. The central arguments of each chapter

[4] Parallel with the definition of similarity, equality of opportunity is attained whenever all distributions conditional on background circumstances coincide (Roemer 1998).

[5] For lower orders of dominance, see Beach and Davidson (1983). The methodology developed here exploits an innovative technique which is dual to Davidson and Duclos (2000). Inverse stochastic dominance at a given order is tested by resorting on the comparison of conditional single-parameter Gini indices calculated for a finite number of quantiles. These coefficients are asymptotically normal and the empirical estimator of their asymptotic covariance matrix is proposed. Hence, inverse stochastic dominance can be tested with similar procedures as in Davidson and Duclos (2000).

are now discussed more in detail in what follows.

Chapter 1 argues that many economic problems involve a comparison of the dissimilarity between sets of (possibly more than two) probability distributions, made conditional on a partition of the population in (possibly more than two) *groups*, and defined over a finite number of *classes*. Hence, these distributions can be represented in a matrix notation. Let for convention the groups be associated with the rows of the matrix, and the classes be associated with the columns of the matrix. An entry of this matrix represents the (relative) frequency distribution of a given group on a given class of realizations. The dissimilarity partial order is a general criterion for ranking sets of distributions according to the degree of dissimilarity between the elements of these sets. This chapter studies the dissimilarity partial order and provides its axiomatic characterization, making use of dissimilarity reducing transfers/exchanges of population masses both across classes and/or groups.

There are two possible classes of transformations of the data, that are considered in the characterization of the dissimilarity partial order. The first class gathers operations of permutation of classes or groups distributions, insertion of classes that are empty (with no population mass on the inside) and the *proportional split* of a class into new adjacent classes. The transformation preserves the degree of dissimilarity, since the new classes do not add or modify the information on the distributions of groups. These transformations allow to make the distributions of groups across classes independent on the labeling of the classes without discarding, if needed, the order of the classes. These operations are therefore suitable for the analysis of problems involving permutable or non permutable classes.

The second class gathers transformations that decrease the dissimilarity. The first transformation is called *merge*, and is suitable only for the analysis involving permutable classes. By merging two classes, one loses information on the differences across the distributions of groups (Blackwell 1953), therefore this operation increases the similarity among the distributions. A similar operation has been already presented in Grant, Kajii and Polak (1998) and Frankel and Volij (2011) in a different context. Similarly to what they have shown, when classes are permutable the merge operation, along with the dissimilarity preserving transformations, identify an equivalent representation for the dissimilarity order that relies on Dahl's (1999) matrix majorization pre order. This result is equivalent to say that one can transform one matrix into another either by applying any sequence of the transformation listed above, or by post-multiplying the matrix by another matrix that is row stochastic (see also Marshall and Olkin 1979, Torgersen 1992).

The second operation is called *exchange*. The exchange, originally introduced by Reardon (2009) to study multi-group segregation measures, defines a Pigou-Dalton transfer of population masses across two adjacent classes, so that the size of the groups, as well as the demographic size of the two classes remain unvaried. Although the merge may lead to counter-intuitive results in the case where classes are permutable, the exchange and the dissimilarity preserving operations allow to correctly exploit exclusively the ordinal information on the configuration of the classes. An innovative result is then proposed: data are transformed according to a sequence of exchanges, split or insertion of empty classes and groups permutations if and only if the distributions of groups can be ordered according to sequential uniform majorization of their cumulative distributions.

This chapter has a second objective: to provide and implementable criterion for assessing changes in dissimilarity on the data. The empirical criterion plays, in some sense, the same role of the Lorenz ordering in inequality. However, the dissimilarity criterion is more general than the Lorenz criterion. If classes are permutable, the dissimilarity order is empirically implemented by an equivalent criterion based on the inclusion of the Zonotopes of the distribution matrices under analysis.[6] Conversely, if classes are ordered, the dissimilarity partial order can be empirically implemented by looking at a finite number of Lorenz majorization comparisons among groups relative conditional frequencies, performed at different cumulation stages of the overall population distribution.

Dahl (1999) provided some intuitive results on the relation between Zonotopes inclusion and matrix majorization in the two groups case. The result proposed in the chapter formalizes these intuitions in a multi-group setting. The criterion for the ordered case has not been studied in the literature.

[6]Zonotopes are centrally symmetric convex bodies. Each point belonging to the Zonotope can be represented by taking all the possible summations of the column vectors of a distribution matrix, or fractional multiples of these vectors.

The relationship of the dissimilarity partial order with related concepts is also discussed. When classes are permutable, the dissimilarity pre-order generalizes in the multi-groups setting the ordering of segregation based on the segregation curves dominance, studied in Duncan and Duncan (1955), Hutchens (1991, 2001) and Flückiger and Silber (1999). When classes are permutable, the dissimilarity order allows not only to define the exchange properties underlying the discrimination curves for two distributions studied by Butler and McDonald (1987), Jenkins (1994) and Le Breton, Michelangeli and Peluso (2012), but also to directly extend these comparisons to the multi-group setting. Finally, it is shown that every inequality comparison entails a dissimilarity comparison, but not the inverse. This is motivated from a multivariate inequality perspective (Kolm 1969, Kolm 1977, Marshall and Olkin 1979, Atkinson and Bourguignon 1982, Koshevoy 1995, Koshevoy and Mosler 1996): in multidimensional inequality, one has to measure the dissimilarity in the distribution of goods with respect to a reference distribution, the one of population weights. This distribution adds a further dimension that, for instance, differentiates the Lorenz Zonotope by Koshevoy and Mosler (1996) from the Zonotopes inclusion ordering studied in chapter 1, but that is not needed to perform dissimilarity comparisons across distributions. This makes dissimilarity an even more interesting framework of analysis for all the problems where the reference distribution is not exogenously provided.

Different types of problems can be represented by exploiting distribution matrices, and can be coherently analyzed by mean of the dissimilarity order. This is shown, for instance, in chapter 2, by noticing that the segregation comparisons at the individual level (i.e. taking into account individual level interaction probabilities with well defined social groups) entail a dissimilarity comparison between interaction likelihoods. However, the dissimilarity partial order is not suitable for comparing the cumulative distribution functions for continuous variables, in all those cases in which parameters such as means or distribution quantiles are of interest for the sake of evaluation. This occurs because the dissimilarity comparisons are grounded on (a minimal number of) operations that only preserve the order of the classes, but eliminate any cardinal interpretation associated to the classes.[7] This limitation is

[7]For instance, if classes are income intervals, splitting one class means loosing the cardinal information

particularly constraining in the case of equality of opportunity comparisons, where economic advantage of some group with respect to the others has to be clearly measured. Chapter 3 deals with these extensions in a normative framework which allows to measure changes in equality of opportunity.

Chapter 2 treats the distributional inequality of social interactions profiles (i.e. individual level probabilities of interact with a well-defined set of groups) across a population. The unequal distribution of the chances of social interactions is often perceived as a leading mechanism for social immobility and a source of inequality of opportunity, that has motivated many desegregation policies at school or neighborhood level (Echenique, Fryer and Kaufman 2006, Frankel and Volij 2011). The analysis in chapter 2 posits that any departure from the rather extreme situation of *equal exposure*, occurring when every individual holds the same interaction profile, is a form of *exposure segregation* (Massey and Denton 1988, Reardon and Firebaugh 2002). The objective of chapter 2 is to derive an index of multi-group segregation that is coherent with this type of arguments.

Suppose that data are available on the distribution of the interaction profiles at the individual level. Then, the data can be represented in a matrix, where each column gives the interaction profiles associated to one individual, and rows are the groups that define the interaction probabilities. A partial order of segregation for these matrices is introduced. This order resorts on an intuitive principle: when two individuals merge their interaction profiles, that is two columns of the matrix are replaced by their resulting convex combination, exposure becomes more equally distributed in the population, and segregation diminishes. Chapter 2 demonstrates that changes in segregation can be equivalently studied by resorting to the notion of dissimilarity in the non-ordered setting. This connection allows to study the exposure dimension of segregation in the two-groups, as well as the multi-group cases indistinctly.

---

on the size of that income interval, unless one is willing to introduce additional assumptions on the relation between the splitting proportion and the income interval splitting.

This chapter proposes an alternative characterization of the dissimilarity order, by identifying the properties of a well-defined family of segregation indicator. One of these indicators, called the Gini Exposure index, is the further studied. This index measures in fact the Zonotope volume. This index is of particular interest because, differently from the other segregation measures (Reardon and O'Sullivan 2004, Frankel and Volij 2011, Alonso-Villar and del Rio 2010) designed to tackle problems from the perspective of the *organizational unit* (that is, the minimal units in which a socioeconomic space of interaction can be partitioned into), the Gini Exposure index is at least consistent with the properties proposed, which are designed to analyze segregation from the perspective of the individuals. An analysis using Italian demographic data on the distribution across municipalities of immigrants groups and the natives reveals the rank correlation between the Gini Exposure index and other segregation indicators. This analysis allows to conclude that the Gini Exposure index is structurally related only with the dissimilarity index.

Chapter 3 proposes a new criterion for measuring the dissimilarity between conditional outcomes distributions that builds on the notion of equality of opportunity developed in Lefranc et al. (2008, 2009). The contribution in this chapter goes beyond Lefranc et al. (2009) in defining a formal criterion for equalization of opportunity. This criterion permits to assess when policy implementation moves a society closer to the Roemer's notion of equality of opportunity, and it allows to evaluate the opportunity equalizing impact of policy implementation consistently with the equality of opportunity prescriptions (Van de gaer 1993, Roemer 1998, Fleurbaey 2008).

The equalization of opportunity criterion is grounded on the notion of consensus among preferences in a given class. Opportunities are equalized whenever, by effect of policy intervention, the size of the class of preferences that unanimously agree on the *existence* of a disadvantaged group shrinks, as well as there is consensus in the same class on the reduction of the *extent* of the disadvantage itself.[8] This procedure amounts to constructing a

[8]The analysis in this chapter is connected to several papers that have recently examined changes over time in inequality of opportunity or differences therein across various national or policy contexts. Ferreira and Gignoux (2011) (for education), Checchi and Peragine (2010) (for income) and Garcia-Gomez et al. (2012) (for mortality) offer some recent examples. In most cases, however, these papers rely on specific cardinal (and often ad hoc) indices of inequality of opportunity.

differences-in-differences comparison between distribution functions conditional on circumstances. This argument evolves in two stages.

In a first stage, differences are taken across distributions *within* each policy regime separately, in order to exploit the direction and distribution of the economic advantage among pairs of circumstances. This is done by imposing sequential restrictions on a class of evaluation preferences until agreement is reached in identifying the disadvantaged circumstance. This is, admittedly, a form of inequality of opportunity. On the contrary, equality of opportunity holds whenever there is lack of consensus on nested preferences sets over the direction of disadvantage.

In a second stage, the differences across conditional distributions are "differentiated out" *between* policy regimes. This is done by combining two criteria. The first criterion is ordinal. It requires that the size of the class of preferences where consensus is reached over the direction of disadvantage (as measured by the sequence of restriction on preferences) among any given pair of conditional distributions, shrinks by effect of the policy. This requires to verify that the "degree" of equality of opportunity is increased by effect of policy implementation, where the "degree" is a measure of the size of the smaller preference class where there is disagreement on the direction of disadvantage. The second is a distance criterion, that is satisfied whenever the extent of the disadvantage (an economic measure of the distance between pairs of distributions) for a given pair of circumstances measured both before and after policy implementation, falls by effect of the policy. The equalization of opportunity criterion holds if and only if the ordinal and cardinal criteria are verified for all pairs of circumstances at all effort levels.

Meaningful alternatives to weaken this demanding criterion for opportunity equalization are also discussed. The first alternative is based upon sequential methods of elimination of dominated circumstances based on agreement on a well-defined set of preferences. A circumstance is eliminated by the set of circumstances whenever it is dominated by another. In this case, if a rational agent were asked to choose a circumstance, she would not choose the dominated one. Further comparisons of this circumstance with the other circumstances are therefore useless. However, it remains unclear in this setting which is the most relevant distance comparison. A second alternative aims at defining the minimal restrictions on the set of preferences that gathers consensus in defining a complete ranking of the distributions before and after policy implementation. Economic disadvantage is measured for distributions (possibly associated to different circumstances) that are ranked in the same position of the ranking both before and after policy implementation. Under precise restrictions, this criterion coincides with the partial order underlying the Gini Opportunity index (Lefranc, Pistolesi and Trannoy 2008).

The equalization of opportunity criterion discussed in chapter 3 is innovative with respect to the methods proposed in the literature (for a survey of theoretical and empirical methods in equality of opportunity, see Ooghe, Schokkaert and Van de gaer 2007, Ramos and Van de gaer 2012). For instance, equalization encompasses robust judgment over a set of preferences and it does not rely on *ad hoc* aggregation principles. Moreover, the equalization criterion does not depend exclusively on the existence of disadvantage (as in Lefranc et al. 2009) but it takes also into account the variation in the extent of the disadvantage, as captured by economic distance measures (Shorrocks 1982). Finally, the equalization criterion is explicitly designed to evaluate policy changes, therefore setting equality of opportunity evaluations in a dynamic perspective.

Implementation issues for the equalization criterion are also discussed. Identification of the equalization criterion relies on the choice of a specific class of preferences and a sequence of restrictions. Results are provided for two very general classes, the class of Yaari's (1987) rank dependent utility functions and the class of preferences admitting the expected utility representation. Restrictions are respectively placed on the derivative of the weighting function and on the Von Neumann-Morgenstern utility functions. These two classes and restriction are not selected by chance. In fact, within these two models, the equalization criterion can be tested by relying on inverse or, respectively, direct stochastic dominance. However, tractable empirical tests for distance comparisons can only be obtained for the Yaari model, which is used to derive the following implementation results, as well as to construct the empirical session of this chapter.

The second empirical issue has to do with unobservable components. The chapter proposes identification conditions for the equalization criterion under unobservability constraints, notably for effort. This result relies, however, on the rank dependent representation of preferences. In that case, only a very special equalization criterion can be implemented, which aggregates conditional distributions with fairly weak requirements. The validity of this criterion remains, however, debatable. An implementation algorithm to test the equalization of opportunity criterion is also proposed.

Finally, new results on the statistical inference for testing inverse stochastic dominance at order higher than the second are proposed and discussed in appendix A. These results are used to implement the opportunity equalization criterion with French data for educational and earning achievements.

The objective of the application consists in evaluating the impact of two simulated educational policies on earnings. This is done making use the *Enquete Emploi*, a large survey of French workers aged 15 to 65. The objective is to detect if a policy that increases access to the secondary education system, or a policy that promotes participation into higher education, foster opportunity equalization. It is not possible to retrieve the impact of these policies from the data, but it is possible to simulate the policies by treating the earnings quantiles of a target group of students (i.e. those who have abandoned the education system too early) with the quantile treatment effects on earnings associated to achieving secondary/higher education levels. Identification assumptions and models for recovering quantile treatment effects are discussed in appendix B. The most appropriate technique for this analysis is the instrumental variables approach. Identification stems from an exogenous change in the underlying *institutional background*, inducing discontinuities in the schooling age profiles for students in the secondary education system (the introduction of the *Loi Berthoin* in France, see Grenet 2012), or by providing a quasi-natural experimental design for changes in the higher education enrollment rules (the *May 1968* events in France, see Maurin and McNally 2008). Although the second policy has sizable distributional effects on earnings, only the first policy fulfills the opportunity equalization criterion, providing strong evidence on the equalizing impact of policies taking place early in the educational career of students.

This result should encourage the use of the equalization criterion to assess other type of policies taking place along the educational career of students, leaving space for future empirical contributions.

Chapter 1

The measurement of multi-group dissimilarity and related orders

Contents

1.1 Introduction

Since the seminal work by Kolm (1977) and Atkinson and Bourguignon (1982), the comparison of multidimensional distributions has received substantial attention in the economic literature on inequality and social welfare. In such a framework, the main objective consists in capturing inequalities in the multivariate distribution of relevant economic measures such as income, wealth, assets, goods, among the units of analysis that usually coincide with individuals or their aggregations. Assessments over alternative distributions are often made by resorting to multivariate stochastic orders and to related empirically implementable dominance tests (Koshevoy 1995, Koshevoy and Mosler 1996, Shaked and Shanthikumar 2006, Marshall, Olkin and Arnold 2011).

Alternative forms of multidimensional assessments have received much less attention in the literature. Here, we focus on "inequalities" that stem from the distribution of a population divided into two or more *groups* across non-overlapping *classes*. In this setting, groups are predetermined by a given partition of the population, while classes correspond to the realizations of a generic discrete outcome variable that can be either ordered (e.g., health, education achievements or income classes) or, alternatively, non-ordered (e.g., residential location or type of occupation).

Concepts such as *inequality*, *polarization* and *diversity* are related to the pattern of *distributional heterogeneity* of each group's population across classes (Rao 1982). However, these notions are not suitable, alone, to analyze and to model more complex and relevant social phenomena like school/occupational/residential segregation, intergenerational mobility, equality of opportunity or discrimination. The evaluations of each of these phenomena should be based on comparisons, across groups, of each group's distributional heterogeneity.

This chapter is concerned with the conceptualization, characterization and implementation of multi-group *dissimilarity* comparisons of groups' distributions across classes.

Dissimilarity comparisons have a long history in the statistical literature, which dates back to the earliest work of Gini (1914, 1965). Gini (1914, p. 189) defines two distributions (α and β, addressed to as "groups", evaluated at modality, or "class" x of variate X) as *similar* when "the overall populations of the two groups take the same values with the same [relative] frequency. If n is the size of group α, m is the size of group β, n_x the size of group α which is assigned to class x and m_x the size of group β assigned to the same class, then it should hold [under similarity] that, for any value of x, $\frac{n_x}{m_x} = \frac{n}{m}$."[1] Moreover Bertino et al. (1987), referring to the work of Gini, extend this notion by defining two or more distributions of the same variate to be similar if "for any modality [...] the absolute frequencies [of the distributions] are proportional". An obvious consequence is that "if two distributions are similar they can have different sizes but their syntheses which are based on relative frequencies are equal."

A configuration under evaluation is given by a set of groups distributions that can be formalized through distribution matrices where rows and columns denote respectively groups and classes and each cell's entry corresponds to the frequency of the population of a given group in a given class. The distribution matrix that embodies perfect similarity satisfies the definition in Gini (1914, 1965) only if its rows are proportional one to the others. Every configuration that does not admit this similarity representation displays some degree of dissimilarity. Various indicators have been proposed in the literature to qualify the degree of dissimilarity. There is however discordance on the properties that these indicators should satisfy to produce a ranking of configurations coherent with decreasing dissimilarity.

A century after the seminal work by Gini we propose a systematic framework to answer the following question: Does configuration B display at most as much dissimilarity as does configuration A? This question is particularly relevant, for instance, in evaluating policy intervention that aims at alleviating the incidence of segregation, intergenerational immobility or discrimination across groups. In this chapter, we single out well defined transformations of distribution matrices based on split, merge or exchange transformations of population masses both across groups and/or classes. When applied to the data, these transformation allow to move from a configuration A to a less dissimilar configuration B, towards perfect similarity. Some of the operations that we consider are related to different streams of literature (Grant et al. 1998, Frankel and Volij 2011, Reardon 2009), here we

[1] The text is translated from the original in Italian.

analyze their combined effect and we clarify substantial differences in the concepts when applied to ordered or permutable classes.

Making use of combinations of these operations we characterize dissimilarity partial orders. Only configurations obtained from sequences of the dissimilarity preserving or reducing transformations can be unambiguously ranked. We show that, when this is the case, the dissimilarity partial orders can be formalized in terms of matrix majorization operations, and that ordered or non-ordered dissimilarity comparisons can be empirically implemented and tested using intuitive criteria.

To illustrate these criteria we consider the case of groups with equal size. Take a set of individuals that corresponds to a proportion p of the overall population. Among them those of group i correspond to a proportion p_i in their group. Dissimilarity assessments are based on the evaluation of the dispersion of the values of p_i across all groups.

When *classes are ordered*, the evaluation is made taking into consideration groups proportions related to the *first classes* that cover the proportion p of the overall population, while the dispersion assessment is based on the Lorenz dominance criterion applied to the vector of p_i's, that on average should sum to p. Configuration B is considered unambiguously less dissimilar than configuration A if the Lorenz dominance of the groups population shares is verified for any p.

For *non-ordered classes* the evaluation is made taking *any combination* of classes, or proportional splits of them (with associated proportional shares of the groups populations), that cover the proportion p of the overall population. For a given p, these combinations lead to a (convex) set of the vectors of p_i's. Given p, groups shares are less disperse under configuration B if the associated set of vectors of p_i's is included in the analogous set derived for configuration A. An unambiguous reduction in dissimilarity is obtained if the inclusion test holds for any p.

The role of the transformations underlying the dissimilarity concept can be illustrated with simple examples that draw from the segregation or the discrimination literature.

Segregation occurs when groups are unevenly distributed across the organizational units in which a social or economic space is partitioned into (Massey and Denton 1988). Sociologists and economists have highlight the importance of desegregation policies to achieve social inclusion goals[2] and have developed, for the two groups case, the appropriate apparatus for measuring segregation consistently with a simple notion of Pigou-Dalton transfer of population masses across sections.[3] Segregation involves the notion of dissimilarity across *non-ordered classes* in a multi-group setting. Consider for instance many ethnic groups of students and three schools. Half of the students or each group are concentrated in one school, while the others are unevenly distributed across the two remaining schools. There is segregation, and it is preserved if, for instance, one considers also schools with no students, or if the labeling of schools is modified, or even if one school is split into two new smaller institutes, while preserving the initial social composition. If the policymaker merges the two latter schools to form a unique institute, then groups proportions are equalized. Frankel and Volij (2011) motivate that segregation should always reduce when data are transformed by merge operations. We take a similar stance to construct a dissimilarity order for permutable classes: by merging two classes, the differences across groups distributions are partially smoothed and dissimilarity is reduced.

Also the study of labor market discrimination patterns involves the analysis of the distribution of population masses across earnings intervals associated with *ordered classes*. A configuration where the proportions of groups in a given class are equalized across groups, and therefore the associated cumulative distribution functions coincide, displays no discrimination. This is in fact a situation of perfect similarity. On the contrary, discrimination is maximized whenever each group is concentrated on a series of adjacent earnings intervals, and only that group occupies the intervals. All the remaining cases display a certain degree of discrimination. We suggest that these cases can be ordered making use of sequences of dissimilarity preserving operations and exchanges of populations from the most represented group in one given class to the less represented group in the same class. This operation,

[2]See Echenique et al. (2006) and Borjas (1992, 1995), for instance.

[3]See Hutchens (1991, 2001), Reardon and Firebaugh (2002), Reardon (2009), Flückiger and Silber (1999), Chakravarty and Silber (2007), Alonso-Villar and del Rio (2010), Frankel and Volij (2011) and Silber (2012), for a survey on the methodology.

which is equivalent to perform Pigou-Dalton transfers of realizations of the cumulative distribution functions, fills the gap between groups' cumulative distribution functions, thus reducing the impact of discrimination. This conclusion relies, in fact, on a dissimilarity comparison and the exchange transformation is an appealing criterion for assessing reductions in dissimilarity.

The notion of dissimilarity can be seen as logically separated from the notion of inequality. In the discrete setting, the overall population inequality can be decomposed into within group and between groups components. The within group component is determined by the degree of heterogeneity of groups' distributions across classes, the between groups component captures dissimilarity concerns. Following this perspective then inequality and dissimilarity can move in different directions. Every equal allocation, where all groups population masses are concentrated on the same class, display no dissimilarity across groups. But if the classes where population masses are concentrated differ across groups then dissimilarity can be maximal. On the other hand, there are configurations characterized by sizable but similar groups heterogeneity that cannot be judged as equal but fulfill the perfect similarity representation.

However, taking a different perspective, we show that inequality comparisons can be interpreted as dissimilarity comparisons but not the reverse. Take the traditional univariate inequality measurement grounded on the Lorenz curves comparisons, in this case we can interpret the classes as the n sampled income units (e.g. individuals or households) and consider two "groups" distributions: the income share owned by each of these income units, and the weighting scheme assigning weight $1/n$ to each of them. There is no inequality in the sense of Lorenz whenever each class/unit income share is equal to its demographic/social weight. This is a similarity requirement, that can be straightforwardly extended to the multidimensional inequality analysis. In the next section we show that the well known Pigou-Dalton (rich to poor) transfer principle is consistent with more general dissimilarity decreasing operations that we use in this chapter.

The rest of the chapter is organized as follows. Section 1.2 presents the notation, as well as an overview of the majorization and geometric ordering exploited throughout this chapter. In section 1.3 we discuss the axiomatic structure and the data transformations underlying the dissimilarity comparisons. In sections 1.4 and 1.5 we illustrate our first contribution: the dissimilarity pre-order relies on well known majorization orderings (Marshall et al. 2011) both in the permutable (Blackwell 1953, Torgersen 1992, Dahl 1999) and the ordered setting (Hardy, Littlewood and Polya 1934, Marshall and Olkin 1979, Le Breton et al. 2012). Section 1.6 proves necessary and sufficient conditions for testing the dissimilarity pre-order according to the ranking produced by Zonotopes inclusion for the non ordered classes case[4] and by Path Polytopes inclusion in the case of ordered classes.[5] This innovative results permits the policymaker to answer questions such as: Is society B less segregated/more mobile/less discriminant than society A? The final section formalizes in which sense inequality comparisons are always nested within dissimilarity comparisons, and proposes possible extensions toward complete orders of dissimilarity, coherently with the axiomatic model that we have introduced.

Example 1.1 (An motivating example for multi-groups comparisons) This example illustrates an application of the dissimilarity concept to the assessment of segregation. We motivate the importance of a multi-groups setting (and transformations) by showing that two-groups comparisons may lead to wrong evaluations. Consider a population partitioned into three groups $\{1, 2, 3\}$. The population in each group is divided across two classes, which can be interpreted as two types of occupations, $\{\text{Class } 1, \text{Class } 2\}$. The value a_{i1} denotes the *proportion* of group i in class/occupation 1, for $i \in \{1, 2, 3\}$ under configuration $\mathbf{A}$, with analogous interpretation for a_{i2}. Thus $a_{i1} + a_{i2} = 1$ for all $i \in \{1, 2, 3\}$.

We compare two alternative configurations $\mathbf{A}$ and $\mathbf{B}$ in terms of segregation/dissimilarity

[4] We contribute by generalizing to the multi-group setting the equivalence between matrix majorization and Zonotopes inclusion for the bi-dimensional setting in Dahl (1999). Zonotopes are in fact extensions of the segregation curve (Hutchens 1991), a plot of the overall dispersion across groups' conditional distributions in a given configurations. Our result links Zonotopes inclusion with the existence of a sequence of dissimilarity preserving/reducing operations, as the Lorenz curve is related to the existence of a sequence of Pigou-Dalton transfers.

[5] This innovative result extends the literature on two-groups discrimination depicted by the comparisons of discrimination curves (Butler and McDonald 1987, Jenkins 1994, Le Breton et al. 2012) to the multi-group setting. It is the first attempt to recover the equivalence between sequences of preserving/decreasing dissimilarity transformations, sequential uniform majorization (Marshall and Olkin 1979) and dominance orders for multi-groups discrimination curves.

between the distribution of the three groups across the two classes/occupations.

The two configurations are formalized as follows:

A :

	Class 1	Class 2
Group 1	0.9	0.1
Group 2	0.1	0.9
Group 3	0.8	0.2

; **B** :

	Class 1	Class 2
Group 1	0.4	0.6
Group 2	0.6	0.4
Group 3	0.45	0.55

.

In order to assess the occupational segregation ranking of the two configurations, we can make use of segregation curves (Hutchens 1991). Consider a partition of the two configurations that takes into account only groups 1 and 2, denoted respectively as $\mathbf{A}(1,2)$ and $\mathbf{B}(1,2)$. The segregation curve of $\mathbf{A}(1,2)$ is obtained by (i) evaluating the ratios a_{21}/a_{11} and a_{22}/a_{12}, (ii) ordering the regions in increasing order with respect to these ratios, i.e., the order of Class 2 precedes Class 1 only if $a_{22}/a_{12} \geq a_{21}/a_{11}$; (iii) plotting for the first class in the order indexed, say, by $j \in \{1,2\}$, the point (a_{1j}, a_{2j}) and connecting it with the origin (0,0) and the upper extreme (1,1).

If $a_{22}/a_{12} = a_{21}/a_{11} = 1$ we get for $\mathbf{A}(1,2)$ a segregation curve coinciding with the 45 degrees line, thus identifying perfect similarity. As the curve moves below this line the degree of dissimilarity between the two groups distributions increases. Thus, if the curve of $\mathbf{B}(1,2)$ lies above the one of $\mathbf{A}(1,2)$, we can make a "robust" statement concerning the fact that $\mathbf{A}(1,2)$ exhibits larger dissimilarity than $\mathbf{B}(1,2)$.

It is possible to use the segregation curve to compare all subsets of the two distributions that consider pairs of groups. Repeated application of these comparisons lead to the following statement: for any $i,j \in \{1,2,3\}$ s.t. $i \neq j$ distribution $\mathbf{B}(i,j)$ dominates $\mathbf{A}(i,j)$ in terms of the segregation curve, that is $\mathbf{B}(i,j)$ is less dissimilar/segregated than $\mathbf{A}(i,j)$.

If groups 1 and 2 are merged so that they are considered as a unique group and then compared to Group 3, will the new configuration made of only two groups exhibit the same pattern of dissimilarity when comparing $\mathbf{A}$ and $\mathbf{B}$?

Suppose that the relative population weights of the two groups are respectively 0.875

and 0.125 and that we denote the new group with index 4. The two new configurations $\mathbf{A}'$ and $\mathbf{B}'$ obtained respectively from $\mathbf{A}$ and $\mathbf{B}$ by merging group 1 and group 2 are:

$\mathbf{A}'$:	Class 1	Class 2
Group 3	0.8	0.2
Group 4	0.8	0.2

;

$\mathbf{B}'$:	Class 1	Class 2
Group 3	0.45	0.55
Group 4	0.425	0.575

.

Clearly distribution $\mathbf{A}'$ exhibits less dissimilarity than $\mathbf{B}'$, in fact the degree of dissimilarity in $\mathbf{A}'$ is zero being the shares of the two groups identical across the two regions.

This result conflicts with the fact that any pairwise comparison of groups in $\mathbf{A}$ and $\mathbf{B}$ shows that $\mathbf{A}$ is more dissimilar.

Analogous mathematical examples with different underlying explanations can be constructed to highlight the theoretical difficulties to move from the established setting of two groups dissimilarity comparisons to the multi-group case. The general dominance conditions we will derive will be robust to these considerations.

For instance, consider the problem of assessing the degree of gender segregation that is induced by the social group of origin across classes represented in matrices $\mathbf{A}$ and $\mathbf{B}$. This can be done by mixing the three groups according to a fixed row stochastic weighting matrix that depicts a groups mixing scheme constant across jobs positions of the three groups, and returns the male-female composition. Thus, we obtain matrices $\mathbf{A}''$ and $\mathbf{B}''$. If $\mathbf{B}$ displays less dissimilarity than $\mathbf{A}$, robustness requires that $\mathbf{B}''$ should display less dissimilarity than $\mathbf{A}''$ for all weighting schemes. Consider the following vector of female shares of working force in each group: $\left(\frac{47}{80}, \frac{41}{80}, \frac{32}{80}\right)$ and suppose that the three groups have the same populations. The resulting distributions of males and females across jobs are given by:

$\mathbf{A}''$:	Class 1	Class 2
Female	0.6	0.4
Male	0.6	0.4

;

$\mathbf{B}''$:	Class 1	Class 2
Female	0.481	0.519
Male	0.485	0.515

.

Again, the inversion in the dissimilarity ranking position of $\mathbf{A}''$ and $\mathbf{B}''$ suggests that two groups comparisons cannot be consistently used to make assessments on multi-group dissimilarity phenomena.

1.2 Setting

1.2.1 Notation

This chapter deals with comparisons of $d \times n$ *distribution matrices*, depicting the absolute frequencies distribution[6] of d groups (indexed by rows) across n disjoint classes (indexed by columns), where $d, n \in \mathbb{N}$ are natural numbers, such that $n \geq 2$ and $d \geq 2$. The set of distribution matrices with d rows is:

$$\mathcal{M}_d := \{\mathbf{A} := (\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) : \ \mathbf{a}_j \in \mathbb{R}^d_+, \ n \in \mathbb{N}\}.$$

Each element of $\mathcal{M}_d$ represents a set of d distributions across n classes. Thus, a_{ij} is the population of group i observed in class j. We will compare matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ representing sets of distributions with fixed d groups and with possibly variable number of classes, denoted respectively n_A and n_B. The set of *all* distribution matrices with possibly different d is denoted $\mathcal{M}$. The perfect similarity matrix matrix $\mathbf{S} \in \mathcal{M}_d$ represents the case in which groups' frequencies distributions are proportional one to the other, that is:

$$\mathbf{S} \ := \begin{pmatrix} \lambda_1 a_1 & \cdots & \lambda_1 a_n \\ \vdots & \ddots & \vdots \\ \lambda_d a_1 & \cdots & \lambda_d a_n \end{pmatrix}.$$

For $\mathbf{A} \in \mathcal{M}_d$, the *cumulative distribution matrix* $\overrightarrow{\mathbf{A}} \in \mathcal{M}_d$ is constructed by sequentially cumulating the classes of $\mathbf{A}$. The column k of $\overrightarrow{\mathbf{A}}$, for all $k = 1, \ldots, n_A$, is therefore $\overrightarrow{\mathbf{a}}_k := \sum_{j=1}^{k} \mathbf{a}_j$ for $j \leq k$.

Let $\mathbf{e}_n := (1, \ldots, 1)^t$ and $\mathbf{0}_n := (0, \ldots, 0)^t$ be n-dimensional column vectors of ones and zeroes. With $\mathbf{\Pi}_n$ we define an element in the set $\mathcal{P}_n$ of all $n \times n$ permutation matrices.

[6] For convenience we use matrices whose entries are real numbers.

The standard simplex in $\mathbb{R}^n_+$ is denoted by $\Delta^n := \{\mathbf{x} \in \mathbb{R}^n_+ : \mathbf{e}^t_n \cdot \mathbf{x} = 1\}$, while $\mathcal{R}_{n,m}$ denotes the set of all *row stochastic* $n \times m$ matrices such that each of the n rows lies in Δ^m. The set $\mathcal{R}_{n,m}$ describes a polytope in $\mathbb{R}^{n,m}_+$. Each matrix $\mathbf{X} \in \mathcal{R}_{n,m}$ can be written as the convex combination of its vertices, given by all the m^n (0,1)-matrices of dimension $n \times m$ with exactly one nonzero element (of value one) in each row, hereafter denoted as $\mathbf{X}(1), \ldots, \mathbf{X}(h), \ldots, \mathbf{X}(m^n)$. The elements of $\mathcal{R}_{n,m}$ can be interpreted as *migration matrices* where the entry x_{ij} gives the probability for the mass of individuals in class i in the distribution of origin to migrate to the class j in the distribution of destination.

The set $\mathcal{C}_{n,m}$ denotes all *column stochastic* matrices such that each of the m columns lies in the Δ^n simplex. The set of row (column) stochastic matrices such that $m = n$ is denoted by $\mathcal{R}_n$ ($\mathcal{C}_n$). The set $\mathcal{D}_n = \mathcal{R}_n \cap \mathcal{C}_n$ contains the *doubly stochastic* matrices.

In the next subsections we review partial orders based on majorization and on comparisons of geometric bodies. The readers who are already familiar with the matrix majorization orders presented in Marshall et al. (2011) and with Zonotopes and Monotone Paths definitions can move to section 1.3.

1.2.2 Orders based upon majorization

Multivariate majorization theory suggests elementary algebraic transformations of data that involve row, column or bistochastic matrices. These transformations have a relevant economic interpretation, which has been exploited to construct multivariate inequality orders.

Definition 1.1 ***(Multivariate Majorization)*** *Given two matrices* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$*:*

1. $\mathbf{B}$ *is uniformly majorized by* $\mathbf{A}$ ($\mathbf{B} \preccurlyeq^U \mathbf{A}$) *provided that* $n_A = n_B = n$ *and there exists a doubly stochastic matrix* $\mathbf{X} \in \mathcal{D}_n$ *such that* $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$.

2. $\mathbf{B}$ *is directionally majorized by* $\mathbf{A}$ ($\mathbf{B} \preccurlyeq^D \mathbf{A}$) *provided that* $n_A = n_B = n$ *and* $\boldsymbol{\ell}^t \cdot \mathbf{B} \preccurlyeq^U \boldsymbol{\ell}^t \cdot \mathbf{A}$*, for every* $\boldsymbol{\ell} \in \mathbb{R}^d$.

3. $\mathbf{B}$ *is column majorized by* $\mathbf{A}$ ($\mathbf{B} \preccurlyeq^C \mathbf{A}$) *provided there exists a column stochastic matrix* $\mathbf{X} \in \mathcal{C}_{n_A,n_B}$ *such that* $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$.

4. $\mathbf{B}$ *is (matrix) majorized by* $\mathbf{A}$ ($\mathbf{B} \preccurlyeq^R \mathbf{A}$) *provided there exists a row stochastic matrix* $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ *such that* $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$.

Uniform majorization has been extensively discussed in Marshall and Olkin (1979) (see also Marshall et al. 2011). Welfare implications of directional majorization have been studied by Koshevoy (1995) and by Kolm (1977), who restricts attention to dominance for vectors $\boldsymbol{\ell} \in \mathbb{R}^d_+$ that can be interpreted as prices. Column majorization is the *weak majorization* in Martínez Pería, Massey and Silvestre (2005), while matrix majorization has been originally proposed by Dahl (1999) as an alternative to uniform majorization for ranking matrices with different number of columns.

In the analysis of univariate distributions, uniform majorization has an interpretation in terms of Pigou Dalton transfers and corresponds to dominance according to the Lorenz order. This interpretation can be extended to the multivariate case, although this type of operations do not account for multivariate structure of correlation across dimensions.

The multivariate majorization order can be weakened in two interesting directions. Directional dominance reduces the number of dimensions of the problem to a univariate comparison of "budget" distributions obtained by a system of weights, positive and/or negative.

Matrix majorization is weaker than uniform majorization. It is obtained via multiplication of row stochastic matrices, therefore it preserve the total dimension of each group and it appears to be an appropriate candidate to represent the dissimilarity order. In fact, matrix majorization has been already investigated (under different names) in other fields such as linear algebra and majorization orders (Dahl 1999, Hasani and Radjabalipour 2007), in inequality analysis (see Chapter 14 in Marshall et al. 2011), in the comparison of statistical experiments (Blackwell 1953, Torgersen 1992) or in a-spatial two groups (Hutchens 1991, Chakravarty and Silber 2007) and multi-group segregation (Frankel and Volij 2011).

1.2.3 Orders based upon polytopes inclusion

This section reviews the orderings of distribution matrices induced by the inclusion of geometric bodies derived by matrices in $\mathcal{M}_d$. We focus on two inclusion orderings.

The Zonotope inclusion order

For matrix $\mathbf{A} \in \mathcal{M}_d$, the associated *Zonotope* $Z(\mathbf{A}) \subseteq \mathbb{R}^d_+$ can be written as the convex set of point-vectors obtained by mixing the columns of $\mathbf{A}$ with a system of weights lying in the unitary interval:

$$Z(\mathbf{A}) = \left\{ \mathbf{z} := (z_1, \dots, z_d)^t \;:\; \mathbf{z} = \sum_{j=1}^{n_A} \theta_j \mathbf{a}_j, \quad \theta_j \in [0,1] \; \forall j = 1, \dots, n_A \right\}.$$

This representation is particularly convenient to prove our results (for an extensive treatment, see McMullen 1971).[7]

The *Dissimilarity Zonotope* $Z_D(\mathbf{A})$ associated to matrix $\mathbf{A}$ is a d-dimensional parallelogram whose edges have size $\mathbf{A} \cdot \mathbf{e}_{n_A}$. When $d = 3$, $Z_D(\mathbf{A})$ is a parallelepiped. Throughout this chapter, we restrict attention to comparisons of Zonotopes that lie inside the same Z_D, hence generated by matrices $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ such that $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \boldsymbol{\mu}$. If, moreover, $\boldsymbol{\mu} = \mathbf{e}_d$, then Z_D coincides with the unit hypercube.

The *Similarity Zonotope* $Z_S(\mathbf{A})$ associated to matrix $\mathbf{A}$ corresponds to the diagonal of Z_D, connecting the origin $\mathbf{0}_d$ and the point with coordinates $\mathbf{A} \cdot \mathbf{e}_{n_A}$. The Z_S coincides with the d-dimensional Zonotope associated to the distribution matrix $\mathbf{S} \in \mathcal{M}_d$ displaying perfect similarity.

The operations that can be used to reshape Z_D toward Z_S, while preserving its convexity and central symmetry, are equivalently characterizing the operations that reduce dissimilarity. For matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$, the inclusion $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$

[7] The *Zonotope* $Z(\mathbf{A}) \subseteq \mathbb{R}^d_+$ is a centrally symmetric convex body defined by the Minkowski sum of a finite number of closed line segments connecting the points generated by the columns of $\mathbf{A}$ with the origin.

indicates that the set of distributions in $\mathbf{B}$ is closer to similarity than is the set of distributions in $\mathbf{A}$. Our main results for dissimilarity comparisons with permutable classes involve thus dominance relations.

The Path Polytope inclusion order

The *Monotone Path* $MP^*(\mathbf{A}) \subseteq \mathbb{R}^d_+$ is an arrangement of n_A line segments connecting the origin and the points with coordinates given by the columns of $\overrightarrow{\mathbf{A}}$. It defines a path *inside* the Zonotope, which connects the origin with the point corresponding to $\mathbf{A} \cdot \mathbf{e}_{n_A}$. The vertices of $MP^*(\mathbf{A})$ are ordered monotonically with respect to the columns of matrix $\mathbf{A}$, such that $\mathbf{v}_j \in MP^*(\mathbf{A})$ if and only if $\mathbf{v}_j = \overrightarrow{\mathbf{a}}_j$ for all j, $\mathbf{v}_0 = \mathbf{0}_d$ and $\mathbf{v}_{n_A} = \mathbf{A} \cdot \mathbf{e}_{n_A}$.[8]

Similarly to the Zonotope, any point on $MP^*(\mathbf{A})$ can be defined as the weighted sum of the columns of matrix $\mathbf{A}$, up to a nonlinear restriction on weights. Let $\mathbf{1}_{j<k}$ and $\mathbf{1}_{j=k}$ be the indicator functions, taking value one when their respective arguments are verified, and zero otherwise. Then:

$$MP^*(\mathbf{A}) := \left\{ \mathbf{p} = (p_1, \ldots, p_d)^t : \mathbf{p} = \sum_{j=1}^{n_A} \theta_j \mathbf{a}_j, \quad \theta_j = \mathbf{1}_{j<k} + \theta\, \mathbf{1}_{j=k}, \ \theta \in [0,1] \ \forall k = 1, \ldots, n_A \right\}.$$

Building on $MP^*(\mathbf{A})$ it is possible to derive the *Path Polytope* $Z^*(\mathbf{A}) \subseteq \mathbb{R}^d_+$:

$$Z^*(\mathbf{A}) := \left\{ \mathbf{z}^* := (z^*_1, \ldots, z^*_d)^t \ : \ \mathbf{z}^* = \text{conv} \left\{ \mathbf{\Pi}_d \cdot \mathbf{p} | \ \mathbf{\Pi}_d \in \mathcal{P}_d \right\}, \ \mathbf{p} \in MP^*(\mathbf{A}) \right\}.$$

The Path Polytope consists in a d-dimensional expansion of the unidimensional ordered set MP^* in the d-variate space. Hence, contrary to the Monotone Path, the Path Polytope has a volume with a nonzero measure. The origin and the ending vertices of the Path Polytope coincide with the ones of the Monotone Path.

We consider the Path Polytopes associated to matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \mathbf{e}_d$. In this case, all points $\mathbf{z}^*$ belonging to the convex hull created from $\mathbf{p} \in MP^*(\mathbf{A})$

[8] See Shephard (1974) and Ziegler (1995) for a definition of the *f-monotone path* and its applications to the study of Zonotopes.

also lie on the same hyperplane supporting the standard simplex Δ^d, properly scaled by a factor $\lambda \in [0, d]$.[9]

The Dissimilarity Path Polytope and the Similarity Path Polytope associated to $Z^*(\mathbf{A})$ coincide with $Z_D(\mathbf{A})$ and $Z_S(\mathbf{A})$, respectively. The inclusion $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ indicates an alternative perspective for assessing that the set of distributions depicted in $\mathbf{B}$ is more close to similarity than is the set in $\mathbf{A}$. In this chapter we characterize this relation and highlight the differences with the Zonotopes inclusion order.

An example

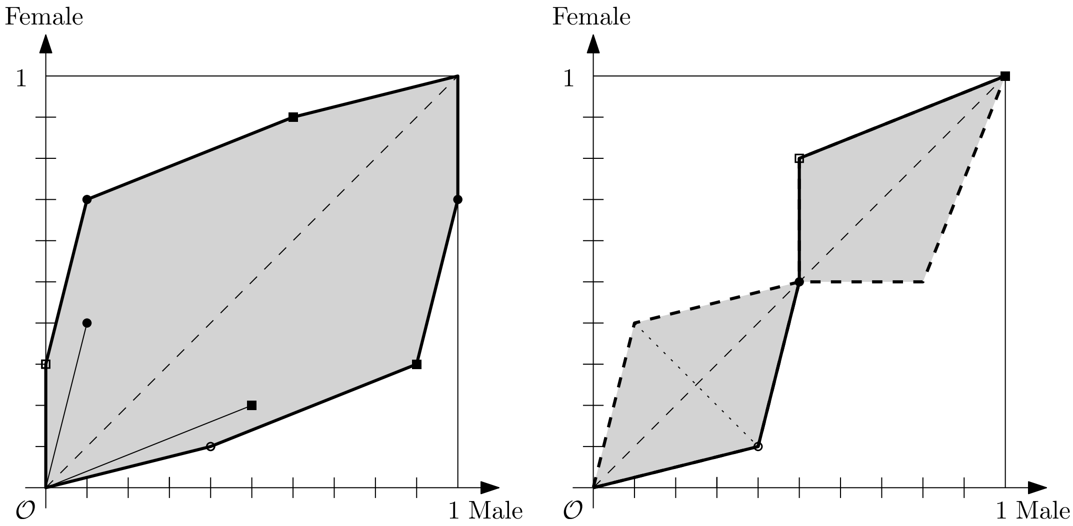
Matrix $\mathbf{A} \in \mathcal{M}_2$ collects the data on the distribution of male (first row) and female (second row) across four classes:

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.1 & 0 & 0.5 \\ 0.1 & 0.4 & 0.3 & 0.2 \end{pmatrix}.$$

The Zonotope of matrix $\mathbf{A}$ is delimited by the grey area in figure 1.1(a). Each column of $\mathbf{A}$ is a vector in the two dimensional space (we draw a small symbol associated to each vector). Consider the case where classes are interpreted as occupations (and therefore are non-ordered). Matrix $\mathbf{A}$ may well represent a segregated distributions of sexes across occupations. The $Z(\mathbf{A})$ is therefore the area between the segregation curve, corresponding to its lower bound, and the dual of the segregation curve.

The Path Polytope of matrix $\mathbf{A}$ (figure 1.1(b)) corresponds to the grey area between the Monotone Path (solid line) and its symmetric projection (dashed line) with respect to the diagonal. If classes are interpreted as ordered non-overlapping income intervals, then matrix $\mathbf{A}$ may well represent a gender based discrimination pattern and the Path Polytope corresponds to the area between the discrimination curve (the lower boundary of the Path Polytope) and its dual discrimination curve.

[9] The hyperplane supporting the simplex has slope $\mathbf{e}_d$. Since we make use of distribution matrices satisfying $\mathbf{A} \cdot \mathbf{e}_n = \mathbf{e}_d$, then the value associated to the hyperplane crossing the Path Polytope in the point $\mathbf{A} \cdot \mathbf{e}_n$ is equal to $\mathbf{e}_d^t \cdot \mathbf{e}_d = d$. We derive a procedure to test the Path Polytopes inclusion which exploits this feature.



(a) $Z(\mathbf{A})$ (b) $Z^*(\mathbf{A})$

Figure 1.1: The Zonotope and the Path Polytope (with the monotone path in solid line)

1.3 An axiomatic approach to dissimilarity

We formalize the normative content of dissimilarity by resorting to an axiomatic structure. The axioms characterize the dissimilarity order by depicting the transformations between classes or between groups that, when applied to matrices in $\mathcal{M}$, either preserve or reduce the degree of dissimilarity embodied in the distribution matrices. Along with the axioms, we define the implied transformations on data matrices.

When we write that the relation "**B** is at most as dissimilar as **A**" satisfies a set of dissimilarity preserving/reducing axioms we mean that there exists a finite sequence of transformations underlying these axioms that allows to move from **A** to **B**. Thus, we assume that the dissimilarity pre-order is fully characterized by these operations and therefore they are not only sufficient to guarantee that **A** and **B** can be compared according to the pre-order but they are also necessary.

For convenience, and without loss of generality, we specify the axioms in the form of transfers of population masses across classes, thus defining the *direct* axiomatic approach to dissimilarity. These operations change name, size and number of the *classes*, while keeping the groups as fixed. These axioms will be at the core of our analysis. Alternatively, we propose a similar structure of transfers of population masses across groups, thus defining the *dual* setting. These operations change the name, the size and the number of the *groups*, while keeping the classes as fixed. In practice, the transformations underlying the dual and direct axioms coincide, provided that the former are applied to the *transpose* of the distribution matrices. We replace the "C" (which stands for classes) with a "G" (which stands for groups) to distinguish the dual from the direct axioms. We keep the two frameworks separated, and we highlight possible incoherences when the two are combined.

1.3.1 Dissimilarity preserving axioms

Let $\preccurlyeq$ be a binary relation in the set $\mathcal{M}$ with symmetric part $\sim$.[10] The relation defines the *dissimilarity order*. We write $\mathbf{B} \preccurlyeq \mathbf{A}$ to say that the distribution of groups in $\mathbf{B}$ are *at most as dissimilar as* the ones in $\mathbf{A}$. We assume from the outset that the dissimilarity order induces a *pre-order* on the set of distribution matrices.[11]

The first axiom defines an anonymity property of the dissimilarity order, by requiring that the name of the classes does not have to be taken into account in dissimilarity comparisons. The underlying operations defines the independence from transformations involving the permutation of columns of a distribution matrix.

Axiom *IPC (Independence from Permutations of Classes)* *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$, *if* $\mathbf{B} = \mathbf{A} \cdot \mathbf{\Pi}_n$ *for a permutation matrix* $\mathbf{\Pi}_n \in \mathcal{P}_n$ *then* $\mathbf{B} \sim \mathbf{A}$.

One direct implication of *IPC* is that by cumulating frequencies across classes, one cannot derive any additional information that can be exploited in the dissimilarity comparison. Hence, admitting *IPC* means restricting attention to a specific class of problems, for this reason we treat the case of permutable versus non-permutable classes separately.

The following two dissimilarity preserving axioms characterize the independence of the

[10] $\mathbf{B} \sim \mathbf{A}$ if and only if $\mathbf{B} \preccurlyeq \mathbf{A}$ and $\mathbf{A} \preccurlyeq \mathbf{B}$.

[11] That is, for any $\mathbf{A},\ \mathbf{B},\ \mathbf{C} \in \mathcal{M}$ the relation $\preccurlyeq$ is *reflexive* ($\mathbf{A} \preccurlyeq \mathbf{A}$) and *transitive* (if $\mathbf{C} \preccurlyeq \mathbf{B}$ and $\mathbf{B} \preccurlyeq \mathbf{A}$ then $\mathbf{C} \preccurlyeq \mathbf{A}$). The assumptions are maintained throughout the chapter.

dissimilarity order from operations that do not add (or eliminate) information on the distribution of groups across classes.

Distributional information is preserved when a new empty class is created. We call the underlying transformations *insertion/elimination of empty classes*.

Axiom *IEC (Independence from Empty Classes)* *For any* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D} \in \mathcal{M}_d$ *and* $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$*, if* $\mathbf{B} = (\mathbf{A}_1, \mathbf{0}_d, \mathbf{A}_2)$, $\mathbf{C} = (\mathbf{0}_d, \mathbf{A})$, $\mathbf{D} = (\mathbf{A}, \mathbf{0}_d)$ *then* $\mathbf{B} \sim \mathbf{C} \sim \mathbf{D} \sim \mathbf{A}$.

Similarly, the splitting of a class into two new classes preserves dissimilarity, when groups frequencies are proportionally split into the two new classes. As a result, one ends up with two proportional classes, each with an smaller population weight. This transformation is a *split of classes*, and it corresponds to a sequence of *linear bifurcations* for a probability distribution, introduced by Grant et al. (1998). If the same bifurcation is applied to *all* conditional distributions (expressed by the rows of a distribution matrix), the ranking of distribution matrices is preserved.

Axiom *SC (Independence from Split of Classes)* *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *with* $n_B = n_A + 1$*, if* $\exists j$ *such that* $\mathbf{b}_j = \beta \mathbf{a}_j$ *and* $\mathbf{b}_{j+1} = (1-\beta)\mathbf{a}_j$ *with* $\beta \in (0,1)$*, while* $\mathbf{b}_k = \mathbf{a}_k$ $\forall k < j$ *and* $\mathbf{b}_{k+1} = \mathbf{a}_k$ $\forall k > j$*, then* $\mathbf{B} \sim \mathbf{A}$.

Alternatively, using similar arguments it is possible to show that the degree of dissimilarity is preserved by the transformations that permute groups or that add/eliminate empty groups to the comparisons, or by applying proportional linear bifurcations of distributions across *classes*, as well as merging classes where the distribution of population across groups are proportional. The corresponding dissimilarity preserving axioms define independence from permutations of groups (*IPG*), from empty groups (*IEG*) and from split of groups (*SG*).

The *dual axioms* can be better understood in the light of the dual concept of dissimilarity, which points at reducing the difformity in the groups composition across classes. Moreover, direct and dual axioms can be combined. In particular, the axiom *IPG* states that the dissimilarity entails a symmetric comparison of groups distributions, and for this reason we retain the groups permutation along with operations involving classes. We formalize *IPG* because we will use it explicitly for some of our characterizations.

Axiom *IPG (Independence from Permutations of Groups)* *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$*, if* $\mathbf{B} = \mathbf{\Pi}_d \cdot \mathbf{A}$ *for a permutation matrix* $\mathbf{\Pi}_d \in \mathcal{P}_d$ *then* $\mathbf{B} \sim \mathbf{A}$*.*

1.3.2 Dissimilarity decreasing axioms

The Merge axiom

The Merge axiom states that the dissimilarity between two or more distributions is reduced whenever any two contiguous classes are mixed together. If the dissimilarity order satisfies *IPC*, the merge can be extended to any pair of classes.

The rationale of the merge axioms is that by mixing together two classes one looses information, in the sense that it becomes more difficult to distinguish the distributions of frequencies associated to different groups. As a result, distributions are more similar. The transformation behind the axiom involve summations of pairs of adjacent columns of a distribution matrix.

Axiom *MC (Dissimilarity Decreasing Merge of Classes)* *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B$*, if* $\mathbf{b}_i = \mathbf{0}_d$*,* $\mathbf{b}_{i+1} = \mathbf{a}_i + \mathbf{a}_{i+1}$ *while* $\mathbf{b}_j = \mathbf{a}_j,\ \forall j \neq i, i+1$*, then* $\mathbf{B} \preccurlyeq \mathbf{A}$*.*

Along with *MC*, one can define a dual merge axiom, *MG*. In this case dissimilarity is reduced as a consequence of the loss of information related to the groups mixture.

The transformations underlying the axioms *IEC*, *IPC*, *SC* and *MC* can be combined into sequences, defining more complex forms of transfers of population masses across classes. When combined together, these operations allow to transform one distribution matrix **A** into the distribution matrix **B** while reducing dissimilarity. The sequence of operations involves classes, and therefore groups are split or merged with equal proportions. Thus, the operations involve a symmetric treatment of groups.

Nevertheless, *MC* entails some operations that preserve the overall row sum of the matrices, while changing completely the size of the sections. Conversely, the dual axioms require the size of the classes (and not the one of the groups) to be fixed across comparison matrices. It follows that, if the dissimilarity order satisfies both types of axioms, then the transformations cannot be independently used.[12]

The Exchange axiom

We formulate an alternative dissimilarity reducing axiom based upon the notion of *exchange* discussed in Reardon (2009) and Fusco and Silber (2011). An exchange transformation entails a movement of individuals across groups but within the same class. It can be applied only if some conditions are verified. Firstly, the exchange can be performed conditionally on a precise order of the columns of a distribution matrix. Secondly, it has to take place only between groups of the same size. Hence, the exchange is meaningful if and only if we consider distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, with $n_A = n_B = n$ and satisfying: $\mathbf{A} \cdot \mathbf{e}_n = \mathbf{B} \cdot \mathbf{e}_n = \lambda \mathbf{e}_d$, with $\lambda \in \mathbb{R}_{++}$.

We say that (the distribution of) group h dominates (the distribution of) group ℓ in class k if $\overrightarrow{a}_{hk} < \overrightarrow{a}_{\ell k}$ and $\overrightarrow{a}_{h,k+1} \leq \overrightarrow{a}_{\ell,k+1}$. According to the exchange principle, if h dominates ℓ in k, and if a *small enough amount* $\varepsilon > 0$ of the population in the ordered class k is moved from group ℓ to group h, while an equally small amount ε of the population in the ordered class $k+1$ is moved from group h to group ℓ, then dissimilarity is reduced. By small enough we mean that, after the population transfer, there is *no re-ranking* across groups: if group h dominates ℓ before the exchange, than group h should dominate (in a weak sense) ℓ even after the exchange, for all ℓ and for all h. The Exchange axiom formulates this principle in a more compact way:

Axiom *E* (Exchange) *For any* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$ *and* $\mathbf{A} \cdot \mathbf{e}_n = \mathbf{B} \cdot \mathbf{e}_n = \lambda \mathbf{e}_d$, *with* $\lambda \in \mathbb{R}_{++}$, *let* h *dominates* ℓ *in* k *in matrix* $\mathbf{A}$. *For* ε *small enough, if* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by an exchange such that (i)* $b_{hk} = a_{hk} + \varepsilon$, *(ii)* $b_{\ell k} = a_{\ell k} - \varepsilon$, *(iii)* $b_{ik} = a_{ik}\ \forall i \neq h, \ell$

[12] To make more explicit the link between the dissimilarity order based on direct/dual axioms and the general notion of dissimilarity, it suffices to see that the transformations induced by direct and dual axioms are the unique transformations that, when applied to a similarity matrix like **S** return another similarity matrix, with the same characteristics that the columns/rows of the matrix are one proportional to the other.

and (iv) $\overrightarrow{\mathbf{b}}_j = \overrightarrow{\mathbf{a}}_j$ $\forall j \neq k$ *then* $\mathbf{B} \preccurlyeq \mathbf{A}$.

The exchange axiom points out that the dissimilarity comparisons are meaningful only when groups sizes are fixed, not only among the matrices under comparison, but also across groups within the same distribution matrix. Hence, it is also possible to interpret the exchange of ε units as an exchange of an *absolute* population measure either across groups or across classes. This assumption is implicit in Fusco and Silber (2011).

By construction, the *MG* and the *E* axioms are natural candidates for defining the dissimilarity order when the classes are non permutable.

1.4 Characterization of dissimilarity orders: permutable classes

The assessment of segregation or socioeconomic mobility are related to non-ordered dissimilarity comparisons of distribution matrices where classes are not ordered.

In the non-ordered setting, one can construct any possible cumulative absolute frequency distribution by permuting the order of the classes of the distribution matrix. Hence, the analysis should focus on comparisons of frequencies distributions rather than on their cumulations. Consider the two groups case ($d = 2$). The set of direct axioms induces sequences of operations on the data that reduce the total variational distance of the two distributions: a requirement already stated in Gini (1914, 1965). The most common dissimilarity index satisfies indeed these axioms.

Remark 1.1 For $\mathbf{A} \in \mathcal{M}_2$ and $\mathbf{A} \cdot \mathbf{e}_n = \mathbf{e}_d$, the *Dissimilarity Index* (Duncan and Duncan 1955, Gini 1965) $D(\mathbf{A}) := \frac{1}{2}\sum_{j=1}^{n} |a_{1,j} - a_{2,j}|$ induces a complete order that satisfies the axioms *IEC*, *IPC*, *SC* and *MC*.

Axioms *IEC*, *IPC* and *SC* define a set of equivalent conditions for the dissimilarity order. By applying any sequence of the transformations underlying these axioms we obtain a set of matrices that are equally "dissimilar". We characterize such a class and then we show in a more general theorem that, by adding *MC* it is possible to state the equivalence between row stochastic majorization and the partial order induced by the elementary operations of merging and splitting. The *IEC* axiom plays a central role in the proof: it allows to modify the number of classes while preserving the order, thus generating empty slots where proportional splits can be reallocated without effects. In fact, the split transformation entails a merge between a proportional split and an empty class. Moreover, under *IEC*, the operations involved by axioms *SC* and *MC* admit a representation through a row stochastic matrix. Finally, the permutation axiom *IPC* is crucial for analyzing the case where classes are non-ordered.

1.4.1 The equivalence with Matrix Majorization

We first prove that any sequence of operations underlying the axioms *IEC, IPC, SC* is equivalent to adopt a specific class of row stochastic matrices $\widehat{\mathcal{R}}_{n_A,n_B}$, generating indifference sets.

Definition 1.2 *The set* $\widehat{\mathcal{R}}_{n_A,n_B} \subset \mathcal{R}_{n_A,n_B}$ *with* $n_A \leq n_B$ *contains all the row stochastic matrices with at most a non-zero entry in each column.*

In order to investigate the dissimilarity equivalence relation we consider matrices that can index the indifference sets.

Definition 1.3 *The set* $\mathcal{M}_d^I \subset \mathcal{M}_d$ *contains all matrices that neither exhibit empty classes nor pairs of adjacent classes that are proportional.*

Next lemma provides the characterization of dissimilarity equivalence sets when only *IEC*, *IPC* and *SC* are assumed to hold.

Lemma 1.1 *Let* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *with* $\mathbf{A} \in \mathcal{M}_d^I$ *and* $n_A \leq n_B$*, the dissimilarity order* $\preccurlyeq$ *satisfies IEC, IPC, SC if and only if*

$$\mathbf{B} \sim \mathbf{A} \;\Leftrightarrow\; \mathbf{B} = \mathbf{A} \cdot \widehat{\mathbf{X}} \text{ for some matrix } \widehat{\mathbf{X}} \in \widehat{\mathcal{R}}_{n_A,n_B}.$$

Proof. See appendix 1.A.1. ■

The result can be generalized to hold for comparisons of matrices $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$, not necessarily belonging to $\mathcal{M}_d^I$.

Corollary 1.1 *Let* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$*, the dissimilarity order* $\preccurlyeq$ *satisfies IEC, IPC, SC if and only if* $\mathbf{B} \sim \mathbf{A}$*, which is equivalent to having that there exists* $\mathbf{A}' \in \mathcal{M}_d^I$ *where* $n_{A'} \leq n_B$*, and* $n_{A'} \leq n_A$ *such that* $\mathbf{B} = \mathbf{A}' \cdot \widehat{\mathbf{X}}$ *and* $\mathbf{A} = \mathbf{A}' \cdot \widehat{\mathbf{X}}'$ *where* $\widehat{\mathbf{X}} \in \widehat{\mathcal{R}}_{n_{A'},n_B}$ *and* $\widehat{\mathbf{X}}' \in \widehat{\mathcal{R}}_{n_{A'},n_A}$.

This derivation is obtained by exploiting the transitivity of the indifference relation in Lemma 1.1 and the fact that by construction a matrix cannot belong to the equivalence class indexed by two different matrices $\mathbf{A}'', \mathbf{A}' \in \mathcal{M}_d^I$.

Making use of Axiom *MC* we introduce a new type of operation that allows to characterize the dissimilarity pre-order in terms of matrix majorization. This result allows to decompose the operations via row stochastic matrices in a series of splits and merges of population masses involving only two classes at a time.

Theorem 1.1 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *with* $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$*, the dissimilarity order* $\preccurlyeq$ *satisfies IEC, IPC, SC and MC if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \ \Leftrightarrow\ \mathbf{B} = \mathbf{A} \cdot \mathbf{X} \text{ for some matrix } \mathbf{X} \in \mathcal{R}_{n_A,n_B}.$$

Thus:

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad \mathbf{B} \preccurlyeq^R \mathbf{A}.$$

Proof. See appendix 1.A.2. ■

The theorem states that the operations underlying the axioms *MC* and *SC*, performed without requiring any particular order of their sequence, allow to transform $\mathbf{A}$ into $\mathbf{B}$ while reducing dissimilarity, and that these operations admit an equivalent representations trough Dahl's (1999) matrix majorization order. Hence, requiring group independent proportional transfers across classes amounts to require that the dissimilarity order respects the informativeness criterion in Blackwell (1953), which is taken as the motivating notion behind the concept of decreasing dissimilarity.[13]

[13]This equivalence provides strong support for interpreting segregation as a form of dissimilarity when classes are non ordered. The multi-group segregation ordering in Frankel and Volij (2011) is indeed the

The dissimilarity order characterized by matrix majorization has very useful properties. The indifference class contains all matrices that can always be obtained one from the other through multiplication by a row stochastic matrix.

Remark 1.2 By exploiting Theorem 1.1, $\mathbf{B} \sim \mathbf{A}$ if and only if $\exists \mathbf{X} \in \mathcal{R}_{n_A,n_B},\ \mathbf{X}' \in \mathcal{R}_{n_B,n_A}$ such that $\mathbf{B} = \mathbf{A}\cdot\mathbf{X}$ and $\mathbf{A} = \mathbf{B}\cdot\mathbf{X}'$. This is the case only if $\mathbf{A}$ and $\mathbf{B}$ satisfy the conclusions in Corollary 1.1.

The perfect similarity is achieved without posing any restriction on the distributional heterogeneity of each single group, but rather by equalizing distributional heterogeneity across groups. We can in fact obtain the matrix $\mathbf{S}$ from $\mathbf{A}$ by a sequence of splits, insertion of empty classes, permutations and merges operations involving classes. Matrix $\mathbf{C}$ is obtained from $\mathbf{A}$ by merging all classes and splitting them according to a sequence of λs.

Remark 1.3 Let $\mathbf{A} \in \mathcal{M}_d$ and consider $\mathbf{C} := (\lambda_1 \mathbf{A}\cdot\mathbf{e}_{n_A}, \ldots,\ \lambda_{n_A}\mathbf{A}\cdot\mathbf{e}_{n_A})$, with $\lambda_j \geq 0$ $\forall j$ and $\sum_j \lambda_j = 1$, then, $\mathbf{C} \preccurlyeq \mathbf{A}$.

Univariate comparisons in the dissimilarity order are meaningless. Moreover, any two different matrices that display perfect similarity among rows are ranked as indifferent by the dissimilarity order.

Remark 1.4 If $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_1$, then $\mathbf{B} \preccurlyeq \mathbf{A}$ if and only if $\mathbf{A} \preccurlyeq \mathbf{B}$. This is because there always exists a matrix $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ with $\mathbf{X} = \frac{1}{\mathbf{B}\cdot\mathbf{e}_{n_B}}(\mathbf{e}_{n_A}b_{1,1},\ldots,\mathbf{e}_{n_A}b_{1,n_B})$ such that $\mathbf{B} = \mathbf{A}\cdot\mathbf{X}$ and there always exists a matrix $\mathbf{Y} \in \mathcal{R}_{n_B,n_A}$ with $\mathbf{Y} = \frac{1}{\mathbf{A}\cdot\mathbf{e}_{n_A}}(\mathbf{e}_{n_B}a_{1,1},\ldots,\mathbf{e}_{n_B}A_{1,n_A})$ such that $\mathbf{A} = \mathbf{B}\cdot\mathbf{Y}$.

1.4.2 Extensions

The result in Theorem 1.1 applies to any pair of matrices with fixed number of rows. In general, matrices in $\mathcal{M}_d$ may well represent absolute frequencies distributions across

result of splitting and merging operations between permutable classes without making the transfer operation sensitive to the name of the groups. This is of practical use to the policymaker that cannot target (or give priority) to some particular groups over the others. Moreover, Theorem 1 is related to results on the analysis of intrinsic attitudes toward information and risk (Grant et al. 1998) and provides insights on the construction of bivariate dependence orderings for unordered categorical variables, as discussed in Giovagnoli, Marzialetti and Wynn (2009).

classes. Nevertheless, in many economic applications we are interested in matrices representing conditional (relative) distributions of frequencies. It seems natural to argue that the dissimilarity should be invariant to *proportional* replications of the overall population under analysis, or even more, dissimilarity should be independent from the relative size of *each group*. That is, one can always freely scale the population of one group while leaving the others unchanged, such that the overall dissimilarity is not affected, provided that the relative distribution across classes remains unchanged.[14]

In the dual setting the perspective is shifted on operations defined over groups rather than classes. Dissimilarity should remain constant if the population of each class is proportionally replicated by the same factor, or even more, it should be independent to the relative size of *each class*. That is, the overall dissimilarity does not change if one scales the population in each class, provided that the relative distribution of groups within classes remains unchanged.[15] The two different standardization concepts are resumed in the following *Normalization* axiom for groups (NG) and classes (NC) axioms.

For $\mathbf{c} \in \mathbb{R}^d_{++}$, the operator diag($\mathbf{c}$) generates a $d \times d$ identity matrix whose elements along the diagonal are replaced by the corresponding elements of $\mathbf{c}$.

Axiom *NG/NC (Normalization of Data)* *Let* $\mathbf{A} \in \mathcal{M}_d$, $\mathbf{c} \in \mathbb{R}^d_{++}$ *and* $\mathbf{d} \in \mathbb{R}^n_{++}$. *Let* $\mathbf{C} := diag(\mathbf{c})$, $\mathbf{D} := diag(\mathbf{d})$ *then:*

$$(\mathbf{NG}) \quad [diag(\mathbf{c})]^{-1} \cdot \mathbf{A} \sim \mathbf{A} \qquad \textit{and} \qquad (\mathbf{NC}) \quad \mathbf{A} \cdot [diag(\mathbf{d})]^{-1} \sim \mathbf{A}.$$

The axiom NG implies that the assessments of dissimilarity are neutral with respect to the differences in the groups overall population size. The axiom NC implies an analogous conclusion concerning the size of classes. By assuming normalization, it is possible to compare sets of distributions with different demographic size. This enforces the idea that dissimilarity is a relative concept boosting indifference with respect to structural changes

[14]This is the equivalent of the Composition Invariance axiom in Frankel and Volij (2011).
[15]This is the equivalent of the Group Division Property in Frankel and Volij (2011).

in the demographic composition of groups or classes that leave unchanged the overall distribution of population across groups or across classes. The following corollary states that when the dissimilarity comparison rests upon the direct axioms, the matrices that differ in size can be made comparable through the axiom *NG*.

Corollary 1.2 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *with* $\boldsymbol{\mu}_A = \mathbf{A} \cdot \mathbf{e}_{n_A}$, $\boldsymbol{\mu}_B = \mathbf{B} \cdot \mathbf{e}_{n_B}$ *and* $\boldsymbol{\mu}_A, \boldsymbol{\mu}_B \in \mathbb{R}^d_{++}$, *the dissimilarity order* $\preccurlyeq$ *satisfies IEC, IPC, SC, MC, and NG if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad [diag(\boldsymbol{\mu}_B)]^{-1} \cdot \mathbf{B} \preccurlyeq^R [diag(\boldsymbol{\mu}_A)]^{-1} \cdot \mathbf{A}.$$

Proof. See appendix 1.A.3. ■

The dissimilarity comparisons can also be made independent on the groups labels, while only the groups conditional distributions should matter. The *IPG* axiom points in this direction by enlarging the indifferent class induced by Theorem 1.1 to all the groups permutations of the distribution matrices under analysis.

Corollary 1.3 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *and a permutation matrix* $\boldsymbol{\Pi}_d \in \mathcal{P}_d$ *(different form the identity matrix) such that* $\mathbf{B} \cdot \mathbf{e}_{n_B} = \boldsymbol{\Pi}_d \cdot \mathbf{A} \cdot \mathbf{e}_{n_A}$, *the dissimilarity order* $\preccurlyeq$ *satisfies IEC, IPC, SC, MC, and IPG if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad \exists \boldsymbol{\Pi}_d : \quad \mathbf{B} \preccurlyeq^R \boldsymbol{\Pi}_d \cdot \mathbf{A}.$$

Proof. See appendix 1.A.4. ■

By reversing the role of rows and columns in the distribution matrices, it is possible to use the previous results to characterize the dissimilarity order based solely on dual axioms, while maintaining the permutability of classes given by *IPC*. Not surprisingly, the dual axioms altogether induce Dahl's (1999) matrix majorization order for the transpose of the distribution matrices. In this case, the operations of mixing of groups can be interpreted as proportional movements of populations masses between groups occurring within the same class. The information dispersion is reduced by making classes look more similar with respect to their relative group composition.

Corollary 1.4 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$*, let* $\boldsymbol{\nu}_A = \mathbf{A}^t \cdot \mathbf{e}_n$ *and* $\boldsymbol{\nu}_B = \mathbf{B}^t \cdot \mathbf{e}_n$*, the dissimilarity order* $\preccurlyeq$ *satisfies IEG, IPG, SG, MG, NC and IPC if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad \exists \boldsymbol{\Pi}_n \in \mathcal{P}_n : \quad [diag(\boldsymbol{\nu}_B)]^{-1} \cdot \mathbf{B}^t \preccurlyeq^R \boldsymbol{\Pi}_n \cdot [diag(\boldsymbol{\nu}_A)]^{-1} \cdot \mathbf{A}^t.$$

Proof. By applying Theorem 1.1 and Corollary 1.2 and 1.3 to matrices with $n_A = n_B$. ■

1.4.3 Robustness to lower dimensional comparisons

Hasani and Radjabalipour (2007) described the linear operator preserving matrix majorization. We make use of their main result to show that the dissimilarity order is preserved when some of the d groups in matrices $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ are mixed together with fixed weights thus generating $d' < d$ groups. The result also suggests that the dominance among d' groups does not guarantee the dominance for the larger set of d groups from whom they are obtained.

Remark 1.5 If the dissimilarity order $\preccurlyeq$ satisfies *IPC, IEC, SC, MC, NG*, then any mixing of groups (rows) preserves $\preccurlyeq$. Let $\widehat{\mathbf{X}} \in \widehat{\mathcal{R}}_{d',d}$ with $d' < d$, if $\mathbf{B} \preccurlyeq \mathbf{A}$, then $\widehat{\mathbf{X}} \cdot \mathbf{B} \preccurlyeq \widehat{\mathbf{X}} \cdot \mathbf{A}$.

This remark can be verified immediately, by exploiting the example reported in the introduction. The traditional analysis based on two groups comparisons (extensively exploited in empirical literature on segregation measurement, see for instance Flückiger and Silber 1999) may well indicate $\mathbf{B}$ as less dissimilar than $\mathbf{A}$ for any pair of groups, although this is not sufficient to guarantee that $\mathbf{B}$ is obtained by $\mathbf{A}$ through a sequence of dissimilarity reducing operations. This result reinforces the idea that dissimilarity is a global construct and partial comparisons may at most serve to determine the direction of dissimilarity within the distributions involved in the comparisons.

The remark may alternatively be exploited to assess the causes of dissimilarity. Suppose that one is interested in assessing the degree of dissimilarity between the distribution of male and female workers (groups) across n occupations (that is, occupational segregation). Consider the case where the population can be split into $d = 3$ ethnic groups. If a policymaker implements a reduction of dissimilarity in ethnic segregation on the labor marker, while leaving unaffected the male/female participation rate by ethnic group (although rates may differ between groups), which is also constant between occupations, one can additionally forecast the effect of the policy in terms of reduction of the gender based dissimilarity in occupational access.

1.5 Characterization of dissimilarity orders: non-permutable classes

In this section we study how dissimilarity comparisons can be constructed in the *ordinal setting*, that is when classes are meaningfully ordered and thus are not permutable. This is the case for instance when classes identify educational or health achievements or even contiguous income intervals.

Our results will allow to deal with comparisons between distribution matrices that differ in the number of classes and also in their interpretation. For instance, one will be able to compare the dissimilarity in the distribution of groups across health statuses between two countries, even if the health scales differ across the two countries. Alternatively, the policymaker guided by dissimilarity concerns may assess the priority of intervention between competing policies for health or schooling by assessing whether the distributions of health across social groups are more or less dissimilar than the distributions of educational achievements.

Within the ordered setting we will maintain the assumptions that the split and the insertion/elimination of empty classes preserve dissimilarity, as stated in the *SC* and *IEC* axioms, while obviously we will disregard the independence from permutation (*IPC*) property. The retained assumptions are associated to transformations of the distribution matrices that preserve the ordinal information, given that a proportional rescaling of some classes would not induce additional distributive information.

In this setting, one can construct cumulative distribution matrices and exploit the underlying information. Moreover, the splitting of classes allows to represent each row i of any cumulative distribution matrix in $\mathcal{M}_d$ by a continuous piecewise linear cumulative distribution function F_i. To see this, note that infinitely splitting a class is equivalent to assume that each group i is uniformly distributed within that class.

More generally, any monotonic continuous function can be derived as the limit of a sequence of step functions (see ch. 1 in Asplund and Bungart 1966). In our case the limit construction involves simultaneously all distribution functions of the groups. Considering the partition in n classes, by letting $(x_{k-1}; x_k]$ denote the interval related to class k, and $F_i(x_k) := \overrightarrow{a}_{i,k}/\overrightarrow{a}_{i,n}$ denote the value of the *cdf* of group i in x_k, we can construct the set of all *cfd*s associated with matrix $\mathbf{A}$ by setting $F_i(x) = F_i(x_k)$ for all $x \in [x_k; x_{k+1})$ with $F_i(x) = 0$ for $x \leq x_0$ and $F_i(x) = 1$ for $x \geq x_n$.

The sequence of splitting operations that leads to the desired result requires that for each splitting involving class k in matrix $\mathbf{A}$ such that each group is split into two adjacent classes k' and k with proportions λ and $(1-\lambda)$ respectively, it is identified a value $x_{k'}$ that partitions the associated interval into $(x_{k-1}; x_{k'}]$ and $(x_{k'}; x_k]$. The value of $x_{k'}$ should be set such that $\frac{x_{k'}-x_{k-1}}{x_k-x_{k-1}} = \lambda = \frac{\overrightarrow{a}_{i,k'}-\overrightarrow{a}_{i,k-1}}{\overrightarrow{a}_{i,k}-\overrightarrow{a}_{i,k-1}} = \frac{F_i(x_{k'})-F_i(x_{k-1})}{F_i(x_k)-F_i(x_{k-1})}$ where the equivalences on the right hand side hold by construction. In order to construct the sequence leading to the uniform *cdf*s within each class it suffices to apply a $\lambda = 0.5$ split in each class, then relabel all obtained $2n$ classes and reiterate the procedure.

In the next section we exploit this representation to introduce the dissimilarity test that we characterize later on.

1.5.1 The rationale behind the dissimilarity comparison

Consider the case where $\mathbf{A}, \mathbf{B} \in \mathcal{M}_3$. Each distribution matrix generates a set of three *cdf*s, denoted by $F_1,\ F_2,\ F_3$ for $\mathbf{A}$ and $F_1',\ F_2',\ F_3'$ for $\mathbf{B}$. The number of classes may vary between $\mathbf{A}$ and $\mathbf{B}$. These *cdf*s are represented in figure 2.4 with solid lines, respectively in the left and right panel of the figure. The dashed lines represent the graph of the *cdf*s of the overall populations, denoted respectively $\overline{F}$ and $\overline{F}'$ and obtained by the arithmetic mean of the rows of distribution matrices $\mathbf{A}$ or $\mathbf{B}$. Thus for instance $\overline{F} = \frac{1}{3}(F_1 + F_2 + F_3)$.

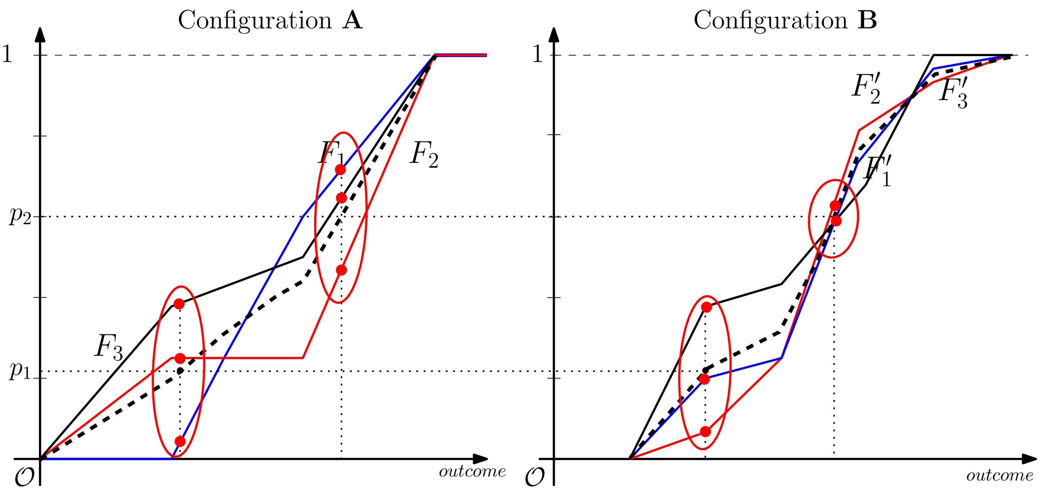


Figure 1.2: *Cdf*s F_1, F_2 and F_3 and an illustration of the dissimilarity test when classes are non-permutable.

The dissimilarity order with non-permutable classes entails the evaluation, based on the Lorenz dominance criterion, of the dispersion of the *cdf*s F_1, F_2, F_3 and F_1', F_2', F_3' around their respective averages $\overline{F}$ and $\overline{F}'$, at any fixed share $p \in (0, 1)$ of the overall population.

To understand the mechanics of the dissimilarity comparison, lets consider two population percentiles, denoted by p_1 and p_2 in figure 2.4. At p_1, we consider the values of F_1, F_2, F_3 at the quantile corresponding to $\overline{F} = p_1$ and of F_1', F_2', F_3' at the quantile corresponding to $\overline{F}' = p_1$. These values are identified with a marked dot in the figure. The dispersion between the dots corresponding to p_1 in configuration **A** is larger than the dispersion of the dots associated to configuration **B**, evaluated for the respective values corresponding to p_1. Recalling that the average of the values of the *cdf*s in the dots is by construction the same in both graphs, this conclusion can be reached by checking that the dots in configuration **B** Lorenz dominate those of configuration **A** at p_1.

A similar conclusion applies for analogous comparisons made at p_2, where the reduction in dispersion from the first to the second configuration is even more evident.

Extending the comparison to any $p \in (0, 1)$, it is possible to check that the dispersion between *cdf*s F_1', F_2', and F_3' evaluated at p is lower than the dispersion of F_1, F_2, and F_3 at the same p.

Because the cumulative distribution functions are continuous and piecewise linear, the dissimilarity test can be performed by looking only at a *finite* number of points, notably those corresponding to cases where either there is a movement from a class to the adjacent in one or both the distribution matrices **A** and **B** (as for p_1), or where two or more *cdf*s cross for (at least) one of the matrices (as for **B** in p_2). We show that the direct dissimilarity preserving axioms together with the exchange property provide a full characterization of the dissimilarity order.

1.5.2 The controversial role of the merge axiom in the ordinal setting

The Axiom *MC* may lead to problematic and counterintuitive results if it is maintained in the ordinal setting. To see this, consider the following distribution matrix for groups 1 and 2 across classes, representing for instance four ordered categories of health status (from bad to good).

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.1 & 0.4 & 0.1 \\ 0.1 & 0.4 & 0.1 & 0.4 \end{pmatrix}.$$

Group 1 is always disadvantaged compared to group 2, because the share of population with health status equal or lower than j, with $j = 1, \ldots, 4$, is always higher in groups 1 than it is in groups 2, that is the distribution of group 2 first order stochastically dominates the distribution of group 1. However, in class two these differences are somehow compensated. In fact, in this class the proportion of individuals with lower or equal health status is equal to $0.5 = 0.4 + 0.1$ for both groups.

Suppose now that the central classes two and three are merged together and then splitted proportionally to obtain again four classes, giving matrix $\mathbf{A}'$. According to the axioms *MC* and *SC* this operation leads to an unambiguous reduction in dissimilarity. However, the operation has a main drawback: while it leaves unaffected the stochastic dominance relation between groups, it eliminates any form of compensation taking place in the classes two and three. This aspect becomes evident if we compare the matrices obtained by cumulating the elements of $\mathbf{A}$ and $\mathbf{A}'$, that is:

$$\overrightarrow{\mathbf{A}} = \begin{pmatrix} 0.4 & 0.5 & 0.9 & 1 \\ 0.1 & 0.5 & 0.6 & 1 \end{pmatrix} \quad \text{and} \quad \overrightarrow{\mathbf{A}'} = \begin{pmatrix} 0.4 & 0.65 & 0.9 & 1 \\ 0.1 & 0.35 & 0.6 & 1 \end{pmatrix}.$$

It then appears, by comparing the column associated to the second class, that the distance between the cumulated populations has increased in $\mathbf{A}'$ with respect to $\mathbf{A}$. Therefore, it can hardly be argued that $\mathbf{A}'$ shows less dissimilarity than $\mathbf{A}$ as implied by the MC and SC axioms.

We propose alternatives to overcome the implications of the MC axiom by developing our arguments as follows. We firstly limit the analysis to a subset of distribution matrices with fixed number of classes, fixed average population distribution across these classes and given ranking of the groups distributions. We define these matrices as *ordinal comparable*. This class can be extended to all distribution matrices in $\mathcal{M}_d$ by resorting on a set of transformations that only preserve ordinal information of the data, but which allow to construct a more formal definition of the dissimilarity order presented in the previous section. Secondly, we show a full characterization of the dissimilarity order that relies on the transformations underlying the Exchange, rather than the Merge axiom.

1.5.3 Dissimilarity preserving "ordinal" information: definition

We say that the *rank* of groups is *preserved* across the classes of $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ if $\overrightarrow{a}_{\ell,k} \geq \overrightarrow{a}_{h,k}$ implies $\overrightarrow{a}_{\ell,k+1} \geq \overrightarrow{a}_{h,k+1}$ as well as $\overrightarrow{b}_{\ell,k} \geq \overrightarrow{b}_{h,k}$, which in turn implies $\overrightarrow{b}_{\ell,k+1} \geq \overrightarrow{b}_{h,k+1}$ for any pair of groups h, ℓ and for any class k.[16] This notion is incorporated in the following definition of ordinal comparability of distribution matrices:

Definition 1.4 (Ordinal comparability) *The matrices* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *are ordinal comparable if and only if (i)* $n_A = n_B = n$, *(ii)* $\mathbf{e}_d^t \cdot \mathbf{A} = \mathbf{e}_d^t \cdot \mathbf{B}$, *(iii)* $\mathbf{A} \cdot \mathbf{e}_n = \mathbf{B} \cdot \mathbf{e}_n = \lambda \cdot \mathbf{e}_d$ *with* $\lambda \in \mathbb{R}_{++}$ *and (iv) the rank of groups is preserved across classes.*

[16] Two groups ℓ and h in configuration $\mathbf{A}$ may swap positions in the rank defined by groups cumulative masses when moving from class $k-1$ to $k+1$, but this occurs if and only if $\overrightarrow{a}_{\ell,k} = \overrightarrow{a}_{h,k}$. A similar arguments holds for configuration $\mathbf{B}$.

Ordinal comparability narrows the set of comparison matrices, as well as the number of admissible transformations. By using the operations of split and insertion/elimination of empty classes underlying the axioms *SC* and *IEC*, for any pair of matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ that may not be ordinal comparable, it is possible to construct pairs of distribution matrices $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ with $n_A^* = n_B^* = n^*$ that are ordinal comparable. The process involves separate transformations for $\mathbf{A}$ and $\mathbf{B}$ that, eventually, lead to two *minimal ordinal comparable* matrices $\mathbf{A}^*, \mathbf{B}^*$ with equal number and size of classes such that $\mathbf{A} \cdot \mathbf{e}_{n^*} = \mathbf{B} \cdot \mathbf{e}_{n^*}$ and such that the rank of groups is preserved. This is the case if and only if the pair $\mathbf{A}^*$, $\mathbf{B}^*$ satisfies the four conditions in the definition below.

Definition 1.5 (Minimal ordinal comparability) *The matrices* $\mathbf{A}^*$, $\mathbf{B}^* \in \mathcal{M}$, *with* $n_A^* = n_B^* = n^*$ *and classes indexed by* $k = 1, \ldots, n^*$, *is derived from the pair* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}$, *where classes are indexed by* $j = 1, \ldots, n$ *if the following conditions are satisfied for any pair of groups* h, ℓ*:*

(i) $\forall j: \sum_{k=n_{j-1}}^{n_j} \mathbf{a}_k^* = \mathbf{a}_j$ *and* $\sum_{k=n'_{j-1}}^{n'_j} \mathbf{b}_k^* = \mathbf{b}_j$, *where* $n_0 = 1$ *and possibly* $n'_j \neq n_j$*;*

(ii) *if* $\left(\overrightarrow{a^*}_{h,k-1} - \overrightarrow{a^*}_{\ell,k-1}\right) \cdot \left(\overrightarrow{a^*}_{h,k+1} - \overrightarrow{a^*}_{\ell,k+1}\right) < 0$ *then* $\overrightarrow{a^*}_{h,k} - \overrightarrow{a^*}_{\ell,k} = 0$*;*

(iii) *if* $\left(\overrightarrow{b^*}_{h,k-1} - \overrightarrow{b^*}_{\ell,k-1}\right) \cdot \left(\overrightarrow{b^*}_{h,k+1} - \overrightarrow{b^*}_{\ell,k+1}\right) < 0$ *then* $\overrightarrow{b^*}_{h,k} - \overrightarrow{b^*}_{\ell,k} = 0$*;*

(iv) $\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{b}^*}_k = \mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}^*}_k$, $\forall k = 1, \ldots, n^*$.

A mechanical but intuitive algorithm to transform the pair $\mathbf{A}, \mathbf{B}$ into the pair of associated minimal ordinal comparable matrices consists in defining the sequence of split and insertion of empty classes that, for a given pair of groups, allows to satisfy conditions (i) to (iii) above for that pair of groups, and then by reiterating the procedure for any pairs of groups. Having done this, one can eventually split (and increase the number of) classes of the resulting matrices consistently with what required in point (iv). An example with three groups clarifies the type of transformations underlying Definition 1.5. Consider the matrices $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_3$ denoted by:

$$\mathbf{A} = \begin{pmatrix} 0.1 & 0.9 \\ 0.4 & 0.6 \\ 0.5 & 0.5 \end{pmatrix} \qquad and \qquad \mathbf{B} = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0 & 0.7 \\ 0 & 0.2 & 0.8 \end{pmatrix}.$$

Using split and insertion of empty classes operations one can obtain the following minimal ordinal comparable matrices, where $n^* = 4$:[17]

$$\mathbf{A}^* = \begin{pmatrix} \frac{1}{2}0.1 & \frac{1}{3}\frac{1}{2}0.1 & \frac{2}{3}\frac{1}{2}0.1 & 0.9 \\ \frac{1}{2}0.4 & \frac{1}{3}\frac{1}{2}0.4 & \frac{2}{3}\frac{1}{2}0.4 & 0.6 \\ \frac{1}{2}0.5 & \frac{1}{3}\frac{1}{2}0.5 & \frac{2}{3}\frac{1}{2}0.5 & 0.5 \end{pmatrix} \qquad and \qquad \mathbf{B}^* = \begin{pmatrix} 0.2 & \frac{1}{3}0.3 & \frac{2}{3}0.3 & 0.5 \\ 0.3 & 0 & 0 & 0.7 \\ 0 & \frac{1}{3}0.2 & \frac{2}{3}0.2 & 0.8 \end{pmatrix}.$$

As required, $\mathbf{e}_3^t \cdot \mathbf{A}^* = \mathbf{e}_3^t \cdot \mathbf{B}^* = (0.5, 0.17, 0.33, 2)$.

The minimal ordinal comparable matrices $\mathbf{A}^*, \mathbf{B}^*$ have an equal number of classes n^*. The sizes of each class k, measured as the sum of groups frequencies, coincide in $\mathbf{A}$ and $\mathbf{B}$. Hence, it is possible to compare every class' entries in the two distribution matrices by resorting on Lorenz dominance, thus implementing the dissimilarity comparison described in section 1.5.1. Given $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, the *sequential uniform majorization* $\preccurlyeq^*$ (SUM hereafter) defines a partial order on the set of comparison matrices $\mathcal{M}_d$: $\mathbf{B} \preccurlyeq^* \mathbf{A}$ if and only if there exists $\mathbf{A}^*, \mathbf{B}^*$ with equal distribution of the overall population across classes such that the vector $\overrightarrow{\mathbf{b}^*}_k$ Lorenz dominates the vector $\overrightarrow{\mathbf{a}^*}_k$ for all $k = 1, \ldots, n^*$.

Definition 1.6 (The SUM order $\preccurlyeq^*$) *For any $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ such that $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \lambda \mathbf{e}_d$ with $\lambda \in \mathbb{R}_{++}$ and there exists $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ that are minimal ordinal comparable, then:*

$$\mathbf{B} \preccurlyeq^* \mathbf{A} \qquad \Leftrightarrow \qquad \overrightarrow{\mathbf{b}^*}_k^t \preccurlyeq^U \overrightarrow{\mathbf{a}^*}_k^t, \quad \forall k = 1, \ldots, n^*.$$

[17] In the example, $n_B < n_A$. By splitting the first class of $\mathbf{A}$ in two new classes with equal overall size, one obtains two matrices with three classes that accommodate requirement (iv) in Definition 1.5. However, by moving from class two to three of $\mathbf{B}$ there is re-ranking of groups two and one. In order to avoid this, we split class two in $\mathbf{B}$ according to the weight 1/3, such that $\overrightarrow{b^*}_{1,2} = \overrightarrow{b^*}_{2,2}$ as required in point (iii). An identical operation is performed to obtain $\mathbf{A}^*$, thus accommodating requirement (iv). We leave to the reader to verify that the conditions in Definition 1.5 applies to any of the remaining pairs of groups.

The SUM order implements the dissimilarity criterion described in Section 1.5.1 by exploiting the sequential uniform majorization for cumulative distribution matrices.[18] Consider for instance figure 2.4, the SUM pre-order allows to meaningfully compare distributions F_1, F_2 and F_3 with distributions F_1', F_2' and F_2' because it only require to perform a sequence of Lorenz dominance comparisons at equal population percentiles for $\overline{F}$ and $\overline{F}'$ respectively. In our case these percentiles are denoted by $p_k = \frac{1}{d}\mathbf{e}^t \cdot \overrightarrow{\mathbf{b}^*}_k = \frac{1}{d}\mathbf{e}^t \cdot \overrightarrow{\mathbf{a}^*}_k$.

1.5.4 Dissimilarity preserving "ordinal" information: characterization

The dissimilarity order with non-permutable classes rests on the *SC* and *IEC* axioms. Accepting these two axioms leads to relevant consequences. In fact, if there exist sequences of splits and insertions of empty classes that starting from $\mathbf{A}, \mathbf{B}$ allow to obtain $\mathbf{A}^*, \mathbf{B}^*$, then $\mathbf{A}^* \sim \mathbf{A}$ and $\mathbf{B}^* \sim \mathbf{B}$. These sequences can always be found, so that if $\mathbf{B} \preccurlyeq \mathbf{A}$ then equivalently $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$. We propose a formal proof of this in the next theorems. One direct implication of this is that the dissimilarity order allows to make comparisons where only the ordinal information is retained. Any cardinality assessment related to classes is lost by introducing the possibility of reshaping the number and size of the classes. If this aspect is accepted, then what remains is to illustrate an ordinal property that allows to rank distributions according to the SUM order.

The dual axioms, that define transformations of split and merge across groups, characterize an ordering that coincides with a particular case of SUM. These operations characterize the dissimilarity relation $\mathbf{B} \preccurlyeq \mathbf{A}$ in terms of matrix majorization. In fact, this is the result in Corollary 1.4, considering that $\mathbf{\Pi}_n$ can only be the identity matrix.

The associated matrix majorization test preserves the size of the *classes*, while changing their composition. In this case dissimilarity is grounded upon operations that do not, in general, preserve the size and the number of *groups*. However, by restricting the domain of comparison matrices to $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ that are also ordinal comparable, then Corollary 1.4

[18]Given that $\overrightarrow{\mathbf{b}^*}_k$ and $\overrightarrow{\mathbf{a}^*}_k$ are obtained from minimal ordinal comparable matrices then the sum of their elements for each k is the same for both vectors. Therefore uniform majorization for each k is equivalent to Lorenz dominance.

boils down to obtaining that there exists a sequence of operations underlying the axioms *IPG*, *IEG*, *SG* and *MG* that allows to obtain $\mathbf{B}$ from $\mathbf{A}$ if, and only if, $\mathbf{B}^t \preccurlyeq^U \mathbf{A}^t$. This is a particular case of matrix majorization, that implies SUM, but that it is not implied by the latter.[19]

The next result provides a more convincing ground for dissimilarity comparisons for ordered classes that can rank larger sets of dissimilarity matrices.

With the following theorem, we establish that $\mathbf{B} \preccurlyeq^* \mathbf{A}$ if and only if there exists at least one sequence of exchanges between pairs of groups within $\mathbf{A}^*$ that allows to obtain $\mathbf{B}^*$. The operations underlying the Exchange axiom are independent, and can be applied in any class of distribution matrices. However, the exchanges can only be performed on (minimal) ordinal comparable matrices. We show that for matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ that are at least *rank comparable*, transformations involving insertion/elimination of empty classes, split of adjacent classes and groups permutation allows to construct the respective minimal ordinal comparable matrices $\mathbf{A}^*, \mathbf{B}^*$, such that for any group ℓ, h it holds that $\overrightarrow{a^*}_{\ell,k} \geq (\leq) \overrightarrow{a^*}_{h,k}$ if and only if $\overrightarrow{b^*}_{\ell,k} \geq (\leq) \overrightarrow{b^*}_{h,k}$, for any class k. One special case of rank comparability occurs when groups can be ordered according to *stochastic dominance*.[20]

Theorem 1.2 *For any* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$ *that are rank comparable with* $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \lambda \mathbf{e}_d$, $\lambda \in \mathbb{R}_{++}$, *the dissimilarity order* $\preccurlyeq$ *satisfies IEC, SC, IPG, and E if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \qquad \Leftrightarrow \qquad \mathbf{B} \preccurlyeq^* \mathbf{A}.$$

Proof. See appendix 1.A.5. ■

[19] In fact, the SUM weakens the uniform majorization criterion in Corollary 1.4. According to SUM $\mathbf{B} \preccurlyeq^* \mathbf{A}$ if and only if $\forall k$ $\exists \mathbf{Y}_k \in \mathcal{D}_d$ such that $\overrightarrow{\mathbf{b}^*}^t_k = \overrightarrow{\mathbf{a}^*}^t_k \cdot \mathbf{Y}_k$. A special case is when $\mathbf{Y}_k = \mathbf{Y}$, $\forall k$. This gives in short notation $\overrightarrow{\mathbf{B}^*}^t = \overrightarrow{\mathbf{A}^*}^t \cdot \mathbf{Y}$. Recall that $\overrightarrow{\mathbf{A}^*}^t = \mathbf{D} \cdot (\mathbf{A}^*)^t$ where $\mathbf{D}$ denotes a lower triangular matrix. It follows that the dominance condition can be rewritten as $\mathbf{D} \cdot (\mathbf{B}^*)^t = \mathbf{D} \cdot (\mathbf{A}^*)^t \cdot \mathbf{Y}$, that is $(\mathbf{B}^*)^t = \mathbf{D}^{-1} \cdot \mathbf{D} \cdot (\mathbf{A}^*)^t \cdot \mathbf{Y} = (\mathbf{A}^*)^t \cdot \mathbf{Y}$ leading to $(\mathbf{B}^*)^t \preccurlyeq^U (\mathbf{A}^*)^t$. Thus, uniform majorization implies SUM, the reverse implication is, in general, not true.

[20] For $\mathbf{A} \in \mathcal{M}_d$, group h stochastically dominates group ℓ, $\forall h \neq \ell$, if and only if $\overrightarrow{a^*}_{h\ k} \leq \overrightarrow{a^*}_{\ell\ k}$, for all $k = 1, ..., n_A$. In this special case, the minimal ordinal comparable matrices are monotonic, up to a permutation of the groups.

Without additional structure, Theorem 1.2 does not allow to compare pairs of matrices where groups are not ordered in the same way for each class. We propose a novel axiom, at least to our knowledge, called *Interchange* of groups. It states that the interchange/permutation of two groups distributions for all classes $k > j$ preserves overall dissimilarity, provided that in class j the cumulative frequencies of the two groups are identical.

Axiom *I* (*Interchange of Groups*) *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *with* $n_A = n_B = n$*, if* $\exists \mathbf{\Pi}_{h,\ell} \in \mathcal{P}_d$ *permuting only group* h *and* ℓ*, such that* $\mathbf{B} = (\mathbf{a}_1, ..., \mathbf{a}_j, \mathbf{\Pi}_{h,\ell} \cdot \mathbf{a}_{j+1}, ..., \mathbf{\Pi}_{h,\ell} \cdot \mathbf{a}_{n_A})$ *whenever* $\overrightarrow{a}_{h,j} = \overrightarrow{a}_{\ell,j}$*, then* $\mathbf{B} \sim \mathbf{A}$.

The axiom enlarges the class of comparable matrices by eliminating all the concerns related to stochastic dominance relations between groups distributions. This is an appealing requirement, since stochastic dominance at order higher than the first entails a cardinal comparison, here excluded. Axiom *I* implicitly assumes that dissimilarity evaluations are separable across sets of adjacent classes where one group dominates another.

The result of Theorem 1.2 is then generalized as follows.

Theorem 1.3 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *such that* $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \lambda \mathbf{e}_d$*,* $\lambda \in \mathbb{R}_{++}$*, the dissimilarity order* $\preccurlyeq$ *satisfies IEC, SC, IPG, I and E if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad \mathbf{B} \preccurlyeq^* \mathbf{A}.$$

Proof. See appendix 1.A.6. ■

Finally, the result in Theorem 1.3 can be extended to all distribution matrices by exploiting the normalization axiom. The dissimilarity order is therefore defined as a comparison of relative distributions of groups across ordered classes.

Corollary 1.5 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *such that* $\boldsymbol{\mu}_A = \mathbf{A} \cdot \mathbf{e}_n$*,* $\boldsymbol{\mu}_B = \mathbf{B} \cdot \mathbf{e}_n$*, the dissimilarity order* $\preccurlyeq$ *satisfies IEC, SC, NG, IPG, I and E if and only if*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad [diag(\boldsymbol{\mu}_B)]^{-1} \cdot \mathbf{B} \preccurlyeq^* [diag(\boldsymbol{\mu}_A)]^{-1} \cdot \mathbf{A}.$$

Proof. See appendix 1.A.7. ■

1.6 Equivalent tests for the dissimilarity orders

The characterization of the dissimilarity order strongly relies on the matrix majorization order or, alternatively, on the sequential uniform majorization order when classes are ordered. However, given two distribution matrices, no algorithm is available to check the majorization relations (Marshall et al. 2011). In this section we determine equivalent tests for the matrix majorization pre-orders underlying the dissimilarity comparisons in the setting where classes are ordered or, alternatively, non ordered. We use test to indicate a pre-order based on the inclusion of Zonotopes or Monotone Paths, provided that this inclusion can be verified empirically. For instance, dominance in the sense of Lorenz curves is a test for uniform majorization, the partial ordering underlying inequality comparisons. Nevertheless, our analytical setting is more general than the Lorenz ordering.

1.6.1 Testing the dissimilarity order with permutable classes

We make use of the matrix majorization order by Dahl (1999) to characterize the Zonotopes inclusion order. We exploit this result to construct a test for the dissimilarity criterion when classes are permutable.

A general test for matrix majorization

The following theorem states that the order based on inclusion of the Zonotopes of the distribution matrices under comparison is equivalent to matrix majorization for those matrices with fixed size of the populations. Our result extends to the multi-group case the result in Dahl (1999), only valid for $d = 2$.

Theorem 1.4 *Let* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *such that* $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$*:*

$$\mathbf{B} \preccurlyeq^R \mathbf{A} \quad \Leftrightarrow \quad Z(\mathbf{B}) \subseteq Z(\mathbf{A}).$$

Proof. See appendix 1.A.8. ■

Remark 1.6 Note that the condition $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$ is implied by $\mathbf{B} \preccurlyeq^R \mathbf{A}$. The condition posits that $Z_D(\mathbf{A}) = Z_D(\mathbf{B})$, which is necessary to prove that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ implies $\mathbf{B} \preccurlyeq^R \mathbf{A}$.

The identification of Zonotopes inclusion with matrix majorization allows to depict properties of the majorization ordering directly from the analysis of the Zonotopes. The projection of the Zonotope on a lower dimensional space allows to reduce a d-variate problem (where $d \geq 3$) to a bivariate comparison that can be analyzed by mean of common instruments such as the Lorenz curve or the segregation curve. Zonotopes inclusion in the d-variate space is sufficient for inclusion of the Zonotope projections (which are indeed Zonotopes, see McMullen 1971) in a lower dimensional space, and therefore to matrix majorization of the projected data matrices. Interesting two groups comparisons include one-against-one or one-against-other groups projections. Nevertheless, the Zonotopes inclusion is not necessary for the inclusions of the Zonotopes' projections. The following example with $\mathbf{A}, \mathbf{B} \in \mathcal{M}_3$ confirms this point.

A Zonotope projection is obtained by *premultiplying* the initial distribution matrix by a row stochastic matrix $\mathbf{P} \in \mathcal{R}_{2,3}$ such that $\mathbf{P} \cdot \mathbf{B} \preccurlyeq^R \mathbf{P} \cdot \mathbf{A}$ if and only if $\mathbf{P} \cdot \mathbf{B} = \mathbf{P} \cdot \mathbf{A} \cdot \mathbf{X}$ with $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$. Excluding the cases where a group is projected against himself, any relevant matrix in the projection class of 2×3 matrices can be written as a convex combination of six zero-one row stochastic matrices, called $\mathbf{P}_1, \ldots, \mathbf{P}_6$. Suppose that it is possible to verify $\mathbf{P}_i \cdot \mathbf{B} \preccurlyeq^R \mathbf{P}_i \cdot \mathbf{A}$ for all is trough the inclusion of all bivariate Zonotopes. As a result, there exist a set $\mathbf{X}_i \in \mathcal{R}_{n_A,n_B}$ of majorization matrices, for $i = 1, \ldots, 6$, such that $\mathbf{P}_i \cdot \mathbf{B} = \mathbf{P}_i \cdot \mathbf{A} \cdot \mathbf{X}_i$. By taking a convex combination of both sides of the relation with $\alpha_i \in [0, 1]$ $\forall i$, we can check wether *any* projection of the Zonotope fulfills the inclusion by writing:

$$\sum_i \alpha_i \mathbf{P}_i \cdot \mathbf{B} = \sum_i \alpha_i \mathbf{P}_i \cdot \mathbf{A} \cdot \mathbf{X}_i.$$

Unless matrix $\mathbf{A}$ has some very particular properties (for instance, it is an identity matrix augmented by some empty columns) or there exist an $\mathbf{X}_i = \mathbf{X}\ \forall i$ that gives matrix majorization, it is not possible to infer Zonotopes inclusion in the d-variate space by looking at bivariate comparisons for a finite set of mixing weights. Multivariate Zonotopes inclusion is therefore a majorization test extremely robust to two groups comparisons when, for instance, aggregation weights differs across comparison matrices or are unknown to the researcher.[21]

The dissimilarity test

Based on Theorem 1.4, it is possible to construct a *test* for the dissimilarity partial orders presented in Theorem 1.1 and in Corollary 1.4. In both cases, we maintain the *IPC* axiom, given that comparisons are made in a setting where classes are not ordered. The following corollary resumes the equivalences in two distinct propositions, whose proofs directly follow as an application of Theorem 1.4.

Corollary 1.6 *Let* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$*, the dissimilarity order* $\preccurlyeq$ *is such that:*

(i) $\preccurlyeq$ *satisfies IPC, IEC, SC, MC, NG, IPG if and only if:*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad \exists \mathbf{\Pi}_d \in \mathcal{P}_d : \qquad Z\left([diag(\boldsymbol{\mu}_B)]^{-1} \cdot \mathbf{B}\right) \ \subseteq \ Z\left(\mathbf{\Pi}_d \cdot [diag(\boldsymbol{\mu}_A)]^{-1} \cdot \mathbf{A}\right).$$

(ii) *For* $n_A = n_B = n$*,* $\preccurlyeq$ *satisfies IPC, IPG, IEG, SG, MG, NC if and only if:*

$$\mathbf{B} \preccurlyeq \mathbf{A} \quad \Leftrightarrow \quad \exists \mathbf{\Pi}_n \in \mathcal{P}_n : \qquad Z\left([diag(\boldsymbol{\nu}_B)]^{-1} \cdot \mathbf{B}^t\right) \ \subseteq \ Z\left(\mathbf{\Pi}_n \cdot [diag(\boldsymbol{\nu}_A)]^{-1} \cdot \mathbf{A}^t\right).$$

Proof. The equivalence between direct (respectively, dual) axioms and matrix majorization is given by Corollary 1.3 (respectively, Corollary 1.4), while the result is a direct application of Theorem 1.4. ■

[21] Zonotopes inclusion is also a robust test with respect to the comparison of distributions obtained under a different grouping criterion. This is a direct implication of Remark 1.5 and Theorem 1.4.

If we do not consider axiom *IPG* in Corollary 1.6 (i) the result holds for $\mathbf{\Pi}_d$ coinciding with the identity matrix.[22] Given $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with groups of equal size such that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$, then $\mathbf{B}$ can be obtained from $\mathbf{A}$ by a sequence of splits, merges, insertions of empty classes and permutations of classes. If the distribution matrices $\mathbf{A}', \mathbf{B}' \in \mathcal{M}_d$ exhibit groups of different size, taking for granted *NG* is equivalent to consider the associated normalized matrices $\mathbf{A}, \mathbf{B}$ whose rows sum up to one. We develop the next argument within this setting.

To construct a parallel with the arguments developed in section 1.5, note that the condition $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ is equivalent to the inclusion of every section of the Zonotope of $\mathbf{B}$ into the respective section of the Zonotope of $\mathbf{A}$. This result holds, for instance, when sections are obtained from the hyperplane perpendicular to the perfect similarity Zonotope. This hyperplane's slopes coincide with a set of weights equal to $1/d$ and identifies sections of the Zonotopes where the overall population proportion is held constant and equal to $p \in [0, 1]$.

The test $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ is therefore equivalent to check that for every proportion p of the overall population, the corresponding groups' populations proportions are less dispersed in configuration $\mathbf{B}$ than they are in $\mathbf{A}$. In the permutable setting, dispersion is measured by the inclusion of the convex hull obtained by all possible splits and merges of the classes, corresponding to all the configurations of groups' shares that sum up to the same proportion p of the overall population. This convex hull is the section of the Zonotope delimited by the hyperplane at level p.

1.6.2 Testing the dissimilarity order with non-permutable classes

The dual Zonotopes inclusion test

The dissimilarity pre-order for non-permutable classes can be tested by Zonotopes inclusion, if it is characterized by the dual axioms. In fact, if $\mathbf{B} \preccurlyeq \mathbf{A}$ satisfies only the dual axioms, along with the requirement that the size of the groups is fixed among comparison matrices, then equivalently should hold that $\mathbf{B}^t \preccurlyeq^U \mathbf{A}^t$. This dominance relation can also be expressed

[22] A similar argument holds for the result in Corollary 1.6 (ii) if we drop *IPC*.

in terms of the following row stochastic majorization condition $(\mathbf{B}, \mathbf{e}_d)^t \preccurlyeq^R (\mathbf{A}, \mathbf{e}_d)^t$. This is the case only if the class of row stochastic matrices involved in the operation is restricted to those that are also doubly stochastic and belong to $\mathcal{D}_d$, as required by the uniform majorization condition. By Corollary 1.6 part (ii), it is possible to determine whether such doubly stochastic matrix exists by checking that $Z\left((\mathbf{B}, \mathbf{e}_d)^t\right) \subseteq Z\left((\mathbf{A}, \mathbf{e}_d)^t\right)$.

However, in many circumstances $d < n$, and the test is very likely to be rejected. This happens because one has to check the inclusion of Zonotopes, that in this case are d-dimensional bodies, in the n-dimensional space.

The test for dissimilarity based on SUM partial order

Classical results in majorization theory (Hardy et al. 1934, Marshall et al. 2011) allow to test the dissimilarity order characterized in Theorem 1.3 making use of sequential Lorenz dominance of the cumulative groups shares across the classes of the minimal ordinal comparable matrices.

For $\mathbf{A}^*,\ \mathbf{B}^* \in \mathcal{M}_d$, let $\mu(k) = \mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}^*}_k = \mathbf{e}_d^t \cdot \overrightarrow{\mathbf{b}^*}_k$ denote the class k sum of cumulated groups' populations, for all columns $k = 1, \ldots, n^*$. The SUM entails a sequence of univariate dissimilarity comparisons of the actual distribution of the cumulative groups frequencies (normalized by $\mu(k)$) and the uniform distribution (with all elements equal to $1/d$) of groups weights, reflecting the case in which groups are similarly distributed and their cumulative shares coincide. To avoid cumbersome notation, we assume that groups size is fixed and equal to one for all groups and across distribution matrices.[23]

Lemma 1.2 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *such that* $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \mathbf{e}_d$*,*

$$\mathbf{B} \preccurlyeq^* \mathbf{A} \quad \Leftrightarrow \quad Z\left(\left(\frac{\overrightarrow{\mathbf{b}^*}_k}{\mu(k)},\ \frac{\mathbf{e}_d}{d}\right)^t\right) \subseteq Z\left(\left(\frac{\overrightarrow{\mathbf{a}^*}_k}{\mu(k)},\ \frac{\mathbf{e}_d}{d}\right)^t\right), \quad \forall k = 1, \ldots, n^*.$$

Proof. See appendix 1.A.9. ■

[23] This restriction can be easily relaxed by introducing the *NC* Axiom, thus reflecting the result of Corollary 1.5.

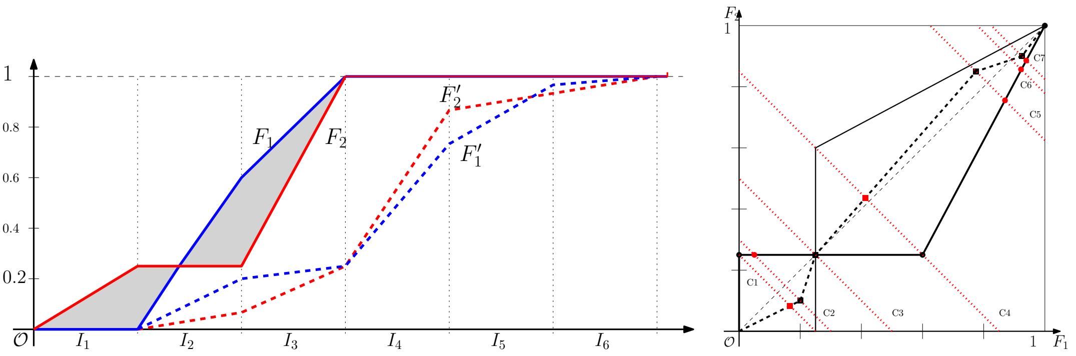
As shown in the proof, for n^* sufficiently large, the Lemma 1.2 may require to perform a long sequence of Lorenz dominance tests. Alternatively, we show that the dissimilarity order can be tested by checking the Path Polytopes inclusion order, which does not rely on the computation of the partitions underlying $\mathbf{A}^*$ (and $\mathbf{B}^*$).

This can be seen in an example involving only two groups. The distribution functions of these two groups in configuration (F_1, F_2) are represented by the continuous lines in figure 2.3, panel (a). These two distributions can be compared with the pair of distributions (F_1', F_2') represented with dashed lines in the same figure. To verify that configuration (F_1', F_2') is less dissimilar than (F_1, F_2), one has to derive the associated minimal ordinal comparable distributions and test the SUM order. These comparisons, however, can be directly assessed by looking at the inclusion of the Monotone Path of configuration (F_1', F_2') into the Path Polytope associated to (F_1, F_2). In panel (b) of figure 2.3 the Monotone Path is represented by the dashed line, while the Path Polytope coincides with the area between the two continuous lines. The verification of this inclusion is necessary and sufficient for the SUM criterion to hold. In fact, the (red) dotted parallel lines in the figure represent the population shares where SUM has to be tested. In this example, Lorenz dominance (and equivalently also uniform majorization) consists in verifying that the point on the dotted Monotone Path corresponding to a given population share is closer to the diagonal than it is the associated point on the continuous Path Polytope. Inclusion is therefore equivalent to test Lorenz dominance for all overall populations shares, and therefore also for those required for the SUM test. As shown in the proof of next theorem, when dominance is verified for the n^* population shares underlying the SUM test this also implies dominance for all populations shares. It follows also that the Path Polytope associated to (F_1', F_2') is included in the one of (F_1, F_2) as it is the case in the figure.

Theorem 1.5 *For any* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}$ *such that* $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B} = \mathbf{e}_d$*,*

$$\mathbf{B} \preccurlyeq^* \mathbf{A} \qquad \Leftrightarrow \qquad Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A}).$$

Proof. See appendix 1.A.10. ■



(a) *cdf*s (b) Monotone Path inclusion

Figure 1.3: *Cdf*s, Monotone Paths and the class division for fixed population masses.

Theorem 1.5 can be used to derive an alternative, but equivalent, representation of the comparison underlying figure 2.4. The information embedded in the *cdf*s $F_1,\ F_2,\ F_3$ and $F_1',\ F_2',\ F_3'$ is equivalently represented by their respective Monotone Paths in the three dimensional unitary hypercube. The hypothesis that the groups are uniformly distributed within classes plays no role in determining the shape of the Monotone Path, which is indeed generated under the assumption that any split of the classes and addition/elimination of empty classes preserve the degree of dissimilarity between *cdf*s.

The average population distributions $\overline{F}$ and $\overline{F}'$ are now measured by the value of the hyperplane orthogonal to the hypercube diagonal. Each hyperplane corresponds to a population percentile, held constant on the hyperplane surface. For one given population percentile there is a unique hyperplane that intersects the (monotonically increasing) Monotone Paths associated to the *cdf*s $F_1,\ F_2,\ F_3$ and the *cdf*s $F_1',\ F_2',\ F_3'$ only once, thus identifying a pair of points on the same hyperplane.

The dissimilarity order is verified if and only if, for any population percentile p, the point associated to the Monotone Path of *cdf*s $F_1',\ F_2',\ F_3'$ on the hyperplane of measure p, lies in the *Kolm triangle* constructed from the point associated to the Monotone Path of *cdf*s $F_1,\ F_2,\ F_3$ on the same hyperplane. This is an equivalent characterization of the Lorenz order in the case of three units.[24]

[24] In fact, for the case $d = 3$, the Lorenz dominance in class k can be equivalently checked by a test of inclusion of the vector $\overrightarrow{\mathbf{b}^*}_k$ into the hexagon generated by all the permutation of $\overrightarrow{\mathbf{a}^*}_k$, which lies in the

By construction, the boundaries of the Kolm triangle associated to any population percentile p defines the contour of the Path Polytope, when intersected with the hyperplane associated to the same percentile.[25] Therefore, the sequential inequality comparison at any population percentile can be equivalently represented by the inclusion $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$.

1.7 Related orders

1.7.1 Less dissimilar vs less spread out

Consider two n-variate vectors of data $\mathbf{a}, \mathbf{b} \in \mathcal{M}_1$ with $\mathbf{a} \cdot \mathbf{e}_n = \mathbf{b} \cdot \mathbf{e}_n = c > 0$. These vectors may well represent any type of distribution across n classes (for instance income distributed across n individuals). The univariate inequality order ranks vector $\mathbf{b}$ better that $\mathbf{a}$ if and only if the elements of $\mathbf{b}$ are "less spread out" than the elements of $\mathbf{a}$.

The notion of progressive (Pigou-Dalton) transfers among vectors classes is a well-known equity criterion invoked in univariate comparisons. It posits that, for $a_j > a_k$, inequality is reduced by operations involving a reduction of a_j by a quantity $\epsilon > 0$ and an equal increase of a_k by the same quantity, therefore preserving the overall amount c.

In the univariate settings, Marshall and Olkin (1979) showed that any Pigou-Dalton transfer occurring between two classes can be formalized through a matrix (i.e. linear) operation involving a *T-transform* $\mathbf{T}(\lambda, k, j)$. The vector $\mathbf{b}$ has been obtained by $\mathbf{a}$ through a Pigou-Dalton transfer between class j and k if and only if $\mathbf{b} = \mathbf{a} \cdot \mathbf{T}(\lambda, k, j)$ and $\mathbf{T}(\lambda, k, j) := \lambda \mathbf{I}_n + (1 - \lambda)\mathbf{\Pi}_{j,k}$, where $\lambda \in [0, 1]$ and $\mathbf{\Pi}_{j,k} \in \mathcal{P}_n$ is a permutation matrix of columns j and k. If this is the case, the degree of inequality in $\mathbf{b}$ is lower than the degree of inequality in $\mathbf{a}$.

simplex with vertices $(\mu(k), 0, 0)$, $(0, \mu(k), 0)$ and $(0, 0, \mu(k))$, as proposed by Kolm (1969). By considering all $k = 1, ..., n^*$, one obtains a sequence of hexagons (one for each value $\mu(k)$), which corresponds to the "contour curves" of the Path Polytope of $\mathbf{A}$, $Z^*(\mathbf{A})$, and calculated with respect to the class cumulative population, thus moving along the diagonal.

[25] For instance, in figure 1.3(b) the hyperplane in two dimension is represented by dotted lines perpendicular to the diagonal. These lines identifies only two points on the boundary of the Path Polytope: one associated to the Monotone Path and the other with its permutation. The Kolm triangle in this case coincides with the segment of the dotted line that lies within the Path Polytope. All points in this segments are clearly closer to the diagonal (represented similarity) than the two extremes.

In the univariate case, it is possible to represent any sequence of T-transforms transforming $\mathbf{a}$ into $\mathbf{b}$ by the order $\mathbf{b} \preccurlyeq^U \mathbf{a}$ (see ch.2, Lemma B.1 in Marshall et al. 2011). Unfortunately, a similar argument does not hold in the d-variate case (Kolm 1977).

We document the relation between dissimilarity and inequality at the multi-group level by showing that the elementary operations involved by Pigou-Dalton transfers, that characterize the inequality order, can be decomposed in a very particular *sequence* of split and merge operations.

In the dissimilarity framework presented here, a T-transform involves a proportional movement of population masses from two classes, which amounts to repeating twice a sequence of splits and merges. We equivalently represent a sequence of split and merge by the matrix $\mathbf{S}(\lambda, i, j) \in \mathcal{R}_{n_A,n_B}$. Given a matrix $\mathbf{A} \in \mathcal{M}_d$ with n columns, $\mathbf{S}(\lambda, k, j)$ is a row stochastic matrix that splits column k of $\mathbf{A}$ and merges a share $(1 - \lambda)$ of k with column j.[26]

Let assume without loss of generality that $\lambda \in [0, 0.5]$, it follows that any T-transform can be equivalently obtained by an ordered sequence of split and merge transformations concerning the same pair of classes:

$$\mathbf{T}(\lambda, k, j) := \mathbf{S}(\lambda', k, j) \cdot \mathbf{S}(\lambda'', j, k),$$

where the splitting parameters must satisfy $\lambda'' = 1 - \lambda$ and $\lambda' = \frac{1-2\lambda}{1-\lambda}$.

The next corollary shows how some interesting assumptions on $(d + 1)$-variate distributions allow to restrict the set of admissible transformations via row stochastic matrices and to characterize the relation with doubly stochastic matrices more in depth.

[26]Exploiting Theorem 1.1, the matrix can be written as:

$$\mathbf{S}(\lambda, k, j) := [\lambda\, (\mathbf{I}_n, \mathbf{0}_n) + (1 - \lambda)\, (\mathbf{I}_n, \mathbf{0}_n)\, \mathbf{\Pi}_{n+1,k}] \cdot \begin{pmatrix} \mathbf{I}_n \\ \mathbf{i}_{j,\cdot} \end{pmatrix},$$

where $\lambda \in [0, 1]$ and $\mathbf{\Pi}_{n+1,k} \in \mathcal{P}_n$ is a permutation matrix of columns $n + 1$ and k while $\mathbf{i}_{j,\cdot}$ corresponds to row j of the identity matrix.

Corollary 1.7 *Let* $\mathbf{A},\ \mathbf{B} \in \mathcal{M}_d$ *and let* $\preccurlyeq$ *satisfy axioms IEC, IPC, SC and MC. Consider:*

$$\mathbf{A}' = \begin{pmatrix} \frac{1}{n_A}\mathbf{e}_{n_A} \\ \mathbf{A} \end{pmatrix} \quad \preccurlyeq \quad \mathbf{B}' = \begin{pmatrix} \frac{1}{n_B}\mathbf{e}_{n_B} \\ \mathbf{B} \end{pmatrix}.$$

Then $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ *with* $\mathbf{X} \in \mathcal{R}_{n_A,n_B}$ *and* $\mathbf{e}^t_{n_A} \cdot \mathbf{X} = \frac{n_A}{n_B}\mathbf{e}^t_{n_B}$*. Moreover, if* $n_A = n_B = n$ *then* $\mathbf{X} \in \mathcal{D}_n$*.*

Proof. Note that each entry in the first row of $\mathbf{A}'$ is a constant equal to $1/n_A$. It can be transformed into the corresponding element in $\mathbf{B}'$, $1/n_B$, only by multiplying each single entry by n_A/n_B. The result is a consequence of Theorem 1.1. ■

For $d = 1$ and $n_A = n_B = n$, the doubly stochastic matrix $\mathbf{X} \in \mathcal{D}_n$ can be equivalently decomposed in a finite sequence of T-transforms, and therefore in a sequence of merge and split operations of classes. Hence, one can use $\mathbf{A}',\ \mathbf{B}'$ to study inequality comparisons.

Univariate equality is therefore a sufficient, but not necessary, condition to increase similarity. In fact similarity implies equalization of elements within each column of a distribution matrix and is achieved only by equalizing entries also between columns. Therefore, the dissimilarity order is constructed on more complex set of independent operations than the ones characterizing the dissimilarity comparison entailed by the univariate inequality order. What turns out from the Corollary 1.7 is that inequality comparisons can be interpreted as spacial cases of dissimilarity comparisons.

Remark 1.7 Let $\mathbf{A}'$ and $\mathbf{B}'$ in Corollary 1.7 be such that $\mathbf{A}, \mathbf{B} \in \mathcal{M}_1$ with $n_A = n_B = n$. This case correspond to vector majorization extensively studied in economic inequality, where for instance $\mathcal{M}_1$ may represent the set of allocation of shares of average income across a population of n individuals, each weighted $\frac{1}{n}$. As already shown by Koshevoy (1995), the Zonotope $Z(\mathbf{A}') \in [0,1] \times [0,1]$ corresponds to the area between the *Lorenz Curve* $L(p)$ and its *dual* $\overline{L}(p) = 1 - L(1-p)$, where p is a given percentile of the population. Making use of Theorem 1.4, it can be shown that the well known result in Lemma 2.B.1 by Hardy et al. (1934) is nested in our framework, that is: $\mathbf{B}' \preccurlyeq \mathbf{A}'$ if and only if $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ with $\mathbf{X} \in \mathcal{D}_n$.

On the contrary, when $d \geq 2$ any sequence of T-transforms induces the multivariate order in Corollary 1.7, while the converse is not true. To see this, note that $\mathbf{B}'$ is matrix majorized by $\mathbf{A}'$ if and only if it is obtained by any possible sequence of merge and split operations.

Nevertheless, it turns out that the dissimilarity comparisons for matrices $\mathbf{A}'$ and $\mathbf{B}'$ in Corollary 1.7 based on uniform majorization is equivalent to the multivariate majorization order based on *Lorenz Zonotopes* $LZ(.) \in \mathbb{R}_+^{d+1}$ (the d-variate generalization of the single attribute Lorenz Curves) studied by Koshevoy (1995, 1997) and Koshevoy and Mosler (1996).

Remark 1.8 The first row of $\mathbf{A}'$ in Corollary 1.7 defines a distribution over classes, then $LZ(\mathbf{A}) \equiv Z(\mathbf{A}')$. It follows from Theorem 1.4 that the ordering of matrices in $\mathcal{M}_d$ with fixed n based upon LZ is equivalent to order such matrices according to the uniform majorization criterion. In fact, the within and between rows type of equalization implied by the Lorenz Zonoids is a particular case of our dissimilarity order. Hence, the Lorenz Zonotope inclusion order implies the dissimilarity order for permutable classes.

1.7.2 Dissimilarity, segregation and discrimination

This chapter organizes into a common analytical framework a set of sparse results that have been proposed in the segregation and discrimination literature. We are able to define and characterize a generalization of the segregation curve and the discrimination curve.

The Zonotopes corresponds to the multidimensional generalization of the segregation curve in Hutchens (1991). For distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, the lower boundary of $Z(\mathbf{A})$ and $Z(\mathbf{B})$ are the Segregation Curve of group one versus group two, as constructed in the introductory example. Hence, the upper boundary of the Zonotope is the dual representation the segregation curve. The curve has an appealing interpretation: it plots vectors according to increasing degree of concentration of one group with respect to the other across classes. It follows that the literature on two groups segregation orderings and measures, which is based upon segregation curves comparisons, entails a sequence of transformation that prove to be dissimilarity-reducing.

For $\mathbf{A}, \mathbf{B} \in \mathcal{M}_2$, let $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$. Consider the situation in which the monotone path $MP^*(\mathbf{B})$ always lie under the Similarity Path Polytope, and that $n_A = n_B$ such that row two of $\mathbf{A}$ coincide with row two of $\mathbf{B}$ ($\mathbf{a}_2^t = \mathbf{b}_2^t$). In this case, the order based on Path Polytopes inclusion coincide with the dominance relation induced by the discrimination curves, studied in Butler and McDonald (1987), Jenkins (1994) and recently in Le Breton et al. (2012). In fact, the lower boundary of $Z^*(.)$ coincides with the discrimination curve, while the upper boundary coincides with the *dual* discrimination curve, obtained by permuting the name of the distributions under analysis.

We have shown that the discrimination curve entails a comparison according to the degree of overlapping between distribution functions. We also show that the ordinal information behind the discrimination rests on the sequence of transformations implied by the exchange and interchange axiom.

1.7.3 Dissimilarity and distance measures in the ordinal setting

We conclude this section by investigating a two groups *dissimilarity measure* inspired by the criterion illustrated in section 1.5 for the ordinal setting.

Note that, in general, any distribution matrix in $\mathcal{M}_d$ can be equivalently represented by d cumulative distribution functions defined on a outcomes domain $\mathcal{X}$ and associated to the d groups. This can be done, as argued in section 1.5, by assuming that population masses are uniformly distributed within classes.

Given two sets of distribution functions $F_1,\ F_2, \dots$ and $F'_1,\ F'_2, \dots$ with average distributions $\overline{F}$ and $\overline{F}'$ (determined respectively by $\overline{F}(x) = \frac{1}{2}F_1(x) + \frac{1}{2}F_2(x)$ in the case of only two distributions), the dissimilarity criterion presented in section 1.5 entails a robust comparison of the degree of inequality (making use of Lorenz dominance) between the two sets of groups population shares at any fixed *overall* population share p, but evaluated at quantiles $\overline{F}^{-1}(p)$ and $\overline{F}'^{-1}(p)$ respectively. When only two distributions are compared, the degree of inequality at p is measured by the *distance* function $\Delta_{1,2}(p) = \left| F_1(\overline{F}^{-1}(p)) - F_2(\overline{F}^{-1}(p)) \right|$.

An index of dissimilarity consistent with the dissimilarity criterion can be constructed by taking the *average* of $\Delta_{1,2}(p)$ across all population shares $p \in [0,1]$. The index $D^*(F_1, F_2)$ can be formalized as follows:

$$D^*(F_1, F_2) = \int_0^1 \left| F_1(\overline{F}^{-1}(p)) - F_2(\overline{F}^{-1}(p)) \right| dp.$$

By changing the variable of integration, we can derive an alternative formalization:

$$D^*(F_1, F_2) = \int_{\mathcal{X}} |F_1(x) - F_2(x)| \, d\overline{F}(x).$$

The functional form of this new index of ordinal dissimilarity is closely related to the Manhattan distance index $D(F_1, F_2)$ between two distribution functions (e.g. Bertino et al. 1987) and is often used as a measure of discrimination in the two groups case:

$$D(F_1, F_2) := \int_{\mathcal{X}} |F_1(x) - F_2(x)| \, dx.$$

There are, however, sharp differences between the two measures $D^*(F_1, F_2)$ and $D(F_1, F_2)$ in the notion of dissimilarity/discrimination they rely on.

Remark 1.9 The index $D^*(F_1, F_2)$ is invariant to monotone transformations of the variable defined on the domain $\mathcal{X}$.

This makes the index suitable for working in the general ordinal setting, while $D(F_1, F_2)$ embodies both ordinal and cardinal concerns.

Remark 1.10 The index $D^*(F_1, F_2)$ is proportional to the area of the Path Polytope.

To see this, let $p \in [0,1]$ denote population fractions and $F_i^{-1}(p)$ the associated quantile, for group $i = 1, 2$. Using a similar notation as in Le Breton et al. (2012) (although we accept that F_2 may not first order stochastically dominates F_1 as assumed there), the two Monotone Paths defining the Path Polytope boundaries can be represented by the functional forms $\phi(p) := F_2(F_1^{-1}(p))$ and $\psi(p) := F_1(F_2^{-1}(p))$. The area A_ϕ and A_ψ between the diagonal representing perfect similarity and the two Monotone Paths represented by $\phi(.)$ and $\psi(.)$

are:

$$A_\phi \;=\; \int_0^1 |p-\phi(p)|\,dp \quad \text{and} \quad A_\psi \;=\; \int_0^1 |p-\psi(p)|\,dp.$$

By construction the two Monotone Paths are symmetric w.r.t. the diagonal of perfect similarity, and therefore $\phi \circ \psi(p) = p = \psi \circ \phi(p)$ at any p. It follows that $A_\phi = A_\psi$.[27] There are two possibilities to perform a change in variables transformation: either by setting $p = F_1(x)$ or by choosing $p = F_2(x)$. This gives the next two alternative definitions of the areas:

$$A_\phi \;=\; \int_{\mathcal{X}} |F_1(x)-F_2(x)|\,dF_1(x) \quad \text{and} \quad A_\psi \;=\; \int_{\mathcal{X}} |F_2(x)-F_1(x)|\,dF_2(x).$$

The distance measure is equal to half of the Path Polytope area (equal to $A_\phi + A_\psi$). In fact:

$$\frac{1}{2}(A_\phi \;+\; A_\psi) \;=\; \int_{\mathcal{X}} |F_1(x)-F_2(x)|\,d\frac{1}{2}(F_1(x)+F_2(x)) \;=\; D^*(F_1,F_2).$$

It follows that given $\mathbf{A},\mathbf{B} \in \mathcal{M}_2$, the condition $Z^*(\mathbf{B}) \subseteq Z^*(\mathbf{A})$ is sufficient (but not necessary) for having $D_1^*(F_1^B,F_2^B) \le D_1^*(F_1^A,F_2^A)$.

The index $D^*(F_1,F_2)$ embodies concerns on the degree of distance and *overlapping* between the distributions F_1 and F_2.

Remark 1.11 The index $D^*(F_1,F_2)$ is maximal when there is no overlapping and, in this case, independent on the distance between distributions.

To see this, note that when two distribution functions F_1 and F_2 do not overlap the associated Path Polytope reaches its maximal extension and coincides with the unitary square. Given that $D^*(F_1,F_2)$ measures this area, it follows that the index is maximal

[27] To see this, denote by $\psi^{-1}(t) := \inf\{p : \ \psi(p) \ge t\}$ the left continuous inverse of $\psi(p)$. The function $\psi^{-1}(t)$ has the same properties of a left continuous quantile function. By changing the variable of integration from p to t, it follows that the area A_ψ coincides with the Lebesgue integral of the function $\psi^{-1}(t)$ on a bounded support $[0,1]$. Thus:

$$A_\psi = \int_0^1 |\psi^{-1}(t)-t|dt = \int_0^1 |\phi(t)-t|dt = A_\phi$$

where the second equality comes from the symmetry of ψ and ϕ, giving $\phi(t) = \psi^{-1}(t)$.

when there is no overlapping between the two distributions. The index is, however, not affected by the distance between the two non-overlapping distributions.

We leave the characterization of the index $D^*(F_1, F_2)$, as well as its multi-group extensions, for future research.[28]

1.8 Conclusions

We study multivariate pre-orders based upon the concept of dissimilarity. Dissimilarity is conceptualized as a form of *exclusion*: the distributions of groups along ordered or non ordered classes are dissimilar whenever some groups are prevented (i.e. excluded) from enjoying some realizations (represented by the classes of the distribution matrix), while other groups are not.

This interpretation opens the dissimilarity comparisons to a variety of applications which concern the measurement and comparison of changes in the patterns of exclusion of social groups along a meaningful partition of a domain of realizations. These realizations may either represent outcomes or, alternatively, a partition of a space in which socioeconomic interactions take palace. The two frameworks motivates the whole chapter, that deals with the characterization of the dissimilarity both in the ordered and in the permutable classes context, and provides an equivalent representation of the dissimilarity ranking though geometric bodies inclusion. The advantage of these representations lies in their empirical testability.

Future extensions of our findings go in three directions. Firstly, we have left uncovered the potential relation between the concept of dissimilarity and the corresponding welfare order. Dahl (1999) proposes a class of evaluation functions whose order is coherent with the dissimilarity in the case of permutable classes, which can be interpreted as loss measures in

[28]We propose only a possible generalization of the index, called $D^*_{G,\omega}(F_1, F_2)$. Let $G, \omega : [0,1] \to [0,1]$ be two strictly increasing and surjective (onto) functions, the general version of the dissimilarity index is:

$$D^*_{G,\omega}(F_1, F_2) = G^{-1}\left[\int_{\mathcal{X}} G\left(|F_1(x) - F_2(x)|\right) d\omega\left(\overline{F}(x)\right)\right],$$

where G is a transformation of the distance and ω can be interpreted as a distortion functions on overall population weights.

the information settings. One can build on this framework to derive economic implications of the dissimilarity order.

Secondly, a promising direction of our research points to the definition of a family of complete orders that are implied by the dissimilarity order. A first example is given by the family of Gini type indices, based on the Zonotope or Path Polytope volume comparison. This objective points at extending the results in Frankel and Volij (2011) in particular to what concerns the cases related to ordered classes.

1.A Appendix: Proofs

1.A.1 Proof of Lemma 1.1

The proof of the lemma consists in showing that a row stochastic matrix in $\widehat{\mathcal{R}}_{n_A,n_B}$ can be constructed through a product series of row stochastic matrices identifying the operations involved by the axioms. We first define two sub classes of row stochastic matrices corresponding to the operations invoked by *IEC* and *SC*, namely $\mathcal{R}^{IEC}_{n_A,n_B}$ and $\mathcal{R}^{SC}_{n}$ respectively. Then, we show that a sequence of such operations always generates a matrix in a larger set $\mathcal{R}^{IEC,SC}_{n_A,n_B} \supset (\mathcal{R}^{IEC}_{n_A,n_B} \cup \mathcal{R}^{SC}_{n})$. Matrices in $\mathcal{R}^{IEC,SC}_{n_A,n_B}$ are defined by blocks $\mathbf{D}_h$ for $h = 1, 2, ..., H$ of matrices of dimensions $(n_A \times n_h)$ such that each matrix $\mathbf{D}_h$ is made of all zeros except for row h whose elements d_{hi} are such that $d_{hi} \geq 0$ and $\sum_{i=1}^{n_h} d_{hi} = 1$, and $\sum_{h=1}^{H} n_h = n_B$.

Therefore $\mathbf{X} \in \mathcal{R}^{IEC,SC}_{n_A,n_B}$ if and only if $\mathbf{X} = (\mathbf{D}_1, \ldots, \mathbf{D}_h)$.

The proof of the lemma can be established by using permutability on the columns of the matrices in the class $\mathcal{R}^{IEC,SC}_{n_A,n_B}$ to generate the class $\widehat{\mathcal{R}}_{n_A,n_B}$.

An operation satisfying *IEC* applied to matrix $\mathbf{A} \in \mathcal{M}_d$ generates a matrix $\mathbf{B} \in \mathcal{M}_d$ with $n_B > n_A$ that is obtained by augmenting $\mathbf{A}$ by $n_B - n_A$ columns with zero entries. It can be formalized in terms of matrix multiplication operations involving identity matrices. Let $\mathbf{i}_j$ be a column vector of zeroes where element j is replaced by a one, such that $\mathbf{I}_n = (\mathbf{i}_1, \ldots, \mathbf{i}_n)$. We have the following definition:

Definition 1.7 *The set* $\mathcal{R}^{IEC}_{n_A,n_B} \subset \mathcal{R}_{n_A,n_B}$ *with* $n_A \leq n_B$ *is such that:*

$$\mathcal{R}^{IEC}_{n_A,n_B} := \left\{\mathbf{Y} \in \mathcal{R}_{n_A,n_B} \ : \ if \ \mathbf{y}_i = \mathbf{i}_j \ then \ \mathbf{y}_{i+1} = \mathbf{i}_{j+1} \ or \ \mathbf{y}_{i+1} = \mathbf{0}_{n_A}, \ otherwise \ \ \mathbf{y}_i = \mathbf{0}_{n_A}\right\}.$$

Let $\mathcal{M}^0_d \subset \mathcal{M}_d$ define the set of matrices exhibiting at least one column of zeroes. For $\mathbf{A} \in \mathcal{M}^0_d$, let $\mathcal{J}^0_A$ denote the index set of all zero columns in $\mathbf{A}$ and $\mathcal{J}_A$ denote the index set of all non-zero columns of $\mathbf{A}$. Let $j \in \mathcal{J}_A$ such that $j + 1 \in \mathcal{J}^0_A$. The matrix $\mathbf{Z}_{[j]}$ (thus depending on the columns distribution in $\mathbf{A}$) is a $n \times n$ identity matrix whose element 1 in position (j, j) is replaced by $z_{j,j} = \lambda$ and the element 0 in position $(j, j + 1)$ is replaced by $z_{j,j+1} = (1 - \lambda)$. The matrix is thus row stochastic.

An operation satisfying *SC* applied to matrix $\mathbf{A} \in \mathcal{M}^0_{d,n_A}$ leads to matrix $\mathbf{B} \in \mathcal{M}^0_{d,n_B}$ with $\mathbf{b}_j = \lambda \mathbf{a}_j$ and $\mathbf{b}_{j+1} = \mathbf{a}_{j+1} + (1 - \lambda)\mathbf{a}_j = (1 - \lambda)\mathbf{a}_j$ with $j \in \mathcal{J}_A$ and $j + 1 \in \mathcal{J}^0_A$.

Formally: $\mathbf{B} = \mathbf{A} \cdot \mathbf{Z}_{[j]}$.

Definition 1.8 *Let* $\mathbf{A} \in \mathcal{M}^0_{d,n_A}$. *The set* $\mathcal{R}^{SC}_{\mathbf{A}} \subset \mathcal{R}_n$ *is the set of all matrices* $\mathbf{Z}_{[j]}$ *such that for* $j, k \in \mathcal{J}_A$, $j+1 \in \mathcal{J}^0_A$, *for all* $k \neq k' \neq j$ *and for* $\lambda \in \mathbb{R}_{++}$*:*

$$\mathcal{R}^{SC}_A := \left\{ \mathbf{Z}_{[j]}(\mathbf{A}, \lambda) \in \mathcal{R}_n \;:\; z_{j,j} := \lambda \;,\; z_{j,j+1} := (1-\lambda),\; z_{k,k} := 1,\; z_{k,k'} := 0 \right\}.$$

Finally, consider a sequence of n random numbers $\{x_i\}_{i=1}^n$ with support in $[0,1]$ satisfying $\sum_i x_i = 1$. For any ordered sub-sequence of $\{x_i\}_{i=1}^n$ given by numbers $x_1, \ldots, x_{i-1}$ with $i \leq n$, the i-th element can be written as:

$$\begin{aligned} & x_1 = \lambda_1 \;\in\; [0,1] \\ & x_i = \lambda_i \left(1 - \textstyle\sum_{k=1}^{i-1} x_k\right) \quad \text{with} \quad \lambda_i \in [0,1]\ \forall i = 2, \ldots, n. \end{aligned} \tag{1.1}$$

The set of elements λ_i obtained from (1.1) describes the full sequence of elements $\{x_i\}_{i=1}^n$. Although each element is independent from the others, the sequence has to be constructed by incorporating the constraint on the unitary sum in the definition of each element. It turns out that in order to satisfy the sum constraint there should exist only an index i such that $\lambda_i = 1$. If $\lambda_i = 1$, then the series is completed and $\lambda_j = 0 = x_j$ for any $j > i$. Note that $x_i = 0$ also if $\lambda_i = 0$, thus the sequence of x_i's may also include elements equal to 0 even if it is not yet completed.

Solving the sequence with backward substitution of elements, and after some algebra, it can be shown that the element x_i can be written as:

$$\begin{aligned} & x_1 = \lambda_1 \;\in\; [0,1], \\ & x_i = \lambda_i \cdot \textstyle\prod_{k=1}^{i-1} (1-\lambda_k) \quad \text{with} \quad \lambda_k \in [0,1]\ \forall k \text{ and } \lambda_i \in [0,1]\ \forall i = 2, \ldots, n, \end{aligned} \tag{1.2}$$

where there exists only one element k such that $\lambda_k = 1$.

Following the same line of reasoning, a row stochastic matrix $\widehat{\mathbf{X}} \in \widehat{\mathcal{R}}_{n_A,n_B}$ with $n_A < n_B$ has generic elements $x_{j,i}$ that are either 0 or correspond to a positive number that can be written as in (1.2) for any fixed j. Given the definition of $\widehat{\mathcal{R}}_{n_A,n_B}$, if $x_{j,i} > 0$ for some i then, by construction, it should be that that $x_{j',i} = 0$ for all $j' \neq j$. These considerations are summarized in the following remark:

Remark 1.12 The entry element $x_{j,i}$ in position (j,i) of any row stochastic matrix $\mathbf{X} \in$

$\mathcal{R}_{n_A,n_B}$ can be written as:

$$\begin{aligned} x_{j,1} &= \lambda_{j,1} \in [0,1] \\ x_{j,i} &= \lambda_{j,i} \cdot \prod_{k=1}^{i-1} (1-\lambda_{j,k}) \;\; \forall j \quad \text{with} \;\; \lambda_{j,k} \in [0,1] \; \forall k \text{ and } \lambda_{j,i} \in [0,1], \end{aligned}$$

where there exists only one element k such that $\lambda_{j,k} = 1$.

We can now identify the class of row stochastic matrices involved in the transformations underlying axioms *IEC* and *SC*, without assuming permutability of classes.

Lemma 1.3 *Let* $\mathbf{A}$, $\mathbf{B} \in \mathcal{M}_d$, *with* $\mathbf{A} \in \mathcal{M}_d^I$ *and* $n_A \leq n_B$, *the dissimilarity order* $\preccurlyeq$ *satisfies IEC and SC if and only if*

$$\mathbf{B} \sim \mathbf{A} \Leftrightarrow \mathbf{B} = \mathbf{A} \cdot \widehat{\mathbf{X}} \textit{ for some matrix } \widehat{\mathbf{X}} \in \mathcal{R}^{IEC,SC}_{n_A,n_B}.$$

Proof. We show that a sequence of matrix transformations derived through the application of operations underlying axioms *IEC* and *SC* generates indeed a row stochastic matrix in $\mathcal{R}^{IEC,SC}_{n_A,n_B}$ ($\Rightarrow$ part), and that the whole class of matrices in $\mathcal{R}^{IEC,SC}_{n_A,n_B}$ can be identified by means of sequences of such operations ($\Leftarrow$ part), making use of Remark 1.12.

($\Rightarrow$ part). Consider matrix $\mathbf{A} \in \mathcal{M}_d^I$. For each column $j \leq n_A$ we augment the matrix by a set of n_j empty columns $\mathbf{0}_d$. We obtain a new matrix

$$\mathbf{A}' := (\mathbf{a}_1, \underbrace{\mathbf{0}_d, \ldots, \mathbf{0}_d}_{n_1 \text{ times}}, \ldots, \mathbf{a}_{n_A}, \underbrace{\mathbf{0}_d, \ldots, \mathbf{0}_d}_{n_{n_A} \text{ times}}),$$

with n_B columns such that $n_B = n_A + \sum_j n_j$. A sequence of matrix operations involving row stochastic matrices allow us to write: $\mathbf{A}' = \mathbf{A} \cdot \mathbf{Y}$ where $\mathbf{Y} \in \mathcal{R}^{IEC}_{n_A,n_B}$. By *IEC* it follows that $\mathbf{A}' \sim \mathbf{A}$.

Consider a *split transformation* that splits a class k with non-zero elements of matrix $\mathbf{A}'$ in two adjacent classes, k and $k+1$. Given that, by construction, there exists a j such that $\mathbf{a}'_k = \mathbf{a}_j$, then $k+1$ is the first of n_j classes following k that are empty. Hence, we use k to refer to a specific class j in $\mathbf{A}$. A share $\lambda_{j,k}$ of each group in class k is left in k while the remaining share $1 - \lambda_{j,k}$ is displaced from column k to column $k+1$. The matrix operation incorporating this splitting is given by $\mathbf{Z}_{[k]} \in \mathcal{R}^{SC}_{A'}$ such that the new distribution matrix obtained is $\mathbf{A}'_{[k]} := \mathbf{A}' \cdot \mathbf{Z}_{[k]}$. By SC and *IEC* we get $\mathbf{A}'_{[k]} \sim \mathbf{A}$.

Following the previous step, consider a *split transformation* involving the entry in column $k+1$, that corresponds to $(1-\lambda_{j,k})\mathbf{a}_j$. We leave a share $\lambda_{j,k+1}$ of the entry in column $k+1$ and move a fraction $1-\lambda_{j,k+1}$ from column $k+1$ to column $k+2$. The matrix incorporating this splitting is $\mathbf{A'}_{[k+1]} := \mathbf{A'}_{[k]} \cdot \mathbf{Z}_{[k+1]}$ with $\mathbf{Z}_{[k+1]} \in \mathcal{R}^{SC}_{\mathbf{A'}_{[k]}}$. By SC and *IEC* it follows that $\mathbf{A'}_{[k+1]} \sim \mathbf{A}$.

For any column k in $\mathbf{A'}$, corresponding to a column j in $\mathbf{A}$, the procedure can be iterated sequentially through all classes $k+1$ to $k+n_j$ of matrix $\mathbf{A'}$ to obtain the matrix $\mathbf{A'}_{[k+n_j]}$. A given class $k < h < k+n_j$ of $\mathbf{A'}_{[k+n_j]}$ can be written as a function of $\mathbf{a}_j$ alone and a weighting coefficient that depends upon the iteration procedure, that is:

$$\mathbf{a'}_h := \lambda_{j,h} \cdot (1-\lambda_{j,h-1}) \cdot \ldots \cdot (1-\lambda_{j,k}) \cdot \mathbf{a}_j.$$

The result has been obtained through a sequence of splitting operations. For a given column j of the original distribution matrix we can rewrite such a sequence by using row stochastic matrices. We have that:

$$\mathbf{B} = \mathbf{A}_{[k+n_j]} := \mathbf{A} \cdot \mathbf{Y} \cdot \mathbf{Z}_{[k]} \cdot \ldots \cdot \mathbf{Z}_{[k+n_j-1]}.$$

The matrix multiplying $\mathbf{A}$ is a product of row stochastic matrices and therefore it is row stochastic. This matrix has at most only one non-zero element by column by construction, moreover by combining the sequence of transformations with addition of empty classes through $\mathbf{Y} \in \mathcal{R}^{IEC}_{n_A,n_B}$ operations we obtain the matrices in $\mathcal{R}^{IEC,SC}_{n_A,n_B}$, thus explaining the *sufficiency* part of the lemma.

($\Leftarrow$ part). Note that each of the elements of the series rewrites as an element of the series in (1.2). In fact, repeating the same procedure for all $j \in \mathcal{J}_A$, it is possible to obtain a product of matrices giving the row stochastic matrix $\mathbf{X}$. Let use k_j to underly the relation between the class j in $\mathbf{A}$ and class k in $\mathbf{A'}$. It follows that:

$$\mathbf{X} = \mathbf{Y} \cdot \prod_{j=1}^{n_A} \prod_{h=k_j}^{k_j+n_j-1} \mathbf{Z}_{[h]}. \tag{1.3}$$

The elements of a matrix $\mathbf{X} \in \mathcal{R}^{IEC,SC}_{n_A,n_B}$ can be written by exploiting Remark 1.12. We now show the necessary condition by proving that $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ for $\mathbf{X} \in \mathcal{R}^{IEC,SC}_{n_A,n_B}$ implies (1.3). In general it holds that $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_{n_A})$ for all the matrices in $\mathcal{R}^{IEC,SC}_{n_A,n_B}$, where each matrix $\mathbf{X}_j$ has a size $n_A \times (n_j+1)$ and is everywhere zero apart from the entries in row j that have to sum up to one. Hence:

$$\mathbf{A}\cdot\mathbf{X} \;=\; (\mathbf{A}\cdot\mathbf{X}_1,\ldots,\mathbf{A}\cdot\mathbf{X}_{n_A})\,.$$

The product of matrices $\mathbf{A}\cdot\mathbf{X}_j$ defines operations that only involve the column j of the matrix $\mathbf{A}$, and can be represented by using the following notation:

$$\begin{aligned}
\mathbf{A}\cdot\mathbf{X}_j &= (x_{j,k_j}\mathbf{a}_j,\ \ldots,\ x_{j,k_j+n_j-1}\mathbf{a}_j,\ x_{j,k_j+n_j}\mathbf{a}_j)\\
&= \left(\lambda_{j,k_j}\mathbf{a}_j,\ldots,\lambda_{j,k_j+n_j-1}\prod_{k_j\le h<k_j+n_j-1}(1-\lambda_{j,h})\mathbf{a}_j,\prod_{k_j\le h\le k_j+n_j-1}(1-\lambda_{j,h})\mathbf{a}_j\right)\\
&= \left(\lambda_{j,k_j}\mathbf{a}_j,\ldots,\prod_{k_j\le h<k_j+n_j-1}(1-\lambda_{j,h})\mathbf{a}_j,\mathbf{0}_d\right)\cdot\mathbf{Z}_{[k_j+n_j-1]}\\
&= (\mathbf{a}_i,\mathbf{0}_d,\ldots,\mathbf{0}_d)\cdot\prod_{k_j\le h\le k_j+n_j-1}\mathbf{Z}_{[h]},
\end{aligned}$$

where the second line uses the definition of the entries of a row stochastic matrix as in the Remark 1.12, the third line follows by the definition of a split operation involving columns k_j+n_j-1 and k_j+n_j, here captured by the $(n_j+1)\times(n_j+1)$ matrix $\mathbf{Z}_{[k_j+n_j-1]}$, and finally the last line develops iteratively the result in the third line. Each $(n_j+1)\times(n_j+1)$ matrix $\mathbf{Z}_{[h]}$ has been defined above. Hence, the previous list of equalities rewrites (using a matrix $\mathbf{Y}$ to add empty columns as before):

$$\begin{aligned}
\mathbf{A}\cdot\mathbf{X} &= \mathbf{A}\cdot\mathbf{Y}\cdot\operatorname{diag}\left(\prod_{k_1\le h\le k_1+n_1-1}\mathbf{Z}_{[h]},\ldots,\prod_{k_{n_A}\le h\le k_{n_A}+n_{n_A}-1}\mathbf{Z}_{[h]}\right)\cdot\mathbf{Y}'\\
&= \mathbf{A}\cdot\mathbf{Y}\cdot\prod_{j=1}^{n_A}\prod_{h=k_j}^{k_j+n_j-1}\widetilde{\mathbf{Z}}_{[h]}\cdot\mathbf{Y}'.
\end{aligned}$$

where $\widetilde{\mathbf{Z}}_{[h]} := \operatorname{diag}\left(\mathbf{I},\ \mathbf{Z}_{[h]},\ \mathbf{I}'\right)$ and $\mathbf{I}$ and $\mathbf{I}'$ are two identity matrices of size $(j-1)+\sum_{k=1}^{j-1}n_k$ and $n_B-(k_j+n_j)$ respectively.

The first line follows by combining the expression derived for each $\mathbf{A}\cdot\mathbf{X}_j$ to define the product $\mathbf{A}\cdot\mathbf{X}$, while the second equality comes from a property of the block diagonal matrix. The block diagonal matrix can be equivalently represented as the product of the matrices associated to each block, obtained substituting the remaining blocks with identity matrices. The matrix $\widetilde{\mathbf{Z}}_{[h]}$ is obtained in the same way, and its size is $n_B \times n_B$. By standard properties of matrix algebra the block diagonal of a product of matrices as $\text{diag}\left(\prod_{k_1 \le h \le k_1+n_1-1} \mathbf{Z}_{[h]}, \dots, \prod_{k_{n_A} \le h \le k_{n_A}+n_{n_A}-1} \mathbf{Z}_{[h]}\right)$ is the product of the block diagonals, given by $\prod_{j=1}^{n_A} \prod_{h=k_j}^{k_j+n_j} \widetilde{\mathbf{Z}}_{[h]}$. The matrix $\widetilde{\mathbf{Z}}_{[h]}$ is comparable in size to the matrices used to construct the sufficient conditions. Altogether, these elements give the second equation, showing that, starting from the definition of the elements of the class $\mathcal{R}^{IEC,SC}_{n_A,n_B}$, we obtain (1.3). Note that $\mathcal{R}^{IEC,SC}_{n_A,n_B}$ is closed with respect to matrix multiplication. Moreover, we have enough degree of freedom in the proof to show that *any* matrix in $\mathcal{R}^{IEC,SC}_{n_A,n_B}$ could be decomposed according to the sequence in (1.3), which establishes the Lemma. ■

The previous result is used for the proof of Lemma 1.1.

Proof. For any pair $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, with $\mathbf{A} \in \mathcal{M}^I_d$, $\mathbf{B} \sim \mathbf{A}$ with $\preccurlyeq$ satisfying *IEC* and *SC* whenever $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ for $\mathbf{X} \in \mathcal{R}^{IEC,SC}_{n_A,n_B}$ by Lemma 1.3. If moreover $\preccurlyeq$ satisfies also *IPC*, then $\mathbf{B} = \mathbf{A} \cdot \mathbf{X} \cdot \mathbf{\Pi} = \mathbf{A} \cdot \widehat{\mathbf{X}}$ where $\mathbf{\Pi}$ is a permutation matrix implies $\mathbf{B} \sim \mathbf{A}$. By using any $n_B \times n_B$ permutation matrix, one gets the whole set of $n_A \times n_B$ matrices with at most one nonzero element by row, that is $\widehat{\mathbf{X}} \in \widehat{\mathcal{R}}_{n_A,n_B}$. ■

1.A.2 Proof of Theorem 1.1

Before moving to the proof, it is worth noting that a *merge transformation* in combination with permutation of classes is equivalently represented by a matrix product involving a row stochastic matrix. An operation satisfying *MC* is defined up to a permutation of columns of the distribution matrix $\mathbf{A}'' = (\mathbf{A},\ \mathbf{0}_d, \dots, \mathbf{0}_d)$, where the vectors $\mathbf{0}_d$ are repeated $n - n_A$ times. As described in Axiom *MC*, when combined with permutation of classes, the operation is such that from a matrix $\mathbf{A}$ it is possible to obtain a new matrix $\mathbf{B}$ where $\mathbf{b}_j = \mathbf{a}_j + \mathbf{a}_{j'}$ and $\mathbf{b}_{j'} = \mathbf{0}_d$ for $j, j' \le n_A$. The operation can be written in matrix product form as: $\mathbf{B} = \mathbf{A}'' \cdot \mathbf{X}_{[j,j']}$ such that $\mathbf{X}_{[j,j']}$ is a $n \times n$ identity matrix whose j'-th row is replaced by row j.

Definition 1.9 *The set $\mathcal{R}^{MC}_n \subset \mathcal{R}_n$ is such that for all j, j', k, k':*

$$\mathcal{R}^{MC}_n := \left\{\mathbf{X}_{[j,j']} \in \mathcal{R}_n \ : \ x_{j',j} = x_{j,j} = 1,\ x_{k,k} = 1\ \forall k \neq j',\ x_{k,k'} = 0\ \forall k \neq k'\right\}.$$

According to Definition 1.9, a sequence of *merge transformations* migrating masses from classes h corresponding to the subset $\mathcal{H}_j$ of columns to class j admits an equivalent representation through a sequence of matrix products with elements in $\mathcal{R}_n^{MC}$: $\prod_{h\in\mathcal{H}_j}\mathbf{X}_{[j,h]}$. By performing the necessary number of matrix products such that all elements of $\mathcal{H}_j$ are merged with class j, we obtain a row stochastic matrix $\mathbf{M}_j$. It corresponds to a transformation of an $n\times n$ identity matrix whose rows $h\in\mathcal{H}_j$ have all been replaced by row j. By performing the same procedure for all j we obtain the matrix $\mathbf{M}=\prod_j^{n_A}\mathbf{M}_j$ such that $\bigcup_j\mathcal{H}_j\cup\mathcal{J}_A=\{1,\ldots,n\}$.

Proof. ($\Rightarrow$ part). If the dissimilarity pre-order satisfies axioms *IEC*, *SC, PC* and *MC* then there exist a sequence of insertion of empty classes, splits and permutations that allows to transform $\mathbf{A}$ into $\mathbf{B}$ such that $\mathbf{B}=\mathbf{A}\cdot\mathbf{X}$ for some $\mathbf{X}\in\mathcal{R}_{n_A,n_B}$.

It has been extensively argued in the proof of Lemma 1.1 that each of the transformations underlying axioms *IEC*, *SC* and *PC* involves a row stochastic matrix operation. The Axiom *MC* induces a merge operation between two or more classes that can be represented by a matrix that is row stochastic. Hence, $\mathbf{B}\preccurlyeq\mathbf{A}$ implies that there exist a sequence of row stochastic matrices transforming $\mathbf{A}$ into $\mathbf{B}$. A product of row stochastic matrices gives a row stochastic matrix, and therefore $\mathbf{B}=\mathbf{A}\cdot\mathbf{X}$ with $\mathbf{X}\in\mathcal{R}_{n_A,n_B}$, which establishes the desired implication.

($\Leftarrow$ part). We show now that if matrix $\mathbf{B}=\mathbf{A}\cdot\mathbf{X}$ for $\mathbf{X}\in\mathcal{R}_{n_A,n_B}$, then it also holds that $\mathbf{B}\preccurlyeq\mathbf{A}$, where the dissimilarity pre-order is characterized by *IEC*, *SC*, *IPC* and *MC* axioms.

Exploiting Lemma 1.1, one can verify that for any row stochastic matrix $\mathbf{X}\in\mathcal{R}_{n_A,n_B}$ such that $\mathbf{B}=\mathbf{A}\cdot\mathbf{X}$ there exists a permutation matrix $\mathbf{\Pi}$ and a $\widehat{\mathbf{X}}\in\widehat{\mathcal{R}}_{n_A,n_A\cdot n_B}$ such that $\widehat{\mathbf{X}}\cdot\mathbf{\Pi}\in\mathcal{R}^{IEC,SC}_{n_A,n_A\cdot n_B}$, hence:

$$\mathbf{X}=\widehat{\mathbf{X}}\cdot\mathbf{\Pi}\cdot\tilde{\mathbf{M}}=\mathbf{Y}\cdot\mathbf{\Pi}'\cdot\prod_{j=1}^{n_A}\prod_{h=k_j}^{k_j+n_j}\mathbf{Z}_{[h]}\cdot\tilde{\mathbf{M}},$$

where $\mathbf{Y}\in\mathcal{R}^{IEC}_{n_A,n_A\cdot n_B}$, with $\mathbf{Y}\cdot\mathbf{\Pi}'$ such that $\mathbf{A}\cdot\mathbf{Y}\cdot\mathbf{\Pi}'=\mathbf{A}':=(\mathbf{a}_1,\mathbf{0}_d,\ldots,\mathbf{a}_{n_A},\mathbf{0}_d)$, $\mathbf{Z}_{[h]}\in\mathcal{R}^{SC}$. The first equality is an algebraic result that holds for any matrix $\mathbf{X}$: for a matrix $\widehat{\mathbf{X}}$ in $\widehat{\mathcal{R}}_{n_A,n_A\cdot n_B}$, one can find a permutation matrix that arranges the terms in $\widehat{\mathbf{X}}$ in a way that the first n_B entries of the first row of the matrix coincide with the first row of $\mathbf{X}$, and then the remaining entries are zero; the entries in the second row between classes $n_B + 1$ and $2n_B$ coincide with the second row of $\mathbf{X}$ and so forth. The matrix $\tilde{\mathbf{M}} = (\mathbf{I}_{n_B}, \ldots, \mathbf{I}_{n_B})^t$ is row stochastic, it is related to the square matrix $\mathbf{M}$ that represents sequences of merge transformations. The matrix $\mathbf{M}$ is of dimension $(n_A \cdot n_B) \times (n_A \cdot n_B)$ and is constructed such that $\mathbf{M} = (\tilde{\mathbf{M}}, \mathbf{0}_{n_B, n_A \cdot n_B})$. Thus, $\tilde{\mathbf{M}}$ represents sequences of merge transformations and eliminations of empty classes. The second equality is a direct consequence of Lemma 1.3. Hence, any row stochastic matrix $\mathbf{X}$ can be decomposed into a sequence of insertions/eliminations of empty classes, splits, merges and permutations (formalized by the operations $\widehat{\mathbf{X}} \cdot \mathbf{\Pi} \cdot \tilde{\mathbf{M}}$). This verification concludes the proof. ∎

1.A.3 Proof of Corollary 1.2

Proof. ($\Rightarrow$ part). If $\mathbf{B} \preccurlyeq \mathbf{A}$ satisfies *NG*, then $\mathbf{B} \sim [diag(\boldsymbol{\mu}_B)]^{-1} \cdot \mathbf{B} := \mathbf{B}'$ and $\mathbf{A} \sim [diag(\boldsymbol{\mu}_A)]^{-1} \cdot \mathbf{A} := \mathbf{A}'$ and, by transitivity of the pre-order $\preccurlyeq$ one gets $[diag(\boldsymbol{\mu}_B)]^{-1} \cdot \mathbf{B} \preccurlyeq [diag(\boldsymbol{\mu}_A)]^{-1} \cdot \mathbf{A}$. Given that $\preccurlyeq$ satisfies the other axioms underlying the result in Theorem 1.1 and that $\mathbf{A}'$ and $\mathbf{B}'$ satisfy the constraints of Theorem 1.1 $\mathbf{A}' \cdot \mathbf{e}_{n_A} = \mathbf{B}' \cdot \mathbf{e}_{n_B}$, then the dissimilarity pre-order can be equivalently represented by the matrix majorization.

($\Leftarrow$ part). Suppose that $\mathbf{B}' \preccurlyeq^R \mathbf{A}'$, then by Theorem 1.1 it holds that $\mathbf{B}' \preccurlyeq \mathbf{A}'$. Moreover, it is possible to move from $\mathbf{A}'$ to $\mathbf{A}$ and from $\mathbf{B}'$ to $\mathbf{B}$ making use of *NG* transformations. It then follows that $\mathbf{B}' \sim \mathbf{B}$ and $\mathbf{A}' \sim \mathbf{A}$. Thus by Theorem 1.1 and the transitivity of $\preccurlyeq$, we obtain that $\mathbf{B} \preccurlyeq \mathbf{A}$ for $\preccurlyeq$ satisfying *NG*. ∎

1.A.4 Proof of Corollary 1.3

Proof. ($\Rightarrow$ part). If $\mathbf{B} \preccurlyeq \mathbf{A}$ satisfies *IPG*, then $\mathbf{A} \sim \mathbf{\Pi}_d \cdot \mathbf{A}$ and, by transitivity of the pre-order $\preccurlyeq$, one gets $\mathbf{B} \preccurlyeq \mathbf{\Pi}_d \cdot \mathbf{A}$. Given that $\preccurlyeq$ satisfies the other axioms underlying the result in Theorem 1.1, the dissimilarity pre-order can be equivalently represented by the matrix majorization.

($\Leftarrow$ part). Suppose that $\mathbf{B} \preccurlyeq^R \mathbf{\Pi}_d \cdot \mathbf{A}'$, then by Theorem 1.1 it holds that $\mathbf{B} \preccurlyeq \mathbf{\Pi}_d \cdot \mathbf{A}'$. Moreover, it is possible to move from $\mathbf{A}$ to $\mathbf{\Pi}_d \cdot \mathbf{A}$ making use of *IPG* transformations. It then follows that $\mathbf{A} \sim \mathbf{\Pi}_d \cdot \mathbf{A}$. Thus by Theorem 1.1 and the transitivity of $\preccurlyeq$, we obtain that $\mathbf{B} \preccurlyeq \mathbf{A}$ for $\preccurlyeq$ satisfying *IPG*. ∎

1.A.5 Proof of Theorem 1.2

To prove the theorem, we make use of two lemmas. The first lemma shows that the operation needed to obtain the minimal ordinal comparable matrices are the same underlying the *IEC* and *SC* axioms. We restrict the domain of admissible matrices to all pairs of matrices satisfying $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$.

Lemma 1.4 *For any* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *there exists* $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ *with* $n_A^* = n_B^* = n^*$ *that are minimal ordinal comparable such that* $\preccurlyeq$ *satisfies IEC and SC if and only if* $\mathbf{B} \preccurlyeq \mathbf{A} \;\;\Leftrightarrow \mathbf{B}^* \preccurlyeq \mathbf{A}^*$.

Proof. ($\Rightarrow$ part). We show that if $\preccurlyeq$ satisfies *IEC* and *SC*, then $\mathbf{B} \preccurlyeq \mathbf{A} \;\Rightarrow \mathbf{B}^* \preccurlyeq \mathbf{A}^*$.

If $\preccurlyeq$ satisfies *IEC* and *SC*, then empty classes can be added, or existing classes can be proportionally split to generate contiguous, new classes. These operations are sufficient to construct the minimal ordinal comparable matrices $\mathbf{A}^*$ and $\mathbf{B}^*$. Therefore it follows that $\mathbf{A} \sim \mathbf{A}^*$ and $\mathbf{B} \sim \mathbf{B}^*$. By transitivity of $\preccurlyeq$, it follows that $\mathbf{B} \preccurlyeq \mathbf{A}$ implies $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$.

($\Leftarrow$ part). We show that whenever $\mathbf{B}^*, \mathbf{A}^*$ are minimal ordinal comparable to $\mathbf{B}$ and $\mathbf{A}$ respectively, then they can be derived from $\mathbf{B}$ and $\mathbf{A}$ through a finite sequence of operations underlying IEC and SC axioms, and therefore $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ implies $\mathbf{B} \preccurlyeq \mathbf{A}$.

We show the result for $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d^I$. Note that following the definition of minimal ordinal comparability, it is always possible to write $\mathbf{A}^* = \mathbf{A} \cdot \mathbf{X}$ and $\mathbf{B}^* = \mathbf{B} \cdot \mathbf{Y}$ with $\mathbf{X}$ and $\mathbf{Y}$ appropriate row stochastic matrices. In fact by construction $\mathbf{X}$ and $\mathbf{Y}$ belong to a subset of $\mathcal{R}_{n_A,n^*}^{IEC,SC}$ and $\mathcal{R}_{n_B,n^*}^{IEC,SC}$ respectively (see the definition in proof of Lemma 1.1), given that the matrices $\mathbf{A}^*, \mathbf{B}^*$ can by construction be obtained from $\mathbf{A}, \mathbf{B}$ only using additions of empty classes and splits of classes. This implies, by Lemma 1.3, that $\mathbf{A} \sim \mathbf{A}^*$ and $\mathbf{B} \sim \mathbf{B}^*$ for $\preccurlyeq$ that satisfies *IEC* and *SC*. Thus, by transitivity of $\preccurlyeq$ we get that $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ implies $\mathbf{B} \preccurlyeq \mathbf{A}$. Note also that this condition extends to $\mathbf{A}', \mathbf{B}'$ not necessarily in $\mathcal{M}_d^I$. In fact, take for instance $\mathbf{A}'$, there exists a matrix $\mathbf{X} \in \mathcal{R}_{n_A,n_{A'}}^{IEC,SC}$ and $\mathbf{A} \in \mathcal{M}_d^I$ such that $\mathbf{A}' = \mathbf{A} \cdot \mathbf{X}$, and thus $\mathbf{A}' \sim \mathbf{A}$ for $\preccurlyeq$ that satisfies *IEC* and *SC*. Similar reasoning holds for $\mathbf{B}'$. ∎

The second lemma shows that any exchange transformation of a minimal ordinal preserving matrix is equivalently mapped into a *rank preserving progressive transfer* of population masses on the correspondent cumulative distribution matrices (the definition of progressive transfer will be given in the proof).

Lemma 1.5 *For* $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ *that are ordinal and rank comparable,* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *form a finite sequence of exchange operations if and only if for all* $k = 1, \ldots, n$, $\overrightarrow{\mathbf{b}}_k$ *is* *obtained from $\overrightarrow{\mathbf{a}}_k$ by a finite sequence of progressive transfers that preserve the ranking of the elements $i = 1, \ldots, d$ of the vectors.*

Proof. Consider $\mathbf{A}, \mathbf{B}$ that are ordinal comparable, with $k = 1, \ldots, n$ classes. For a given k, $\mathbf{b}_k$ is obtained from $\mathbf{a}_k$ by an exchange between group h and ℓ if and only if group h dominates group ℓ in k.

Given that both distribution matrices are rank comparable, not only it should hold that $\overrightarrow{a}_{hk} \leq \overrightarrow{a}_{\ell k} \Rightarrow \overrightarrow{b}_{hk} \leq \overrightarrow{b}_{\ell k}$, but the same implication should also hold for *any* pair of groups. Moreover, the exchange operation implies that there exists δ such that $\overrightarrow{b}_{hk} = \overrightarrow{a}_{hk} + \delta$ and $\overrightarrow{b}_{\ell k} = \overrightarrow{a}_{\ell k} - \delta$ with $\overrightarrow{b}_{ik} = \overrightarrow{a}_{ik}$ for all groups $i \neq h, \ell$ and $\overrightarrow{\mathbf{b}}_j = \overrightarrow{\mathbf{a}}_j$ for all classes $j \neq k$. This is by definition a rank preserving progressive transfer (for a formal definition, see Fields and Fei 1978), which is independently implemented among entries of one class of the distribution matrix.

Conversely, every rank preserving transfer of cumulative population masses can be mapped into an exchange operation, provided that the matrices $\mathbf{A}, \mathbf{B}$ are both ordinal comparable. ■

The proof of Theorem 1.2 is as follows:

Proof. To prove the *sufficiency* part ($\Rightarrow$), consider $\mathbf{B} \preccurlyeq \mathbf{A}$ with $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ that are rank comparable up to a permutation of the groups. It follows that there exists $\mathbf{A}^*$, $\mathbf{\Pi}_d \cdot \mathbf{B}^* \in \mathcal{M}_d$ that, by construction, are also rank comparable. If $\preccurlyeq$ satisfies *IEC*, *SC* and *IPG*, then $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ by Lemma 1.4. If moreover, $\preccurlyeq$ satisfies E, then $\mathbf{B}^*$ could be obtained from $\mathbf{A}^*$ through a sequence of exchange operations or, equivalently, (by Lemma 1.5) it should hold that $\forall k$ $\overrightarrow{\mathbf{b}^*}_k$ is obtained from $\overrightarrow{\mathbf{a}^*}_k$ by a sequence of rank preserving "progressive transfers" of population masses. Classical theorems on univariate majorization (see for instance ch.2, Lemma B.1 in Marshall et al. 2011) show that the latter is equivalent to $\overrightarrow{\mathbf{b}^*}_k \preccurlyeq^U \overrightarrow{\mathbf{a}^*}_k$.

The proof of the *necessity* part ($\Leftarrow$), requires to show that if $\mathbf{B} \preccurlyeq^* \mathbf{A}$ then there exist a sequence of transformations underlying axioms *IEC*, *SC, IPG* and *E*, that can lead from $\mathbf{A}$ to $\mathbf{B}$ and therefore $\mathbf{B} \preccurlyeq \mathbf{A}$.

The proof makes use of the Theorem 2.1 in Fields and Fei (1978) and the Lemma 2.B.1 by Hardy et al. (1934) to get that $\overrightarrow{\mathbf{b}^*}_k \preccurlyeq^U \overrightarrow{\mathbf{a}^*}_k$ for all $k = 1, \ldots, n^*$ implies that there exists a finite sequence of rank preserving "progressive transfers" of population masses, defined up to a permutation of the groups (the entries of the vectors) that leads from $\overrightarrow{\mathbf{a}^*}_k$ to $\overrightarrow{\mathbf{b}^*}_k$. By Lemma 1.5, the "progressive transfers" can be equivalently formalized as a finite sequence of exchange operations. Consequently these transformations underlying axiom E allow to move from $\mathbf{A}^*$ to $\mathbf{B}^*$. Therefore $\mathbf{B}^* \preccurlyeq^* \mathbf{A}^*$ implies $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ where the dissimilarity pre-order $\preccurlyeq$ satisfies axioms *E* and *IPG*. Next, consider Lemma 1.4, the fact that $\mathbf{A}^*$ and $\mathbf{B}^*$ are minimal ordinal comparable matrices and transitivity of $\preccurlyeq$ gives that $\mathbf{B} \preccurlyeq \mathbf{A}$ for $\preccurlyeq$ satisfying also *IEC* and *SC*, which establishes the result. ■

1.A.6 Proof of Theorem 1.3

Proof. The theorem holds for matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$. ($\Rightarrow$ part). Suppose $\mathbf{B} \preccurlyeq \mathbf{A}$, if $\mathbf{A}$ and $\mathbf{B}$ are not ordinal comparable but $\preccurlyeq$ satisfies Axiom *I*, then there exists $\mathbf{A}'$ and $\mathbf{B}'$ obtained by a sequence of interchange operation such that $\mathbf{A}' \sim \mathbf{A}$ and $\mathbf{B}' \sim \mathbf{B}$ and $\mathbf{A}'$ and $\mathbf{B}'$ are ordinal and rank comparable. Therefore $\mathbf{B}' \preccurlyeq \mathbf{A}'$ and Theorem 1.2 applies leading to $\mathbf{B}' \preccurlyeq^* \mathbf{A}'$ and therefore also to $\mathbf{B} \preccurlyeq^* \mathbf{A}$ given that interchange operations do not affect the ranking produced by $\preccurlyeq^*$ and as a consequence $\mathbf{B} \sim^* \mathbf{B}'$ and $\mathbf{A} \sim^* \mathbf{A}'$. Thus it follows that the transitivity of $\preccurlyeq^*$ leads to $\mathbf{B} \preccurlyeq^* \mathbf{A}$.

The proof of the *necessity* part ($\Leftarrow$), requires to show that if $\mathbf{B} \preccurlyeq^* \mathbf{A}$ then there exist a sequence of transformations underlying axioms *IEC*, *SC, IPG, E* and *I*, that can lead from $\mathbf{A}$ to $\mathbf{B}$ and therefore gives $\mathbf{B} \preccurlyeq \mathbf{A}$.

If $\mathbf{B} \preccurlyeq^* \mathbf{A}$ for matrices that are rank comparable then we are back to the proof of the implication ($\Leftarrow$) in Theorem 1.2.

Suppose that $\mathbf{B} \preccurlyeq^* \mathbf{A}$ for matrices $\mathbf{B}$ and $\mathbf{A}$ that *are not* necessarily rank comparable. Then, consider the minimal ordinal comparable matrices $\mathbf{B}^*$ and $\mathbf{A}^*$ that are also by construction not rank comparable, given that $\mathbf{B}$ and $\mathbf{A}$ are not.

It is then possible to transform $\mathbf{B}^*$ and $\mathbf{A}^*$ into matrices $\mathbf{B}'$ and $\mathbf{A}'$ that are rank comparable through a finite sequence of interchanges and permutation of groups. The algorithm requires to first permute the groups of one of the two matrices such that they are both rank comparable for the first class. Then, consider in sequence next classes and apply the interchange operation for each pair of groups that happens to violate the rank comparability assumption between the matrices. By construction of the matrices $\mathbf{B}^*$ and $\mathbf{A}^*$ the interchange operation can be applied because whenever for one minimal ordinal comparable distribution matrix (say $\mathbf{A}^*$) the rank of two groups i, j is modified between two classes l, h, that is $(\overrightarrow{a^*}_{il} - \overrightarrow{a^*}_{jl}) \cdot (\overrightarrow{a^*}_{ih} - \overrightarrow{a^*}_{jh}) < 0$ then there exists an intermediate class k where $\overrightarrow{a^*}_{ik} = \overrightarrow{a^*}_{jk}$ (see property (ii) in the definition of minimal ordinal comparability). Thus, starting from class k the distribution for all higher classes can be interchanged between groups i and j.

A finite sequence of such operations will lead to matrices $\mathbf{B}'$ and $\mathbf{A}'$ that are rank comparable. Recall by the proof of Theorem 1.2 that according to Lemma 1.4, the fact that $\mathbf{A}^*$ and $\mathbf{B}^*$ are minimal ordinal comparable matrices and transitivity of $\preccurlyeq$ gives that $\mathbf{B} \preccurlyeq \mathbf{A}$ is equivalent to $\mathbf{B}^* \preccurlyeq \mathbf{A}^*$ for $\preccurlyeq$ satisfying *IEC* and *SC.*

Thus by applying the interchange transformations, given that $\preccurlyeq$ satisfies I we obtain $\mathbf{B}^* \sim \mathbf{B}'$ and $\mathbf{A}^* \sim \mathbf{A}'$. Thus, (i) $\mathbf{B} \preccurlyeq \mathbf{A} \Leftrightarrow \mathbf{B}^* \preccurlyeq \mathbf{A}^* \Leftrightarrow \mathbf{B}' \preccurlyeq \mathbf{A}'$.

Because $\mathbf{B} \preccurlyeq^* \mathbf{A}$ is not affected by permutations of elements in each matrix within the same column it follows that $\mathbf{B} \preccurlyeq^* \mathbf{A} \Leftrightarrow \mathbf{B}' \preccurlyeq^* \mathbf{A}'$. Moreover, as shown in the proof of Theorem 1.2 if $\mathbf{B}' \preccurlyeq^* \mathbf{A}'$, by Lemma 1.5, the transformations underlying Axiom E allow to move from $\mathbf{A}'$ to $\mathbf{B}'$, therefore $\mathbf{B}' \preccurlyeq^* \mathbf{A}'$ implies $\mathbf{B}' \preccurlyeq \mathbf{A}'$. Thus, (ii) $\mathbf{B} \preccurlyeq^* \mathbf{A} \Leftrightarrow \mathbf{B}' \preccurlyeq^* \mathbf{A}' \Rightarrow \mathbf{B}' \preccurlyeq \mathbf{A}'$.

To summarize, making use of sequences of transformations underlying the *IEC*, *SC*, *IPG,* *E* and *I* we obtain from (ii) that $\mathbf{B} \preccurlyeq^* \mathbf{A} \Rightarrow \mathbf{B}' \preccurlyeq \mathbf{A}'$ and from (i) that $\mathbf{B}' \preccurlyeq \mathbf{A}' \Leftrightarrow \mathbf{B} \preccurlyeq \mathbf{A}$, it then follows the desired result that $\mathbf{B} \preccurlyeq^* \mathbf{A} \Rightarrow \mathbf{B} \preccurlyeq \mathbf{A}$. ■

1.A.7 Proof of Corollary 1.5

Proof. The corollary holds for matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ with possibly $\mathbf{A} \cdot \mathbf{e}_{n_A} = \boldsymbol{\mu}_A \neq \boldsymbol{\mu}_B = \mathbf{B} \cdot \mathbf{e}_{n_B}$. ($\Rightarrow$ part). Suppose $\mathbf{B} \preccurlyeq \mathbf{A}$, if $\preccurlyeq$ satisfies Axiom *NG*, then there exists $\mathbf{A}' = [\text{diag}(\boldsymbol{\mu}_A)]^{-1} \cdot \mathbf{A}$ and $\mathbf{B}' = [\text{diag}(\boldsymbol{\mu}_B)]^{-1} \cdot \mathbf{B}$ such that $\mathbf{A}' \sim \mathbf{A}$ and $\mathbf{B}' \sim \mathbf{B}$. Therefore $\mathbf{B}' \preccurlyeq \mathbf{A}'$ and $\boldsymbol{\mu}_{A'} = \boldsymbol{\mu}_{B'} = \mathbf{e}_d$, thus Theorem 1.2 applies. Then it follows that $\mathbf{B}' \preccurlyeq^* \mathbf{A}'$, as required in the corollary.

($\Leftarrow$ part). Suppose that $\mathbf{B}' \preccurlyeq^* \mathbf{A}'$, then by construction Theorem 1.2 holds and therefore $\mathbf{B}' \preccurlyeq \mathbf{A}'$. Moreover, it is possible to move from $\mathbf{A}'$ to $\mathbf{A}$ and from $\mathbf{B}'$ to $\mathbf{B}$ making use of *NG* transformations. It then follows that $\mathbf{B}' \sim \mathbf{B}$ and $\mathbf{A}' \sim \mathbf{A}$. Thus by Theorem 1.2 and the transitivity of $\preccurlyeq$, we obtain that $\mathbf{B} \preccurlyeq \mathbf{A}$ for $\preccurlyeq$ satisfying *NG.* ■

1.A.8 Proof of Theorem 1.4

Proof. We prove the *sufficiency* part ($\Rightarrow$) by construction. Recall that $\mathbf{B} \preccurlyeq^R \mathbf{A}$ implies that matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ are such that $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ for $\mathbf{X} \in \mathcal{R}_{n_A, n_B}$. Given the set of composition matrices $\mathbf{X}(h)$ indexed by $h \in \{1, \dots H\}$, where $H := n_B^{n_A}$,[29] we have $\mathbf{B} = \sum_h \lambda_h \mathbf{A} \cdot \mathbf{X}(h)$

[29] H is the total number of permutations of n_A ones in a matrix with $n_A \times n_B$ entries that are either zeroes or ones.

with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_H)^t \in \Delta^H$, such that $\lambda_h \geq 0\ \forall h$ and $\sum_h \lambda_h = 1$. Any column vector $\mathbf{b}_k$ of $\mathbf{B}$ can be written as $\mathbf{b}_k = \sum_h \lambda_h \mathbf{A} \cdot \mathbf{x}_k(h)$. Therefore, the Zonotope of $\mathbf{B}$ can be written as:

$$\begin{aligned}
Z(\mathbf{B}) &= \left\{ \mathbf{z} := (z_1, \ldots, z_d)^t \ : \ \mathbf{z} = \sum_{k=1}^{n_B} \theta_k \mathbf{b}_k, \quad \theta_k \in [0,1]\ \forall k = 1, \ldots, n_B \right\} \\
&= \left\{ \mathbf{z} = \sum_{k=1}^{n_B} \theta_k \left(\sum_h \lambda_h \sum_j \mathbf{a}_j \cdot x_{jk}(h) \right), \quad \theta_k \in [0,1],\ \boldsymbol{\lambda} \in \Delta^H,\ x_{jk}(h) \in \{0,1\}, \forall k, i, j \right\} \\
&= \left\{ \mathbf{z} = \sum_{j=1}^{n_A} \mathbf{a}_j \left(\underbrace{\sum_h \lambda_h \sum_k \ \theta_k x_{jk}(h)}_{\widetilde{\theta}_j \in I} \right), \quad \theta_k \in [0,1],\ \boldsymbol{\lambda} \in \Delta^H,\ x_{jk}(h) \in \{0,1\}, \forall k, i, j \right\} \\
&= \left\{ \mathbf{z} = \sum_{j=1}^{n_A} \widetilde{\theta}_j \mathbf{a}_j, \quad \widetilde{\theta}_j \in \mathcal{I} \subset [0,1]\ \forall j \right\}.
\end{aligned}$$

The last line comes from the fact that if $x_{jk}(h) = 1$ then $x_{jk'}(h) = 0$ for all $k' \neq k$. Therefore, $\sum_k \theta_k x_{jk}(h)$ takes values on the $[0,1]$ real interval, for each h. The convex combination with weights $\boldsymbol{\lambda}$ necessary lies, at most, in the same interval. We fix such interval to be $\mathcal{I}$ and its elements are the new weights $\widetilde{\theta}_j$, as long as $\boldsymbol{\lambda}$ is considered to be fixed. As a result, matrix majorization implies that any point in $Z(\mathbf{B})$ can be written as a point in $Z(\mathbf{A})$ or equivalently $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$. When $\mathcal{I} = [0,1]$, $Z(\mathbf{B})$ *coincide* with $Z(\mathbf{A})$ and $\mathbf{B}$ is an equivalent representation of $\mathbf{A}$.

For the *necessity* part ($\Leftarrow$), we prove that $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ implies $\mathbf{B} \preccurlyeq^R \mathbf{A}$. We assume $\mathbf{A} \cdot \mathbf{e}_{n_A} = \mathbf{B} \cdot \mathbf{e}_{n_B}$ to show that if the columns of matrix $\mathbf{B}$ (indexed by k) lie in the Zonotope of $\mathbf{A}$: $\mathbf{b}_k \in Z(\mathbf{A})\ \forall k$, this is equivalent to matrix majorization, and a necessary condition for Zonotopes inclusion. Consider a set of n_B vectors $\mathbf{b}_k$ with $k \leq n_B$, which lie in $Z(\mathbf{A})$ *and* satisfy the condition $\sum_k \mathbf{b}_k = \mathbf{A} \cdot \mathbf{e}_{n_A}$. They can be written as follows (where vector k' is written in a way that satisfies the stochasticity constraint):

$$\begin{aligned}
\mathbf{b}_k :=& \quad \sum_j \theta_j(k) \mathbf{a}_j, \quad \text{for all } k \in \{1, \ldots, n_B\} \backslash k' \\
\mathbf{b}_{k'} :=& \quad \sum_j \theta_j(k') \mathbf{a}_j \ = \ \mathbf{A} \cdot \mathbf{e}_{n_A} - \sum_{k \neq k'} \sum_j \theta_j(k) \mathbf{a}_j = \sum_j \left(1 - \sum_{k \neq k'} \theta_j(k) \right) \mathbf{a}_j.
\end{aligned}$$

Given that $\theta_j(k) \in [0,1]$ and $\theta_j(k') := \left(1 - \sum_{k \neq k'} \theta_j(k)\right) \in [0,1]$, this implies that $\sum_k \theta_j(k) = 1$ with $\theta_j(k) \geq 0$ for all k including k'. We define the vector $\boldsymbol{\theta}_j = (\theta_j(1), \ldots, \theta_j(n_B)) \in \Delta^{n_B}$. The matrix $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^t, \ldots, \boldsymbol{\theta}_{n_A}^t)^t$ is a row stochastic matrix. It follows that matrix $\mathbf{B}$ can be written as $\mathbf{B} = \mathbf{A} \cdot \boldsymbol{\Theta}$ with $\boldsymbol{\Theta} \in \mathcal{R}_{n_A,n_B}$, which is $\mathbf{B} \preccurlyeq^R \mathbf{A}$. ■

1.A.9 Proof of Lemma 1.2

Proof. To prove this result, it is worth noting that for any pair of vectors $\mathbf{x} = (x_1, \ldots, x_i, \ldots, x_d)^t$ and $\mathbf{y} \in \mathbb{R}^d_{++}$ whose elements are ranked in increasing order and are such that $\mathbf{e}_d^t \cdot \mathbf{x} = \mathbf{e}_d^t \cdot \mathbf{y} = \mu > 0$, the area between the Lorenz curve $L_{\mathbf{x}}(i) = \sum_{j=1}^{i} \frac{x_j}{\mu}$ and its *dual* $\overline{L}_{\mathbf{x}}(i) = 1 - L_{\mathbf{x}}(n-i)$ (obtained by ordering the elements of $\mathbf{x}$ from the largest to the smallest) coincides with the area of the zonotope $Z\left(\left(\frac{\mathbf{x}}{\mu}, \frac{\mathbf{e}_d}{d}\right)^t\right)$ (see Koshevoy and Mosler 1996). Therefore:

$$L_{\mathbf{x}}(i) \;\geq\; L_{\mathbf{y}}(i), \; \forall i = 1, \ldots, d \quad \Leftrightarrow \quad Z\left(\left(\frac{\mathbf{x}}{\mu}, \frac{\mathbf{e}_d}{d}\right)^t\right) \subseteq Z\left(\left(\frac{\mathbf{y}}{\mu}, \frac{\mathbf{e}_d}{d}\right)^t\right). \tag{1.4}$$

For $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$, let $\mathbf{A}^*, \mathbf{B}^* \in \mathcal{M}_d$ be the pair of associated minimal ordinal comparable matrices. In this case, for all $k = 1, \ldots, n^*$, the size of vectors $\overrightarrow{\mathbf{a}^*}_k$ and $\overrightarrow{\mathbf{b}^*}_k$ is fixed to d and $\mathbf{e}_d^t \cdot \overrightarrow{\mathbf{a}^*}_k = \mathbf{e}_d^t \cdot \overrightarrow{\mathbf{b}^*}_k$.

The conditions of the well known Lemma 2.B.1 by Hardy et al. (1934) are satisfied and therefore $\mathbf{B} \preccurlyeq^* \mathbf{A}$ if and only if for each $k = 1, \ldots, n^*$, $L_{\overrightarrow{\mathbf{b}^*}_k}(i) \geq L_{\overrightarrow{\mathbf{a}^*}_k}(i)$, for all $i = 1, \ldots, d$. Using identity (1.4), the result is established. ■

1.A.10 Proof of Theorem 1.5

Proof. We first prove that $MP^*(\mathbf{A}^*) = MP^*(\mathbf{A})$ for $\mathbf{A}, \mathbf{A}^* \in \mathcal{M}_d$, where $\mathbf{A}^*$ is obtained from $\mathbf{A}$ and satisfies conditions (i) to (iv) in Definition 1.5. By construction, it follows that for any $k^* = 1, \ldots, n^*$ there exists a $k = 1, \ldots, n_A$ and $\theta \in [0,1]$ such that:

$$\overrightarrow{\mathbf{a}^*}_{k^*} := \sum_{j=1}^{k} \sum_{j^*=1_j}^{n_j} \mathbf{a}^*_{j^*} + \sum_{j^*=n_1^*+\ldots+n_k^*}^{k^*} \mathbf{a}^*_{j^*} = \overrightarrow{\mathbf{a}}_k \; + \theta \, \mathbf{a}_{k+1}, \tag{1.5}$$

and similarly for $\mathbf{B}$. For any k^*, $\overrightarrow{\mathbf{a}^*}_{k^*} \in MP^*(\mathbf{A}^*)$, and by (1.5), $\mathbf{z}^* := \overrightarrow{\mathbf{a}}_k \; + \theta \, \mathbf{a}_{k+1} \in MP^*(\mathbf{A}^*)$. Given that, by definition, $\mathbf{z}^* \in MP^*(\mathbf{A})$ and (1.5) holds for any k^*, it must follow that $MP^*(\mathbf{A}^*) = MP^*(\mathbf{A})$. A similar argument holds for $\mathbf{B}$. Hence, the inclusion of the Path Polytopes of $\mathbf{A}, \mathbf{B}$ can be equivalently studied as a problem of inclusion of the Path Polytopes of $\mathbf{A}^*, \mathbf{B}^*$.

By Lemma 1.2, if $\mathbf{B} \preccurlyeq^* \mathbf{A}$ then $\overrightarrow{\mathbf{b}^*}_k \in conv\{\mathbf{\Pi}_d \cdot \overrightarrow{\mathbf{a}^*}_k,\ \forall \mathbf{\Pi}_d\}$ for every $k = 1, \dots, n^*$. Given that $\overrightarrow{\mathbf{a}^*}_k \in MP^*(\mathbf{A}^*)$, then $\overrightarrow{\mathbf{b}^*}_k \in Z^*(\mathbf{A}^*)$ by definition.

To conclude the proof, it is necessary to extend the inclusion argument over the entire domain of the Path Polytope. We exploit the rank preserving property of the partition $k = 1, \dots, n^*$.

To show the *sufficiency* part ($\Rightarrow$), note that for any pair k and $k+1$ of contiguous classes, by construction the ranking of the groups within each class (defined by increasing magnitude of cumulative groups population masses within the class) is preserved in both classes and for both configuration $\mathbf{A}^*$ and $\mathbf{B}^*$.

The comparisons have to be made at "fixed mean", so that one can exploit the test proposed in Lemma 1.2 to check whether the Lorenz curve of $\theta\overrightarrow{\mathbf{b}^*}_k + (1-\theta)\overrightarrow{\mathbf{b}^*}_{k+1}$ lies above the Lorenz curve of $\theta\overrightarrow{\mathbf{a}^*}_k + (1-\theta)\overrightarrow{\mathbf{a}^*}_{k+1}$, for any $\theta \in [0, 1]$. This comparison preserves the means, since $\mathbf{e}_d^t \cdot \left(\theta\overrightarrow{\mathbf{b}^*}_k + (1-\theta)\overrightarrow{\mathbf{b}^*}_{k+1}\right) = \mathbf{e}_d^t \cdot \left(\theta\overrightarrow{\mathbf{a}^*}_k + (1-\theta)\overrightarrow{\mathbf{a}^*}_{k+1}\right)$. Given any two ordered Lorenz curves, a sufficient condition for having that a third Lorenz curve lies in the area between the two initial curves is that the two distributions underlying the two curves are obtained one from the other by a finite sequence of Pigou-Dalton transfers that preserve the rank of the (population of d) individuals in both distributions. This particular structure applies to comparisons involving contiguous sections k and $k+1$ with fixed means (because $\mathbf{A}^*$ and $\mathbf{B}^*$ are rank comparable).

Following Lemma 1.2, if the Lorenz curve of $\overrightarrow{\mathbf{b}^*}_k$ lies above the one of $\overrightarrow{\mathbf{a}^*}_k$, and the Lorenz curve of $\overrightarrow{\mathbf{b}^*}_{k+1}$ lies above the one of $\overrightarrow{\mathbf{a}^*}_{k+1}$, then the Lorenz curve associated to the convex combination of the initially less disperse configurations $\overrightarrow{\mathbf{b}^*}_k$ and $\overrightarrow{\mathbf{b}^*}_{k+1}$, lies above the Lorenz curve associated to the convex combination of the initially more disperse configurations $\overrightarrow{\mathbf{a}^*}_k$ and $\overrightarrow{\mathbf{a}^*}_{k+1}$. The Lorenz test can be alternatively written by $\theta\overrightarrow{\mathbf{b}^*}_k + (1-\theta)\overrightarrow{\mathbf{b}^*}_{k+1} \in \text{conv}\left\{\mathbf{\Pi}_d \cdot \left(\theta\overrightarrow{\mathbf{a}^*}_{j^*} + (1-\theta)\overrightarrow{\mathbf{a}^*}_{j^*+1}\right) \mid \mathbf{\Pi}_d \in \mathcal{P}_d\right\}$. As a consequence, the SUM order is equivalently represented by this sequence of inclusions, holding for all $k \in 1, \dots, n^*$ and for all $\theta \in [0, 1]$, which implies $MP^*(\mathbf{B}^*) \subseteq Z^*(\mathbf{A}^*)$.

The *necessity* part ($\Leftarrow$) is easier to prove, because $MP^*(\mathbf{B}^*) \subseteq Z^*(\mathbf{A}^*)$ implies that any given $\mathbf{p} \in MP^*(\mathbf{B}^*)$ can be written as a convex combination of the permutations of $\mathbf{z}^* \in Z^*(\mathbf{A}^*)$, such that $\mathbf{e}_d \cdot \mathbf{p} = \mathbf{e}_d \cdot \mathbf{z}^*$. By taking $\mathbf{p} = \overrightarrow{\mathbf{b}^*}_k$ and $\mathbf{z}^* = \overrightarrow{\mathbf{a}^*}_k$, for all $k = 1, \dots, n^*$, one gets the desired result. $\blacksquare$

1.B Appendix: An algorithm to test Zonotopes inclusion

We exploit the link between Zonotopes and their *central representations* reported in McMullen (1971) to obtain empirical conditions for Zonotopes inclusion. After introducing some concepts of geometric analysis of Zonotopes and Zonohedrons (Zonotopes with $d = 3$), we show that a reduced set of hyperplane slopes need to be computed in order to check for inclusion, by exploiting the method based upon the supporting hyperplane. Throughout this section we assume that the distribution matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_d$ satisfy the following requirements: (i) matrices do not contain empty columns, (ii) the columns which are proportional one to another by a positive scalar are merged together, (iii) $d < n_A$ and $d < n_B$. If (i) and (ii) hold and $d = n_A = n_B = n$, then $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ if and only if there exists $\mathbf{X} = \mathbf{A}^{-1} \cdot \mathbf{B}$ which is row stochastic. If $d > n_A$, than $Z(\mathbf{A})$ is not a d-dimensional polytope in $\mathbb{R}^d_+$, and the inclusion algorithm cannot be successfully applied.

Given a matrix $\mathbf{A} \in \mathcal{M}_d$ with $n_A = n$, and a permutation π of its n columns, a vertex (either internal or external) of the Zonotope ($\mathbf{v} \in Z(\mathbf{A})$) can be written as $\mathbf{v} = \varepsilon_1 \mathbf{a}_1 \oplus \ldots \oplus \varepsilon_n \mathbf{a}_n$ with $\varepsilon_j \in \{0, 1\}$ for all $j \leq n$. There are 2^d of such vertices.

An r-dimensional face F_A of $Z(\mathbf{A})$ (with $r \leq d - 1$) can always be written as:

$$F_A := \mathbf{S}_{1\pi} \oplus \ldots \oplus \mathbf{S}_{r\pi} \;\oplus\; \varepsilon_{(r+1)\pi} \mathbf{a}_{(r+1)\pi} \oplus \ldots \oplus \varepsilon_{n\pi} \mathbf{a}_{n\pi},$$

where we use the index $j\pi$ to identify a permutation of the columns of the distribution matrix $\mathbf{A}$, and $\mathbf{S}_{r\pi}$ stands for the convex hull between the origin and the vector $\mathbf{a}_{r\pi}$.

By adapting Theorem 1 and 2 in McMullen (1971) to our case and considering only $r = d - 1$, it is possible to verify whether F_A is a face of $Z(\mathbf{A})$ by checking that the point $(\frac{1}{2}, \ldots, \frac{1}{2})$ belongs to the convex hull of the *central representation* generated by F_A. This occurs if and only if there exist a vector $\mathbf{u} \in \mathbb{R}^d$ that solves the following program:[30]

$$\mathbf{u}^t \cdot \frac{1}{2}(\mathbf{e}_d \oplus \mathbf{a}_{j\pi}) = 0 \;\; (j\pi \leq r\pi) \;\text{ and }\; \mathbf{u}^t \cdot \frac{1}{2}(\mathbf{e}_d \oplus \mathbf{a}_{j'\pi}) > 0 \;\; ((r+1)\pi \leq j'\pi \leq n\pi). \quad (1.6)$$

The program can be solved with standard optimization methods. If a solution is found, F_A is indeed a face of $Z(\mathbf{A})$ and the slopes of the hyperplane supporting F_A can be recovered.

An hyperplane in $\mathbb{R}^d$ is identified by d points, although only $d - 1$ points are needed

[30] We provide conditions for testing with Zonotopes whose symmetry point is $\frac{1}{2}\mathbf{e}_d$. McMullen (1971) consider Zonotopes that are centrally symmetric with respect to the origin.

to identify its slopes, since the hyperplane corresponds to a translation of the isomorphic hyperplane passing through the origin. Therefore, we identify the slopes of the hyperplane supporting the faces of the Zonotope making use of $d-1$ columns of the distribution matrix generating it.[31] Let $\mathcal{H}$ be the set of all possible $\frac{n!}{(d-1)!(n-d-1)!}$ permutations of $d-1$ columns of $\mathbf{A}$. We define $\mathbf{A}(H)$, $H \in \mathcal{H}$, the corresponding $d \times (d-1)$ matrix. Let $\boldsymbol{\beta}(H) \in \mathbb{R}^d$ be the column vector of slopes of the hyperplane generated by vectors in H.

Slopes can be now identified by solving the following program (the $\mathbf{0}_d$ has to be included since the hyperplane passes through the origin):

$$(\mathbf{0}_d,\, \mathbf{A}(H))^t \cdot \boldsymbol{\beta}(H) \;=\; \mathbf{0}_d, \quad \forall H \in \mathcal{H}. \tag{1.7}$$

Assume, without loss of generality, that $\beta_1(H) = 1$. Since $\mathbf{A}$ satisfies the minimal structure requirements listed above, the vector of slopes is identified. Let $\widehat{\mathbf{A}}(H)$ be a $(d-1) \times (d-1)$ matrix corresponding to the matrix $\mathbf{A}(H)$ diminished of the first *row* $\mathbf{a}_1^t(H)$, a $1 \times (d-1)$ vector. Then:

$$\boldsymbol{\beta}(H) := \begin{pmatrix} 1 \\ -\left(\mathbf{a}_1^t(H) \cdot \widehat{\mathbf{A}}^{-1}(H)\right)^t \end{pmatrix}. \tag{1.8}$$

For a point $\mathbf{z} \in F_A \subseteq Z(\mathbf{A})$, the set of slopes associated to the hyperplane supporting the Zonotope face F_A is $\boldsymbol{\beta}(H_{F_A})$, and the level of this hyperplane can be calculated by: $\mathbf{z}^t \cdot \boldsymbol{\beta}(H_{F_A})$.

The following algorithm allows to test for Zonotopes inclusion, for two distribution matrices $\mathbf{A}$ and $\mathbf{B}$. The algorithm allows to identify the faces and the slopes of the associates hyperplane and, finally, it permits to compare the associated values.

Algorithm 1.1 *The algorithm checks inclusion under the assumption that* $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$*:*

1. *Using (1.8), identify the set of all slopes of hyperplane generated by the* $(d-1)$ *columns of* $\mathbf{A}$*, given by the set* H*, and the origin. This set is* $\mathcal{B} := \{\boldsymbol{\beta}(H) \; : H \in \mathcal{H}\}$ *whose norm is equal to the number of permutations of* $(d-1)$ *columns of* $\mathbf{A}$*.*

2. *Identify the faces of* $Z(\mathbf{A})$ *that contain the origin. To do so, construct the set* $\mathcal{H}(0) \subset \mathcal{H}$ *of columns of* $\mathbf{A}$ *that, together with the origin (*$\mathbf{0}_d$*), solve problem (1.6), with*

[31] For instance, if the Zonotope is a polyhedron in $\mathbb{R}^3$, the supporting hyperplane for its faces are planes in the 2-dimensional space.

$r = (d-1)+1$. *For* $H \in \mathcal{H}(0)$ *it must hold that* $\mathbf{z}^t \cdot \boldsymbol{\beta}(H) = 0$ *for all* $\mathbf{z} \in F_H$.

3. *Check that the columns of* $\mathbf{B}$ *lie in the cone generated by the columns of* $\mathbf{A}$. *It is sufficient to check that:*

$$\left\{\mathbf{b}_j \;:\; \mathbf{b}_j^t \cdot \boldsymbol{\beta}(H) \;>\; \mathbf{z}^t \cdot \boldsymbol{\beta}(H),\ \forall j \le n_B,\ \mathbf{z} \in F_H\right\} = \emptyset, \quad \forall H \in \mathcal{H}(0). \tag{1.9}$$

4. *If (1.9) is not verified then* $Z(\mathbf{B}) \nsubseteq Z(\mathbf{A})$. *Repeat steps 1, 2, and 3 under the assumption that* $Z(\mathbf{A}) \subseteq Z(\mathbf{B})$. *If step 3 rejects this hypothesis, then* $\mathbf{A}$ *and* $\mathbf{B}$ *are not comparable in terms of dissimilarity. Otherwise, proceed.*

5. *Let* $\overline{\mathcal{H}(0)} = \mathcal{H}/\mathcal{H}(0)$ *be the set of all* $(d-1)$ *columns of* $\mathbf{A}$ *that does not generate an hyperplane passing from the origin. For any* $H \in \overline{\mathcal{H}(0)}$ *there exists a subset of a permutation of the columns of* $\mathbf{A}$, *indexed by* $1\pi(H), \ldots, j\pi(H)$, *such that when merged to obtain* $\overrightarrow{\mathbf{a}}_{j\pi(H)}$, *these columns give the position in the space of the face defined by* H. *These columns must satisfy: (i)* $\overrightarrow{\mathbf{a}}_{j\pi(H)} \neq \mathbf{0}_d$, *(ii)* $\mathbf{a}_{1\pi(H)}, \ldots, \mathbf{a}_{j\pi(H)}$ *do not belong to* H *and (iii) that* $\overrightarrow{\mathbf{a}}_{j\pi(H)}$ *together with* H *determine a solution of problem (1.6), with* $r = d$. *The* d *columns generate the face* F_H *of* $Z(\mathbf{A})$. *Moreover, let* $\mathcal{V}(\mathbf{B})$ *the set of all vertices of* $Z(\mathbf{B})$. *To check* $Z(\mathbf{B}) \subseteq Z(\mathbf{A})$ *it is sufficient to check that:*

$$\left\{\mathbf{v} \;:\; \mathbf{v}^t \cdot \boldsymbol{\beta}(H) \;>\; \mathbf{z}^t \cdot \boldsymbol{\beta}(H),\ \mathbf{v} \in \mathcal{V}(\mathbf{B}),\ \mathbf{z} \in F_H\right\} = \emptyset, \quad \forall H \in \overline{\mathcal{H}(0)}. \tag{1.10}$$

To compute the value of the hyperplane supporting F_H, *it suffice to see that one of the vertices of the face* $\mathbf{z} \in F_H$ *is given by* $\mathbf{z} = \varepsilon_{(r+1)H}\mathbf{a}_{(r+1)H} \oplus \ldots \oplus \varepsilon_{n_A H}\mathbf{a}_{n_A H}$ *with* $r = d$. *Hence, the vector* $\mathbf{z} = \overrightarrow{\mathbf{a}}_{j\pi(H)}$ *can be used to compute the value of the hyperplane supporting face* F_H.

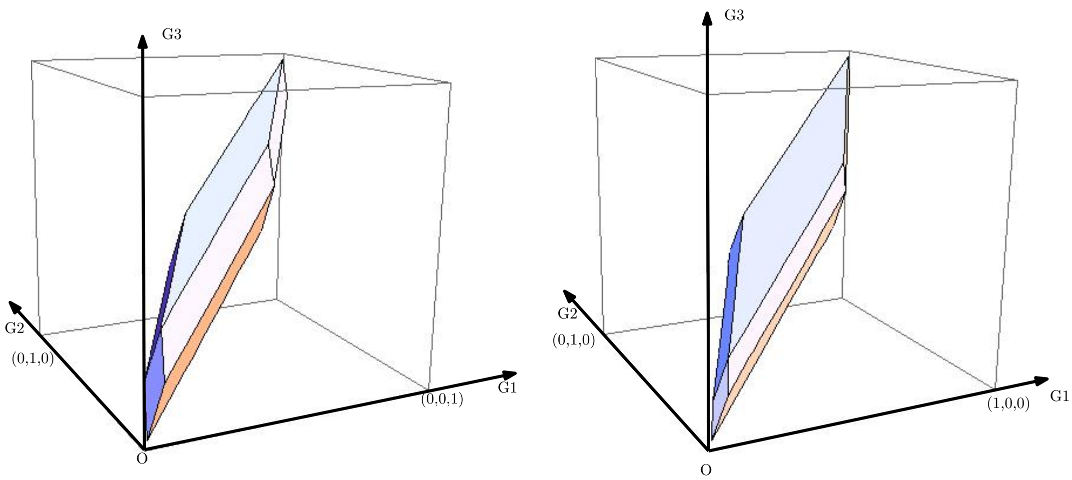
1.C Appendix: An application to the case of residential segregation in an Italian city

In this appendix we use the dissimilarity partial order to rank segregated distributions of d groups across n residential units according to the criterion of *multi-group a-spatial residential segregation* (see for instance Frankel and Volij 2011). An allocation A is said to be at least as segregated as B if in B the groups share more common space than they do in

A (Massey and Denton 1988). This can be measured by the distribution matrices $\mathbf{A}$ and $\mathbf{B}$. We conclude that "A is at least as segregated as B" if and only if $\mathbf{B} = \mathbf{A} \cdot \mathbf{X}$ with $\mathbf{X} \in \mathcal{R}_n$.

We consider the problem of immigrants segregation in a representative Italian city, Verona, for the comparison years 2000 and 2005. The problem of segregation of natives and immigrants (partitioned in Africans, Asians, East Europe and other classes) has been studied by Andreoli (2012), who relates immigrants groups distributions across city neighborhoods in a dynamic perspective, and shows an empirical methodology to depict distributional changes and to correlate them with social-demographic and structural attributes of the city neighborhoods. Similarly to Andreoli (2012), we partition the Verona's residential area into three major areas, involved in the internal migration process: the city historic center (HC), two peripheral areas characterized by poor residential environments and characterized by intense migration flows between 2000 and 2005 (*A1* and *A2* respectively) and the remaining neighborhoods (R). We partition the population in three major groups: natives (*G1*), Sub-Saharian Africa Immigrants (*G2*) and the remaining immigrants groups (*G3*). The choice is motivated by the fact that the second group concentration is significantly positively associated with the concentration of all the other groups, controlling for local neighborhood characteristics and, more importantly, this group has experienced major reductions in distance from the natives distribution by the year 2000.

We compare the 3×4 distribution matrices $\mathbf{V}_{00}$ and $\mathbf{V}_{05}$, whose entries are represented in bold characters in table 1.1. The example allows to represent graphically the Zonotopes in the three-dimensional space and to test for their (or the bivariate segregation curves) inclusion. Previous results showed that from 2000 to 2005 there have been a massive movement of African groups from area *A1* to *A2* thus decreasing their degree of segregation with respect to natives. The rest of the immigration groups experienced a sustained growth in the quinquennium (the weight of *G2* almost tripled in the period), with section *A2* experiencing the major increase. Again, partial analysis on segregation curves showed that the increase was not sufficient to increase segregation with respect to natives. Two groups segregation indices (Dissimilarity, Gini and Entropy) give a similar result. The Zonotope analysis in figure 1.4 suggests, however, a different picture. In fact, the inclusion of $Z(\mathbf{V}_{00})$ and $Z(\mathbf{V}_{05})$ is not verified. This proves the relevance of the multi-group dissimilarity analysis when more than two groups have to be compared, and traditional two groups comparisons may provide misleading results.



(a) $\mathbf{V}_{00}$ (b) $\mathbf{V}_{05}$

Figure 1.4: Zonotopes for segregated distributions

Table 1.1: Distribution matrices and population frequencies for $\mathbf{V}_{00}$ and $\mathbf{V}_{05}$.

	$\mathbf{V}_{00}$					$\mathbf{V}_{05}$				
(a) Distribution Matrices										
	HC	A1	A2	R	*Pop (Group)*	HC	A1	A2	R	*Pop (Group)*
G1	**0.081**	**0.037**	**0.228**	**0.653**	*0.951*	**0.079**	**0.035**	**0.220**	**0.666**	*0.911*
G2	**0.061**	**0.117**	**0.313**	**0.509**	*0.021*	**0.066**	**0.077**	**0.376**	**0.481**	*0.025*
G3	**0.143**	**0.130**	**0.275**	**0.452**	*0.028*	**0.101**	**0.086**	**0.345**	**0.468**	*0.064*
Pop (Section)	*0.083*	*0.041*	*0.231*	*0.645*	*1.000*	*0.080*	*0.040*	*0.232*	*0.649*	*1.000*
(b) Absolute Frequencies										
	HC	A1	A2	R	*Pop (Group)*	HC	A1	A2	R	*Pop (Group)*
G1	19901	9068	55871	159981	*244821*	18637	8367	51863	157385	*236252*
G2	326	629	1685	2739	*5379*	432	505	2469	3162	*6568*
G3	1038	948	2003	3288	*7277*	1673	1422	5715	7750	*16560*
Pop (Section)	*21265*	*10645*	*59559*	*166008*	*257477*	*20742*	*10294*	*60047*	*168297*	*259380*

Source: The 2000 Census (ISTAT) for the municipality of Verona and the Vital Statistics Official Register for 2005.

Note: The four panels depict the conditional distribution of Verona population conditional on natives (G1), African immigrants (G2) and other immigrants (G3), in four areas: Historic City Center (HC) peripheral areas (A1, A2) and the rest (R). The upper panels (case *(a)*) depict the distributional matrices $\mathbf{V}_{00}$ and $\mathbf{V}_{05}$, while the lower panels (case *(b)*) provide the absolute frequencies distributions.

Chapter 2

The measurement of multi-group interaction and segregation

Contents

2.1 Introduction

One leading mechanism of the persistence of differences in outcomes prospects across individuals is determined by the unequal distribution of the chances that these individuals have to interact with the relevant socio-economic groups in which the population can be partitioned into. Unequal access to interaction across networks, segregation and social isolation stem from this type of inequalities, that can be studied by resorting to models of interaction specified at individual level.

In their seminal analysis on segregation measurement in he context of spatial interaction models, Massey and Denton (1988) define five dimensions of analysis for segregation: evenness, exposure, clustering, centralization and concentration. These dimensions can be analyzed in the case where the population is partitioned in two groups, or alternatively in the multi-group setting. We focus our attention on multi-group measures of segregation for the exposure dimension.

Consider a population exogenously partitioned into social groups. Following Massey and Denton (1988), exposure indices should capture the differences across groups in the likelihood that any randomly selected individual from one of these groups interacts with another person in the population. Segregation is zero when the chances that two individuals interact are made independent on their respective groups of origin. On the contrary, segregation is maximized whenever the likelihood to interact with a given individual in the population is different from zero only for those coming from one given social group.

These models gather together different concerns about segregation. Consider a large population that can be partitioned into two groups of equal size, the "Reds" and the "Greens". Suppose that data on networks are available, then it is possible to assess segregation by looking at individual interaction patterns. Let consider the case in which each individual interacts with half of the remaining individuals, the degree of segregation depends on how different types interact among them, given the same observed network. Three possible configurations are of particular interest. In the first configuration, each individual interacts with half of the Reds and half of the Greens. Admittedly, there is no segregation. In a second configuration, every individual of the Reds interacts with all the Reds and exclusively with them, and *vice versa* for the Greens. Admittedly, an highly segregated society. The third configuration is such that every individual of the Reds interacts with all the Greens and exclusively with them, and *vice versa* for the Greens. Hence, the likelihood that a Red interacts with individual i is positive if i is Green, it is zero whenever i is Red.

As far as Reds and Greens are treated symmetrically, the last configuration is as much segregated as the second one. Virtually all the indices of segregation that can be adapted to measure exposure from the individual perspective are grounded on this symmetry property, and they agree on ranking the last two configurations as equally more segregated than the first (Massey and Denton 1988, Hutchens 1991, Flückiger and Silber 1999, Reardon and Firebaugh 2002, Reardon and O'Sullivan 2004, Frankel and Volij 2011). This issues arises because segregation indices are constructed on properties of transfers or exchanges of individuals across groups and units in which the interaction space is partitioned into that are not informative of the structure of interaction patterns across individuals/groups of the population.

Moreover, none of these indices has been designed to deal with problems of segregation that use individual level data and the axiomatization of these indices, where it exists, cannot be meaningfully adapted to capture segregation patterns across individuals interaction profiles. We fill this gap by proposing the appropriate axiomatic setting to study segregation at individual level, and we fully characterize a class of segregation indicators.

Developing the idea that segregation at the individual level can be assessed by measuring the degree of inequality in the distribution of the *interaction profiles* across the population, our analysis bridges the gap between two research fields: segregation measurement and multivariate inequality analysis (Koshevoy 1995, Koshevoy and Mosler 1996). In our stylized model, every individual is endowed with an interaction profile, specifying the probability that this individual interacts with a set of groups in which the population is partitioned into. We show that measuring the inequality in the distribution of interaction profiles across the population is equivalent to order distributions coherently with the exposure dimension of segregation.

We propose an axiomatic structure to characterize the exposure segregation order. The axioms that we introduce define operations on interaction profiles that preserve or decrease segregation. Our analysis is grounded in a simple but powerful principle: if two individual's interaction profiles are merged together, so that each of the new interaction profiles corresponds to a convex combination of the old profiles, then segregation is reduced, and exposure equalized across individuals. If all individuals profiles are merged, everybody experiences the same interaction profile, and equal exposure is reached.

The implementation of this principle does not require a symmetric treatment of the groups, and the resulting methodology is then able to solve the issues raised by configurations two and three in the example with the Reds and the Greens. In fact, the merge of the profiles of a Red and a Green in the second configuration reduces segregation as much as the merge of the profiles of a Red and a Green in the third configurations. This is an intuitive result, since in our example either a merge takes place between individuals with identical profiles, and nothing changes, or a merge occurs between individuals with very different profiles, and then segregation is reduced. This allows to rank all configurations obtained by merging individual profiles as less segregated than configurations two or three in the example. However, configurations two and three are ranked as equally segregated only if the segregation order is equipped with indifference with respect to permutations of the groups. The result is a segregation partial order that treats the two groups symmetrically, and allows to construct a family of indicators that are consistent with the segregation literature.

However, one may exploit the merge axiom to construct more sophisticated and non-symmetric partial orders of segregation. This can be done, for instance, by requiring that the merge of a Red with a Green always reduces segregation, while merges within the Reds or the Greens increase, or at most preserve, the degree of segregation. We leave the study of these alternative configurations for future research.[1].

In this chapter we show that the segregation comparisons grounded in our normative

[1] An example on how these measures can be derived from the data is given in appendix 2.A

setting can be equivalently represented by the ranking produced by the dissimilarity order studied in chapter 1. In addition, we prove that the segregation order can be also represented by a family of segregation measures satisfying regular properties (Mosler 1994, Dahl 1999), and we study one the indices in this family, the *Gini Exposure index*.

Our segregation model can be analyzed by resorting on the dissimilarity partial order, or equivalently by exploiting the Zonotopes inclusion criterion. Zonotopes are constructed from matrices reporting by row groups and by column individuals. Each row represents the likelihood of interacting with one given individual in the distribution, made conditional on the groups under analysis. From this matrix, we construct a Zonotope, called the Segregation Zonotope, the multi-group counterpart of the segregation curve (Duncan and Duncan 1955, Hutchens 1991, Silber 1989) for the exposure dimension. The Gini Exposure index is the volume of the Segregation Zonotope. We look at differences across groups in the likelihood of interacting with individuals, as Koshevoy and Mosler (1996) look at the distribution of goods shares across the population. However, we do not impose a reference distribution to construct similarity comparisons, while in Koshevoy and Mosler the reference distribution coincides with the population weighting scheme. As a result, we only consider heterogeneity in the distribution of interaction profiles across individuals.

In fact, our analysis of segregation allows to treat separately two forms of heterogeneity in interaction profiles. The heterogeneity *between* interaction profiles is the source of segregation patterns. Segregation is zero only when interaction profiles are equalized across the population. This type of equalization does not eliminate the heterogeneity *within* profiles, a characteristic of the composition of an interaction pattern. There is no within profile heterogeneity whenever individuals have the same probability to interact with all groups. Clearly, no within profiles inequality implies no between profiles inequality, but the inverse is not true. As far as the within heterogeneity may reflect structural components, the size of the groups, or individuals' preferences for interaction, it should not be accounted for in the evaluation process. To eliminate the within component, we resort to a normalization axiom that allows to compare distributions differing in the composition of the *expected* interaction profile.

Compared to the other indices of segregation, the Gini Exposure index has an advantage: it can be easily applied, decomposed and interpreted to assess the degree of social exposure in a variety of problems the involve information at the individual level, and interaction profiles may vary substantially among individual in the same group, and even among individuals attributed to the same organizational units in which the interaction space is partitioned into. We propose to use the Gini Exposure index to asses the dynamics of segregation across Italian municipalities, and we use a *spatial* model of interaction to determine the interaction profiles associated to each individual in each municipality.

We use Italian data by ISTAT to study the degree of spatial segregation of immigrant groups across municipalities in Italy. We use a spatial model to identify interaction probabilities across Italian municipalities (nearly 8400), for each of the 101 Italian provinces separately, in an interval of eight years by 2003 to 2010. We consider three groups: the groups of immigrants coming from low HDI and high HDI countries and the natives group. We apply our Gini Exposure index to study multi-group spatial segregation patterns.

In this setting, we treat municipalities as the basic units of our analysis. Unfortunately, we cannot exploit the heterogeneity in interaction profiles within each municipality. We associate the same profile to each individual living in the same municipality. The municipalities weights are then proportional to their demographic sizes. One expects that, if there is no segregation, the interaction patterns with the three groups do no systematically vary across the municipalities in the same province.[2]

How does the Gini Exposure index behave compared to traditional segregation indicators for exposure? The other indicators can only be partially analyzed according to our simple axiomatic structure. Our empirical application allows to tackle this issue. We resort therefore to an empirical analysis on a large sample of observations to assess the differences in the patterns and ranking of segregation produced by the Gini Exposure index compared to the other measures. We identify two types of families of indices, and the Gini Exposure index behaves coherently with other distance based indices of segregation, like the spatial,

[2]We consider a province as the reference metropolitan area. The province administrative center usually coincides with the province larger municipality, and we study the dispersion of social groups around this central municipality.

multi-group index of dissimilarity.

In the rest of the chapter, we axiomatically define the normative content of the exposure segregation ordering (section 2.2) and we study an empirical, multi-group test for the exposure segregation ordering based on Segregation Zonotopes inclusion, as well as a class of consistent segregation measures (section 2.3). Our main result in section 2.3.3 demonstrates the equivalence of these tools. The Gini Exposure index is presented in section 2.4. An empirically implementable version of the index is proposed and used to discuss the issue of comparisons of spatial segregation patterns (section 3.5), and we conclude with the empirical application to Italian data (section 2.6).

2.2 Definition of the exposure segregation ordering

2.2.1 Notation

In this chapter we analyze and compare *allocations* of interaction profiles across individuals. Let $\mathcal{A}(G)$ be the set of all allocations characterized by a fixed number of groups but variable population size, then:

Definition 2.1 (Allocation) *An allocation $A \in \mathcal{A}(G)$ is represented by a triplet*

$$\left[\mathcal{N}(A),\ \mathcal{G},\ ((\pi_{gi}(A))_{g\in\mathcal{G}}, \xi_i(A))_{i\in\mathcal{N}(A)}\right]$$

where $\mathcal{N}(A)$ is a finite, nonempty set of individuals with size $N(A)$, $\mathcal{G}$ is a finite, nonempty set of G population groups, with variable demographic size denoted by $N_g(A)$. For each individual $i \in \mathcal{N}(A)$ and group $g \in \mathcal{G}$, the variable $\pi_{gi} \in [0,1]$ represents the probability that individual i interacts with a randomly selected individual from group g, conditional on the probability of observing i (that is its demographic weight), denoted by $\xi_i(A)$.[3]

[3]The weight $\xi_i(A)$ may vary across individuals if the population of interest is heterogeneous in some respects. The weighting scheme is useful when data have to be scaled by some demographic factors, for instance if one observes interaction probabilities for households and individuals, or if one wants to assign different priorities to children with respect to adults when accounting for the distribution of interaction probabilities. One particular case is the uniform weighting scheme, where $\xi_i(A) = 1/N(A)$ for all $i \in \mathcal{N}(A)$.

We define an ordering of allocations within $\mathcal{A}(G)$, or, equivalently, of the triplets representing the allocations.

For each individual $i \in \mathcal{N}(A)$, an *interaction profile* is a column vector

$$\boldsymbol{\pi}_i(A) := (\pi_{1i}(A), \ldots, \pi_{Gi}(A))^t \in [0,1]^G,$$

representing the distribution of probabilities that individual i interacts with each of the groups in $\mathcal{G}$. The interaction probabilities are defined up to an aggregation constraint: $\sum_{g \in \mathcal{G}} \pi_{gi}(A) = 1$. An interaction profile can be alternatively interpreted as a *lottery*, assigning to each outcome (groups) a probability or realization (probability of interaction). In this chapter, we deal with mixtures of such lotteries.

For each group $g \in \mathcal{G}$, the *group profile* is a row vector $\boldsymbol{\pi}_g(A) := \big(\pi_{g1}(A), \ldots, \pi_{gN(A)}(A)\big) \in [0,1]^{N(A)}$ denoting the distribution in the population of the interaction probabilities with group g.

The *interaction matrix* $\boldsymbol{\pi}(A)$ is a $G \times N(A)$ matrix which collects by row all the G groups profiles or, alternatively, by column all the $N(A)$ interaction profiles. In general, the interaction profiles may vary consistently across individuals and they may well depend on other factors than population counts or groups shares.

Denote $\mathbb{E}[\pi_g(A)] = \sum_{i \in \mathcal{N}(A)} \xi_i(A)\, \pi_{gi}(A)$ the *expected interaction* probability with group g in the population. For allocation A, the $G \times N(A)$ *distribution matrix* $\mathbf{A}$ represent, by column, the entire distribution of the *normalized* interaction profiles across the population, such that the element gi in $\mathbf{A}$ is defined as $a_{gi} := \frac{\pi_{gi}(A)}{\mathbb{E}[\pi_g(A)]}$, a ratio measure of relative distributional inequality.

We stress that the distribution matrix may be constructed from interaction profiles that gather information on interaction at individual level, and can be observed, calculated from network relations, or estimated by econometric models. This feature has never been explored in segregation literature.

In shorthand notation, we indicate the interaction probability conditional on group by the $G \times N(A)$ matrix $\mathbf{A} \cdot \boldsymbol{\xi}(A)$, where $\boldsymbol{\xi}(A) = \text{diag}(\xi_1(A), \ldots, \xi_{N(A)}(A))$ and $\text{diag}(\mathbf{x})$ indicates an identity matrix with the elements along the diagonal replaced by the corresponding elements of the vector $\mathbf{x}$. The element gi of $\mathbf{A} \cdot \boldsymbol{\xi}(A)$ is written as:

$$a_{gi}\xi_i(A) = \frac{\pi_{gi}(A)\xi_i(A)}{\mathbb{E}[\pi_g(A)]},$$

which gives a probability that any randomly draw individual from group g interacts with individual i. Therefore $\sum_{i \in \mathcal{N}(A)} a_{gi}\xi_i(A) = 1$.

Finally, let $\mathbf{s}(A) = (s_1(A), \ldots, s_G(A))$ the vector of groups weights in the population, satisfying $s_g(A) := \frac{N_g(A)}{N(A)}$, for all $g \in \mathcal{G}$.

In the following, we motivate that exposure can be analyzed as a form of multidimensional inequality.

2.2.2 The exposure segregation ordering: a normative approach

We study and characterize here partial and complete orderings of segregation. We use the term *segregation* to indicate any departure from the situation of equal exposure, occurring when individuals share the same interaction profile.[4]

A *segregation ordering for the exposure dimension* $\preccurlyeq$ on a class of allocations $\mathcal{A}(G)$ is a complete and transitive binary relation (that is, a *quasi-order*) on that set of allocations, with symmetric part $\sim$.[5] For $A, B \in \mathcal{A}(G)$, we interpret $B \preccurlyeq A$ to mean that "allocation A is at least as much segregated in the exposure dimension as allocation B."

Differently from other works (such as Alonso-Villar and del Rio 2010, Reardon and

[4]This notion of exposure coincides with some alternative notions of segregation presented in literature. Following Massey and Denton (1988), exposure requires that the degree of potential contact, or the possibility of interaction between majority and minority members (or in general any group) within organizational units is maximal. Reardon and Firebaugh (2002) an Reardon and O'Sullivan (2004) underline that segregation can be conceptualized as a form of "disproportionality in groups proportions", when $\pi_{gi}(.)$ and the expected degree of interaction with group g are compared across individuals and groups. Alternatively, our notion of perfect exposure coincide with the least segregated allocation according to the idea that "an individual is more segregated the more segregated are the individuals with whom she interacts with" endorsed by Echenique and Fryer (2007).

[5]We use $B \sim A$ if and only if $B \preccurlyeq A$ and $A \preccurlyeq B$, and $A \prec B$ if and only if $A \preccurlyeq B$ but not $B \preccurlyeq A$.

Firebaugh 2002, Reardon and O'Sullivan 2004, Hutchens 2001, Echenique and Fryer 2007) we propose a minimal set of axioms that only have an ordinal interpretation (as in Frankel and Volij 2011). The axioms are based on meaningful transformations of the interaction profiles at individual level that either preserve or decrease segregation.

The first group of axioms is technical: we rule out the possibility that all allocations are indifferent and we introduce a continuity property (thus defining the indifference set) for the segregation order.

Axiom *Nontriviality* *There exists allocations $A, B \in \mathcal{A}(G)$ such that $B \prec A$.*

Axiom *Continuity* *Given $C \in \mathcal{A}(G)$, the sets $\{A \in \mathcal{A}(G) : \ C \preccurlyeq A\}$ and $\{B \in \mathcal{A}(G) : B \preccurlyeq C\}$ are closed.*

The second group of axioms posits that the segregation ranking is preserved if population is replicated, permuted or if one focuses on ratio scale measures of interaction. We assume for all axioms, apart form Normalization, that $\mathbb{E}[\pi_g(B)] = \mathbb{E}[\pi_g(A)]$ for all groups $g \in \mathcal{G}$, and that $\mathcal{N}(A) = \mathcal{N}(B)$ if not differently stated.

Axiom *Normalization* *For $A, B \in \mathcal{A}(G)$, if $\boldsymbol{\pi}(B) \neq \boldsymbol{\pi}(A)$ but $\mathbf{B} = \mathbf{A}$ and $\xi_i(B) = \xi_i(A)$ for all i, then $B \sim A$.*

Axiom *Units Anonymity* *If B is obtained from A by permuting the name of units in the population, then $B \sim A$.*

Axiom *Groups Anonymity* *If B is obtained from A by permuting the name of groups, then $B \sim A$.*

Axiom *Individual Replication Invariance* *If one individual's interaction profile in A is replicated to obtain B so that $\mathcal{N}(B) = \mathcal{N}(B) \cup \{i'\}$ but $\xi_i(B) + \xi_{i'}(B) = \xi_i(A)$, then $B \sim A$.*

The Normalization axiom states that any comparison of allocations should be based on distribution matrices, which are independent on the expected degree of exposure in the distribution. The latter is assumed to be fixed in the remaining axioms. The axiom points that only distributional information is important for assessing exposure, and that allocations that differ in the expected interaction profiles can be made comparable.

Units and Groups Anonymity axioms state the independence of the segregation ordering on the *identity* of individuals or groups. Individual Replication Invariance simply states that replicating only one individual, while keeping as fixed the overall individual weights, preserves the segregation ranking.[6] The well known *Population Replication Invariance* axiom results from repeating a finite sequence of operations underlying individual replication invariance, thus introducing a form of anonymity with respect to the *size* of the population.[7]

Another intuitive axiom is the *Interaction Profile Merge.* It says that if two individuals at some point *share* their interaction profiles, thus generating a common *compound profile* according to their relative demographic weights, then exposure should increase and segregation decrease. This operation smooths the differences in interaction profiles across individuals, thus reducing the differences between the interaction profiles and the expected profile. This is the unique axiom that implies a (weak) dominance condition. It plays a similar role as the Pigou-Dalton transfer principle in inequality literature.

Axiom *Interaction Profiles Merge* *For $A, B \in \mathcal{A}(G)$ and $N(A) = N(B)$, if the interaction profiles of $i, j \in \mathcal{N}(B)$ are such that*

$$\boldsymbol{\pi}_i(B) = \boldsymbol{\pi}_j(B) = \alpha\, \boldsymbol{\pi}_i(A) + (1-\alpha)\, \boldsymbol{\pi}_j(A) \quad for \quad \alpha = \frac{\xi_i(A)}{\xi_i(A) + \xi_j(A)},$$

while $\boldsymbol{\pi}_h(B) = \boldsymbol{\pi}_h(A),\ \forall h \neq i, j$ and $\xi_h(B) = \xi_h(A)\ \forall h$, then $B \preccurlyeq A$.

[6]For instance, suppose that i in A represents a *couple* of individuals with a common exposure profile. Considering the couple i or two "equivalent" individuals that, when joined, have the same demographic weight, is irrelevant for segregation comparisons.

[7]Formally, the segregation measured in the exposure dimension is the same in A and B whenever $N(B) = \lambda N(A)$ and $\xi_i(B) = \frac{1}{\lambda}\xi_i(A)$ for $\lambda \in \mathbb{N}_{++}$.

Finally, we focus our attention on an exposure segregation ordering that satisfies exclusively these axioms.

Definition 2.2 (Exposure Segregation Ordering) *For any $A, B \in \mathcal{A}(G)$, $B \preccurlyeq A$ are ranked according to the exposure segregation ordering if and only if $\preccurlyeq$ satisfies axioms Non-triviality, Continuity, Normalization, Units and Groups Anonymity, Individual Replication Invariance and Interaction Profiles Merge.*

A direct implication of this definition is that if $B \preccurlyeq A$, then there exists a sequence of operations underlying the axioms considered that allows to transform the data associated to allocation A into the data associated to allocation B. In fact, the definition states that a pre-order is the exposure segregation ordering only if it satisfies the axioms provided above, which is equivalent to say that two distributions of interaction profiles can be compared according to this partial order if and only if there exists a sequence of operations that transform the data. These operations provide necessary and sufficient conditions for the orderings studied in the following section, based on the inclusion of geometric bodies or on segregation indicators. The advantage of these orderings is that they can be empirically verified, while it is no possible to retrieve from the data the sequence of operations described above. By adding more structure, one obtains a complete order of segregation that, at most, is coherent with the structure that we propose.

2.3 Characterization of new tools for measuring exposure

In this section, we proposes alternative characterizations of the exposure segregation ordering, inspired by the famous results by Marshall et al. (2011) on univariate inequality. We show that $B \preccurlyeq A$ is equivalent to the ranking produced by (i) the inclusion of the Segregation Zonotopes, a multi-group adaptation to the study of exposure of the two groups *segregation curve* (see Hutchens 1991), and (ii) a particular class of exposure indices.

2.3.1 The Segregation Zonotope

As shown in chapter 1, traditional segregation orderings rest on a dissimilarity comparison of the distribution of groups across organizational units. Bivariate comparisons in this setting resort on the ranking produced by the *segregation curves*, introduced by Duncan and Duncan (1955) and further studied by Hutchens (1991) and generalized by Carrington and Troske (1997) and Silber (1989).

There is, however, less agreement on which partial ordering can be used in the multi-group framework. Frankel and Volij (2011) proposed an ordering based on information theory (Blackwell 1953). Alonso-Villar and del Rio (2010) study a new concept of dominance based on *local segregation curves*, that are Lorenz curves for the distribution of group profiles (estimated with count data) for a specific group g across the entire population. The resulting multi-group partial order is an intersection of single group partial orders.

Much less has been done to characterize a partial order for the exposure dimension of segregation.[8] Exposure patterns are characterized by distribution matrices, measuring the inequality in the distribution across the population of vectors of individual-specific interaction probabilities. A similar type of comparisons have been largely exploited in the multidimensional inequality literature (Kolm 1977, Koshevoy 1995). We propose to combine the ranking criterion based on segregation curves or by the comparison of Zonotopes inclusion ordering in the dissimilarity setting, with the appropriate framework that is needed to study interaction segregation from the perspective of the individual. This is done by introducing the *Segregation Zonotope*, the multi-group counterpart of the segregation curve in the exposure segregation setting.

The *Segregation Zonotope* $SZ(A)$ is a G-dimensional convex geometric body, centrally symmetric around the point $(\frac{1}{2}, \ldots, \frac{1}{2})$. It consists in a representation of the distribution matrix $\mathbf{A}$.

Definition 2.3 *Let* $A \in \mathcal{A}(G)$ *with* $G \geq 2$ *to be represented by a distribution matrix*

[8] Reference works on the topic of exposure/interaction measurement are Reardon and Firebaugh (2002), Reardon and O'Sullivan (2004) and Echenique and Fryer (2007).

$\mathbf{A} \in \mathbb{R}_{+}^{G,N(A)}$. *The Segregation Zonotope $SZ(A)$ associated to allocation A is a centrally symmetric polytope in the G-dimensional space defined by the set:*

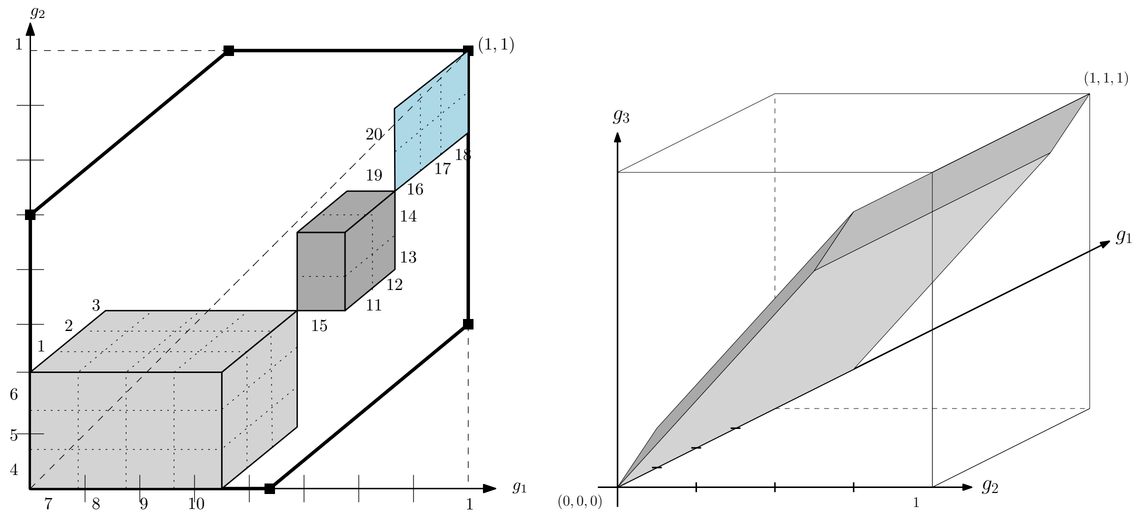
$$SZ(\mathbf{A}) := \left\{ \mathbf{z} := (z_1, \ldots, z_G) : \ \mathbf{z} = \sum_{i=1}^{N(A)} \theta_i \cdot (\xi_i(A)\, \mathbf{a}_i), \ \theta_i \in [0,1] \ \ \forall i \in \mathcal{N}(A) \right\},$$

where individual weights $\xi_i \in [0,1]$ are such that $\sum_i \xi_i = 1$.

The Segregation Zonotope depicts a graphical representation of the inequality in the distribution of the interaction profiles. This inequality can be alternatively represented by the degree of dissimilarity between the posterior interaction probabilities. Any Segregation Zonotope is included in the G-variate unit hypercube, thus making multi-group exposure comparisons well defined on a similar scale. Moreover, the hypercube coincides with the Segregation Zonotope associated to the lowest degree of exposure. This is the case whenever the interaction profiles assign probability one of interacting with one of the G groups, and coincide with maximal segregation.

On the contrary, if every individual in the population is associated with the same interaction profile, then exposure is maximized. In this case the Segregation Zonotope coincides with the diagonal of the unit hypercube, connecting the origin with the upper vertex. Examples of Segregation Zonotopes for the two and three groups case (with three possible types of interaction profiles) are reported in figure 2.1, panel (a) and (b) respectively.

We study the partial order generated by the inclusion of Segregation Zonotopes. Given two allocations $A, B \in \mathcal{A}(G)$, we interpret $SZ(B) \subseteq SZ(A)$ as indicating that the interaction profiles are more equally distributed in allocation B than they are in allocation A. The underlying idea is that the Segregation Zonotope associated to B is closer to the hypercube diagonal, i.e. the Segregation Zonotope associated with the highest degree of exposure, than it is the Segregation Zonotope of allocation A.



(a) Case g_1 and g_2

(b) Case g_1, g_2 and g_3

Figure 2.1: Segregation Zonotope and the Gini Exposure index. In the first graph, the population of 20 individuals is partitioned according to the group of origin: $\mathcal{N}_{g_1}(A) = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\mathcal{N}_{g_2}(A) = \{11, 12, 13, 14, 15\}$ and $\mathcal{N}_{g_3}(A) = \{16, 17, 18, 19, 20\}$. The share of overall segregation as experienced exclusively by the members of the three groups is given by the area of the three polytopes identified in panel (a).

2.3.2 A family of exposure segregation indicators

A *segregation index* is a function $I : \mathcal{A}(G) \to \mathbb{R}_+$ that assigns to each allocation a number that is interpreted as the allocation's segregation level. Any pair of allocations can be compared by virtue of a segregation index, thus defining a complete order. To characterize $B \preccurlyeq A$ we study the ranking produced at unanimity in a *class* of segregation measures. We study here the properties of this class.

First of all, exposure is measured by comparing the posterior interaction likelihoods associated to different groups, and has to be evaluated relatively to the overall social composition of the allocation. A function $\phi : [0, 1]^{G \times N} \to \mathbb{R}_+$ is a *relative* and *population weighted* multi-group exposure index if and only if $B \preccurlyeq A$ with $N(A) = N(B) = N$ implies $\phi(\mathbf{B} \cdot \boldsymbol{\xi}(B)) \leq \phi(\mathbf{A} \cdot \boldsymbol{\xi}(A))$. Hence, we restrict attention to the exposure indices $I(A)$ that can be represented by a function in the same class as ϕ.

Let $\mathcal{P}_N$ the set of $N \times N$ *permutation* matrices and $\mathcal{R}_N$ the set of all $N \times N$ *row stochastic* matrices with at most a nonzero (a one) element by row, and that are not permutation matrices (hence $\mathcal{P}_N \cap \mathcal{R}_N = \emptyset$).

The function ϕ is called *column symmetric* if an only if for any $\mathbf{X} \in \mathcal{P}_N$, $\mathbf{B} \cdot \boldsymbol{\xi}(B) = \mathbf{A} \cdot \boldsymbol{\xi}(A) \cdot \mathbf{X}$ implies $\phi(\mathbf{B} \cdot \boldsymbol{\xi}(B)) = \phi(\mathbf{A} \cdot \boldsymbol{\xi}(A))$. This property says that exposure indicators should be independent on individuals labeling.[9]

The function ϕ is called *additive respondent* if and only if for any $\mathbf{X} \in \mathcal{R}_N$, $\mathbf{B} \cdot \boldsymbol{\xi}(B) = \mathbf{A} \cdot \boldsymbol{\xi}(A) \cdot \mathbf{X}$ implies $\phi(\mathbf{B} \cdot \boldsymbol{\xi}(B)) \leq \phi(\mathbf{A} \cdot \boldsymbol{\xi}(A))$. This property generalizes subadditivity, that can be considered as paying the same role of S-convexity[10] in the study of (multidimensional) inequality. However, the two properties are not related, thus showing that the model used to capture segregation phenomena is not grounded on traditional inequality comparisons.

Finally, ϕ is called *quasiconvex* (Mosler 1994) if and only if, for every sequence $\{(\mathbf{A} \cdot \boldsymbol{\xi}(A))_\ell\}_{\ell=1}^m$ and $(\lambda_1, \ldots, \lambda_m) \in [0,1]^m$ with $\sum_{\ell=1}^m \lambda_\ell = 1$, $\phi\left(\sum_{\ell=1}^m \lambda_\ell (\mathbf{A} \cdot \boldsymbol{\xi}(A))_\ell\right) \leq \max_\ell \phi\left((\mathbf{A} \cdot \boldsymbol{\xi}(A))_\ell\right)$. This is a mild condition of coherence, stating that the overall amount of exposure segregation of a convex combination of a set of distribution matrices must reduce the degree of exposure segregation compared to the more segregated among these allocations.[11]

When all indicators within the family of segregation indices characterized as ϕ register less segregation in the exposure dimension in one allocation with respect to another, one gets a very robust assessment over changes in segregation. In the following, we provide the axiomatic characterization of this ordering.

2.3.3 Main result

Combining together merge and indifference axioms, one gets a rationalization of the Segregation Zonotope inclusion order. Moreover, any partial order results from the intersection

[9] A similar property, *row symmetry* can be constructed with regard to permutations of the rows. However, as stated before, we would like to keep the argument of groups permutability separated in the analysis, and in fact we do not need to use it to obtain our main result.

[10] The function ϕ is *S-convex* if and only if $\mathbf{B} \cdot \boldsymbol{\xi}(B) = \mathbf{A} \cdot \boldsymbol{\xi}(A) \cdot \mathbf{Y}$ for $N(A) = N(B)$ and $\mathbf{Y} \in \mathbb{R}^N$ bistochastic implies $\phi(\mathbf{B} \cdot \boldsymbol{\xi}(B)) \leq \phi(\mathbf{A} \cdot \boldsymbol{\xi}(A))$ (Marshall et al. 2011).

[11] See Mosler (1994) for an interpretation of this axiom in the multidimensional inequality setting.

of complete orders (Donaldson and Weymark 1998), hence we provide a characterization of a class of segregation measures studied above, that are consistent with the Segregation Zonotope inclusion.[12]

Theorem 2.1 *Let* $A, B \in \mathcal{A}(G)$ *with* $N(A) = N(B) = N$*, then* $B \preccurlyeq A$ *is equivalent to each of the following conditions:*

i) $\mathbf{B} \cdot \boldsymbol{\xi}(B) \in conv\{\mathbf{A} \cdot \boldsymbol{\xi}(A) \cdot \mathbf{X} : \mathbf{X} \in \mathcal{P}_N \cup \mathcal{R}_N\}$*;*

ii) $SZ(B) \subseteq SZ(A)$*;*

iii) $\phi(\mathbf{B} \cdot \boldsymbol{\xi}(B)) \leq \phi(\mathbf{A} \cdot \boldsymbol{\xi}(A))$ *for all* ϕ *which are column symmetric, additive respondent and quasiconvex.*

Proof. See appendix 2.B.1. ■

The Theorem can be easily extended to the case where *any* pair of allocations $A, B \in \mathcal{A}(G)$ are compared, with possibly different demographic sizes. This can be done by exploiting the Individual (or Population) Replication Invariance axiom. In fact, there exists $A', B' \in \mathcal{A}(G)$ such that $N(A') = N(B')$, with $A \sim A'$ and $B \sim B'$. Continuity of the exposure ordering gives $B \preccurlyeq A$ if and only if $B' \preccurlyeq A'$, which fulfills the requirements of the theorem.

When Segregation Zonotopes inclusion cannot be verified, it is possible to introduce more structure on the exposure measure, in order to characterize one particular family, or an index, in the class defined by ϕ. In the following we study a multi-group extension of the Gini index called the Gini Exposure index, and we show that it is consistent with the class of functionals ϕ characterized above.

[12] Before proceeding, it is worth noting that the Groups Anonymity axiom states that there exists at least one permutation of the groups for which the equivalence in the following theorem holds, thus enlarging the class of comparable allocations. Nevertheless, for the sake of the validity of our results, not much is missed by overlooking this particular aspects.

2.4 The Gini Exposure index

2.4.1 The index

The *Gini inequality* index of a univariate income distribution, represented by the $N \times 1$ vector $\mathbf{x}$, is defined as the average distance between any pair of incomes in $\mathbf{x}$, scaled by the overall income in the distribution. In symbols:

$$G(\mathbf{x}) \;:=\; \frac{1}{2N^2\left(\sum_i x_i/N\right)} \sum_{i=1}^{N}\sum_{j=1}^{N} |x_i - x_j|.$$

Alternatively, the Gini index can be related to the Lorenz curve: it is equal to twice the area between the Lorenz curve and the diagonal, representing the equal distribution. That is, the area of the Lorenz Zonotope $LZ(\mathbf{x})$.[13]

The Gini index can be decomposed into a sum of areas, capturing the dispersion between each possible pair of observations (a property studied in Shephard 1974). Each of these areas is measured by the determinant of a 2×2 matrix, thus giving the following representation of the Gini index:[14]

$$G(\mathbf{x}) \;:=\; \frac{1}{2} \sum_{\forall \{i,j\} \subseteq \{1,\ldots,N\}} \left| \det \begin{pmatrix} x_i/\left(\sum_i x_i\right) & x_j/\left(\sum_i x_i\right) \\ 1/N & 1/N \end{pmatrix} \right|.$$

This convenient formulation of the Gini index can be readily extended to evaluate the distance between *any* pair of probability distributions defined over a discrete, non ordered support, and represented by the $N \times 1$ vectors $\mathbf{p}$ and $\mathbf{q}$.[15]

---

[13] The Lorenz Zonotope of distribution is defined as the area between the Lorenz curve and its *dual*. It can be written as a Minkowski sum of line segments, hence its area equals the sum of the areas spanned by each pair of bi-dimensional vectors $\left(\frac{x_i}{\sum x_i}, \frac{1}{N}\right)$ and $\left(\frac{x_j}{\sum x_i}, \frac{1}{N}\right)$, for all i, j. This area coincides with a parallelogram and it corresponds to a measure of inequality between incomes shares received by two individuals i, j equally weighted $\frac{1}{N}$ in the population.

[14] The terms $\frac{1}{N^2 \sum_i x_i/N}$ disappears at it is incorporated in the determinant calculation. The comparison is now expressed in relative, rather than absolute, incomes. Moreover, the determinant is a measure of linear dependence, and therefore similarity, between oriented vectors.

[15] For instance, the inequality Gini index is obtained by replacing the vector $\mathbf{p}$ by the vector of income shares $(x_1/\sum x_i, \ldots, x_N/\sum x_i)$ and $\mathbf{q}$ is replaced by the individual weighting scheme $(1/N, \ldots, 1/N)$.

$$G((\mathbf{p},\mathbf{q})) \;:=\; \frac{1}{2}\sum_{\forall\{i,j\}\subseteq\{1,\ldots,N\}}\left|\det\begin{pmatrix} p_i & p_j \\ q_i & q_j \end{pmatrix}\right|. \tag{2.1}$$

The Gini inequality index measure the dissimilarity between two particular distribution vectors: the income shares versus the population weighting scheme.

We use the convenient formulation in (2.1) to derive the *Gini Exposure* index of segregation, $G_E : \mathcal{A}(G) \to [0,1]$. In this case, the vector of *normalized* probability interactions given by the column $\mathbf{a}_i$ of the distribution matrix $\mathbf{A}$, takes the place of bivariate vector of probability masses $(p_i,\, q_i)^t$. Every normalized interaction vector has G components (one probability for each group), and therefore the determinant can be constructed by comparing G-tuples of individuals, instead of pairs as in the Gini index.[16]

The degree of dissimilarity between the G posterior interaction distributions is measured, as in the previous case, by the *determinant* of a $G \times G$ matrix. The correction term $(1/2)$ has now to account for all the $G!$ possible permutations of the groups that leave the dissimilarity unchanged (because the determinant does not change).

Definition 2.4 (The Gini Exposure index) *The index is defined as follows:*

$$G_E(A) := \frac{1}{G!}\sum_{\forall\{i_1,\ldots,i_G\}\subseteq\mathcal{N}(A)}\left|\det\begin{pmatrix} \xi_{i_1}(A)\cdot\mathbf{a}_{i_1} & \ldots & \xi_{i_G}(A)\,\mathbf{a}_{i_G} \end{pmatrix}\right|.$$

The Gini Exposure index generates a complete order of the allocations in $\mathcal{A}(G)$, coherent with the notion of exposure embedded in the exposure segregation ordering. This is demonstrated in the following proposition.

Proposition 2.1 *The Gini Exposure index $G_E(A) = \phi\left(\mathbf{A}\cdot\boldsymbol{\xi}(A)\right)$, where ϕ is column symmetric, additive respondent and quasiconvex.*

Proof. See appendix 2.B.2. ■

[16] At each comparison G degrees of freedom are needed, since G for the dimension of the distribution vector. The classical Gini inequality index involves the comparison of two distributions at a time, and for this reason the summation is constructed over all possible pairs of individuals of the population.

This new multi-group measure for the exposure dimension of segregation stands in sharp contrast with the analogous of the *expected Gini EG* analyzed in Flückiger and Silber (1999) and Alonso-Villar and del Rio (2010), when applied to the exposure dimension. The EG index is an average of local Gini indices G_g, weighted by groups size:

$$EG(A) \ := \ \sum_{g \in G} s_g(A)\, G_g(A).$$

Each local Gini index is meant to capture the inequality in the distribution of any single interaction probability across the population.

$$G_g(A) \ := \ \frac{1}{2} \sum_{\forall \{i_1, i_2\} \subseteq \mathcal{N}(A)} \left| \det \begin{pmatrix} \frac{\xi_{i_1}(A)\, \pi_{gi_1}(A)}{\mathbb{E}[\pi_g(A)]} & \frac{\xi_{i_2}(A)\, \pi_{gi_2}(A)}{\mathbb{E}[\pi_g(A)]} \\ \xi_{i_1}(A) & \xi_{i_2}(A) \end{pmatrix} \right| .$$

The expected Gini index is based only on a partial comparison of inequality across individuals, leaving aside any concern about the composition of the interaction profiles.

2.4.2 Decomposition properties and discussion

In many instances, one would like to assess the degree of exposure as experienced only by some subgroups of the population. This can be done by computing the share of the overall segregation that can be attributed to each group $g \in \mathcal{G}$ in the population, through a suitable decomposition of the overall Gini Exposure index.

Similarly to the traditional Gini inequality index, the Gini Exposure index can be decomposed into a weighted average of the degree of segregation experienced by each subgroup, captured by a group specific Gini index, and an overlapping term. This linear decomposition allows to study separately the dynamics of exposure for each of the G groups.

We consider a repartition in groups given by the subset $\mathcal{N}_g(A)$, for all $g \in \mathcal{G}$. For each group g, the Gini Exposure index $G_E(A|g)$ measures the overall degree of segregation (inequality in the interaction profiles distribution) as experienced only by members of group

g. It is defined as:

$$G_E(A|g) := \frac{1}{G!} \sum_{\forall\{i_1,\ldots,i_G\}\subseteq\mathcal{N}_g(A)} \left| \det \begin{pmatrix} \widetilde{\xi}_{i_1|g}(A)\,\frac{\pi_{1i_1}(A)}{\mathbb{E}[\pi_1(A)|g]} & \ldots & \widetilde{\xi}_{i_G|g}(A)\,\frac{\pi_{1i_G}(A)}{\mathbb{E}[\pi_1(A)|g]} \\ \vdots & & \vdots \\ \widetilde{\xi}_{i_1|g}(A)\,\frac{\pi_{Gi_1}(A)}{\mathbb{E}[\pi_G(A)|g]} & \ldots & \widetilde{\xi}_{i_G|g}(A)\,\frac{\pi_{Gi_G}(A)}{\mathbb{E}[\pi_G(A)|g]} \end{pmatrix} \right|,$$

where $\widetilde{\xi}_{i|g}(A) = \frac{\xi_i(A)}{\sum_{i\in\mathcal{N}_g(A)}\xi_i(A)}$ is the relative weight of individual i in her group g, and $\mathbb{E}[\pi_m(A)|g] = \sum_{i\in\mathcal{N}_g(A)} \widetilde{\xi}_{i|g}(A)\pi_{mi}(A)$ is the expected probability of interaction with group m experienced by individuals in group g. Notice that the multi-group $G_E(A|g)$ index is logically different from the single group index $G_g(A)$ defined above.

Let define the overlapping set $\mathcal{O}$ as composed by all the possible G-tuple of individuals with at least two individuals coming from different groups. It is given by:

$$\mathcal{O} \;:=\; \{\{i_1,\ldots,i_G\} \subseteq \mathcal{N}(A) : \; \nexists\{i_1,\ldots,i_G\} \subseteq \mathcal{N}_g(A) \text{ for any } \; g\in\mathcal{G}\}\,.$$

We are now able to show an additive decomposition of the multi-group Gini Exposure index in a within groups and an overlapping component.

Proposition 2.2 *The Exposure Gini index can be decomposed as follows:*

$$G_E(A) \;=\; \left(\sum_{g\in\mathcal{G}}\alpha_g\right)\sum_{g\in\mathcal{G}}\beta_g\; G_E(A|g) \;+\; G_E(A|\mathcal{O}),$$

where $\beta_g = \frac{\alpha_g}{\sum_{g\in\mathcal{G}}\alpha_g}$ *and* $\alpha_g = (\sum_{i\in\mathcal{N}_g(A)}\xi_i(A))^G \prod_{m\in\mathcal{G}} \frac{\mathbb{E}[\pi_m(A)|g]}{\mathbb{E}[\pi_m(A)]}$.

Proof. See appendix 2.B.3. ■

If there are no systematic differences between groups in the expected interaction profiles (although there exists within group variability), then $\mathbb{E}[\pi_m(A)|g] = \mathbb{E}[\pi_m(A)]$ for all m is expected to hold across all groups $g \in \mathcal{G}$. It follows that the weighting scheme β_g would definitely depend only on groups densities. If, moreover, one compares allocations with little or no variability in groups compositions, the unique sources of variation for the Gini index are given either by the variations in the conditional segregation captured by $G_E(A|g)$ or by changes in the degree of overlapping. Otherwise, differences in the structural composition of the groups populations may play a relevant role in determining the overall degree of segregation.

2.4.3 An example

Consider an allocation A with a population of 20 individuals, partitioned in three non-overlapping groups $\mathcal{G} = \{g_1, g_2, g_3\}$. Out of the 20 individuals, ten belong to group g_1, five belong to group g_2 and the remaining five are of group g_3, such that $\mathcal{N}_{g_1}(A) = \{1, \ldots, 10\}$, $\mathcal{N}_{g_2}(A) = \{11, \ldots, 15\}$ and $\mathcal{N}_{g_3}(A) = \{16, \ldots, 20\}$.[17]

We consider two different frameworks. In the first, we analyze a bivariate model of segregation, based on the interaction profiles with groups g_1 and g_2, as experienced by the whole population. In Figure 2.1 we draw the related Segregation Zonotope, we identify the Gini Exposure index and we depict a graphical representation of the index decomposition. The second example extends the analysis to the multi-groups (three groups) case.

Consider a simplified setting where there are only three possible interaction profiles with groups g_1 and g_2, namely $\boldsymbol{\pi}(A)$, $\boldsymbol{\pi}'(A)$ and $\boldsymbol{\pi}''(A)$, defined as follows:

$$\boldsymbol{\pi}(A) = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}; \quad \boldsymbol{\pi}'(A) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \boldsymbol{\pi}''(A) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The third interaction profile, $\boldsymbol{\pi}''(A)$, is such that all the individuals allocated with that profile do not have any chance to interact with group g_2. The other profiles can be interpreted in a similar way.

In the following table we summarize the distribution of individuals across groups *and* interaction profiles. For instance, individual $i = 3$ in group g_1 is allocated with interaction profile $\boldsymbol{\pi}(A)$, while individual $i = 15$ is the unique individual in group g_2 allocated with

[17] The elements of the three sets represent individuals of the population.

profile $\boldsymbol{\pi}''(A)$.

	$\boldsymbol{\pi}(A)$	$\boldsymbol{\pi}'(A)$	$\boldsymbol{\pi}''(A)$
g_1	$\{1,2,3\}$	$\{4,5,6\}$	$\{7,8,9,10\}$
g_2	$\{11,12\}$	$\{13,14\}$	$\{15\}$
g_3	$\{16,17,18\}$	$\{19,20\}$	$\emptyset$

Assuming a uniform weighting scheme ($\xi_i(A) = \frac{1}{20}$), one can easily construct the expected interaction profiles with the two groups: $\mathbb{E}[\pi_{g_1}(A)] = \frac{8}{20}1 + \frac{7}{20}0 + \frac{5}{20}0.5 = 0.45$ and $\mathbb{E}[\pi_{g_2}(A)] = 0.55$. The distribution matrix obtained from these data defines the underlying information that is necessary to construct the Segregation Zonotope in Figure 2.1(a) (the Segregation Zonotope is defined by the solid contour of the figure). The Gini Exposure index corresponds to its area, and it amounts to $G_E(A) = 0.8383$.

Within this example, it is possible to provide a graphical representation of the decomposition of the Gini Exposure index into a weighted sum of Gini Exposure indices, each measuring the overall inequality in interaction profiles between individuals of the same group. In Figure 2.1(a) we represent three distinct areas (denoted by different scales of grey), that correspond to the contributions of each group to the overall exposure inequality. Each area is decomposed according to the distribution of individuals across the three groups. The values of the group-specific Gini Exposure indices are:

$$G_E(A|g_1) = 0.909, \quad G_E(A|g_2) = 0.8333, \quad G_E(A|g_3) = 0.5714, \quad G_E(A|\mathcal{O}) = 0.53,$$

and the associated weighting scheme is:

$$\alpha_{g_1} = 0.25, \quad \alpha_{g_2} = 0.06, \quad \alpha_{g_3} = 0.053.$$

Using a similar analysis, one can evaluate the multi-group exposure patterns for cases where the interaction profiles are defined on more than two dimensions. We consider the case where interaction takes place with respect to the three groups considered, thus redefining the new interaction patterns $\boldsymbol{\pi}(A)$, $\boldsymbol{\pi}'(A)$ and $\boldsymbol{\pi}''(A)$ as follows:

$$\boldsymbol{\pi}(A) = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix}; \quad \boldsymbol{\pi}'(A) = \begin{pmatrix} 0 \\ 0.7 \\ 0.3 \end{pmatrix}; \quad \boldsymbol{\pi}''(A) = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.5 \end{pmatrix}.$$

The distribution of these interaction profiles across the population is fixed as before. In Figure 2.1(b) it is reported the Segregation Zonotope associated to this distribution of interaction profiles (shaded in gray). The overall Gini Exposure index coincides with the zonotope *volume*, and it is equal to $G_E(A) = 0.1138$. It is possible to replicate the previous exercise to obtain the decomposition of the Gini Exposure index and identify (also graphically) the segregation patterns of the three groups under analysis.

2.5 Comparison with other indices within the spatial interaction model

We study now the issue of spatial segregation from the interaction perspective. Within this framework, the Gini Exposure index can be compared with alternative measures of exposure segregation.

In many applications, data are only available for (i) the demographic size of the groups, (ii) the distribution of groups across a well defined partition of the interaction space into organizational units and (iii) a measure of the proximity between the units, decreasing with their distance or diversity. Within this framework, we calculate the interaction profiles in the context of a spatial model for interaction.

2.5.1 Additional notation for the spatial model

An allocation A defines the distribution of individuals within a territory (for instance, a city). The true interaction profiles are not observable. Consider the case in which the interaction space is partitioned into a set of N_A non-overlapping *organizational units* $i = 1, \ldots, N_A$ and use $\mathcal{N}(A)$ to indicate this set. In the rest of the analysis, organizational units will take the place of the individuals in the original formulation of the segregation index. This simplification rests on the idea that all individuals living in the same section also face the same interaction profile.

Let $n_{gi}(A)$ be the *observed* number of individuals living in the same organizational unit i who are of group g. Each organizational unit is assumed to have a demographic weight $\widehat{\xi}_i(A) = \frac{n_{gi}(A)}{\sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{N}(A)} n_{gi}(A)}$ where the hat symbol is used to denote a weighting scheme estimated under the assumption that all individuals living in the same organizational unit face the same interaction profile.

The second ingredient in our analysis is a measure of *distance*, d, between organizational units. In most of the analysis of residential segregation, the distance measure stands for spatial distance between organizational units centroids, or alternatively a measure of adjacency/negihboring of units. Distance is used to construct spatial measures of interaction. In sociological literature (see for instance White 1983, Reardon and O'Sullivan 2004, Echenique and Fryer 2007), it is assumed that the spatial distance accounts for social distance between individuals, so that the likelihood that two individuals interact is a decreasing function of their spatial distance.

Let introduce $\delta_A(i, h; d)$, a measure of *proximity* of two units $i, h \in \mathcal{N}_A$ which is inversely related to the distance function $d(i, h)$. This function satisfies the regularity properties of a distance function (see for instance Shorrocks 1982). We impose the proximity to be maximal and equal to 1 when i and h coincide according to the distance criterion d (so that $\delta_A(i, i; d) = 1$), while the measure decreases and approaches the value 0 the higher is the distance between the two units.[18]

[18] In general, one can interpret the proximity measure as a tool to identify the degree of *diversity* between any two units. The location in a geographic/social space of an organizational unit represent a possible factor of diversity, which is capture by the distance function. Units are, in general, diverse in a variety of socioeconomic relevant aspects. In residential segregation, any two urban units may differ in the degree of connectedness granted by public transports, on the real estate market, on local quality and supply of public goods. All these factor may affect, which different weights, the distance between the two sections,, which has a negative impact on the measured degree of proximity. Similar extensions can be adapted to study segregation in the labor market or across schools.

We can generalize the classical *spatial* measures of segregation by looking at d as a diversity measure that depends on a L-variate vector of attributes that characterize each unit. For each unit the data can be represented by vectors $\mathbf{x}_k,\ \mathbf{x}_h \in \mathbb{R}^L_+$ where each entry $x_{\ell,k}$ and $x_{\ell,h}$ is a standardized measure of attribute ℓ of sections k and h. A possible candidate for the proximity measure is $\delta_A(k, h; d) = \frac{1}{\max\{1, d(h,k)\}}$ where d is

The proximity-weighted counting indicator $\widehat{n}_{gi}(A) = \sum_{i\in\mathcal{N}_A} n_{gi}(A)\delta_A(i,h;d)$ measures the number of individuals of groups g with whom an individual in unit i may interact with. The overall interaction potential can be measured by the total amount of individuals that can be associated to unit i, $\sum_{g\in\mathcal{G}} \widehat{n}_{gi}(A)$. Combining together these two indicators, one obtains an empirical measure of the probability to interact with group g conditional on the fact that interaction takes place in organizational unit i: $\widehat{\pi}_{gi}(A) = \frac{\widehat{n}_{gi}(A)}{\sum_{g\in\mathcal{G}} \widehat{n}_{gi}(A)}$. An interaction profile is a vector with G entries $\widehat{\pi}_{gi}(A)$, constructed under the assumption that interaction probabilities decay with spatial distance. The spatial model may be replaced by more sophisticated models that account for the impact of other factors in determining the interaction patterns across individuals or organizational units, provided that the relevant information is available. The expected interaction profile is denoted $\mathbb{E}[\widehat{\pi}_g(A)]$ and the interaction matrix $\widehat{\mathbf{A}}$ with entry g,i is equal to $\widehat{a}_{gi} = \frac{\widehat{\pi}_{gi}(A)}{\mathbb{E}[\widehat{\pi}_g(A)]}$.

It is now possible to study the relation between the Exposure Gini index and other segregation indices, using the fact that organizational units are now the reference individuals.

2.5.2 Comparison with other indices: the spatial connection

Reardon and O'Sullivan (2004) systematically analyzed the spatial indices of segregation treated in the literature, and proposed some meaningful properties that these indices should satisfy. The model for data representation used by Reardon and O'Sullivan draws on the implementable spatial model proposed in the previous section. Therefore, many of the spatial indices that they describe can be easily replicated in our framework.

---

the Minkowski distance $M(\mathbf{x}_k, \mathbf{x}_h, \alpha)$, a generalized score that compute the distance among points $\mathbf{x}_k$ and $\mathbf{x}_h$ in the L-dimensional real space, such that:

$$M(\mathbf{x}_k, \mathbf{x}_h, \alpha) := \left[\sum_{\ell=1}^{L} |x_{\ell,k} - y_{\ell,h}|^{\alpha}\right]^{\frac{1}{\alpha}}, \quad \alpha > 0.$$

The Manhattam Distance and the Euclidean Distance correspond to the cases where $\alpha = 1$ and $\alpha = 2$. If $\alpha > 1$ the distance measure is convex. Note that spatial distance is a particular form of the Euclidian distance $M(\mathbf{x}_k, \mathbf{x}_h, 2)$, where $\mathbf{x}_k, \mathbf{x}_h \in \mathbb{R}^2_+$ entries are the latitude and longitude of the centroid of the organizational unit k and h, respectively.

Within the spatial interaction model, the Gini Exposure index can be compared to existing spatial segregation measures of exposure according to the principles listed in Reardon and O'Sullivan (2004).

The first property, *scale interpretability*, is satisfied by construction of the Gini Exposure index. We interpret $G_E(A) = 0$ as the case where exposure is equalized across units, while $G_E(A) = 1$ as the opposite case of perfect segregation, occurring only in the case where the interaction profile allocated to each of the units is degenerate, that is it assigns a probability of interaction with group g equal to one, and zero for the remaining groups.

The implementable model is not exempted from the MAUP problem,[19] and therefore the *arbitrary boundary independence* property is not satisfied. This is a drawback of the identification model that we use, based on a pre-determined partition of the space into organizational units, rather than an issue related to the index itself.

The implementable Gini Exposure index meets the requirements of *location equivalence*. In fact, if two organizational units are associated to the same interaction pattern, the operations of mixing the two together into a new unit preserves the segregation order characterized by the merge axiom, and hence the Gini Exposure index.[20]

Population density invariance is clearly satisfied. On the contrary, *composition invariance* is not satisfied. In fact, the Gini Exposure index captures a form of *relative* inequality in the distribution of interaction patterns across the population, therefore it is independent from the overall expected interaction profiles. However, the Gini Exposure index is not independent from the variations in the *size* of the groups. The convenience of satisfying composition invariance is, nevertheless, debatable (see Frankel and Volij 2011). We consider in our empirical comparison the *Atkinson* multi-group segregation index in Frankel and Volij (2011) that, differently from the other exposure indicators, is composition invariant.

Finally, it is impossible to establish if *transfer* and *exchange* principles are satisfied by the Gini Exposure index. In fact, the merge of interaction profiles is not defined in the form of a movement of population masses across organizational units (transfer) or groups

---

[19] The modifiable areal unit problem, occurring when the partition of the space into organizational units is exogenously fixed.

[20] This is so because all the agents living in the two regions are endowed with the same interaction profiles.

(exchange) but rather as a convex combination of interaction profiles. We compute empirical correlations between indices satisfying the transfer/exchange principle and the Gini Exposure index, to recover a relation between merges, transfers and exchanges.

We compare the multi-group Gini Exposure index with other multi-group measures proposed in the literature. In the class of indices that do not satisfy composition invariance, the first index that we consider is a *spatial* version of the the *Mutual Information* index $M(A)$ characterized (among others) by Frankel and Volij (2011).

The entropy of the discrete probability distribution $(p_1, \ldots, p_G)$ is defined by:

$$E(p_1, \ldots, p_G) \;=\; \sum_{g \in \mathcal{G}} p_g \log_2 \left(\frac{1}{p_g} \right).$$

The Mutual Information index equals the entropy of an allocation's groups distribution minus the average entropy of the groups across its organizational units:

$$M(A) \;=\; E\left(\mathbb{E}[\widehat{\pi}_1(A)], \ldots, \mathbb{E}[\widehat{\pi}_G(A)]\right) \;-\; \sum_{i \in \mathcal{N}(A)} \widehat{\xi}_i(A)\, E\left(\widehat{\pi}_{1i}(A), \ldots, \widehat{\pi}_{Gi}(A)\right).$$

Alternatively, we also consider other spatial indices studied in Reardon and O'Sullivan (2004). All these indices satisfy the transfer and exchange principles, provided that some symmetry requirements are imposed on the proximity measure. The first index is the *Spatial Relative Diversity* index $R(A)$, which is a measure of how much less diverse individuals' local environments are, on average, than is the total population in the allocation as a whole. This can be done by comparing the *interaction* coefficient $I_i(A) := \sum_{g \in \mathcal{G}(A)} \widehat{\pi}_{gi}(A)\,(1 - \widehat{\pi}_{gi}(A))$ for each organizational unit i and for the population as a whole, denoted by the coefficient $I(A) := \sum_{g \in \mathcal{G}} \mathbb{E}[\widehat{\pi}_g(A)]\,(1 - \mathbb{E}[\widehat{\pi}_g(A)])$. The relative diversity amounts to:

$$R(A) \;=\; 1 \;-\; \sum_{i \in \mathcal{N}(A)} \widehat{\xi}_i(A) \frac{I_i(A)}{I(A)}.$$

The *spatial dissimilarity* index $D(A)$ is a measure of how different the composition of individuals' organizational units environments are, on average, from the composition of the population as a whole. It is defined as follows:

$$D(A) \;=\; \frac{1}{2\,I(A)} \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{N}(A)} \widehat{\xi}_i(A)\,|\widehat{\pi}_{gi}(A) - \mathbb{E}[\widehat{\pi}_g(A)]|.$$

The last two indices that we consider are the empirical counterparts of the expected Gini index, denoted $EG(A)$, and the *normalized spatial exposure* index $NE(A)$, which is defined as:

$$NE(A) \;=\; \sum_{i \in \mathcal{N}(A)} \widehat{\xi}_i(A) \sum_{g \in \mathcal{G}} \frac{(\widehat{\pi}_{gi}(A) - \mathbb{E}[\widehat{\pi}_g(A)])^2}{1 - \mathbb{E}[\widehat{\pi}_g(A)]}.$$

This index belongs to the class of variance indicators. In the Reardon and Firebaugh (2002) taxonomy, the two indices fall into the class of the indicators measuring segregation as a form of distributional inequality of the interaction profiles.

2.6 An empirical application to the Italian case

In this section we study the empirical performances of the spatial segregation indices discussed above. We exploit a panel of nearly 8400 Italian municipalities, observed in the period 2003/2010. The municipalities are clustered at province level (110 provinces in 2010, of which 101 remain fixed over time, each made by 74 municipalities on average), covering on average a population of 551,000 inhabitants. Each municipality has an average demographic size of 6,400 individuals, comparable to the dimension of the US MSAs districts. In the analysis, each municipality corresponds to an organizational unit, with $\mathcal{N}_{pt}$ the set of municipalities that belongs to a given province p in time t. We exploit the patterns of segregation of immigrants and natives (for a total of $G = 3$ groups) for each province p in each year t. This can be done by calculating a segregation index for each pair p, t. In this way, we have sufficient time and space variability to construct and analyze segregation patters in Italy, while keeping a sufficiently refined spatial scale.

We propose to study the degree of segregation between three mutually exclusive social groups: Italian natives, immigrants from countries with *high* HDI levels, and immigrants from countries with *low* HDI levels.[21] This multi-group separation (compared to the traditional bivariate analysis of immigrants versus natives) is of particular relevance in Italy, since immigration is a recent and growing phenomenon, and the *type* of the country of origin (as captured by the HDI) is a relevant factor to account for.

Our analysis aims at verifying the degree of consistency across different segregation indicators in ranking Italian provinces according to the within-province degree of exposure segregation, measured at the municipality level. The distribution of the Italian provinces is represented in figure 2.2. For each province we construct a spatial model to measure interaction profiles at the municipality level. Then, we compute the values of the segregation indices G_E, M, NE, R, D, EG and A for each of the provinces, using municipalities as organizational units. These indicators are meant to summarize the information about the distribution of interaction probabilities *within provinces.* We obtain 808 data point for each of the indicators, varying across the 101 provinces and the 8 years considered.

We study the empirical rank correlations of the indices, and we asses the differences in the type of segregation patterns that can be captured according to the indicator used. Then, we apply the decomposition of the Gini Exposure index to the data to study the contribution of each group to the overall exposure.

2.6.1 Data

We build the spatial analysis using ISTAT demographic data[22] at municipality level. We obtain data on the demographic size of the resident population, partitioned according to the nationality. Municipalities are grouped into provinces, according to the official repartition of the Italian territory. Table 2.3 in the appendix collects information on the number, the

[21] The *Human Development Indicator* (HDI) proposed by the UNDP department is a synthetic indicator computed on a country and year bases for evaluating the multivariate distribution of health, resources and educational indicators across the population in that year and country. The UNDP also provides a classification of countries according the their HDI profile.

[22] The municipality level composition (by nationality) of the resident population in Italy from 2003 to 2010 can be freely downloaded from the official ISTAT (the Italian Statistical Institute) webpage at the following link: `http://demo.istat.it/`.

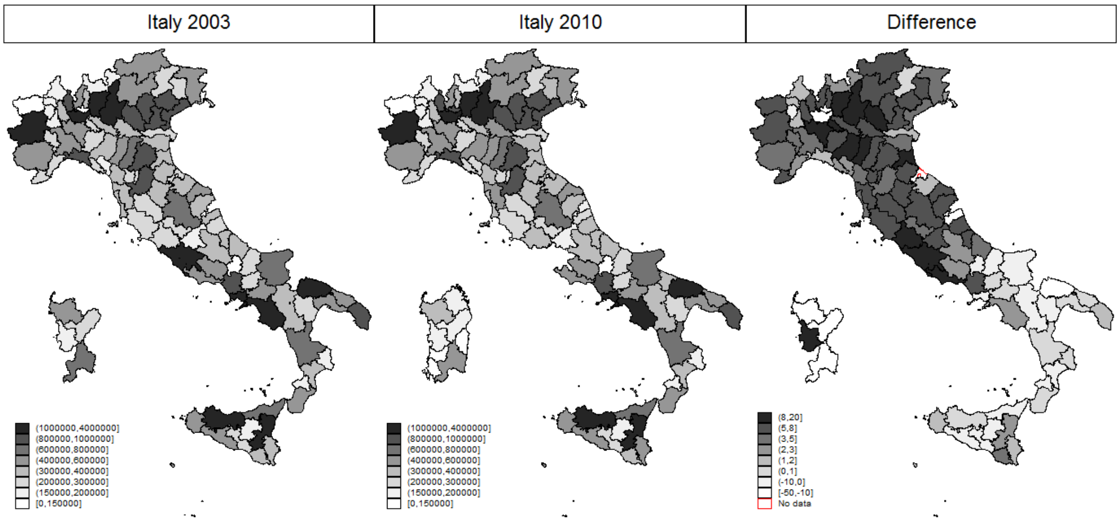


Figure 2.2: Population by province and year, and its growth rate (in %).

demographic size and the characteristics of provinces and municipalities across the period considered.

Two immigration groups have been created according to the definition of low level and medium/high level HDI countries provided by the UNDP for 2011. In 2010, the share of low HDI type immigrants amounts to 6.7% of total population by provinces, on average (table 2.3), and it is particularly relevant in the north of Italy (figure 2.2).

We use a spatial proximity index to identify the interaction probabilities $\widehat{\pi}_{gi}$. We proceed as follows. Each municipality has been geocoded, so that latitude and longitude are now available for each municipality's *centroid*. We assume that the interaction probability decays with spatial distance. We construct, for each pair p, t separately, a set of interaction profiles associated to each municipality $i \in \mathcal{N}_{g,t}$. To do so, we compute $\widehat{n}_{gi}$ for every municipality i, assuming $\delta(i, h; d)$ to be a *biweight kernel* estimator of proximity, and we take d to measure spatial distance between municipalities, censored at 20km threshold.[23] Interaction probabilities and *expected* (by province) probabilities are calculated according

[23]The biweight kernel kernel has a Gaussian-like shape, although it is bounded, so that all the municipalities outside a given radius of length $r = 20km$ are assumed to have no weight in determining interaction

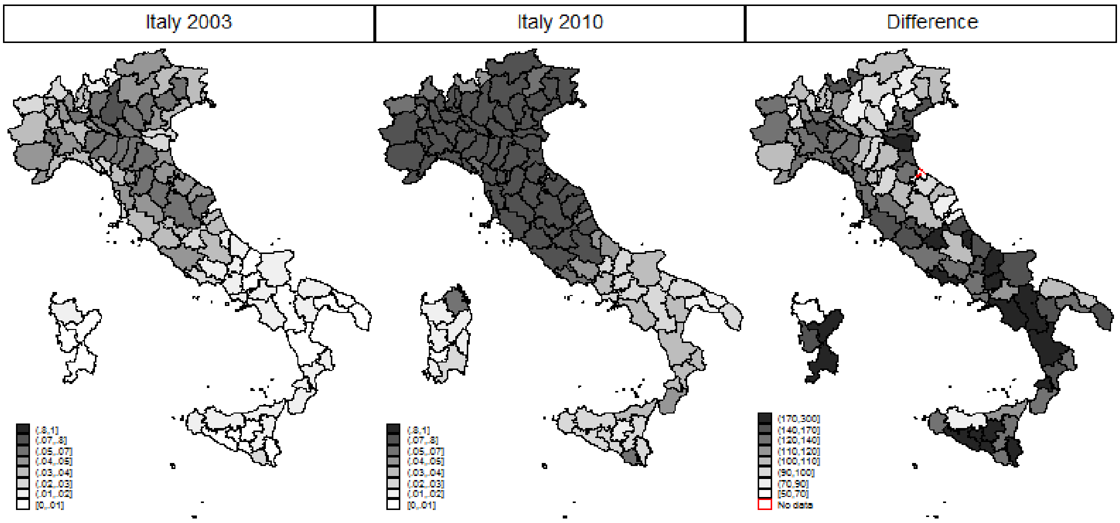


Figure 2.3: Share of immigrants by province and year, and its growth rate (in %).

to their definitions. We use the population weights distribution across municipalities within the same province to identify $\widehat{\xi}_i$.

Interaction probabilities with low HDI type immigrants grew substantially but uniformly over the 8 years span, although the interaction probabilities remain particularly high and concentrated in the north of Italy, both at the level of provinces (figure 2.4) and municipalities (figure 2.5).[24]

The distribution of interaction profiles across municipalities within the same province/year

---

probabilities for the population living in municipality i. The weight decreases according to the spatial distance, although one could have used more refined measures such as transportation time. The proximity weighting function is given by:

$$\delta(i,h;d) := \mathbf{1}(d(i,h) < 20km) \left(1 - \left(\frac{d(i,h)}{20km}\right)^2\right)^2.$$

where $\mathbf{1}(.)$ is the indicator function and $d(i,h)$ is the spatial distance obtained by the *cosine method*, and calculated by using latitude and longitude information for the municipalities' centroids.

[24]It is worth noting that the model is not fully spatial, as long as the kernel model is applied only within each province separately. So, at most, it is informative on the distribution of immigrants within the same province but cannot be generalize to the overall Italian case. This is a problem emerging also in the analysis of spatial models for urban segregation, where *between* cities interaction possibilities are often neglected, and each city (with its neighborhood) is analyzed in isolation.

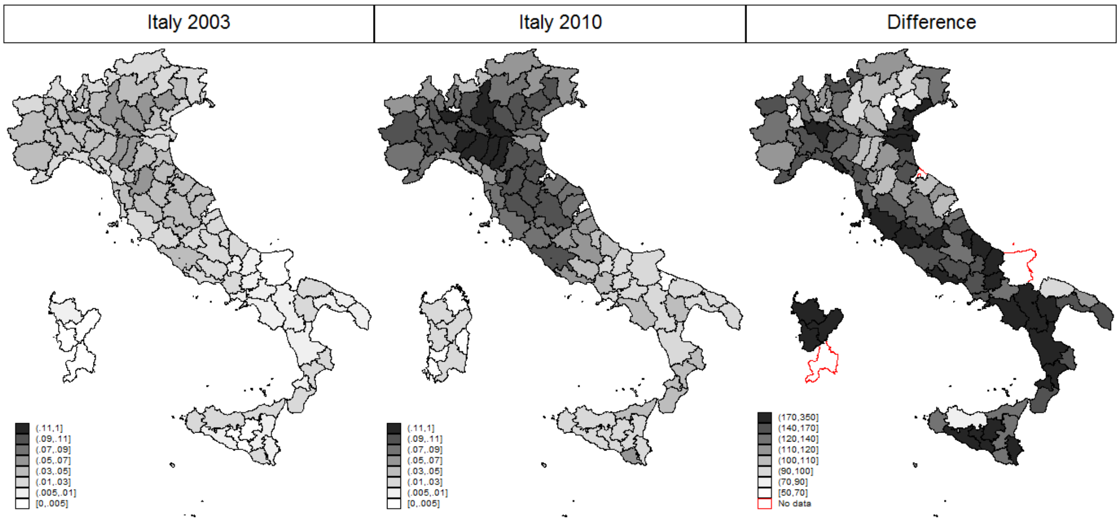


Figure 2.4: Expected interaction probability with *low HDI* immigrants by province and year, and its growth rate (in %).

pair defines the object of our study. Additional measures of disproportionality at municipality level, as well as indices at province level, can be now calculated.

2.6.2 Results

Segregation patterns across Italian provinces

The distribution matrix associated to a given province provides information about the disproportion in interaction probabilities at municipality level versus the expected probabilities at province level. Figure 2.9 reports the spatial distribution of the *disproportionality* coefficient for the low HDI type immigrants, defined as a_i. If a_i is larger than one, then the probability of interacting with low HDI type immigrants in municipality i is larger than what is expected at province level.

In north-east and central Italy it is observed the largest within province variability in interaction disproportionality, which implies higher variability across municipalities in the

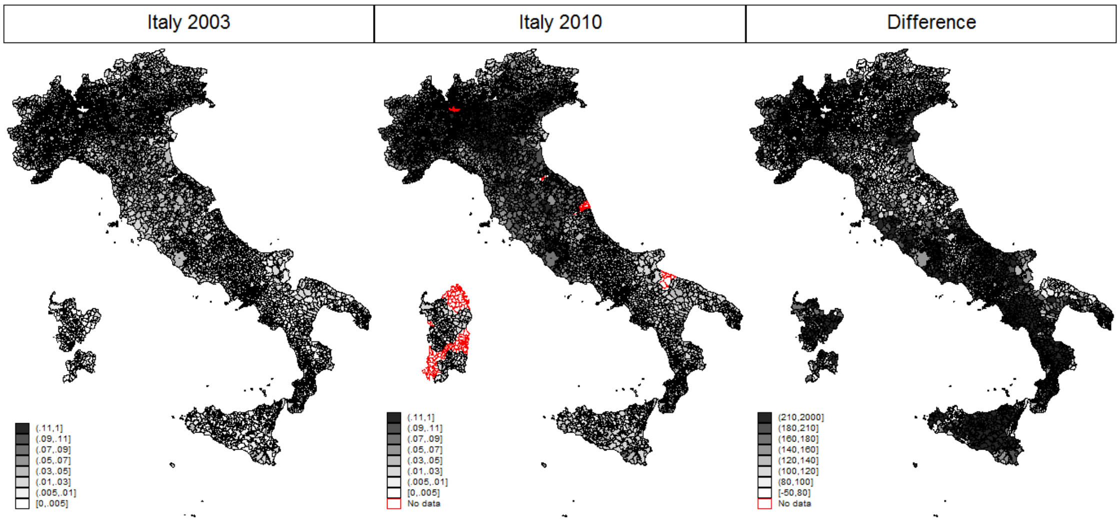


Figure 2.5: Interaction probability with *low HDI* immigrants by municipality and years, and its growth rate (in %).

type of interaction profiles. These macro-regions (the distribution of disproportionality coefficients at municipality level in the north Italy region is reported in figure 2.10) are also characterized by large variability in their ranking position throughout the period, while the expected interaction level remains substantially stable (see figure 2.11 and the figures in table 2.3, where the percentage of municipalities with $a_i > 1$, is shown to be substantially stable in the 2003/2010 period).

This particular pattern of (exposure) segregation across municipalities is captured both by the Gini Exposure index and the Mutual Information index by Frankel and Volij (2011), which we take as a reference for the class of multi-group exposure indices that do not satisfy composition invariance. The rank of all the 101 Italian provinces (for which data are available) produced by the Gini Exposure index is reported in figure 2.6 for the year 2003 and 2010. Provinces are ranked according to increasing segregation. The top 20 segregated provinces are concentrated in the center and the north-east regions of Italy. This outcome is coherent with the fact that the Gini Exposure index captures the within province variability

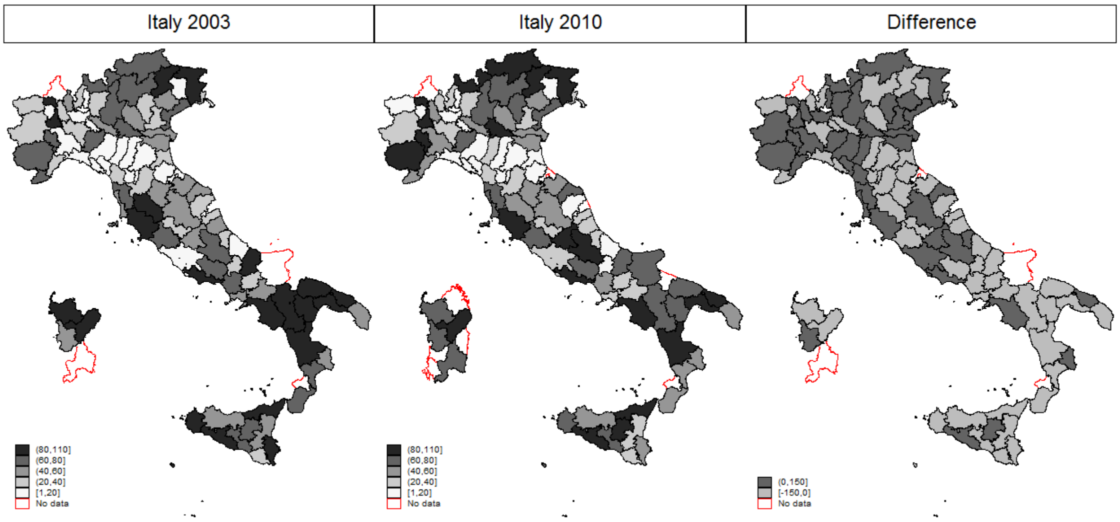


Figure 2.6: Ranking of Italian provinces according to the multi-group Gini Exposure index, by year. Differences are reported for provinces where segregation is increased (positive rank changes, in dark gray) and where segregation is decreased (negative rank changes, in pale gray).

in interaction profiles, which is consistent in the two regions in the period considered.

The Mutual Information index provides a closely related (although not coincident) picture (see figure 2.12 in the appendix). The changes in the ranking of the provinces by 2003 to 2010 (right panel of figures 2.6 and 2.12) generated by the two indices does not coincide. Thus calls for an appropriate analysis of the evolution of the differences between the indices and on the sources of this divergence.

Comparison of segregation indices

The graphs in figure 2.7 suggest two well defined patterns of segregation that distinguish the spatial indices under analysis. For each of the six indices considered (G_E, M, NE, R, EG, D) calculated by province and year, we report three curves, identifying the dynamics of segregation across years associated to the province scoring at the first, median and third quartile of the ranking of provinces defined, for each year, by each one of the indices.

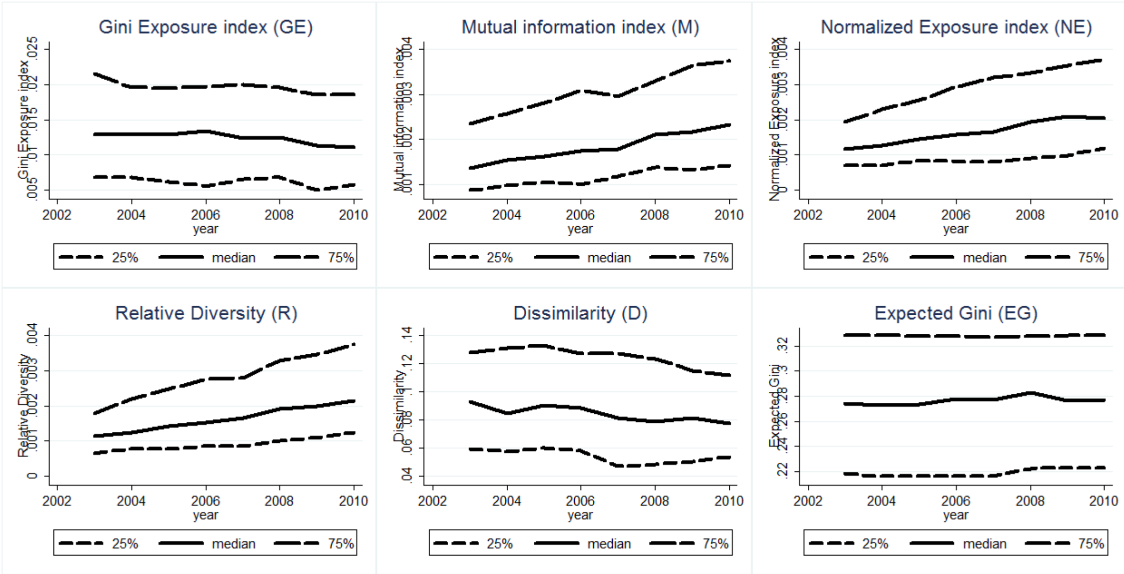


Figure 2.7: Dynamics of the six exposure segregation indices in the period 2003 to 2010, for the 101 Italian provinces. For each index and each year are reported the values of the index associated to the provinces in the first quartile (25%), *median* and third quartile (75%) of the ranking produced by the index in that year.

The Gini Exposure index identifies a slightly decreasing pattern of segregation across years. One can interpret the graph in the following way: the segregation pattern measured for the province scoring among the top 25%, 50% and 75% most segregated provinces in a given year is decreasing across the time interval considered. A similar type of pattern emerges by comparing the ranking produced by the Expected Gini index and the multi-group Dissimilarity index.

On the other hand, the Mutual Information index defines a different pattern: segregation is slightly increasing in time for the moderately (25% and median) segregated provinces, while the growth in segregation of the most segregated provinces is even more evident. Similar patterns can be constructed if the Mutual Information index is replaced by the Normalized Exposure or the Relative Diversity indices. One possible explanation of this divergence in segregation patterns is that the two families of indices obey to different aggregation principles.

Table 2.1: Rank correlation between segregation indices

Index	G_E	M	R	D	NE	EG	G_E2	A
Gini Exposure (G_E)	1	0.484	0.501	0.773	0.506	0.410	0.614	0.407
Mutual Information (M)	0.347	1	0.901	0.498	0.859	0.251	0.335	0.073
Relative Diversity (R)	0.356	0.750	1	0.665	0.988	0.305	0.499	0.121
Dissimilarity (D)	0.593	0.356	0.486	1	0.686	0.311	0.767	0.379
Normalized Exposure (NE)	0.357	0.709	0.919	0.502	1	0.325	0.534	0.109
Expected Gini (EG)	0.285	0.168	0.207	0.222	0.221	1	0.272	0.140
Gini Exposure, pair (G_E2)	0.437	0.230	0.350	0.576	0.377	0.184	1	0.341
Atkinson (A)	0.278	0.050	0.078	0.259	0.070	0.091	0.233	1
Mean (diagonal excluded)	0.379	0.373	0.427	0.506	0.221	0.321	0.341	0.151

Source: Data by ISTAT, demographic statistics, years 2003/2010.

Notes: Kendall (τ_b, below the diagonal) and Spearman (ρ, above the diagonal) rank correlation coefficients of spatial segregation indices calculated at province level, years 2003 to 2010. The total number of observations is 808 (708 for the Atkinson index, year 2010 is chosen to set the index weighting scheme). All coefficients are significantly different from zero at 1% level. Universe is set according to the ISTAT statistical definition of Italian provinces (reduced to 101 here), and indices are computed with information at municipality level. Provinces created or destroyed after 2003 are excluded from the sample. Social groups are mutually exclusive: natives (Italian nationality), immigrants from low HDI countries and immigrants from high HDI countries.

We study more in depth the ordinal relation between the six indices, along with the composition invariant Atkinson index characterized by Frankel and Volij (2011), by resorting on the rank correlations between the indices, reported in table 2.1. The correlations are all positive and significant. As anticipated above, the Gini Exposure index is significantly positively rank correlated with the Dissimilarity index ($\tau_b = 0.593$ and $\rho = 0.773$), although the link with the Expected Gini is less evident. On the other hand, the Mutual Information index, the Relative Diversity and the Normalized Exposure measures generate significantly similar rankings of segregated distributions.

Differently from the majority of the composition invariant measures, the Gini Exposure and the Dissimilarity indices are also correlated with the Atkinson index, thus remarking that the two indices, in part, are affected by the variation in the demographic size of the groups and account for the changes in overall composition.[25].

In figure 2.8(a) we decompose this correlation across years. We identify two distinct

[25] The reported correlation, for the appropriate indices, are comparable to the one computed in Frankel and Volij (2011)

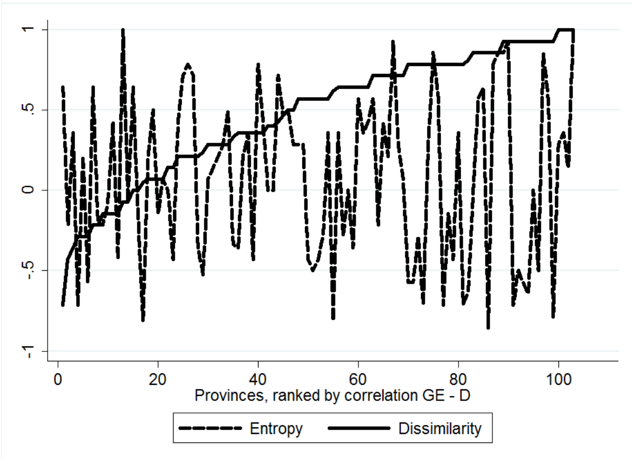
patterns of correlation between the Gini exposure index and the remaining composition invariant indices. The rank correlation between the Gini Exposure index and the Dissimilarity index remains fairly stable across time and persistently high. This pattern is well distinguishable from the patterns of rank correlation between the Gini Index and other indices such as M, NE and R. We also tried to perform the inverse analysis, that is comparing for each province the correlation in ranking between years. We do so by calculating the correlation between the Gini Exposure index and, alternatively, the Dissimilarity or the Mutual Information index for each province separately, exploiting the variability across time. However, it is not possible to disentangle any clear pattern among observed correlations.

Finally, we try to single out the sources of correlation of the indices across periods by looking at the main demographic variables that we have studied, such as the share of groups in a province, the interaction profile associated to a province, or the size of the population. The objective is to assess the impact of the variability in the data on the rank correlation between pairs of indices. To do so, we focus our attention to the correlation of the Gini Exposure with D (first pattern) and with M (second pattern). We use regression models to explain the contribution of each province in determining the Kendall's τ_b correlation measure used to construct figure 2.8. In fact, the Kendall's index of rank correlation is an average of the degree of measured *concordance* associated to each observation.[26] Hence, traditional OLS methods are suitable to assess the association of variability in concordance across provinces with the characteristics of each province.

We perform six regressions and we report the results in table 2.2. Each regression gives a list of coefficients that identify the impact of marginal variations in the independent variable on the rank correlation between the Gini Exposure and the Dissimilarity index (models (1) to (3)) or alternatively the Mutual Information index (models (4) to (6)). Models (3) and (6) control for time dummies (where 2003 is the reference value). As shown before, correlation patterns do not substantially differ across years. Moreover, the association between

[26] Let $\rho_i(I)$ be the rank of province i in the ranking of provinces in a given year produced by the index I. We say that, within a given year, provinces i and j are concordant with respect to indices G_E and D if $\rho_i(G_E) > \rho_j(G_E)$ and $\rho_i(D) > \rho_j(D)$ or $\rho_i(G_E) < \rho_j(G_E)$ and $\rho_i(D) < \rho_j(D)$. The degree of concordance associated to i is equal to the number of provinces that are concordant with i in a given year.

(a) Across periods



(b) Across provinces

Figure 2.8: Kendall's τ_b correlation coefficient of Mutual Information index (M), Relative interaction (R), Dissimilarity (D), Normalized Exposure (NE), Expected Gini (EG) and Gini Exposure for two groups (GE) indices with the multi-group Gini Exposure index. See the note of table 2.1 for further details. Correlations in panel (a) are calculated for each year for the whole set of realizations of the indices across provinces (on average 101 observations per year), while correlations in panel (b) are calculated for each province using the data of the years 2003/2010 (eight years, negative correlations are statistically zero at 5% confidence level). Provinces are ordered by increasing magnitude of the correlation between G_E and D.

common variables entering in the segregation indices is very low. We conclude that the rank association between the Gini Exposure index and the Dissimilarity index does not rely on the variability of the data considered. Moreover, the two indices produce very consistent rankings, and these rankings are not influenced by the structure of the data.

We repeat the same analysis by regressing the contribution of each observation in determining the rank correlation between the Gini Exposure index and the Mutual Information index. Results for the complete specification are reported in Model (6). As in the previous case, variables measuring the population (total or group level) distribution across provinces in absolute or relative terms have no impact in explaining changes in correlation. This is consistent with the normalization of the indices. However, in this case the year dummy captures some significant part of the trend in correlation. This result, along with the fact that the variability in inequality within interaction profiles (captured by the odds of interacting with an immigrant) have a significant negative impact on correlation between G_E and M, let us conclude that the association between G_E and M is in part due to the variability in the data, and decreases sensibly when the odds of interacting with one of the groups are low. Therefore, the two indices may capture different information when faced with substantial within interaction profiles heterogeneity. The Gini Exposure index is, however, robust with respect to these differences.

2.7 Concluding remarks

We have characterized a new partial order for the multi-group exposure dimension of segregation, based on the notion of inequality in the distribution of interaction profiles across individuals. Our characterization rests on the idea that if the interaction profiles of two individuals are merged, then segregation decreases. This rather attractive principle, along with a series of population replication invariance axioms, allows to characterize a well defined family of indicators. We study one of them, the Gini Exposure index. We adopt a spatial model to analyze the behavior of this index compared to other indices, which are

Table 2.2: The impact of demographic factors on the rank correlation

Dependent var.:	Num. of concordances (G_E and D)			Num. of concordances (G_E and M)		
	(1)	(2)	(3)	(4)	(5)	(6)
Pop. total	-0.000	-0.000	-0.000+	0.000	0.000	0.000+
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pop. natives	0.000	0.000	0.000+	-0.000	-0.000	-0.000+
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pop immigrants	0.000	0.000	0.000*	-0.000	-0.000	-0.000+
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share imm.		-124.199	556.237		151.561	-638.273
		(86.62)	(436.70)		(123.50)	(569.17)
Proportion imm.		5.353	-596.959		-164.355	676.974
		(83.73)	(432.16)		(119.84)	(565.11)
Proportion ratio			16.256			-133.832**
			(44.00)			(65.59)
Rank (prop. ratio)			-14.540			125.770**
			(41.89)			(63.28)
Rank by G_E			0.081***			-0.046*
			(0.02)			(0.02)
Year 2004		-0.626	-0.782		6.876***	6.691***
		(2.32)	(2.28)		(2.45)	(2.32)
Year 2005		-0.394	-0.458		1.758	1.160
		(2.37)	(2.31)		(2.46)	(2.26)
Year 2006		0.539	-0.796		9.410***	8.059***
		(2.38)	(2.36)		(2.63)	(2.41)
Year 2007		-0.130	-0.921		8.784***	8.325***
		(2.45)	(2.46)		(2.60)	(2.44)
Year 2008		-0.578	-1.812		7.226***	6.987***
		(2.51)	(2.50)		(2.59)	(2.47)
Year 2009		4.802*	2.222		10.515***	9.260***
		(2.47)	(2.51)		(2.66)	(2.51)
Year 2010		-0.027	-2.109		7.852***	6.243**
		(2.69)	(2.73)		(2.82)	(2.73)
Constant	60.236***	65.082***	55.332***	54.055***	47.755***	54.232***
	(0.82)	(2.15)	(3.69)	(0.95)	(2.34)	(4.07)
Provinces (8 years)	847	847	795	847	847	795
R^2	0.004	0.03	0.03	0.007	0.04	0.06
p-value model	0	0	0	0	0	0

+ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: Data by ISTAT, demographic statistics, years 2003/2010.

Notes: Regression (OLS) of the number of *concordant* observations on predictors, controlling by year effects. An observation is a province in a given year. Let $\rho_i(I)$ be the rank of province i in the ranking of provinces in a given year produced by the index I. We say that, within a given year, provinces i and j are concordant with respect to indices G_E and D if $\rho_i(G_E) > \rho_j(G_E)$ and $\rho_i(D) > \rho_j(D)$ or $\rho_i(G_E) < \rho_j(G_E)$ and $\rho_i(D) < \rho_j(D)$. The dependent variable is obtained by the number of cases of concordance associated to a given province i in a given year, both for the pair (G_E, D) and (G_E, M). These values, normalized by the maximum number of comparisons, give the Kendall's τ_b coefficient. Dependent variables are defined above.

not defined from an individual level perspective. We use Italian data on the distribution of immigrants across municipalities to show our point.

Possible extensions of our application would require to account for *conditional* residential choices, so that to eliminate confounding factors in determining interaction probabilities, as income and housing price dynamic.

We argue that the index proposed here is not limited to the study of spatial segregation phenomena, but it can be adapted to the analysis of others forms of segregation. One interesting case is segregation in networks, measured as inequality in the distribution of interaction profiles with predetermined social groups across the individuals in the network. The Gini Exposure index is designed to evaluate also this dimension of inequality.

2.A Appendix: An illustrative example

Let us introduce, with a simple example, the representation of the data that we use and the type of transformations involved in our analysis.

An interaction profile defines the conditional probability that a given population unit (an individual, a family or even a group of individuals), denoted by i, interacts with each of the social groups, denoted by g, in which the population is divided. This probability is indicated by π_{gi}. Each individual is associated with its own interaction profile, which may depend on her network, location, or demographic attributes. For instance, we consider four individuals l_1, l_2, j and k, equally partitioned into two groups g_1 and g_2. Who belongs to which group has no relevance for assessing segregation at the individual level when individuals are treated symmetrically. The interaction profile specifies, for each of the four individuals, the probability of interacting with groups g_1 and g_2 separately.[27] Let assume for simplicity that individuals l_1 and l_2 share the same interaction profile, which is marked with an l. We can reduce the analysis to three profiles, since thy represent all the distributional information that is needed to exploit segregation.

We use the following data to fix ideas:

$$\begin{pmatrix} \pi_{g_1 l} \\ \pi_{g_2 l} \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}, \quad \begin{pmatrix} \pi_{g_1 j} \\ \pi_{g_2 j} \end{pmatrix} = \begin{pmatrix} 1/8 \\ 7/8 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \pi_{g_1 k} \\ \pi_{g_2 k} \end{pmatrix} = \begin{pmatrix} 3/8 \\ 5/8 \end{pmatrix}.$$

Interaction profiles are treated as if they are bundles of goods allocated to each demographic unit.

To normalize the data and eliminate any form of heterogeneity *within* interaction profiles we use the vector of expected interaction probabilities π_g as the endogenously determined reference interaction profile.

It turns out that segregation can be measured as a form of *dissimilarity* between the likelihood that any randomly drawn individual of group g interacts with individual i, for any group g and any unit i (see for instance chapter 1). This likelihood, denoted by $\Pr[i|g]$, should ideally equate the probability of interacting with unit i, namely $\Pr[i]$ if interaction profiles are equally distributed in the population. That is, $\Pr[i|g] = \Pr[i|g']$ for all is and all groups $g \neq g'$. Any departure from this rather extreme allocation entails a form of exposure

[27] This probability may, in general depend on may factors such as location, unit level characteristics, group level characteristics.

segregation.

The Bayes rule ties interaction probabilities to the likelihood of interaction in the following way:

$$\Pr[i|g] = \frac{\Pr[i] \cdot \pi_{gi}}{\pi_g}.$$

In our example, suppose that weights are defined as follows: $\Pr[l] = 2/4$, $\Pr[j] = 1/4$ and $\Pr[k] = 1/4$. Unit l is weighted double as much as the other. The expected interaction profile can be easily computed as follows:

$$\begin{pmatrix} \pi_{g_1} \\ \pi_{g_2} \end{pmatrix} = \frac{2}{4}\begin{pmatrix} \pi_{g_1 l} \\ \pi_{g_2 l} \end{pmatrix} + \frac{1}{4}\begin{pmatrix} \pi_{g_1 j} \\ \pi_{g_2 j} \end{pmatrix} + \frac{1}{4}\begin{pmatrix} \pi_{g_1 k} \\ \pi_{g_2 k} \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}.$$

It is easy to see in this example that interaction profiles are not equally distributed. In fact, one obtains:

$$\begin{pmatrix} \Pr[l|g_1] \\ \Pr[l|g_2] \end{pmatrix} = \begin{pmatrix} 2/4 \\ 2/4 \end{pmatrix}, \quad \begin{pmatrix} \Pr[j|g_1] \\ \Pr[j|g_2] \end{pmatrix} = \begin{pmatrix} 1/8 \\ 7/24 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \Pr[k|g_1] \\ \Pr[k|g_2] \end{pmatrix} = \begin{pmatrix} 3/8 \\ 5/24 \end{pmatrix},$$

which shows that the sources of exposure are units j and k, given that $\Pr[l|g_1] = \Pr[l|g_2]$. In fact, unit l's interaction profile coincides with the expected profile.

A merge of units j and k is a compounding of their respective interaction profiles, with weights 1/2 (given that j and k have similar demographic weights). This operation gives the new profile π':

$$\begin{pmatrix} \pi'_{g_1 j} \\ \pi'_{g_2 j} \end{pmatrix} = \begin{pmatrix} \pi'_{g_1 k} \\ \pi'_{g_2 k} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1/8 \\ 7/8 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 3/8 \\ 5/8 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix},$$

that corresponds to the expected profile. After the merge, l, j and k hold the same profile, and segregation is eliminated. More complex examples can be constructed with a larger number of groups and larger demographic size. The analytical background in the multi-group setting is similar.

2.B Appendix: Proofs

2.B.1 Proof of Theorem 2.1

Proof. We prove that $A \preccurlyeq B \;\Leftrightarrow\; i) \;\Leftrightarrow ii)$ and $i) \Leftrightarrow iii)$.

Consider an allocation $A \in \mathcal{A}(G)$ and the distribution matrix $\mathbf{A}$. Let define, in a shorthand notation, the population weighted distribution matrix $\widetilde{\mathbf{A}}$ as:

$$\widetilde{\mathbf{A}} \;:=\; \mathbf{A} \cdot \boldsymbol{\xi}(A) \;=\; \mathbf{A} \cdot \mathrm{diag}(\xi_1(A), \dots, \xi_{N(A)}).$$

Moreover, let $\mathbf{e}_{N(A)}$ and $\mathbf{0}_{N(A)}$ to be a $1 \times N(A)$ row vectors with all entries equal to one and zero respectively. The superscript "t" stands for transpose. Note that by construction the sum of the entries od each of the G rows of $\widetilde{\mathbf{A}}$ satisfy $\widetilde{\mathbf{A}} \cdot \mathbf{e}^t_{N(A)} = \mathbf{e}^t_G$.

In the first part of the proof it is shown that $B \preccurlyeq A \;\Leftrightarrow\; i)$, that is the ranking of allocations $A, B \in \mathcal{A}(G)$ produced by the axioms characterizing $\preccurlyeq$ is equivalently represented by a partial order of the matrices $\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}$, denoted by $\widetilde{\preccurlyeq}$, which is the matrix majorization order in Dahl (1999). Thus $B \preccurlyeq A$ if and only if $\widetilde{\mathbf{B}} \;\widetilde{\preccurlyeq}\; \widetilde{\mathbf{A}}$.

The order $\widetilde{\preccurlyeq}$ is rationalized by a set of operations that are the direct counterpart, on normalized distribution matrices, of the operations defined by the axioms. Non-triviality, Continuity and Normalization are satisfied by construction of the ordering. If not, then there are some cases where $\widetilde{\mathbf{A}} = \widetilde{\mathbf{B}}$ implies $B \prec A$ by the fact that Normalization is not satisfied, a contradiction of the assumption that the order does not satisfy Non-triviality or Continuity (and so all matrices are indifferent or non ordered).

Units and Groups Anonymity are both equivalent to state that if $\exists \boldsymbol{\Pi}_G$ and $\exists \boldsymbol{\Pi}_{N(A)}$, which are permutation matrices of size G and $N(A)$ respectively, such that $\widetilde{\mathbf{B}} = \boldsymbol{\Pi}_G \cdot \widetilde{\mathbf{A}} \cdot \boldsymbol{\Pi}_{N(A)}$, then $\widetilde{\mathbf{B}} \;\widetilde{\sim}\; \widetilde{\mathbf{A}}$.

The transformation underlying Population Replication Invariance is equivalent to write (for any finite scalar $\lambda \in \mathbb{N}_{++}$):

$$\widetilde{\mathbf{B}} \;=\; \underbrace{\left(\frac{1}{\lambda}\widetilde{\mathbf{A}}, \dots, \frac{1}{\lambda}\widetilde{\mathbf{A}}\right)}_{\lambda \text{ times}}.$$

Note that, although each individual's interaction profile is replicated λ times, it has to be scale by the factor $1/\lambda$ to be consistent with the requirement that when the population is replicated, then the population weight of the observations has to be proportionally scaled by the same replication factor. This transformation can be equivalently represented by a matrix operation, involving a $N(A) \times \lambda N(A)$ *row stochastic matrix* $\mathbf{X}$ with the following structure:

$$\mathbf{X} = \frac{1}{\lambda} \begin{pmatrix} \mathbf{e}_\lambda & \mathbf{0}_\lambda & \dots & \mathbf{0}_\lambda \\ \mathbf{0}_\lambda & \mathbf{e}_\lambda & \dots & \mathbf{0}_\lambda \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_\lambda & \mathbf{0}_\lambda & \dots & \mathbf{e}_\lambda \end{pmatrix}. \tag{2.2}$$

The Individual Replication Invariance axioms posits a similar transformation of the data. Suppose individual i in allocation A is replicated to obtain individual i', and the population weight of i in A is split according to a parameter $\rho \in [0, 1]$ to obtain the weights associated to i and i' in B. The resulting allocation B is such that $N(B) = N(A) + 1$, $\xi_i(B) = \rho\xi_i(A)$, $\xi_{i'}(B) = (1-\rho)\xi_i(A)$ and the columns od the distribution matrix are such that $\mathbf{b}_i = \mathbf{b}_{i'} = \mathbf{a}_i$ while $\mathbf{b}_j = \mathbf{a}_j \ \forall j \neq i, i'$. In matrix notation, this operation involves a row stochastic matrix:

$$\widetilde{\mathbf{B}} = \left(\widetilde{\mathbf{a}}_1, \dots, \rho\widetilde{\mathbf{a}}_i, (1-\rho)\widetilde{\mathbf{a}}_{i'}, \widetilde{\mathbf{a}}_{i+1}, \dots, \widetilde{\mathbf{a}}_{N(A)}\right) = \widetilde{\mathbf{A}} \cdot \mathbf{X}. \tag{2.3}$$

The $N(A) \times (N(A)+1)$ row stochastic matrix $\mathbf{X}$ is such that the elements satisfy $x_{ii} = \rho$, $x_{ii'} = (1-\rho)$, $x_{jj} = 1$ for all $j \neq i, i'$ and zero otherwise.

The operation of lottery compounding underlying the Interaction Profiles Merge entails a linear transformation of matrices $\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}$ that are related to the merge and split transformations studied in chapter 1. The merge of two interaction profiles $\mathbf{a}_i$ and $\mathbf{a}_j$ generates a new *compounded* interaction profile $\mathbf{b}_i = \mathbf{b}_j = \mathbf{a}^c = \alpha\mathbf{a}_i + (1-\alpha)\mathbf{a}_j$ where $\alpha = \frac{\xi_i(A)}{\xi_i(A)+\xi_j(A)}$. Thus, B is obtained from A by an interaction profiles merge if and only if $\widetilde{\mathbf{b}}_i = \xi_i(A)\mathbf{a}^c$ and $\widetilde{\mathbf{b}}_j = \xi_j(A)\mathbf{a}^c$.

Equivalently, any merge of interaction profiles concerning two individuals i and j can be represented by an operation involving a row stochastic matrix:

$$\widetilde{\mathbf{B}} = \left(\widetilde{\mathbf{a}}_1, \dots, \alpha\left(\widetilde{\mathbf{a}}_i + \widetilde{\mathbf{a}}_j\right), \dots, (1-\alpha)\left(\widetilde{\mathbf{a}}_i + \widetilde{\mathbf{a}}_j\right), \dots, \widetilde{\mathbf{a}}_{N(A)}\right) = \widetilde{\mathbf{A}} \cdot \mathbf{X}. \tag{2.4}$$

The $N(A) \times N(A)$ row stochastic matrix $\mathbf{X}$ defines a merge transformation for columns i and j, and a proportional *split* according to the parameter α. The matrix $\mathbf{X}$ takes the form of an identity matrix with the elements $x_{ii} = x_{ji} = \alpha$ and $x_{ij} = x_{jj} = (1 - \alpha)$ respectively.

The order $\widetilde{\preccurlyeq}$ characterized as before represents the segregation ordering $\preccurlyeq$ and based upon the comparison of population normalized distribution matrices.

The operations studied here are independent, therefore it is possible to move from allocation A towards allocation B by using any sequence of these transformations applied to matrix $\widetilde{\mathbf{A}}$ to obtain $\widetilde{\mathbf{B}}$. Moreover, each operation entails a linear transformation of the population weighted distribution matrix $\widetilde{\mathbf{A}}$ toward $\widetilde{\mathbf{B}}$ which is based upon a row stochastic matrix. In chapter 1 we use similar transformations to define the *dissimilarity order*. Using chapter 1 notation, the *merge* operation can be written as in (2.4), the *split* operation for individuals or of the entire population can be written using (2.3) and (2.2) respectively; the permutation of groups and individuals is instead achieved making use of permutation matrices, which are indeed row stochastic. Finally, the insertion of "empty" individuals (that is having $\widetilde{\mathbf{a}}_{i'} = \mathbf{0}_G$, a special case of the transformation underlying the Individual Replication Invariance axiom) entails a row stochastic matrix $\mathbf{X}$ as in (2.3), with $1 - \rho = 0$, so that the new individual has a population weight $\xi_{i'}(A) = 0$.

Following the arguments in chapter 1, any independent sequence of these operations entails a linear transformation of $\widetilde{\mathbf{A}}$ toward $\widetilde{\mathbf{B}}$ which involves a row stochastic matrix $\mathbf{X}$. Moreover, any row stochastic matrix transforming transforming $\widetilde{\mathbf{A}}$ into $\widetilde{\mathbf{B}}$ can be decomposed into the sequence of operations required by merge, split, permutation and insertion of empty individuals, so that $B \preccurlyeq A$. By adding the Anonymity axioms, we obtain the following equivalence, which concludes the first part of the proof:[28]

$$\widetilde{\mathbf{B}} \; \widetilde{\preccurlyeq} \; \widetilde{\mathbf{A}} \quad \text{if and only if} \quad \exists \mathbf{\Pi}_G \quad \text{such that} \quad \widetilde{\mathbf{B}} = \mathbf{\Pi}_G \cdot \widetilde{\mathbf{A}} \cdot \mathbf{X}.$$

To prove that $i) \;\Leftrightarrow\; ii)$, we use a result in chapter 1. There, we study the *Zonotope* $Z(\widetilde{\mathbf{A}})$ which is equal by construction to $SZ(A)$. Using Theorem 4 in chapter 1, $\widetilde{\mathbf{B}} = \widetilde{\mathbf{A}} \cdot \mathbf{X}$, with $\mathbf{X}$ row stochastic, if and only if $Z(\widetilde{\mathbf{B}}) \subseteq Z(\widetilde{\mathbf{A}})$, and therefore $SZ(B) \subseteq SZ(A)$, which gives the desired result.

To prove $i) \Rightarrow \; iii)$, we restrict attention to the case where $N(A) = N(B) = N$. Let $B \preccurlyeq A$ hold, and let ϕ indicate a member of to the largest class of functions that rank A and

[28]The linear transformation involving a row stochastic matrices is often denoted as *matrix majorization* and it has been introduced by Dahl (1999) and further studied in Marshall et al. (2011).

B coherently with $\preccurlyeq$. The desired implication is a consequence of the fact that the class of column symmetric, addition respondent and quasiconvex functions is a subset of this class of functions ϕ. If $i)$ is satisfied then $\widetilde{\mathbf{B}} = \widetilde{\mathbf{A}} \cdot \mathbf{X}$, where $\mathbf{X}$ is a row stochastic matrix, and by Proposition 3.1 in Dahl (1999), there exist $(\lambda_1, \ldots, \lambda_m) \in [0,1]^m$ with $\sum_{\ell=1}^m \lambda_\ell = 1$ and $\mathbf{X}_1, \ldots, \mathbf{X}_m \in \mathcal{R}_N \cup \{\mathbf{I}_N\}$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \in \mathcal{P}_N$ such that $\widetilde{\mathbf{B}} = \sum_{\ell=1}^m \lambda_\ell \widetilde{\mathbf{A}} \cdot \mathbf{X}_\ell \cdot \mathbf{Y}_\ell$. As ϕ is quasiconvex and additive respondent, $\phi(\widetilde{\mathbf{B}}) \leq \max_\ell \phi\left(\widetilde{\mathbf{A}} \cdot \mathbf{X}_\ell \cdot \mathbf{Y}_\ell\right) \leq \max_\ell \phi\left(\widetilde{\mathbf{A}} \cdot \mathbf{Y}_\ell\right)$. As ϕ is also column symmetric, $\phi(\widetilde{\mathbf{A}} \cdot \mathbf{Y}_\ell) = \phi(\widetilde{\mathbf{A}})$ for all ℓ; hence $\phi(\widetilde{\mathbf{B}}) \leq \phi(\widetilde{\mathbf{A}})$ for all ϕ consistent with $\preccurlyeq$.

Finally, we show that $iii) \Rightarrow i)$ by exploiting Dahl's (1999) Corollary 3.4. It states that:

$$i) \quad \Leftrightarrow \quad \Psi(\widetilde{\mathbf{B}}) = \sum_{i=1}^N \psi(\xi_i(B)\mathbf{b}_i) \leq \sum_{i=1}^N \psi(\xi_i(A)\mathbf{a}_i) = \Psi(\widetilde{\mathbf{A}}), \tag{2.5}$$

holds for all *positively homogeneous* ($\psi(\lambda\mathbf{x}) = \lambda\psi(\mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^G$ and $\lambda \in \mathbb{R}_+$) and *subadditive* ($\psi(\mathbf{x}+\mathbf{y}) \leq \psi(\mathbf{x}) + \psi(\mathbf{y})$) functions $\psi : \mathbb{R}^G \to \mathbb{R}_+$. We argue that every Ψ is a linear additive functional which is: column symmetric, because by permuting the elements of the sum in Ψ, the sum does not change; additive respondent, because ψ is subadditive; quasiconvex, because $\Psi(\sum_\ell \lambda_\ell (\widetilde{\mathbf{A}})_\ell) = \sum_i \psi\left(\sum_\ell \lambda(\xi_i(A)\mathbf{a}_i)_\ell\right)$, that by subadditivity and linear homogeneity is inferior, or at most equal to $\sum_i \sum_\ell \lambda_\ell \psi((\xi_i(A)\mathbf{a}_i)_\ell) = \sum_\ell \lambda_\ell \Psi((\widetilde{\mathbf{A}})_\ell) \leq \max_\ell\{\Psi((\widetilde{\mathbf{A}})_\ell)\}$, where the last inequality holds by construction. Hence, if B is ranked as less segregated than A for all indices in the class of functions ϕ, then the ranking is preserved for all indices in the class of functionals Ψ, while the reverse is not true. From (2.5), this is equivalent to say that $iii) \Rightarrow i)$, which concludes the proof. ■

2.B.2 Proof of Proposition 2.1

Proof. The result is a direct implication of the properties of the determinant of a $G \times G$ matrix. In fact, the Gini Exposure index is the sum of the absolute value of the determinants of $G \times G$ matrices, hence any operation that preserves the determinant or make it closer to zero also reduces the measured segregation. ■

2.B.3 Proof of Proposition 2.2

Proof. The proof is made by construction. We firstly partition the set of all possible G-tuple $\{i_1, \ldots, i_G\}$ of individuals into two groups. There are some G-tuples gathering individuals that exclusively belong to subpopulation $\mathcal{N}_g(A)$, for each of the groups $g \in \mathcal{G}$. The remaining G-tuples belong, instead, to the overlapping set $\mathcal{O}$. This originate the first result: the Exposure Gini index is linearly separable into a within component plus the overlapping term. The latter is representable itself as a Gini index (because the whole population is taken into consideration in calculating it):

$$G_E(A) \;:=\; \text{Within term} \;+\; G_E(A|\mathcal{O}).$$

We now turn to the *within term*. Again, by linearity of the Gini index it is possible to separate the different observations by group, defined by $\mathcal{N}_g(A)$, such that for $i \in \mathcal{N}_g(A)$ it holds that $i \in \{i_1, \ldots, i_G\}$ only if $\{i_1, \ldots, i_G\} \in \mathcal{N}_g(A)$. An obvious requirement, always satisfied by definition of an allocation, is that $\mathcal{N}_g(A) \cap \mathcal{N}_m(A) = \emptyset$ for all groups $g \neq m$. As a result one obtains a comparison of G-tuples for all groups separately, for a total of G factors adding up to the within component.

Each of the G factors can be written as a sum of absolute values of determinants of a squared matrix of size G, which for simplicity is referred to by $\mathbf{D}$. Note that within a chosen group g, $\mathbf{D}$ only depends upon the chosen G-tuple in $\mathcal{N}_g(A)$. The within term can be written as:

$$\text{Within term} \;=\; \sum_g \frac{1}{G!} \sum_{\{i_1,\ldots,i_m,\ldots,i_G\} \in \mathcal{N}_g(A)} |\det(\mathbf{D})|$$

For a chosen group (say g) and a given G-tuple (say the one including i_m), an element of the matrix $\mathbf{D}$ chosen in any position (say the one corresponding to row g and column i_m) is given by $\pi_{gi_m}(A)\mathbf{a}_{i_m}$.

Multiplication and division of the interaction probability vector by an appropriate conversion factor $\frac{\mathbb{E}[\pi_g(A)]}{(\sum_{i\in\mathcal{N}_g(A)} \xi_i(A))\mathbb{E}[\pi_g(A)|g]}$ does not produce any effect. The operation gives a new matrix, where a generic element in row g, column i_m is defined by

$$\frac{(\sum_{i\in\mathcal{N}_g(A)} \xi_i(A))\mathbb{E}[\pi_g(A)|g]}{\mathbb{E}[\pi_g(A)]} \frac{\widetilde{\xi}_{i_m}(A)\pi_{gi_m}(A)}{\mathbb{E}[\pi_g(A)|g]}$$

In compact form, one can substitute $\mathbf{D}$ with $\widehat{\mathbf{D}}$ in the calculation of the within term, to obtain:

$$\text{Within term } = \sum_g \frac{1}{G!} \sum_{\{i_1,\dots,i_m,\dots,i_G\}\in\mathcal{N}_g(A)} |\det(\boldsymbol{\alpha}_g \cdot \widehat{\mathbf{D}})|,$$

where

$$\boldsymbol{\alpha}_g := \operatorname{diag}\left(\frac{(\sum_{i\in\mathcal{N}_g(A)} \xi_i(A))\mathbb{E}[\pi_1(A)|g]}{\mathbb{E}[\pi_1(A)]}, \dots, \frac{(\sum_{i\in\mathcal{N}_g(A)} \xi_i(A))\mathbb{E}[\pi_G(A)|g]}{\mathbb{E}[\pi_G(A)]}\right),$$

and $\widehat{\mathbf{D}} = \boldsymbol{\alpha}_g^{-1} \cdot \mathbf{D}$.

The determinant of the product of two matrices is the product of the respective determinants of the factors. Moreover, the determinant of a diagonal matrix is the product of elements on the diagonal. Few calculations show that $\det(\boldsymbol{\alpha}_g) = \alpha_g$, defined in the proposition. The value α_g only depends on the group index. Hence, the following result applies, which concludes the proof:

$$\text{Within term } = (\sum_{g\in\mathcal{G}} \alpha_g) \sum_{g\in\mathcal{G}} \frac{\alpha_g}{\sum_{g\in\mathcal{G}} \alpha_g} G_E(A|g).$$

■

2.C Appendix: Additional material for the empirical analysis

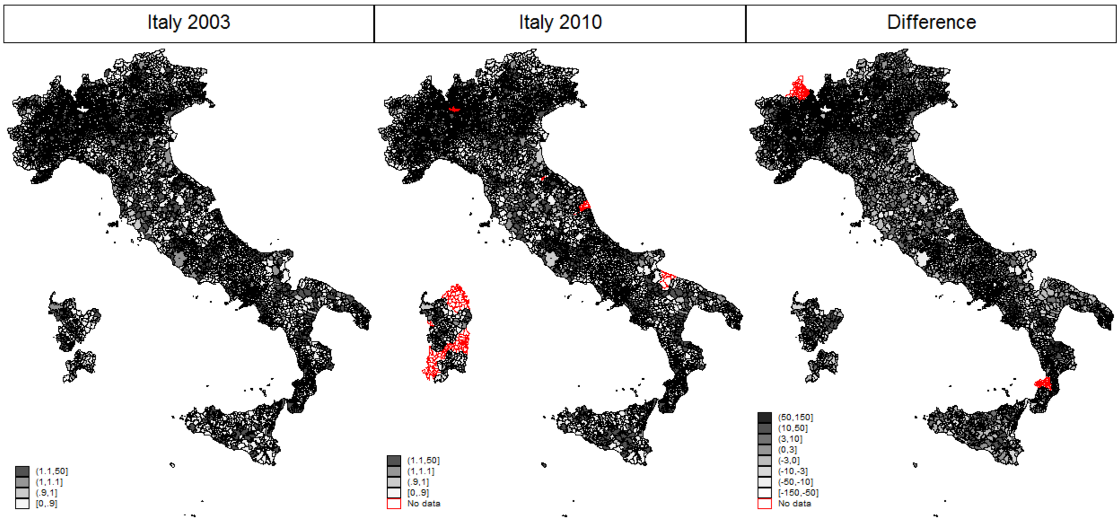


Figure 2.9: Disproportionality in interaction probability (with respect to the expected interaction) with *low HDI* immigrants by municipality and year, and the change in ranking (relative to the position of the municipalities in 2003 within the same province).

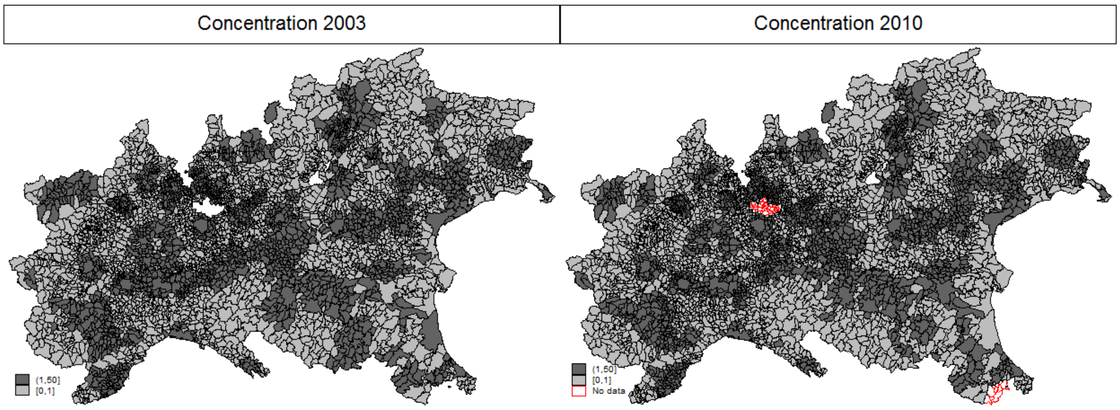


Figure 2.10: Disproportionality in interaction probability (with respect to the expected interaction) with *low HDI* immigrants by municipality in 2010 for the North Italy macro-region. Higher concentration ($a_k > 1$) is interpreted as the disproportion between the interaction with people from low HDI countries with respect to the expected interaction. The expected interaction probability is calculated at province level.

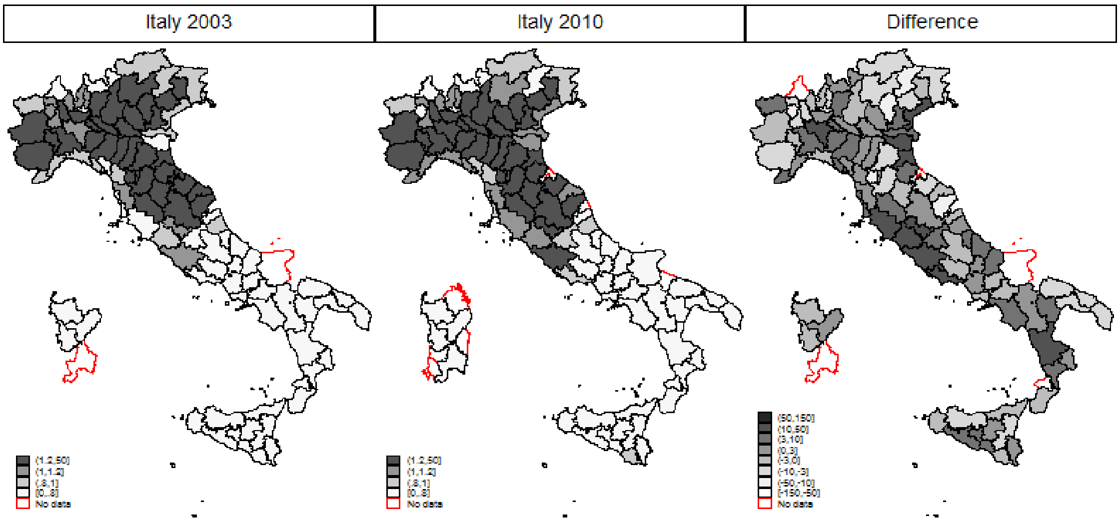


Figure 2.11: Disproportionality in *expected* interaction probability with *low HDI* immigrants by province and year, and it growth rate (in %).

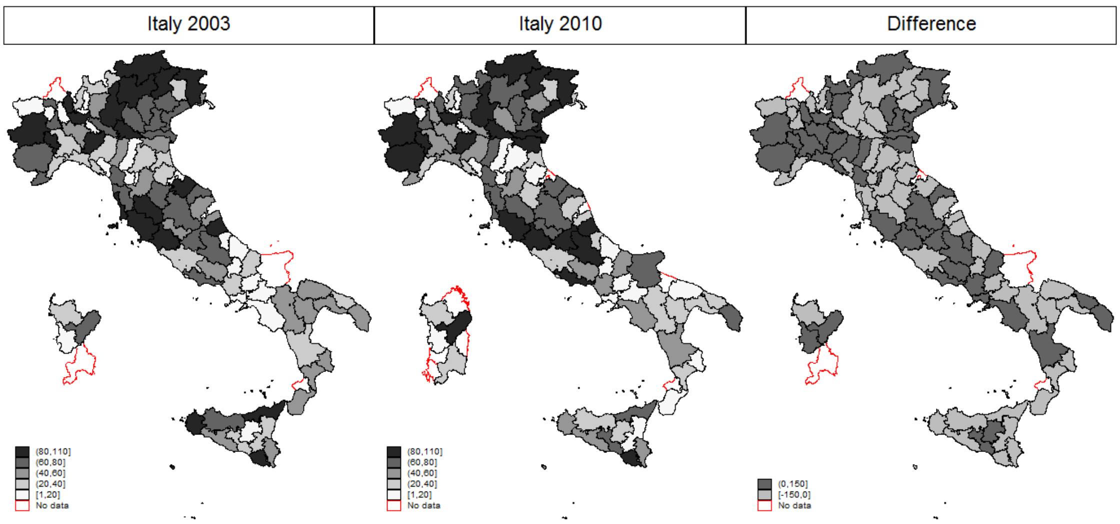


Figure 2.12: Ranking of Italian provinces according to the multi-group Mutual information index, by year. Differences are reported for provinces where segregation is increased (positive rank changes, in dark gray) and where segregation is decreased (negative rank changes, in pale gray).

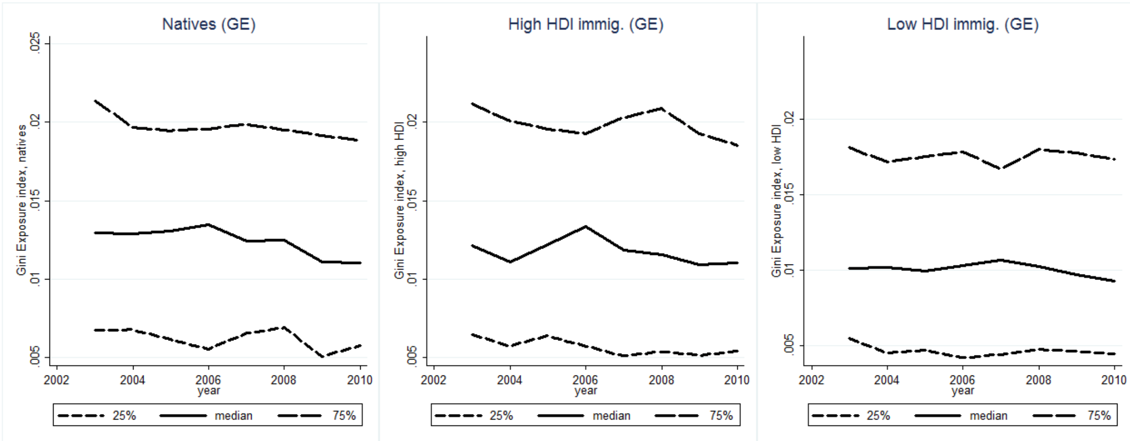


Figure 2.13: Dynamics of the decomposition of the G_E index in the period 2003 to 2010, for the 101 Italian provinces. For each subgroups and each year are reported the values of the index associated to the provinces in the first quartile (25%), *median* and third quartile (75%) of the ranking produced by the index in that year.

Table 2.3: Descriptive statistics

Year	2003	2004	2005	2006	2007	2008	2009	2010
Provinces (N)	103	103	103	107	107	107	107	110
Municipalities (N)	8100	8101	8101	8101	8101	8101	8100	8094
Municipalities by province	79	79	79	76	76	76	76	74
Population (total, mil)	57.9	58.5	58.8	59.1	59.6	60.0	60.3	60.6
Population by municipality	6267	6342	6375	6441	6500	6549	6586	6417
Population by province	562022	567596	570405	552629	557190	561169	563928	551149
Low HDI Immigrants								
Share	0.029	0.035	0.039	0.043	0.050	0.057	0.062	0.067
	0.017	*0.021*	*0.023*	*0.025*	*0.028*	*0.031*	*0.032*	*0.035*
Interaction probability	0.029	0.036	0.040	0.044	0.045	0.052	0.063	0.066
	0.018	*0.022*	*0.025*	*0.027*	*0.030*	*0.033*	*0.034*	*0.037*
High HDI immigrants								
Share	0.005	0.006	0.006	0.007	0.007	0.008	0.008	0.008
	0.003	*0.004*	*0.004*	*0.004*	*0.005*	*0.005*	*0.005*	*0.005*
Interaction probability	0.005	0.006	0.006	0.007	0.006	0.007	0.008	0.008
	0.004	*0.004*	*0.004*	*0.004*	*0.005*	*0.005*	*0.005*	*0.005*
Immigrants concentration as a proportion of cases where $a > 1$:								
By municipality	0.419	0.409	0.407	0.406	0.419	0.419	0.415	0.413
	0.493	*0.492*	*0.491*	*0.491*	*0.493*	*0.493*	*0.493*	*0.492*
By province	0.447	0.456	0.437	0.458	0.570	0.570	0.495	0.536
	0.500	*0.501*	*0.498*	*0.501*	*0.497*	*0.497*	*0.502*	*0.501*
Polarization ($.9 < a < 1.1$)	0.078	0.097	0.107	0.112	0.131	0.140	0.131	0.155
	0.269	*0.298*	*0.310*	*0.317*	*0.339*	*0.349*	*0.339*	*0.363*

Data by ISTAT, demographic statistics, years 2003/2010.

Interaction probabilities constructed with a spatial biweighted quadratic kernel, boundary distance is 20km.

Standard deviations are reported in italic.

Chapter 3

Equalization of opportunity: Definitions and implementable conditions

Contents

3.1 Introduction

Equality of opportunity has gained popularity, in scholarly debates as well as among policy-makers, for defining the relevant egalitarian objective for the distribution, among individuals, of a broad range of social and economic outcomes, such as health, wealth, income, etc. The ethical foundations of the equality of opportunity principle have been extensively discussed and are well-established (for a comprehensive discussion, see Dworkin 1981a, Roemer 1998, Fleurbaey 2008).

Public policy has now often set as its main objective to level the playing field among citizens and to provide equality of opportunity in a variety of areas of intervention such as education, health and, eventually, income. Assessing whether policy intervention indeed succeeds at equalizing opportunities is obviously a key issue for policy evaluation.

Addressing this issue requires to draw on explicit evaluation criteria that are, to a large extent, absent from the existing literature. More specifically, while *equality* of opportunity has now been clearly defined in the recent literature, criteria allowing to assess the (partial) *equalization* of opportunity, understood as a reduction in the extent of inequality of opportunity, are so far absent from the literature.

The objective of this chapter is to formally define a criterion of opportunity equalization, that would be both consistent with theoretical views of equality of opportunity and empirically implementable, to allow for the evaluation of the effect of public policy intervention.

The recent philosophical and economic literature has offered a clear characterization of the requisite of equality of opportunity. The equality of opportunity perspective amounts to draw a distinction between fair and unfair inequality of individual outcomes. This requires to single out two polar sets of determinants of the observed outcomes: on the one hand, *effort* gathers the legitimate sources of inequality among individuals; on the other hand, *circumstances* correspond to the set of morally-irrelevant factors fostering inequalities across individuals that call for compensation. Define a *type* the set of individuals with similar circumstances. In general terms, equality of opportunity will be said to prevail, if, given effort, no set of circumstances yields an advantage over the others. This reflects what is usually referred in the literature as the *compensation principle*.

Provided one explicitly defines the relevant notion of "advantage", this general principle allows to assess in various empirical contexts whether equality of opportunity prevails. However, this leads to a binary criterion (equality of opportunity is satisfied or not) and it does not allow to rank, from the point of view of equality of opportunity, different situations where equality of opportunity is not satisfied. Assessing the equalizing impact of policy intervention obviously calls for such a ranking, especially when policies do not allow to reach full equality of opportunity.

The perspective of Lefranc et al. (2009) (henceforth denoted LPT) breaks down the dichotomy between equality and inequality of opportunity by distinguishing between a strong and a weak form of equality of opportunity. As discussed in LPT, not all the determinants of outcome fall under the two categories of effort and circumstances: there exists a third class of determinants, denoted *luck*, that gathers the morally-irrelevant factors of inequality that do not call for explicit compensation. In this context, given their level of effort and their type, individuals face a distribution of possible outcomes. Comparing the outcomes prospects offered to individuals with similar effort, when their circumstances vary, amounts to compare lotteries of outcomes.

There might be different ways of formalizing the requirement that no type is advantaged over the others. A first possibility is to require that the outcome distributions be identical across individuals with similar effort, i.e. independent of circumstances. This is the strong form of equality of opportunity analyzed in LPT. A weaker condition is to require that it is not possible to unanimously rank the outcome distributions attached to different circumstances within the class of risk-averse Von Neumann - Morgenstern (VNM) preferences under risk. This corresponds to the weak form of equality of opportunity considered in LPT. Overall, this three-tier taxonomy allows for a richer, and least partial, ranking of social states, that could be used for policy evaluation. However, the model of LPT would not allow to rank situations where the weak form of equality of opportunity is satisfied (or violated) both before and after the policy intervention.

Our analysis relies on the model of LPT and develops a definition of opportunity equalization that combines two distinct criteria.

The first criterion is an ordinal criterion that elaborates on the notion of weak equality of opportunity defined in LPT. Instead of restricting attention to the class of risk-averse preferences under risk, we consider a more general approach and assume that individual preferences over lotteries belong to a general class of preferences. For instance, we might consider the class of rank-dependent utility functions or the VNM representation. We further assume that this general class can be partitioned into nested sub-classes, according to the series of restrictions imposed on preferences. When the outcome distributions are not independent of circumstances, it might be possible to find a sub-class of individual preferences within which all preferences consistently rank the type-specific outcome distributions. This should be seen as a case of inequality of opportunity as all preferences in that sub-class consistently agree on a ranking of circumstances. Our first criterion for opportunity equalization is that the class of preferences within which it is possible to unanimously rank circumstances should shrink as a result of the policy implementation. In a sense, this amounts to requiring that the degree of consensus on the ranking of types should fall after the implementation of the policy.

Beyond this ordinal criterion, we impose a second criterion for opportunity equalization. This second criterion is a distance criterion. It essentially requires that the cardinal advantage conferred to the "privileged" types should fall, according to all preferences in the subclass for which it is possible to unanimously rank circumstances (and inducing a form of inequality of opportunity). This criterion is, in its essence, a cardinal criterion, although the final assessment over distance reduction is more general and robust, since it holds for a general subclass of preferences.

As a result, when both criteria are satisfied, the implementation of the policy results (i) in a decrease in the degree of unanimity for the ranking of circumstances, in terms of the advantage that they confer and, (ii) a fall in the size of advantage of the privileged types according to all preferences in the class within which types can be ranked.

In the rest of the chapter, we formalize these two criteria in a general setting with multiple circumstances, and we discuss other relevant, but less partial, criteria of opportunity equalization that aggregate evaluations over the circumstances. We show how to derive empirically the implementable conditions for the dominance and the distance criteria of equalization. Although the two criteria do not depend on a particular representation of the class of preferences, some restrictions on this class are needed to implement the equalization criterion. We provide explicit identification procedures for the Yaari (1987) rank-dependent model,[1] as well as for the expected utility model. However, only the Yaari type preferences allow to construct testable procedures for the equalization criterion. Within the context of the Yaari model, we first show that the ordinal equalization criterion can be formalized in terms of the order of inverse-stochastic dominance at which it is possible to rank the distributions of outcomes conditional on type. Second, we show that the economic distance criterion amounts to require that the opportunity gap between any pair of types should fall, in the sense of stochastic dominance, as a result of the policy implementation.

We use the equalization of opportunity criterion to evaluate the impact of two simulated educational policies whose objective is to increase the accessibility of students in the educational system: the first policy simulates the impact of increasing the mandatory minimum age for school leaving; the second policy extends accessibility to the higher education system to all students that at least complete the secondary education requirements. We use the French Labor Survey (*Enquête Emploi*) to identify the causal impact of few years spent in secondary education and in higher education on the quantiles of the earnings of the target group. Then, we simulate the two policies by attributing the appropriate quantile treatment effect from policy treatment on the observed earnings distribution (before policy implementation), and then we determine the simulated earning profiles (after policy implementation) that has to be used to evaluate opportunity equalization.

Our analysis is connected to several papers that have recently examined changes over time in inequality of opportunity or differences therein across various national or policy

[1] Note that here the rank-dependent utility model is assumed to characterizes the individual preferences: contrary to Aaberge (2009) and Zoli (2002)

contexts. Ferreira and Gignoux (2011) (for education), Checchi and Peragine (2010) (for income), Peragine et al. (2011) (for growth) and Garcia-Gomez et al. (2012) (for mortality) offer some recent examples. In most cases, however, these papers rely on specific cardinal (and often ad hoc) indices of inequality of opportunity. Our combined criterion of opportunity equalization, in contrast, offer more general conditions. Van de gaer et al. (2011) offer the only example that we are aware of, of a policy evaluation based on the equality of opportunity principle. However, their analysis is more focused on the assessment of opportunity improvement (i.e. to what extent does the opportunity set offered to any type improves as a result of the policy) rather than on the analysis of opportunity equalization (i.e. to what extent does the opportunity gap between types fall as a result of the policy).

The rest of the chapter is organized as follows. We describe in section 3.2 the notation and the tools that we exploit to define weak and strong forms of equality and inequality of opportunity. In section 3.3 we combine the definitions of equality and inequality of opportunity in LPT to construct an equalization criterion. We shows the limitations of the test by resorting to a simple framework with only two circumstances and one effort level. Within this framework, we formalize the dominance (section 3.2.4) and the distance (section 3.3.2) criteria for opportunity equalization, which we combine together to obtain the equalization of opportunity criterion. Then, the test is generalized to the multiple effort, multiple circumstances case (section 3.4). We also provide a definition of the test for the general case. Implementation issues and identification of equalization of opportunity, when the relevant determinants of outcome are only partially observable, are discussed in section 3.5. The results from the empirical applications are illustrated in section 3.6, while section 3.7 concludes.

3.2 Equality of opportunity: notation and definitions

3.2.1 Determinants of outcome

Our analysis builds upon the framework developed in Roemer (1998) and Lefranc et al. (2009). Individual outcome, y is determined by four types of factors. Following the terminology of Roemer, *circumstances*, denoted by an element $c \in C$, capture the factors that are not considered a legitimate source of inequality. *Effort* summarized by a scalar e includes the determinants of outcome that are seen as a legitimate source of inequality. Following LPT we also consider *luck*, captured by a scalar l, that comprises the random factors that are seen as a legitimate source of inequality as long as they affect individual outcomes in a neutral way, given circumstances and effort.[2] Lastly, we consider that individual outcome may be affected by a policy variable, denoted π. In the rest of the paper, the policy variable is dichotomous and takes values in $\{0, 1\}$, thus defining two possible states of the world. These two states of the world may define two policy regimes. More generally, they may correspond to two periods or two countries, that one would like to compare. The analysis can be extended to comparisons involving more than two policy regimes.

Notice that the variable π cannot be interpreted as a circumstance, but rather as a particular state of the world. For instance, let $\pi = 1$ indicates the introduction of a taxation scheme coherent with the compensation principle. This taxation scheme may implement a progressive marginal tax rate which depend on the effort exerted by the individuals. Random factors are, however, still in, and accounting for these factors in evaluating the opportunity equalization impact of the tax system requires to deal with non degenerate conditional distributions associates to pre and post tax incomes.[3]

[2]So far, circumstances, effort and luck have only be defined in a formal sense, i.e. by the way they should be taken into account in equality of opportunity judgements. What precise factors should count as circumstances, effort and luck is yet another question that calls in both ethical and political value judgements, as discussed for instance in Roemer (1998) and LPT. Here we take a neutral stance on the question of what factors should count as *circumstances*, *effort* or *luck*.

[3]It is often implicitly assumed that the policymaker constructs optimal policies knowing the whole set of circumstances, thus leaving aside only effort components as a residual. This may not be always the case. In many instances it is too costly to retrieve information on the complete set of individual circumstances, or even not ethically acceptable. A good example is genetic screening. Genetic profiles, largely predetermined

Following Roemer, define a *type* as the set of individuals with similar circumstances. Following LPT, define a *variety* as the set of individuals with similar circumstances and effort. We say that two varieties are *comparable* when they only differ in the circumstances.

These four sets of factors provide a complete partition of the determinants of individual outcome. Consequently, one may write outcome as:

$$y = Y(c, e, l, \pi),$$

where $Y(.)$ denotes the outcome function. By conditioning the outcome to random factors l, we mean that we focus our attention on conditional *distributions* of outcomes, rather than individual realizations. Hence $Y(.)$ can be thought at the outcome quantile at luck level l.

Alternatively, the outcome of each individual may be seen as a draw from a lottery, whenever there is uncertainty or randomness in the determinants of individual outcomes. We let $F(.)$ denote the cumulative distribution function of the outcome, which is assumed to be left-continuous. In the rest of the paper, our analysis will largely involve the comparison of conditional distribution functions. In particular, we will focus on the distribution of outcome conditional on circumstances, effort and policy: $F(y|c, e, \pi)$. Lastly, we define $F^{-1}(p)$ the outcome quantile distribution associated with F, for all population shares p in [0,1].[4]

in one's life, may provide a large amount of information on individual predisposition toward some types of health diseases. However, genetic screening is costly and invasive of individual privacy. Therefore, the policymaker has to design health policies without using genetic screening as a device for better assessing the spectrum of individual circumstances. If circumstances are defined only on a reduced set of information on the family background of origin, it is still possible to observe a large variability in health outcomes for people exerting identical effort that is driven by neglected circumstances such as genetic factors. We treat these components as random factors.

[4]If the cumulative distribution function is only left continuous, we define F^{-1} by the left continuous inverse distribution of F:

$$F^{-1}(p|c, e, \pi) = \inf\{y \in \mathbb{R}_+ : \ F(y|c, e, \pi) \geq p\}, \quad \text{with } p \in [0, 1].$$

3.2.2 Definitions of equality of opportunity

Let us now review the notions of equality of opportunity that are prevalently discussed. In general terms, equality of opportunity is said to prevail, under a given policy regime, if, given effort, no set of circumstances yields an unambiguous advantage over the others. This reflects what is usually referred in the literature as the compensation principle (e.g. Fleurbaey 2008). As discussed in LPT, this suggests at least two notions of equality of opportunity, which we now review.

Strong equality of opportunity

The first conception of equality of opportunity corresponds to the situation where, given effort, the distribution of outcome does not depend on circumstances. This is a strong notion of equality that requires that individuals face similar distributions of outcome, regardless of their type, once their level of effort is known. This criterion can be formalized by the following definition, adapted from LPT.

Definition 3.1 (EOP-S) *For a given policy π, Strong Equality of Opportunity (EOP-S) is satisfied iff:*
$\forall(c,c')\ \forall e, \quad F(.|c,e,\pi) = F(.|c',e,\pi).$

Of course, this definition can be straightforwardly reformulated using quantile functions by requiring that $F^{-1}(.|c,e,\pi) = F^{-1}(.|c',e,\pi)$ for all varieties.

Weak equality of opportunity

The fact that two types are facing different outcome distributions does not necessarily imply that one is advantaged over the other, in terms of outcome. Furthermore, if it is not possible to unambiguously rank circumstances according to the advantage they confer, it may be argued that a weak form of equality of opportunity prevails. LPT introduce the view that a type c can only be said to confer an unambiguous advantage over another type c' if there is agreement within a given set of preferences in assigning a higher utility to the lottery

$F(.|c,e,\pi)$ with respect to the lottery $F(.|c',e,\pi)$. Therefore, $F(.|c,e,\pi)$ is associated with an unambiguous *advantage* compared to $F(.|c',e,\pi)$. Of course, the implementation of this criterion requires to specify the admissible set of preferences under risk.

LPT assume that preferences satisfy the Expected Utility Theory representation, and restrict this class to the preferences displaying risk aversion, such that this view of unambiguous advantage implies comparing the outcome distribution of all varieties using the criterion of second-order stochastic dominance.[5] This leads to the following definition of a weaker form of equality of opportunity.

Definition 3.2 (EOP-W) *For a given policy π, Weak Equality of Opportunity (EOP-W) is satisfied iff:*

(a) $\forall c \neq c'\ \forall e, \quad F(.|c,e,\pi) \not\succ_{SD2} F(.|c',e,\pi)$ where $\succ_{SD2}$ denotes stochastic dominance at the order 2.

and

(b) EOP-S is not satisfied.

This definition differs slightly from the one in LPT. Since equality of all distribution function is a special form of non-dominance, we have imposed that EOP-S should not be satisfied in order to get two mutually exclusive notions of equality of opportunity.

Again, given the equivalence between direct and inverse second-order stochastic dominance, the above definition can be reformulated in terms of quantile functions by requesting the absence of inverse second-order dominance.[6]

Inequality of opportunity

Lastly, we can define the case of inequality of opportunity as the complement of weak and strong equality of opportunity.

[5]Formally, the definition of first order stochastic dominance ($F(.|c,e,\pi) \succ_{SD1} F(.|c',e,\pi)$) requires that $\forall y \in \mathbb{R}_+\ F(y|c,e,\pi) \leq F(y|c',e,\pi)$ and $\exists y$ for which the inequality is strict. Similarly, second order stochastic dominance ($F(.|c,e,\pi) \succ_{SD2} F(.|c',e,\pi)$) is satisfied iff $\forall y \in \mathbb{R}_+\ \int_0^y F(t|c,e,\pi)dt \leq \int_0^y F(t|c',e,\pi)dt$ and $\exists y$ for which the inequality is strict.

[6]See appendix 3.A.1 for a formal definition of inverse stochastic dominance.

Definition 3.3 (IOP) *For a given policy π, Inequality of Opportunity (IOP) prevails iff:*

(a) EOP-S is not satisfied

and

(b) EOP-W is not satisfied.

IOP corresponds to the case where the outcome distributions for at least two varieties with similar effort can be ordered using the criterion of strict second-order stochastic dominance. Hence all risk-averse agents will unanimously agree that one of these two varieties is advantaged over the other.

By definition, the three situations EOP-S, EOP-W and IOP offer a complete partition of all possible allocations of outcomes.

3.2.3 Using equality of opportunity to evaluate policy changes

Definitions EOP-S, EOP-W and IOP allow to establish if, within a policy regime, equality of opportunity prevails or not. The intuitive implication running from EOP-S to EOP-W allows to order EOP-S as a stronger notion than EOP-W, while IOP is an intermediate status occurring when EOP is rejected, but EOP-W is not granted. This trivial ranking defines a taxonomy for comparisons of different policy regimes.

Setting

Our objective is to evaluate the efficiency of a given policy from the point of view of equality of opportunity: we want to be able to assess whether implementing a given policy improves equality of opportunity over the *status quo*. We refer to such an improvement as an *equalization* of opportunities. This requires comparing the allocation of an outcome across types under both $\pi = 0$ and $\pi = 1$. The rest of the chapter is devoted to defining an equalization criterion.

Before discussing possible criteria, it is worth clarifying the type of comparisons involved in the assessment of changes in equality of opportunity. Assessing equality of opportunity in the static context (for a given allocation of outcomes, within a policy regime) requires to evaluating the difference in the distributions associated to different circumstances evaluated at the same effort level. Moving to the dynamic context (comparison of different allocations, between policy regimes), assessing equalization of opportunities requires a *difference-in-differences* approach in order to examine *changes* in the possibility of ranking the distributions offered to different varieties.

To simplify the framework of these comparisons, we start by considering a restricted setting with only two varieties, with a common effort level e and two distinct circumstances c and c'. Both varieties are observed under the two policy regimes $\pi = 0$ and $\pi = 1$. To simplify notations, we let F_π (resp. F'_π) denote the distribution of outcome for variety (c, e) (resp. (c', e)), under policy regime $\pi = 0, 1$. These distributions are identical to $F(.|c, e, \pi)$ and $F(.|c', e, \pi)$.

Possible configurations

A natural first-step to assess changes in equality of opportunity is to resort to the criteria of LPT and examine which of the three situations EOP-S, EOP-W and IOP prevails, under both $\pi = 0$ and $\pi = 1$.

Possible configurations are summarized in the following table.

Table 3.1: Equality of opportunity configurations under $\pi = 0$ and $\pi = 1$

	$\pi = 1$		
$\pi = 0$	EOP-S	EOP-W	IOP
EOP-S	A	C	C
EOP-W	B	F	E
IOP	B	D	G

Cell A corresponds to the case where the strong form of equality of opportunity prevails both before and after policy implementation: in both cases, the outcome distributions are identical across types. Hence, the policy can be considered as neutral from the point of view of equality of opportunity. Note that this does not mean that the policy has no effect: it may well affect the aggregate level of outcome or the degree of inequality within types.

Case B and its symmetric case C are characterized by changes in the nature of "equality" of opportunity that prevails. At this point, one should remark that the three situations of EOP-S, EOP-W and IOP could reasonably be ranked in terms of how successful they are at securing equality of opportunity, as we discuss in greater details in the next section: EOP-S undeniably represents the highest form of equality of opportunity and IOP represents the worst situation. Here, cases denoted by B start with situations where the two varieties are offered different outcome prospects and end up, under $\pi = 1$ in a situation where outcome prospects are identical. Given the intuitive ranking of possible states, B is an improvement in terms of equality of opportunity. In fact, it corresponds to a full equalization of opportunities. Symmetrically, C corresponds to a deterioration from the perspective of equality of opportunity.

To some extent, cases D and E are similar to cases B and C. EOP-W should be considered as a better situation than IOP, from the perspective of equality of opportunity: under EOP-W, there is no consensus, among all risk-averse agents, as to what type is the privileged one, while on the contrary, under IOP, there is a consensus on which type is the advantaged one. Hence, case D corresponds to a move towards equality of opportunity (although in a weak form), while cells denoted by E move away from it. Again, case D can be seen as an improvement, from the perspective of equality of opportunity, although this time, the equalization is not complete, unlike case B. Symmetrically, E corresponds to a deterioration of equality of opportunity.

Lastly, F and G correspond to situations where, according to the notions of equality of opportunity in LPT, the policy has no effect on the nature of equality of opportunity at work. Does it necessarily imply that the policy has no effect at all, from the point of view of equality of opportunity? The answer is, obviously, no. This is a major difference with case A. In cell A, distributions are identical within both policy regimes (although they are not necessarily identical between policy regimes), so that the two varieties are perfectly equal, before and after. This stands in contrast with case F: EOP-W does not require that the two distributions be identical, but simply that they cannot be ranked according to second-order stochastic dominance. However, among all the pairs of distributions that cannot be ranked according to second-order stochastic dominance, some pairs are likely to lie closer together than others. The same holds true among the pairs of distributions that can be ranked according to SD2. Hence, in cases F and G, it is possible that the degree of inequality varies before and after the policy.

The empirical relevance of cases such as F and G is demonstrated in several instances. For example, the analysis of changes over time in equality of opportunity in France, undertaken in Lefranc et al. (2009) concludes to case G: outcome distributions can almost always be ranked by the SD2 criterion, throughout the period they study, although the authors claim that the degree of dissimilarity of the outcome distribution of the different types falls over time. The same seems to hold true in cross country comparisons (Lefranc et al. 2008).

One limitation of existing characterizations of equality of opportunity is that they provide little guidance regarding the criteria that should be used to assert changes in equality of opportunity, in cases such as F and G. To some extent, this critique extends to studies that attempt to measure the degree of inequality of opportunity using scalar indices (Bourguignon, Ferreira and Walton 2007). Our objective is precisely to offer a formal analysis of possible criteria of opportunity equalization. As we will now discuss, two distinct approaches can be taken. The first one is directly inspired by the definitions of LPT and is based on the order of stochastic dominance at which the distributions of the different varieties can be compared. The second criterion investigates the use of distance measures to assess the degree of inequality of opportunity.

Before moving to equalization criteria, let us formalize a more general model for equality of opportunity, which nests EOP-S, EOP-W and IOP notions.

3.2.4 Equality of opportunity: a generalization

Introductory example

A case such as G might look as a *statu quo* situation, from the perspective of the taxonomy of equality of opportunity underlying table 3.1. However, it should be noted that case G still mixes together situations that differ in terms of the nature of the dominance relationships that prevail. IOP is defined in LPT by the occurrence of second-order dominance. Hence it incorporates both first- and second-order stochastic dominance. The criterion of LPT can thus be refined in case G in order to achieve a least partial criterion for asserting equalization of opportunity, by separating first-order from second-order dominance. This leads to splitting cell G in four sub-cases, as shown in table 3.2.

Table 3.2: Equality of opportunity configurations under $\pi = 0$ and $\pi = 1$ - refinement

	$\pi = 1$	
$\pi = 0$	$\succ_2$	$\succ_1$
$\succ_2$	G1	G3
$\succ_1$	G2	G4

Cases G1 and G4 are two situations where the nature of the dominance relationships at work are similar under $\pi = 0$ and $\pi = 1$. But this does not occur in cases G2 and G3. In case G2, the nature of stochastic dominance relationships changes from first-order dominance under $\pi = 0$ to second-order $\pi = 1$. As discussed below, this can be interpreted as a weakening of the (nature of the) advantage conferred to the dominant type over the dominated type. For this reason, one might argue that case G2 corresponds to a partial equalization of opportunities. Conversely, G3 could be considered as a deterioration of equality of opportunity.

This reasoning can be extended to provide a more complete ranking criterion for case F. We now consider a generalization of the definition of equality of opportunity developed by LPT.

Criterion: equality of opportunity as lack of consensus

The taxonomy of equality of opportunity situations introduced in LPT rests on an ordinal ranking of the income distributions of different types. This ranking is based on the existence of a consensus, within a class of preferences, to evaluate one type as disadvantaged compared to another one. In LPT, this class is the one of preference displaying risk aversion, which is equally identified by testing stochastic dominance at order two.

The discussion in the previous paragraph amounts to distinguish first-order from second-order stochastic dominance in the IOP case, on the basis that the class of preferences displaying risk aversion is included in the class of preferences that are only increasing in outcomes. Unanimity in the latter can be verified by resorting on first order stochastic dominance comparisons. A straightforward generalization of this idea is to consider higher orders of stochastic dominance. However, this extension based on direct stochastic dominance posits that the relevant class of preferences where consensus has to be verified admits the expected utility representation, and that this class is restricted according to a sequence of assumptions on risk behavior. These assumptions are in general sufficient to achieve identification of a criterion for equality and inequality of opportunity that can be applied to the data. They are not necessary, however, because there may exist other classes of preferences and other sequences of restrictions that can be also identified by their corresponding stochastic orders.

For this reason, we develop our arguments by referring to a general class of preferences, denoted by $\mathcal{C}$. For a pair of comparable varieties with distributions $F(.|c,e,\pi)$ and $F(.|c',e,\pi)$ and a given preference $W \in \mathcal{C}$, we write $F(.|c,e,\pi) \succcurlyeq_W F(.|c',e,\pi)$ to say that according to W the variety (c,e) confers an economic advantage compared to (c',e) under policy regime π. The choice of a class $\mathcal{C}$ defines a normative criterion which sets the domain of preferences in which consensus has to be verified.

Given two varieties (c,e) and (c',e), each preference $W \in \mathcal{C}$ produces a ranking of the two varieties. Two possible configurations may emerge. In the first configuration, all preferences within $\mathcal{C}$ will concur with the view that the variety (c,e) is advantaged compared to (c',e), and that consequently equality of opportunity does not prevail between (c',e) and (c,e). This is in fact the strongest case of inequality of opportunity, which is grounded on the consensus of all preferences in the largest class $\mathcal{C}$. The consensus within $\mathcal{C}$ makes sure that the judgement over the direction of the advantage is robust with respect to the choice of the evaluation function.

Assessing inequality of opportunity within $\mathcal{C}$ may become very demanding, according to the choice of the class of preferences. However, it always occurs in one of the following two situations. When LPT's EOP-S holds, every preference ranks necessarily (c, e) at least as good as (c', e) and (c', e) at least as good as (c, e). If $\mathcal{C}$ is equipped with the transitivity property, one has that EOP-S implies that all preferences within $\mathcal{C}$ will concur with the view that no variety is advantaged or disadvantaged compared to the others. Our notion of strong equality opportunity, denoted simply as EOP, coincides with a situation where there is agreement in $\mathcal{C}$ on the fact that the two comparable varieties are indifferent. The LPT's EOP-S is sufficient (but not always necessary for all possible choices of $\mathcal{C}$) for EOP.

The second possible case occurs when there is *lack of agreement* among preferences in C in assessing which variety among (c, e) and (c', e) is the advantaged one. However, it might be possible to achieve unanimity in ranking the two varieties by placing further restrictions on the preferences used to compare such varieties. One may, for instance, introduce enough restrictions that identify one preference relation in $\mathcal{C}$. However, the achieved ranking will lack robustness. Alternatively, one may define a minimal sequence of restrictions on the preferences in $\mathcal{C}$ with a normative appealing interpretation, and isolate a subclass of preferences, call it $\mathcal{C}'$ in $\mathcal{C}$ within which (c, e) is unanimously preferred to (c', e).

If this class exists, then a form of inequality of opportunity is verified, since there is consensus among all preferences in $\mathcal{C}' \subseteq \mathcal{C}$ in ranking the pair (c, e) as providing more advantage than the pair (c', e). Motivated by a similar argument as in LPT, we define a new notion of weak equality of opportunity that is verified for all the sets of preferences that include $\mathcal{C}'$ but that are included in $\mathcal{C}$. As a consequence, for a given subclass of preferences in $\mathcal{C}$ given by a particular sequence of restrictions, it is always possible to assess if EOP, IOP or a weak form of equality of opportunity is satisfied. This taxonomy can be used even when introducing an infinite amount of restrictions on $\mathcal{C}$.

We formalize this criterion of equality of opportunity and show that it extends the LPT definitions.

Order-k equality and inequality of opportunity

For a given set $\mathcal{C}$, we use the scalar k to indicate a sequence of *restrictions* on $\mathcal{C}$, and with $\mathcal{C}^k$ the subset of preferences within $\mathcal{C}$ identified by these restrictions. We obtain the second key normative ingredient of our model: a sequence of k restrictions on $\mathcal{C}$ introduces precise restrictions on the behavior of preferences *vis-à-vis* the risk underlying distributions conditional on type, and at the same time it identifies subsets of $\mathcal{C}$ that are included one within the other. In fact, $l \geq k$ if and only if $\mathcal{C}^l \subseteq \mathcal{C}^k \subseteq \mathcal{C}$. The sequence of restrictions embeds a normative choice, and it should not, of course, depend on the data.[7]

Let define $\kappa(c, c', e, \pi)$ the minimal sequence of the restrictions that allows to identify a class of preferences $\mathcal{C}^{\kappa(c,c',e,\pi)} \in \mathcal{C}$ for which F_π and F'_π can be ranked. For short, since in this section we only consider two possible circumstances, c and c' and one effort level e, and concentrate on policy change, we let κ_π denote $\kappa(c, c', e, \pi)$. In the case where one distribution dominates the other, we let, by convention, c denote the dominant circumstance in the class $\mathcal{C}^{\kappa_\pi}$. As a consequence, c will also be the dominant circumstance at all orders greater than κ_π. Hence, we have, by definition for all $k \geq \kappa_\pi$: $F_\pi \succcurlyeq_W F'_\pi$ for all $W \in \mathcal{C}^k$.

The order of restrictions κ_π is well defined, apart from the very particular case in which, even for $k = \infty$, there might not be unanimity in ranking F_π and F'_π within $\mathcal{C}^\infty$ (see for instance Fishburn 1976, for the expected utility case). In this case, there is no inequality of opportunity, and therefore weak equality of opportunity holds for all subclasses of preferences within $\mathcal{C}$.

The taxonomy of LPT can be extended on the basis of the minimal sequence of restrictions on $\mathcal{C}$ allowing to retrieve unanimity in assessments. This leads to define two notions of order-k inequality of opportunity and weak equality of opportunity.

Definition 3.4 (IOP-k) *For a policy π, a preferences class $\mathcal{C}$ and for a sequence of nested sub-sets $\{\mathcal{C}^k\}_{k=1}^{\infty}$, order-k Inequality of Opportunity (IOP-k) prevails between two varieties*

[7] For instance, consider the set of rank dependent preferences $\mathcal{R}$ and the set of preferences satisfying the expected utility model, denoted by $\mathcal{U}$. The set of preferences increasing in the outcome is denoted with $k = 1$, while the set of preferences displaying risk aversion, a particular subset of the first group, is denoted by $k = 2$, since the latter class is identified for at least *two* restrictions on the initial class. Hence $\mathcal{U}^1 \subseteq \mathcal{U}^2$ and $\mathcal{R}^1 \subseteq \mathcal{R}^2$.

(c,e) and (c',e) with outcome distributions F_π *and* F'_π *iff:*
$F_\pi \not\succcurlyeq_W F'_\pi$ *or* $F'_\pi \not\succcurlyeq_W F_\pi$ *for all* $W \in \mathcal{C}^k$.

Definition 3.5 (EOP-W*k*) *For a policy* π*, a preferences class* $\mathcal{C}$ *and for a sequence of nested sub-sets* $\{\mathcal{C}^k\}_{k=1}^{\infty}$*, order-k Weak Equality of Opportunity (EOP-Wk) prevails between two varieties (c,e) and (c',e) with outcome distributions* F_π *and* F'_π *iff:*
$\forall l \leq k :$ $F_\pi \succcurlyeq_W F'_\pi$ *or* $F'_\pi \succcurlyeq_W F_\pi$ *cannot be established for all* $W \in \mathcal{C}^l$.

The relationship between inequality and equality of opportunity at different orders follows from the fact that the restrictions imposed on the class of preferences $\mathcal{C}$ are sequential, and therefore they identify inclusion of its subclasses. It is straightforward to establish the following proposition:

Proposition 3.1 *For all* $l > k$ *: IOP-k* $\Rightarrow$ *IOP-l and EOP-Wl* $\Rightarrow$ *EOP-Wk.*

Definitions 3.4 and 3.5, together with the definition of κ_π, imply that the pair (F_π, F'_π) satisfies IOP-k, for all $\mathcal{C}^k \subseteq \mathcal{C}^{\kappa_\pi}$ $(k \geq \kappa_\pi)$ and satisfies EOP-Wk for all $\mathcal{C}^k \supseteq \mathcal{C}^{\kappa_\pi}$ $(k < \kappa_\pi)$.

The link between these definitions and definitions 3.1, 3.2 and 3.3 is obvious. Let the preferences in $\mathcal{C}$ admit the expected utility representation, then LPT's notion of IOP corresponds to the IOP-k for $k = 2$, which indicates risk averse preferences; EOP-Wk for $k = 2$ gathers LPT's notions of EOP-W and EOP-S. It is also worth stressing that LPT's notion of EOP-W gathers both IOP-k, for $k \geq 3$, and EOP-Wk for $k \geq 2$. This makes it clear that EOP-W is an intermediate situation, the labeling of which is somehow misleading: to some extent, EOP-W, in the definition of LPT, could also be seen as weak form of inequality of opportunity.

3.3 Equalization of opportunity: a simplified setting

3.3.1 The dominance criterion for equalization of opportunity

Criterion

The definitions of EOP-Wk and IOP-k allow for a refinement, in a more general context, of the partition of the configurations in table 3.1 that may occur when moving from $\pi = 0$ to $\pi = 1$. This partition is based on the pair (κ_0, κ_1), which summarizes all the relevant information in the perspective of a dominance approach to equality of opportunity. In what follows, we assume that the sequence of restrictions is well specified, so that for any value of the parameter k, there is only one subset $\mathcal{C}^k$ of $\mathcal{C}$ that is identified by these restrictions.

By definition of κ_π, under policy π, IOP-k is satisfied for all $k \geq \kappa_\pi$ and EOP-Wk is satisfied for all $k < \kappa_\pi$. When κ_1 is greater than κ_0, moving from $\pi = 0$ to $\pi = 1$ leads to satisfy a more stringent form of (weak) equality of opportunity, as established by Proposition 3.1. This leads us to define the following *ordinal* criterion of partial equalization of opportunities, based on the order of dominance:

Criterion 3.1 (Dominance-order criterion of opportunity equalization - O-ezOP) *Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities between varieties (c, e) and (c', e) according to the dominance-order criterion iff $\kappa(c, c', e, 1) > \kappa(c, c', e, 0)$.*

Interpretation

To understand the foundation for dominance-order criterion of opportunity equalization, one first needs to analyze the content of definitions 3.4 and 3.5. Suppose that $\mathcal{C}$ represents the domain of admissible preferences of a given population. Conceptions IOP-k and EOP-Wk define intermediate cases between two polar situations. The first polar case is the EOP situation. In this situation, every agent, regardless of her preferences will be indifferent between the two varieties (c, e) and (c', e). The second polar case is the IOP-1 situation. In this situation, since there is agreement among preferences in the largest class $\mathcal{C}$ on saying that the variety (c, e) is strictly advantaged compared to (c', e), then we are also sure that every agent will prefer the first variety compared to the second regardless of her preferences, as soon as these preferences are increasing in outcome. Hence, there is unanimity across all agents in the evaluation of EOP and of IOP-1, regardless of their preferences.

On the contrary, agents' judgement on all intermediate configurations between EOP-S and IOP-1 will never be unanimous and will be contingent on their preferences. Of course, in this broad set of intermediate cases, all configurations are not identical. Some may lie closer to one of the two polar cases than others. And it is crucial to be able to differentiate among these intermediate cases. One way to do this is by considering the restrictions that need to be placed on individual preferences for agents to concur with the view that inequality of opportunity does prevail or not.

It seems therefore natural to use the notions of EOP-Wk to rank different policy regimes. Suppose that for varieties (c, e) and (c', e) one has that $\kappa_1 > \kappa_0$, then the class of preferences within which (c, e) is unanimously preferred to (c, e) under $\pi = 1$ is a strict subset of the class of preferences within which (c, e) is unanimously preferred to (c', e) under $\pi = 0$: $\mathcal{C}^{\kappa_1} \subset \mathcal{C}^{\kappa_0}$. In this case, all preferences according to which equality of opportunity is violated under $\pi = 1$ will also concur with the view that equality of opportunity is violated under $\pi = 0$. But the reverse is not true. For some preferences, equality of opportunity prevails under $\pi = 1$ but not under $\pi = 0$. This leads to conclude that a more encompassing form of inequality of opportunity prevails under $\pi = 0$ as compared to $\pi = 1$, or equivalently that a less weak form of equality of opportunity is satisfied under $\pi = 1$.

A case where $\kappa_1 > \kappa_0$ corresponds to a situation where one need to put more restrictions on individual preferences under $\pi = 1$ than under $\pi = 0$ to be able to reach the conclusion that inequality of opportunity prevails. This point can be easily established by noticing the following relationships between IOP-k and EOP-Wk:

$$\text{IOP-}k \;\Leftrightarrow\; \begin{array}{c} F \succeq_W F' \\ \forall W \in \mathcal{C}^k \end{array}; \qquad \text{EOP-W}k = \bigcap_{l=1}^{k} \cancel{\text{IOP-}l}; \qquad \text{EOP} = \cancel{\text{EOP-W}k} \cap \bigcap_{l=1}^{\infty} \cancel{\text{IOP-}l}.$$

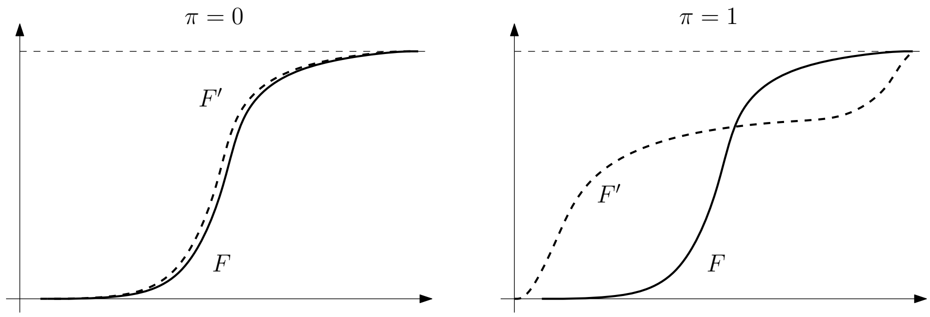


Figure 3.1: Distance comparisons between two types, before and after policy implementation

This leads to conclude, on the base of the ordinal criterion, that moving from $\pi = 0$ to $\pi = 1$ improves equality of opportunity.

Of course this approach can rest on any class of individual preferences and partitioning sequences thereof. We will discuss in section 3.5 the identification of the ordinal criterion in the case of the *rank dependent* or the *expected utility* model for preferences, and the most logical sequences of restrictions on these sets of preferences that are commonly used in literature to describe choice behavior under risk.

3.3.2 The distance criterion for equalization of opportunity

The ordinal criterion offers only a partial criterion for assessing opportunity equalization. In particular, it is unable to assess opportunity equalization or disequalization when the degree of dominance κ remains the same before and after the implementation of the policy. In such situations, the class of preferences according to which (c, e) is advantaged over (c', e) is not affected by the policy. However, it might nevertheless be the case that the *distance* between the distributions of the two varieties decreases as a result of the implementation of the policy. Provided that all agents agree with this view, we would like, in this case, to conclude that opportunities have been equalized by the policy. We now formalize this criterion and we show that it can be also motivated by robustness concerns also for cases in which O-ezOP is satisfied but equalization seems debatable. An example is shown in figure 3.1.

Distance measures

To assess the distance between the distributions attached to two different varieties, one can use the notions of economic distance between two distributions developed in particular by Shorrocks (1982) and Chakravarty and Dutta (1987). These contributions suggest to characterize each distribution by its certainty equivalent and to measure the economic distance between distributions by the gap between their certainty equivalents.

Let $W(F)$ denote the expected utility derived from a distribution with cdf F, where $W \in \mathcal{C}$ defines individual preferences for risk. Let $D(y)$ define the *certain distribution* in which each percentile receives income y. For preferences under risk $W(.)$, we define $CE_W(F)$, the certainty equivalent of distribution F. It is implicitly defined by:

$$W\left(D(CE_W(F))\right) \;=\; W(F).$$

For a pair of distributions F and F', Chakravarty and Dutta (1987) define $\Delta_W(F, F')$ the distance between these two distributions as:

$$\Delta_W(F, F') \;:=\; |CE_W(F) \;-\; CE_W(F')|.$$

When the two distributions are equal, their distance is obviously zero. Otherwise, the measure of distance depends upon the degree of dissimilarity of the two distributions but also on the individual preferences under risk, as captured by W.

Criterion

Opportunity equalization can be assessed by comparing the distance between the outcome distributions of the two varieties before and after the implementation of the policy. If it is the case that $\Delta_W(F_0, F_0') > \Delta_W(F_1, F_1')$, it can be argued that the policy implementation has equalized opportunities between the two varieties.

Of course, we would like this judgement to be robust to the utility function used to evaluate the opportunities. A very strong case for opportunity equalization would be if all utility functions under risk agree with the view that the distance between the two distributions has fallen after the implementation of the policy. A less general requirement is that all utility functions in a particular sub-set of $\mathcal{C}$ conclude that the distance has fallen. This is summarized by the following *distance* criterion.

Criterion 3.2 (Distance criterion of opportunity equalization - D-ezOP) *Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunity between varieties (c, e) and (c', e) according to the distance criterion on the set of preferences $\widetilde{\mathcal{C}}$ iff*
$\forall W \in \widetilde{\mathcal{C}}: \; \Delta_W(F_0, F_0') > \Delta_W(F_1, F_1')$ *or at most equal.*

3.3.3 Equalization of opportunity criterion

The two criteria defined above emphasize two distinct facets of opportunity inequality: the first one is the degree of unanimity with the assessment of the advantage enjoyed by one variety over another variety; the second one is the measure of the extent of the advantage enjoyed by one variety over another variety. In this section, we provide an encompassing definition of opportunity equalization that combines both criteria. We first do so within the restricted setting where we compare only two varieties, but we also extend this definition to the general case where we have multiple circumstances and multiple effort levels.

Definition

Definition 3.6 (ezOP for two varieties) *For a given set of preferences $\mathcal{C}$ and a sequence of restrictions on that set, moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities between varieties (c, e) and (c', e) iff criteria O-ezOP is satisfied and D-ezOP is satisified on the set of preferences $\mathcal{C}^{\kappa(c,c',e,1)}$.*

Discussion

In the ordinal setting, inequality of opportunity is considered to decrease if the class of preferences according to which it is possible to rank circumstances unanimously becomes smaller. In the cardinal setting, inequality of opportunity decreases if the extent of the advantage enjoyed by one variety diminishes.

In principle, these two criteria might contradict each other: for instance the dominance-order might be satisfied and at the same time the distance order might be violated or indicate a deterioration of equality of opportunity, as in figure 3.1. The above definition requires that both criteria must be satisfied. More precisely, it requires that the class of preferences according to which circumstances can be unanimously ranked becomes smaller and that within this class preferences agree that the extent of the advantage of the dominant circumstances has fallen.

We can find at least three arguments for motivating the ezOP criterion. The first motivation is given by robustness of the comparisons. Suppose that EOP has to be verified by using income as the relevant outcome. Income distributions are defined un a support $[a, b]$. Without the policy, the income distribution of circumstance c', F'_0, takes value $(a + b)/2 = \mu$ with probability one, while the income distribution of c, F_0, gives income $\mu + \epsilon$ for sure, with ϵ small compared to $b - a$. There is no within type inequality. After policy implementation, the distributions are $F_1 = F_0$, while now F'_1 gives income equal to a or b with probability one half. Now there is inequality only among type c.

Before policy implementation, every class of preferences $\mathcal{C}$ increasing in income will agree that type c is advantaged with respect to type c'. This is a case of IOP-1. However, not all preferences will agree on assigning the advantage after that the policy is implemented. Suppose that there exists a sequence of restrictions such that one can verify IOP-κ_1 for $\mathcal{C}^{\kappa_1} \subseteq \mathcal{C}$. Hence O-ezOP is satisfied. However, it can be the case that every preference agrees that the advantage conferred to c is small (at the limit is null as ϵ goes to zero) compared to the advantage that one can measure after policy implementation, a clear violation of the D-ezOP criterion. Robustness with respect to these critics is achieved by demanding that both ordinal and cardinal criteria are satisfied at the same time.

This point leads to our second motivation, which has to do with the class in which the D-ezOP criterion has to be verified. The distance evaluation is based on agreement on the reduction of the advantage of c over c'. Suppose that in the previous example one considers agreement over the class of preferences $\mathcal{C}$ that also agree on assessing the advantaged circumstance. If agreement is found, then there is also agreement on the reduction of distance within $\mathcal{C}^{\kappa_1}$, while if agreement cannot be verified, nothing guarantees that agreement cannot be found within $\mathcal{C}^{\kappa_1}$, given that $\mathcal{C}^{\kappa_1} \subseteq \mathcal{C}$. Thus, the risk is that the ordinal criterion is rejected on the ground of the set of preferences that do not agree on determining the advantaged circumstance after policy implementation.

On the other hand, if unanimity is not reached within $\mathcal{C}^{\kappa_1}$, then it cannot be reached in $\mathcal{C}$, which guarantees the robustness of the assessment over economic distance changes.

A third naive explanation is empirically grounded: as explained in section 3.5, it is not possible to retrieve empirically testable conditions for the distance criterion if the ordinal criterion is not satisfied and if the distance is not tested within $\mathcal{C}^{\kappa_1}$.

3.4 Equalization of opportunity: the general case

The two criteria defined in the previous section emphasize two distinct facets of opportunity inequality: the first one is the degree of unanimity within a class of preferences in assessing the *existence* of an advantage enjoyed by one variety over another, comparable, variety; the second one concerns the *extent* of this advantage.

We now examine the case where there are multiple circumstances and multiple effort levels. For each pair of circumstances, Definition 3.6 allows to conclude whether opportunities have been equalized within a specific pair. The overall judgment rests, however, on an aggregation of the judgements over all pairs of circumstances and effort level.

We take an ex post perspective in this section (Fleurbaey 2008), by requiring that aggregation of judgements across varieties should be made within each effort level separately. We do not investigate possible aggregations across effort levels. Hence, all the comparisons performed in this section, and the type of aggregation procedures that we study, should be understood as conditional on a given effort level.

Different mechanisms of aggregation are proposed. The first mechanism, presented in section 3.4.1, builds on a *Paretian* perspective: consensus on the existence of disadvantage and on its extent must follow, by effect of policy intervention, for all possible pairs of circumstances. This stringent perspective for equalization builds on comparisons of pairs of distributions that may not be appropriate. This critics is motivated by the fact that the equalization criteria should look at the changes in consensus over the existence and extent of disadvantage, and not at the labeling of the circumstances that have to be compared. We builds on this principle to construct the very weak ordinal criterion presented in section 3.4.2. This criterion generalizes, and puts in a policy evaluation perspective, the definition of equality of opportunity based on the set of non-dominated circumstances introduced by Lefranc et al. (2008). However, this criterion does not permit to assess the relevant comparisons in terms of extent of the disadvantage. A solution is proposed in section 3.4.3, where anonymity is fulfilled by replacing the label of the circumstances with their position in the ranking of disadvantage, and distance comparisons are constructed by fixing the position of circumstances in this ranking. This criterion is sufficient for the ranking produced by the *Gini Opportunity* index in Lefranc et al. (2008).

3.4.1 Extending the ezOP criterion to multiple circumstances and effort

Equalization of opportunity rests on judgements over pairwise comparisons of circumstances. There might be several ways to combine these judgements. For instance, one may be willing to trade off a mild disequalization of opportunities for some pair of circumstances against a strong equalization for some other. Equality of opportunity indices implicitly undertake this form of aggregation. The grounds for such an aggregation should probably be clarified. If we rule it out, it is reasonable that a situation of overall opportunity equalization be a situation where opportunities are equalized for all pair-wise comparisons. The general version of the ezOP criterion is formalized in the following definition:

Definition 3.7 (ezOP) *Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities if ezOP is satisfied for all varieties (c, e) and (c', e) with $c, c' \in C$ and $e \in E$.*

The ezOP criterion is based on a completely disaggregated assessment ranging on all the possible pairwise comparisons. This criterion has two potential pitfalls. First, the ezOP criterion is extremely demanding, and therefore likely to be rejected by the data. Second, ezOP requires to perform some comparisons that may not be relevant in the logic of advantage/disadvantage. In fact, the ezOP criterion relies considerably on the *status quo*, i.e. the initial situation. In the following, we develop weaker criteria of opportunity equalization that accommodate these critics.

One possibility to eliminate the impact of the status quo is to exploit a form of *anonymity* with respect to the label of the circumstances. We propose two solutions: the first alternative builds on a notion of non-dominance and introduces a sequential mechanism of elimination of dominated circumstances for a *fixed class of preferences*; the second alternative builds on dominance to construct the ranking of circumstances (by ranking their associated distributions conditional on the same effort level), and proposes to compare changes in the extent of disadvantage for a *fixed rank* in the scale of circumstances.

Before proceeding, we introduce some useful notation. For a given effort e (omitted), define the $|C| \times |C|$ matrix[8] $\mathbf{K}(\pi) = \{\kappa_{ij}(\pi)\}$, where i, j are circumstances and

$$\kappa_{ij}(\pi) := \kappa(i, j, e, \pi) \cdot \text{sign}(i, j, \pi),$$

where $\text{sign}(i, j, \pi)$ is equal to $+1$ if and only if there is agreement over $\mathcal{C}^{\kappa(i,j,e,\pi)}$ on the fact that circumstance i provides an advantage over circumstance j, for policy π (and at effort level e), and equal to -1 if i and j are permuted. So $\kappa_{ij} = -2$ if and only if circumstance i is dominated by j in the class $\mathcal{C}^2$. Additional notation will be introduced when needed. Note that $\mathbf{K}(\pi)$ is obtained by constructing all possible pairwise comparisons of circumstance within each policy regime, and it should, in general, depend on effort.

Using this notation, the O-ezOP criterion can be reformulated by requiring that:

$$|\mathbf{K}(0)| \ll |\mathbf{K}(1)| \quad \forall e \in E.$$

[8] We use $|C|$ to indicate the cardinality of the set C. Alternatively, $|c|$ is the module of a scalar or matrix c. Where necessary, we make the difference explicit.

Finally, we use $c \geq_k c'$ as a shorthand notation for $W(F(.|c,e,\pi)) \geq W(F(.|c,e,\pi))$ for all $W \in \mathcal{C}^k$.

3.4.2 Generalization: a non-dominance perspective

The first type of aggregation criteria relies on the notion of non-dominance, or lack of consensus, as does the ezOP criterion. However, not all the comparisons that are necessary to verify ezOP are also relevant for the perspective of non-dominance. If a given class of preferences agrees that one circumstance i confers a disadvantage compared to another circumstance j, then i does not need to be further compared with the other circumstances, since the disadvantage associated to i has already been settled. We say, in brief, that circumstance i can be *eliminated* from the set of circumstances that need to be compared to assess equalization, also with respect to further restrictions on the class of preferences.

The elimination of circumstance i can be motivated by iterating the following conceptual exercise. Consider an hypothetical situation in which a decision maker, who does not know her tastes but only knows that the relevant class of preferences is $\mathcal{C}$, has to choose its own circumstance (given effort). This is done by selecting an outcome distribution. If there is disagreement among preferences on which is the best distribution, the decision maker may exploit some type of aggregation methods to select. However, if all preferences in the class $\mathcal{C}^k$ agree that the circumstance i provides a disadvantage compared to some other circumstance j, then the rational decision maker will not select i, which can be then eliminated from the bundle of available circumstances. In fact, independently on the ranking of i compared to the remaining circumstances, the rational decision maker will always be better off by selecting j instead of i.

As a result, the disadvantage of i with respect to j is also set for all the classes $\mathcal{C}^{k'} \subseteq \mathcal{C}^k$, and therefore it is always possible to find another circumstance that provides unambiguously an advantage over i. Reiterating this process as the class of admissible preferences shrinks, other circumstances can be eliminated because dominated, and therefore they will never be chosen by a rational agent from the pool of circumstances.[9]

[9]Note that this procedure based on $\kappa_i(\pi)$ restrictions is necessary and sufficient to ensure that one

This exercise can be formalized into a new ordinal criterion for opportunity equalization. Define:

$$\kappa_i(\pi) := \min_j\{|\kappa_{ij}(\pi)| \;\; s.t. \text{ sign}(i, j, \pi) = -1\},$$

which indicates the largest class of evaluation functions where there is agreement in saying that circumstance i is disadvantaged compared to at least one other circumstance. The $\kappa_i(\pi)$ may not be always defined. When two distributions coincide, we assume that $\kappa_i(\pi) = \infty$. When $\kappa_i(\pi)$ is missing, it is because i is the best circumstance in the ranking.

We use the values of $\kappa_i(\pi)$ associated to each circumstance to define a new weak ordinal criterion for opportunity equalization, which is formally presented in section 3.4.2. There are, however, many alternative procedure that allow to combine the information provided by $\kappa_i(\pi)$. We construct our argument in favor of a specific criterion by showing with some counterexamples that alternative, and more partial criteria do not satisfy one basic requirement, that is anonymity with respect to the label of the circumstances.

Anonymity requires that the relabeling of circumstances across dominance comparisons should not have an effect on the equalization assessment, while only the class of preferences within which dominance is verified should matter. Anonymity also requires a form of independence with respect to the number of comparisons that allow to eliminate a dominated circumstance. It is in fact sufficient that one given circumstance is disadvantaged compared to another in order to eliminate that circumstance from further comparisons. Let us start by illustrating the more partial criteria.

Criteria based on iterative elimination of dominated circumstances

A natural criterion for opportunity equalization in this setting consists in requiring that the effect of policy intervention is to shrink the class of preferences that agree in assessing the disadvantage procured by circumstance i compared to (at least) another circumstance. We

given circumstance is always compared with circumstances that survive the iterative elimination procedure. In fact, suppose that $j \geq_k i$ and that $i \geq_{\kappa_\ell(.)} \ell$ for $\kappa_\ell(.) > k$. Since i is eliminated because it provides an unambiguous disadvantage with respect to circumstance j for all preferences in $\mathcal{C}^k$, then it would be incorrect to compare i with ℓ for preferences in $\mathcal{C}^{\kappa_\ell(.)} \subset \mathcal{C}^k$. However, by transitivity $j \geq_{\kappa_\ell(.)} \ell$, and therefore ℓ is eliminated for the same class identified by the sequence of restrictions $\kappa_\ell(.)$.

call this the O-ezOP2 criterion of equalization of opportunity, which is satisfied if and only if $\forall i \ \kappa_i(0) \leq \kappa_i(1)$.

The O-ezOP2 criterion refines the O-ezOP criterion by introducing a procedure of iterative elimination of dominated circumstances. It allows to properly deal with cases in which the equalization fails because of comparisons that require to introduce a large amount of restrictions on $\mathcal{C}$, as shown in the following example.

Example 3.1 Consider the following configuration for $C = \{i, j, \ell\}$

$\pi = 0$:	$i \ \geq_2 \ j$;	$i \ \geq_2 \ \ell$;	$j \ \geq_{100} \ \ell$
$\pi = 1$:	$i \ \geq_3 \ j$;	$i \ \geq_3 \ \ell$;	$j \ \geq_{99} \ \ell$

O-ezOP is not satisfied. However $\kappa_j(0) = \kappa_\ell(0) = 2 < 3 = \kappa_j(1) = \kappa_\ell(1)$. Hence O-ezOP2 is satisfied.

In the example, i is the dominating circumstance. The comparison between j and ℓ in $\mathcal{C}^{100}$ is pointless, because the two circumstances are dominated by i according to all the preferences in $\mathcal{C}^3$. However, the O-ezOP2 is not exempt from critics. In particular, it still preserves the status quo, because the criterion relies extensively on the fact that one circumstance that was dominated before policy implementation has to remain a dominated circumstance even after policy implementation. This is well illustrated by inverting the orders of dominance in the previous example:

Example 3.2 (Continued) Consider the following configuration for $C = \{i, j, \ell\}$

$\pi = 0$:	$i \ \geq_2 \ j$;	$i \ \geq_2 \ \ell$;	$j \ \geq_{100} \ \ell$
$\pi = 1$:	$j \ \geq_3 \ i$;	$\ell \ \geq_3 \ i$;	$j \ \geq_{99} \ \ell$

O-ezOP and O-ezOP2 are not satisfied because $\kappa_j(0) = \kappa_\ell(0) = 2$ while j is the dominance circumstance under $\pi = 1$.

A way to restrict further the O-ezOP criterion consists in introducing some anonymity requirements on the ranking of circumstances. We do so by replacing the name of the circumstance with her *position* in a ranking, which is determined by the number of restrictions that one has to impose on a given class $\mathcal{C}$ to identify the i-th circumstance that has to be excluded. We call this criterion O-ezOP3. Let $\widehat{\kappa}_i(\pi)$ be the ordered distribution of elements $\kappa_i(\pi)$ such that $\widehat{\kappa}_i(\pi) \leq \widehat{\kappa}_j(\pi)$ if and only if $i < j$, for all i.

The O-ezOP3 criterion is satisfied if and only if for all i: $\widehat{\kappa}_i(0) \leq \widehat{\kappa}_i(1)$, for each effort levels. In the Example 3.2 O-ezOP3 is satisfied, because $\widehat{\kappa}_1(0) = \widehat{\kappa}_2(0) = 2 < 3 = \widehat{\kappa}_1(1)$ and $\widehat{\kappa}_2(1) = 99$. Before proceeding to define potential flawless of the criteria O-ezOP2 and O-ezOP3, let us reformulate O-ezOP2 and O-ezOP3 in a more conventional notation introduced by Lefranc et al. (2008), based on comparisons of sets of non dominated circumstances.

The set of non-dominated circumstances

The *set of non-dominated circumstances* $C_{ND}(\mathcal{C}^k, \pi)$ gathers all the circumstances (evaluated at the same effort level) under policy π that are not dominated by another circumstance, according to all preferences in $\mathcal{C}^k$. That is:

$$C_{ND}(\mathcal{C}^k, \pi) \ := \ \left\{ c \in C \ : \ \nexists c' \in C \text{ s.t. } W(F(.|c, e, \pi)) \leq W(F(.|c', e, \pi)) \ \forall W \in \mathcal{C}^k \right\}.$$

Obviously, circumstance j provides a disadvantage compared to some other circumstances if and only if $j \not\in C_{ND}(\mathcal{C}^k, \pi)$. We can use the set of non-dominated circumstances to characterize the O-ezOP2 and O-ezOP3 criteria exposed above. A formal proof of this characterization is not given here.

The O-ezOP2 criterion is not anonymous, in the sense that it requires to verify that if a given circumstance is not dominated under $\pi = 0$, then the *same* circumstance should not be dominated under $\pi = 1$. This is modeled by an inclusion:

$$C_{ND}(\mathcal{C}^k, 0) \subseteq C_{ND}(\mathcal{C}^k, 1), \quad \forall\ k \in \mathbb{N}_+.$$

The O-ezOP3 allows to focus only on the order of deletion of circumstances and not on their name. Hence anonymity of circumstances is in part achieved and the criterion can be equivalently represented by comparing the cardinality of the sets of non-dominated circumstances:

$$\left|C_{ND}(\mathcal{C}^k, 0)\right| \leq \left|C_{ND}(\mathcal{C}^k, 1)\right|, \quad \forall\, k \in \mathbb{N}_+.$$

To contextualize these comparisons, note that in the Example 3.2 one has that $C_{ND}(\mathcal{C}^2, 0) = C_{ND}(\mathcal{C}^3, 0) = \{i\}$ while $C_{ND}(\mathcal{C}^2, 1) = C$ and $C_{ND}(\mathcal{C}^3, 1) = \{j, \ell\}$. As shown above, $C_{ND}(\mathcal{C}^3, 0) \not\subseteq C_{ND}(\mathcal{C}^3, 1)$ but $|C_{ND}(\mathcal{C}^3, 0)| \leq |C_{ND}(\mathcal{C}^3, 1)|$. However, very simple counterexamples show that even ezOP3 may not be exempted from potential pitfalls. We provide a solution by introducing a very weak ordinal criterion for opportunity equalization.

The Weak O-ezOP criterion

With the O-ezOP3 criterion we have focused attention on a counting approach for the non-dominated circumstances. Only their number should matter. However, by accepting this method one risks to mix concerns on the direction of disadvantage with concerns on the number of disadvantaged circumstances. However, the information on the count of circumstances relies on the label of a circumstance *across* pairs of circumstances, rather than within each pair. This is a form of lack of anonymity that we would like to attenuate. We would like to further attenuate the impact of the circumstances labeling. Consider the following example, showing a case where there is an important fall in consensus by moving from $\pi = 0$ to $\pi = 1$ for two pairs, and a very marginal increase for the third. However, O-ezOP2 and O-ezOP3 are not verified.

Example 3.3 (Continued) Consider the following configuration for $C = \{i, j, \ell\}$:

$\pi = 0$:	$i \geq_{100} j$;	$i \geq_2 \ell$;	$j \geq_2 \ell$
$\pi = 1$:	$j \geq_{99} i$;	$\ell \geq_{99} i$;	$\ell \geq_{99} j$

O-ezOP, O-ezOP2 and O-ezOP3 are not satisfied because $C_{ND}(\mathcal{C}^{99}, 0) = \{i, j\}$ and $C_{ND}(\mathcal{C}^{99}, 1) = \{\ell\}$.

To cope with these cases, the O-ezOP3 has to be further weakened in terms of the anonymity of circumstances when constructing pairwise comparisons. The most meaningful way to do so is to resort on the original notion of EOP-Wk, and extending it to the case where there are multiple circumstances (for fixed effort levels).

When there are many circumstances, the notion of EOP-Wk can be generalized by requiring it to hold for all pairs of circumstances: EOP-Wk is satisfied if and only if no circumstance is advantaged compared to another for all preferences in $\mathcal{C}^k$, and therefore the set of non dominated circumstances has to coincide with the whole set of circumstances C. When it is not the case, at least one circumstance confers a disadvantage and a form of inequality of opportunity occurs.

In assessing O-ezOP, one has to check that *if* EOP-Wk is satisfied under $\pi = 0$ for all pairs of varieties, *then* EOP-Wk has to be satisfied also under $\pi = 1$ for all pairs of varieties. This new criterion, which we call Weak O-ezOP, is formalized in terms of sets of non-dominated circumstance:

Criterion 3.3 (Weak O-ezOP) *Moving from* $\pi = 0$ *to* $\pi = 1$ *equalizes opportunities in the sense of Weak O-ezOP iff*

$$C_{ND}(\mathcal{C}^k, 0) = C \quad \Rightarrow \quad C_{ND}(\mathcal{C}^k, 1) = C, \quad \forall\ k \in \mathbb{N}_+.$$

An alternative characterization of this criterion rests on the definition of EOP-Wk in the multivariate setting. Within a policy regime π, equality of opportunity is not verified in $\mathcal{C}^{\min_i\{\kappa_i(\pi)\}}$, because it is the largest class for which at least one circumstance can be deemed as disadvantaged compared to another. The Weak ezOP criterion is satisfied if and only if this class of preferences shrinks by effect of the policy, that is:

$$\min_i\{\kappa_i(0)\} \quad \leq \quad \min_i\{\kappa_i(1)\}.$$

The following proposition illustrates the role of the anonymity restrictions introduced, and organizes in a sequence of implications the criteria presented in this section.

Proposition 3.2 *$O - ezOP$ such that $sign(i, j, 0) = sign(i, j, 1)\,\forall i, j \ \Rightarrow\ O - ezOP2 \ \Rightarrow\ O - ezOP3 \ \Rightarrow\ O - ezOP4$.*

Proof. See appendix 3.A.2. ■

This section has illustrated the restrictions that generalize, and operationalized in a policy evaluation context, the notion of set of non-dominated circumstances in Lefranc et al. (2008). Their analysis rests on the class of risk averse individuals, and, in general, they cannot identify changes in EOP.

It seems difficult to argue which distance comparison is meaningful in this context. In fact, the evaluation of the extent of advantage can be made only for comparable and well identified varieties. Weak O-ezOP, on the contrary, is grounded on a strong notion of anonymity. Possible alternative solutions are surveyed in the following section.

3.4.3 Generalization: a distance perspective

This section discusses the appropriate distance comparisons for the cases in which EOP-Wk in a multivariate setting cannot be verified both before and after policy implementation. In fact, the extent of the disadvantage can be meaningfully measured when the disadvantage exists. When there are only two circumstances, the distance comparison is well defined. It requires to compare only the two circumstances within the class of preference for which it is possible to assess the existence of a disadvantaged circumstance among the two, both before and after policy implementation. However, in a setting with multiple circumstances, it is not clear which circumstances should be taken into account to measure the changes in the size of the disadvantage. The relevant distance comparison across policy regimes may not be the one involving a fixed pair of circumstances, but rather the circumstances sitting on two fixed positions in the raking of disadvantaged circumstances, both before and after policy implementation.

This ranking scores in the lowest position the circumstance that is always disadvantaged compared to all the others, and in the first position the circumstance that is never disadvantaged. By transitivity of the ordering based on sequential restriction of the set of preferences, these circumstances always exist. Hence, distance equalization can be assessed without looking at all pairwise comparisons, but only by fixing the positions in the ranking and by comparing across policy levels the circumstances that occupy these positions.

This procedure introduces a form of anonymity with respect to the name of the circumstances, by replacing it with the position of each circumstance in the ranking. Moreover, the position is endogenously determined by the comparison of all the circumstances, and it is not imposed (as it is the labeling). Let us now survey some alternative ways to construct the circumstances ranking.

The Weak D-ezOP criterion

To construct the possible extensions of the O-ezOP criterion, we have introduced the minimal sequence of restriction such that one given circumstance can be judged as disadvantaged compared to the others. This is not, however, sufficient to construct the full ranking of circumstances.

One way to determine a ranking of circumstances consists in identifying the set of preferences for which a given circumstance i can be correctly associated to position (i) in the ranking of circumstances, and mechanically extending this procedure to all circumstances. For a given circumstance i, her position in the ranking of disadvantaged circumstances can be found by introducing enough restrictions on $\mathcal{C}$ so that the set of circumstances dominating i is identified. The following example illustrates the procedure.

Example 3.4 (Continued) Consider the circumstances $C = \{i, j, \ell\}$. In Example 3.1 above, under $\pi = 0$, for $\mathcal{C}^{100}$ circumstance ℓ is disadvantaged compared to i and j, hence $(\ell) = 3$. For $\mathcal{C}^2$ circumstance j is disadvantaged compared to i, hence $(j) = 2$. Circumstance i is ranked above the others, so that $(i) = 1$. For $\mathcal{C}^{100}$, $i \geq_{100} j \geq_{100} \ell$ both before and after the policy implementation.

Let us define a new indicator, denoted $\kappa_{(i)}(\pi)$, which measures the minimal number of restrictions on $\mathcal{C}$ that are needed to recover the whole set of circumstances *dominating* circumstance i. With this set, one identifies also the score (i) of i in the ranking of disadvantage. The indicator $\kappa_{(i)}(\pi)$ can be identified from the matrix $\mathbf{K}(\pi)$ as follows:

$$\kappa_{(i)}(\pi) \ := \ \max_j \{|\kappa_{ij}(\pi)| \ s.t. \ \text{sign}(i, j, \pi) = -1\}.$$

For $\mathcal{C}^{\kappa_{(i)}(\pi)}$, the position of circumstance i compared to all circumstances providing more advantage is identified, and it coincides with (i). For the class of preferences $\mathcal{C}^{\max_i\{\kappa_{(i)}(\pi)\}}$, it is possible to exactly rank all circumstances with respect to the others. Hence, within the same class one can compare symmetrically the distance between the advantaged and disadvantaged circumstances, according to their position.

Within this ranking, advantaged and disadvantaged circumstances are always well identified for any comparison. The result is dual to the principle used to construct the Weak O-ezOP criterion, although the two remain grounded on two different notions of anonymity. In the former case, anonymity is achieved by considering the restrictions that allow to preserve the whole set of non-dominated circumstances, while here we look for the amount of restrictions needed to construct a complete ranking of circumstances. If, in the former case, distance comparisons may not have a meaningful interpretation, in this new setting distance can be assessed by fixing the rank of the circumstances across policy levels.

One natural candidate to construct distance comparisons requires to reduce the set of admissible distance comparisons to circumstances that are sequential in the ranking, and requires to compare every pair in sequence under policy $\pi = 0$ with the corresponding pair that occupies the same position under policy $\pi = 1$. The Weak Distance Equalization of Opportunity criterion (Weak D-ezOP) is satisfied if distance is reduced between all sequential pairs of circumstances in the ranking. Of course, this criterion is limited by an implementation constraint: the ranking is defined under both policy regimes only by exploiting $\max\{\max_i\{\kappa_{(i)}(0)\},\ \max_i\{\kappa_{(i)}(1)\}\}$ restrictions on $\mathcal{C}$.

Criterion 3.4 (Weak D-ezOP) *For a circumstance $c \in C$, let $c', \widetilde{c}, \widetilde{c}' \in C$ be the circumstances respectively in position $(c) - 1$ under policy $\pi = 0$ and in position (c) and $(c) - 1$ under $\pi = 1$. Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities in the sense of Weak D-ezOP iff $\forall W \in \mathcal{C}^{\max\{\max_i\{\kappa_{(i)}(0)\},\ \max_i\{\kappa_{(i)}(1)\}\}}$:*

$$\Delta_W(F(.|c,e,0), F(.|c',e,0)) \geq \Delta_W(F(.|\widetilde{c},e,1), F(.|\widetilde{c}',e,1)) \quad \forall c \in C,\ \forall e \in E.$$

Along with this criterion, concerns on reduction of agreement on dominance can also be taken into account by demanding that $\max_i\{\kappa_{(i)}(0)\} \leq \max_i\{\kappa_{(i)}(1)\}$. The interpretation is that the class of preferences for which it is possible to identify the disadvantaged circumstance in any pairwise comparison, shrinks by effect of policy implementation. A very weak equalization criterion can be constructed by combining this latter requirement with the Weak D-ezOP criterion, thus giving a partial order.

The Weak D-ezOP criterion exploits a reduction in distance across adjacent circumstances, thus involving $|C|-1$ distance comparisons. Other types of comparisons are meaningful in this context. For instance, one possibility consists in comparing each circumstance with the circumstances ranked as more advantaged, or directly with the most advantaged. These sequences entail different distance comparisons that are all verified whenever the Weak D-ezOP criterion is verified. A more partial criterion can be obtained by exploiting only a minimal sequence of restrictions necessary to determine the place of one circumstance in the overall ranking. As shown in the following section, weakening in this direction makes some distance comparisons unfeasible.

Alternative configurations

The Example 3.2 shows that it is not necessary to restrict the set of preferences to $\mathcal{C}^{100}$ to obtain a complete ranking. In fact, circumstance i is ranked under $\pi = 1$ as the worse-off circumstance for the class $\mathcal{C}^2$. Hence, by demanding to verify the distance comparisons for for all preferences in $\mathcal{C}^{100}$ (as is done by the Weak D-ezOP criterion) one may introduce useless restrictions on the class of preferences for which circumstance i can be properly placed in the ranking.

The minimal amount of restrictions on the class of preferences $\mathcal{C}$ that is needed to associate circumstance i with her position (i) in the ranking of disadvantage is identified by $\kappa_{(i)}(\pi)$. Corresponding to this value, it is possible to identify all the $(i)-1$ circumstances that dominate i. Another example clarifies the logic behind the procedure used to construct the ranking, and the relation between consecutive dominated circumstances.

Example 3.5 In the previous examples, i, j, ℓ were ranked under $\pi = 0$ as: $i \geq_2 j \geq_{100} \ell$.

Let $i, j, r, \ell \in C$, a general ranking of circumstances can be represented as follows:

$$i \geq_{\kappa_2(\pi)} j \geq_{\kappa_3(\pi)} \cdots \geq_{\kappa_{(r)}(\pi)} r \geq_{\kappa_{(r)+1}(\pi)} \cdots \geq_{\kappa_{(\ell)}(\pi)} \ell. \tag{3.1}$$

The mechanism underlying the construction of the ranking in (3.1) determines a ranking that may not be appropriate to exploit all possible distance comparisons. In fact, the distance comparison of circumstance i with the circumstance scoring in position $(i) - 1$ is meaningful, because the disadvantage of i with respect to the adjacent circumstance has been established according to the class of preferences $\mathcal{C}^{\kappa_{(i)}(\pi)}$. However, nothing guarantees that for the same class of preferences there is agreement on the disadvantage of the circumstance in position $(i) + 1$ when compared to i. This is the case when $\mathcal{C}^{\kappa_{(i)+1}(\pi)} \subseteq \mathcal{C}^{\kappa_{(i)}(\pi)}$.[10]

These simple examples illustrate that the distance comparison based on this ranking procedure does not meet the basic anonymity requirements: it may not be possible to assess, within the same class of preference, if the utility gap between circumstances in position (i) and the circumstance in position $(i) - 1$ (respectively $(i) + 1$) has decreased/increased by effect of the policy, while at the same time the utility gap with the circumstance in position $(i) + 1$ (respectively $(i) - 1$) has decreased/increased. To do so, one is obliged to change the class of evaluation functions. Given i, this reasoning applies also to circumstances that are in position $(i) \pm 2$, and so on.

Equality of opportunity indicators

The Weak D-ezOP criterion requires that the distance among every pair of adjacent circumstances, occupying the same rank position before and after policy implementation, is reduced. There are cases, however, where this partial order fails to compare policy regimes. One possibility is to use an *Index of Opportunity* (IO) to obtain (always) conclusive results on policy impact, although the final assessment would not be robust to changes in the

[10] By exploiting the example, although for the class $\mathcal{C}^{\kappa_{(r)}(\pi)}$ it is possible to verify the disadvantage of i compared to the precedent elements of the ranking (for instance i and j), for the same class of preferences it is not (in general) possible to verify the advantage of r compared to circumstances scoring below it in the ranking (like ℓ).

specification of the indicator.

Consider the class $\mathcal{C}^{\max_i\{\kappa_{(i)}(1)\}}$ for which circumstances can be completely ranked both before and after policy intervention. If Weak D-ezOP is not verified, one can pick up a specific preference, represented by a function W in that class, and use the distance constructed on W to rank policy regimes. This can be done, for instance, by taking the *weighted average* of the utility gaps, denoted by $W\left(F(.|j,e,\pi)\right) - W\left(F(.|i,e,\pi)\right)$ for $(j) < (i)$, associated to all possible comparisons of a circumstance with the remaining circumstances that dominate it, according to the weighting scheme p_i. We obtain a new indicator of opportunity inequality:

$$IO_k(\pi) \;=\; \sum_{i\in C}\sum_{j:(j)<(i)} p_i p_j \left[W\left(F(.|j,e,\pi)\right) - W\left(F(.|i,e,\pi)\right)\right].$$

We are not aware of any similar generalization in the literature. However, we are well aware of the *Gini Opportunity* index proposed by Lefranc et al. (2008).[11] Their index is, in fact, a very particular case of the IO_k that can be only applied to data under the (possibly empirically rejected) assumption that both before and after policy intervention all preferences displaying risk aversion agree on the ranking of all circumstances. This class of preferences can be identified by resorting on the rank dependent model. The *GO* index is obtained by selecting one particular utility function in this class, which coincides with the Gini social welfare function (Zoli 2002):

$$W\left(F(.|i,e,\pi)\right) \;=\; \mu_i(1-G_i),$$

where μ_i is the average outcome associated to the distribution conditional on circumstance i, and G_i is the Gini coefficient of i's distribution. Then the *GO* index is:

$$GO(\pi) \;=\; \frac{1}{\mu}\sum_{i\in C}\sum_{j:(j)<(i)} p_i p_j \left[\mu_j(1-G_j) \;-\; \mu_i(1-G_i)\right].$$

[11] For a complete survey of Gini-type indices for Equality of Opportunity sets, see Weymark (2003).

Similarly to the $GO(\pi)$ index, the $IO_k(\pi)$ index can only be applied when all circumstances can be ranked, and it relies on a specific evaluation out of a set of possible alternatives. However, when Weak D-ezOP is satisfied, then distance is reduced for every pair of circumstances and therefore the $IO_k(1)$ index will be lower than $IO_k(0)$. However, differently from the $GO(\pi)$ index, the $IO_k(\pi)$ index is always determined, because it is grounded on a flexible definition of the class of functions for which all circumstances can be order.[12]

3.4.4 The Weak ezOP criterion: a proposal

Let us conclude this section by introducing a new criterion that takes into account the ordinal and the distance comparisons when there are multiple circumstances (and multiple effort levels). The Weak O-ezOP criterion requires that the class of preferences for which there is agreement on the existence of some disadvantaged circumstance, shrinks by effect of the policy. The Weak D-ezOP criterion requires that the extent of the disadvantage between adjacent pairs in the ranking of disadvantaged circumstances, as measured by the class of preferences where consensus is reached on the ranking of the circumstances, diminishes by effect of the policy. Putting the two criteria together, one obtains the Weak ezOP criterion, which demands to verify the Weak O-ezOP and the Weak D-ezOP at all effort levels:

Definition 3.8 (Weak ezOP) *Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities in the weak sense if and only if Weak O-ezOP and Weak D-ezOP are satisfied $\forall e \in E$.*

In the case with only two circumstances (and one effort level) the Weak ezOP criterion boils down to the ezOP criterion.

[12] Consider for instance the rank dependent model for preferences, and the family of *S-Gini* functionals (see Maccheroni, Muliere and Zoli 2005). The utility function in this family obtained by introducing k constraints on preferences is defined, for circumstance i, as $W(F(.|i,e,\pi)) = \mu_i(1 - G_i^k)$, where:

$$G_i^k = \int_0^1 \left[1 - k(1-p)^{k-1}\right] \frac{F^{-1}(.|i,e,\pi)}{\mu_i} dp.$$

This straightforward generalization of the $GO(\pi)$ index allows to consistently define a ranking of policy levels. We leave for further research the discussion, characterization and application of the $IO_k(\pi)$ index.

3.5 Implementation of the equalization of opportunity criterion

The empirical implementation of the criteria of opportunity equalization defined in the previous sections requires two main ingredients. The first is a model for preferences, along with a definition of a sequence of restrictions on the class of preferences identified by the model. In section 3.5.1 and in section 3.5.2 we provide identification conditions respectively for the rank dependent models and the expected utility model for preferences, and we highlight potential pitfalls.

The second ingredient is a testing procedure that can be easily adapted to cases in which unobservability constraints only permit ex ante evaluations of opportunity sets. We discuss these issues in detail in sections 3.5.3 and 3.5.4.

3.5.1 Identification of ezOP under the rank dependent model

We now formalize the ezOP criterion in the setting of the rank dependent model for preferences. The rest of the analysis relies on the notion of inverse stochastic dominance, introduced by Muliere and Scarsini (1989) and further studied by Aaberge (2009), Maccheroni et al. (2005), and Zoli (1999, 2002).

Before proceeding, let us introduce some further notation. We use $F \succ_{ISDk} F'$ and $F \succeq_{ISDk} F'$ to indicate respectively the strict and the weak inverse stochastic dominance at order k. We refer the reader to appendix 3.A.1 for a formal discussion on the topic. First- and second-order inverse stochastic dominance are equivalent to their direct dominance counterparts. Inverse dominance at higher orders is defined on the basis of integrals of the Generalized Lorenz curve, so that at orders higher than the second direct and inverse stochastic dominance do not coincide.

Characterization of the ordinal criterion for equalization

Let restrict attention to the case in which there is only a pair of varieties (c, e) and (c', e), the analysis can be easily generalized to the multivariate setting. The ordinal equalization criterion is satisfied if, for a class of preferences $\mathcal{C}$ and a sequence of restrictions that identify the class of preferences $\mathcal{C}^{\kappa(c,c',e,\pi)}$ for which (c, e) is advantaged with respect to (c', e), one has that $\mathcal{C}^{\kappa(c,c',e,1)} \subseteq \mathcal{C}^{\kappa(c,c',e,0)}$.

Of course, implementing the equalization criterion requires to resort to a partitioning of the set of admissible individual preferences. The analysis of this paper rests upon the class $\mathcal{R}$ of rank dependent utility functions introduced by Yaari (1987, 1988). The equalization criterion can be reformulated by substituting $\mathcal{C} = \mathcal{R}$. A utility function $W \in \mathcal{R}$ can be written as:

$$W(F(y|c, e, \pi)) \;=\; \int_0^1 w(p)F^{-1}(p|c, e, \pi)dp,$$

where $w(p) \geq 0 \; \forall p \in [0, 1]$ and $\widetilde{w}(p) = \int_0^p w(t)dt \in [0, 1]$ is such that $\widetilde{w}(1) = 1$. The function $w(p)$ is interpreted as a weight assigned to rank p in the outcome conditional distribution, while the function $\widetilde{w}(p)$ defines the *cumulative weighting scheme* applied to the percentiles domain (see Zoli 2002). Restrictions on the sign of the derivatives of $w(.)$ (and $\widetilde{w}(p)$) induce a partition on the set $\mathcal{R}$. For instance, $\mathcal{R}^1$ is the largest class of interest, associated to the utility function assigning positive weights to all the quantiles. We restrict attention to this set of preferences by imposing that $\mathcal{R}^1 = \mathcal{R}$.

The class $\mathcal{R}^2$ is the class of utility functions where $w(p)$ is positive, decreasing over the domain of p and such that $w'(1) = 0$. More generally, let $\mathcal{R}^k \subseteq \mathcal{R}$ define the subset of rank dependent utility functions $W(.)$ with $w(p) \geq 0$ that satisfy:

$$\mathcal{R}^k \;=\; \left\{ W \in \mathcal{R} : \; (-1)^{i-1} \cdot \frac{d^i \widetilde{w}(p)}{dp^i} \geq 0, \quad \frac{d^i \widetilde{w}(1)}{dp^i} = 0 \; \forall p \in [0, 1] \text{ and } i = 1, 2, 3, \ldots, k \right\}.$$

Obviously, these subsets are nested: $\mathcal{R}^k \subset \mathcal{R}^{k-1} \subset \ldots \subset \mathcal{R}^1$. Imposing restrictions on higher-order derivatives of the weighting function implies putting more weight on least favorable outcomes.[13]

Aaberge (2009) has established the logical equivalence between inverse stochastic dominance at order k and unanimity in ranking a pair of distribution functions within the class $\mathcal{R}^k$. This result establishes the link between EOP-Wk and IOP-k, and the equivalent characterization based upon inverse stochastic dominance, that is:

$$\begin{aligned}\mathcal{R}^{\kappa(c,c',e,1)} \subseteq \mathcal{R}^{\kappa(c,c',e,0)} \quad &\Leftrightarrow \quad \kappa(c,c',e,1) \geq \kappa(c,c',e,0) \\ &\Leftrightarrow \quad \begin{array}{c} F(.|c,e,1) \succ_{ISD\kappa(c,c',e,1)} F(.|c',e,1) \;\; \text{implies} \\ F(.|c,e,0) \succ_{ISD\kappa(c,c',e,1)} F(.|c',e,0). \end{array}\end{aligned}$$

Therefore, $\kappa(c,c',e,\pi)$ now represent the minimum order of inverse stochastic dominance according to which $F(|c,e,\pi)$ and $F(|c',e,\pi)$ can be ranked and we have, by definition, for all $k \geq \kappa_\pi$: $\quad F_\pi \succ_{ISDk} F'_\pi$.

The characterization based on the Yaari model can always be identified, and allows to recover the LPT's EOP-S as a limiting case. In fact, in order to characterize inequality of opportunity on the basis of κ_π, we need to be sure that κ is defined for all pairs of distributions. This is established in the following proposition.

Proposition 3.3 *For any pair of distributions with bounded support, with inverse cumulative distribution functions denoted by $F^{-1}(.)$ and $F'^{-1}(.)$ satisfying:*

$\exists p_\beta > 0 | \forall p \in [0, p_\beta)$ $F^{-1}(p) \geq F'^{-1}(p)$ *and the strict inequality holds on a positive mass interval* $[p_\beta - \epsilon, p_\beta)$ *with* $\epsilon > 0$,

we have:

$\exists \kappa \in \mathbb{R}_+$ *and finite such that* $F \succ_{ISDk} F'$ $\forall k \in \mathbb{N}_+$ *such that* $k > \kappa$.

Proof. See appendix 3.A.3. ■

In the simple model with two comparable varieties, the consequence of the proposition is that if stochastic dominance cannot be verified (in any direction) at a finite degree, then the two distributions have to coincide. This is LPT's EOP-S. Within the Yaari model

[13] Note that k is a measure of the effect of a precise sequence of restrictions on all possible cumulative weighting schemes $\widetilde{w}(p)$ defined on $\mathcal{R}$. Hence, k embodies information on the risk attitude of preferences isolated by the class $\mathcal{R}^k$.

it is therefore possible to obtain EOP-S as a limiting case of non dominance, while it is still possible to have agreement on indifference in the class $\mathcal{R}$ without having EOP-S. All these separate results are gathered together to construct table 3.3, which gives the possible configurations of the ezOP test for a pair of comparable varieties.

This property is not featured by the expected utility model. In fact, for bounded distributions one can always determine an order of direct stochastic dominance that may not be finite (Fishburn 1976), and therefore non dominance cannot be verified empirically. Hence, LPT's EOP-S is no more a limiting case.

Within the rank dependent model the O-ezOP criterion is implemented by looking at inverse stochastic dominance relations among pairs of distributions, and it is tested empirically as shown in the appendix A. Conversely, the condition defining the distance criterion is not easily implementable in practice, since it requires computing distance measures over a set of utility functions. Hence, it may be useful to identify some conditions on the quadruple (F_0, F_0', F_1, F_1') under which distance falls as a result of the policy. We work out such conditions in what follows.

The rank dependent model under first- and second-order stochastic dominance

We consider first the distance measures based on the rank dependent utility functions.[14] Define $G(F_\pi, F'_\pi, p)$ the income gap between distributions F_π and F'_π at each quantile: $G(F_\pi, F'_\pi, p) = F_\pi^{-1}(p) - F_\pi'^{-1}(p)$.

For $W \in \mathcal{R}$, the certainty equivalent is defined by:

$$\int_0^1 w(p) CE_W(F_\pi) dp = CE_W(F_\pi) = \int_0^1 w(p) F_\pi^{-1}(p) dp$$

[14] An advantage of the Yaari-type rank dependent utility function is that the certain equivalent is linear in its argument and therefore *distributionally homogeneous*: if $\widetilde{F}_\pi^{-1}(p) = \alpha F_\pi^{-1}(p) + \beta$, $\forall p \in [0,1]$, then $CE_W(\widetilde{F}_\pi) = \alpha CE_W(F_\pi) + \beta$, $\forall \alpha,\ \beta \in \mathbb{R}$. Chakravarty and Dutta (1987) showed that this is a necessary and sufficient condition for a distance measure based on certain equivalents to satisfy the homogeneity (from the previous example, $\Delta_W(\widetilde{F}_\pi, \widetilde{F}'_\pi) = \alpha \Delta_W(F_\pi, F'_\pi)$) and translation invariance ($\Delta_W(\widetilde{F}_\pi, \widetilde{F}'_\pi) = \Delta_W(F_\pi, F'_\pi)$ when $\alpha = 1$) properties in Ebert (1984).

Table 3.3: Implementation of equalization of opportunity: two types and one responsibility level

		$\pi = 1$							
$\pi = 0$		EOP-S		LPT's EOP-W1				LPT's IOP	
		$\not\succ_{ISD\infty}$	$\cdots$	$\succ_{ISDk}$	$\cdots$	$\succ_{ISDh}$	$\cdots$	$\succ_{SD2}$	$\succ_{SD1}$
EOP-S	$\not\succ_{ISD\infty}$	No effect	$\cdots$	Disequalization	$\cdots$	Disequalization	$\cdots$	Disequalization	Disequalization
LPT's EOP-W	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	$\succ_{ISDk}$	ezOP-S	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^k$	$\cdots$	Disequalization	$\cdots$	Disequalization	Disequalization
	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$
	$\succ_{ISDh}$	ezOP-S	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^k$	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^h$	$\cdots$	Disequalization	Disequalization
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
LPT's IOP	$\succ_{SD2}$	ezOP-S	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^k$	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^h$	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^2$	Disequalization
	$\succ_{SD1}$	ezOP-S	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^k$	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^h$	$\cdots$	ezOP iff D-ezOP $\forall W \in \mathcal{R}^2$	ezOP iff D-ezOP $\forall W \in \mathcal{R}$

Notation: LPT's EOP-S (def. 3.1), EOP-W (def. 3.2) IOP (def. 3.3); ezOP is partial equalization of opportunity; ezOP-S is strong equalization of opportunity and D-ezOP is the distance criterion associated to the class of preferences $\mathcal{C}$.

and the distance can be written as:

$$\Delta_W(F_\pi, F'_\pi) = |\int_0^1 w(p)[F_\pi^{-1}(p) - F_\pi'^{-1}(p)]dp| = |\int_0^1 w(p)G(F_\pi, F'_\pi, p)dp|.$$

Under first-order stochastic dominance, proposition 3.4 establishes a necessary and sufficient condition for D-ezOP:

Proposition 3.4 *Let* $F_\pi \succ_{SD1} F'_\pi \ \forall\pi$*, then* $\Delta_W(F_0, F'_0) - \Delta_W(F_1, F'_1) \geq 0$ *for all* $W \in \mathcal{R}$ *iff* $G(F_0, F'_0, p) - G(F_1, F'_1, p) \geq 0$ *for all* $p \in [0, 1]$.

Proof. See appendix 3.A.4. ■

In other words, there is unanimity within the largest class of utility functions $\mathcal{R}$ in saying that the distance between the income prospects of two comparable varieties has decreased by effect of policy intervention if and only if the income gap between the distributions at each quantile is smaller in the treated pair of distributions than it is in the non treated one. This induces a form of dominance at the first order of the gaps distributions, once that gaps have been ordered according to their percentile (and not according to the gaps magnitudes).

One can also note that $G(F_0, F'_0, p) - G(F_1, F'_1, p) = G(F_1, F_0, p) - G(F'_1, F'_0, p)$. Requiring that $G(F_1, F_0, p) - G(F'_1, F'_0, p) \geq 0$ for all p is requiring that the income gains due to policy treatment be greater at all percentiles of luck for the most disadvantaged type c'.

Consider now the case in which distributions within each pair can only be ordered according to second-order stochastic dominance. It is now interesting to evaluate if, within the class of utility displaying risk aversion, it is still possible to rank the two pairs of distributions according to the distance criterion. Under second-order stochastic dominance, Proposition 3.5 establishes a necessary and sufficient condition for D-ezOP:

Proposition 3.5 *Let* $F_\pi \succ_{SD2} F'_\pi \ \forall\pi$*, then* $\Delta_W(F_0, F'_0) - \Delta_W(F_1, F'_1) \geq 0$ *for all* $W \in \mathcal{R}^2$ *iff* $\int_0^p G(F_0, F'_0, t)dt - \int_0^p G(F_1, F'_1, t)dt \geq 0$ *for all* $p \in [0, 1]$.

Proof. See appendix 3.A.5. ■

In other words, there is unanimity within the largest class of utility functions $\mathcal{R}^2$ in saying that the distance between the income prospects of two comparable varieties has decreased by effect of policy intervention if and only if the income *cumulated* gaps between the distributions at each quantile is smaller in the treated pair of distributions than it is in the non treated one. This induces a form of dominance at the second order of the gaps distributions, once that gaps have been ordered according to their percentile (and not according to the gap magnitude itself).

Consider writing the gap difference as $G(F_1, F_0, p) - G(F_1', F_0', p)$. Requiring that $\int_0^p [G(F_1', F_0', t) - G(F_1, F_0, t)]\, dt \geq 0$ for all p is requiring that the cumulated income gain due to policy treatment be greater at all percentiles for the most disadvantaged type c'.

Following a similar scheme, it is possible to derive a dominance condition for all classes of utility functions, provided that outcomes prospects can be ranked according to the corresponding order of *inverse* stochastic dominance.

The rank dependent model under k-th order inverse stochastic dominance

Consider the k-th integral of the quantile functions as defined in appendix 3.A.1. Under k-th order inverse stochastic dominance, Proposition 3.6 establishes a necessary and sufficient condition for D-ezOP:

Proposition 3.6 *Let $F_\pi \succ_{ISDk} F_\pi'$ $\forall\pi$, then $\Delta_w(F_0, F_0') \geq \Delta_w(F_1, F_1')$ for all $W \in \mathcal{R}^k$ iff $G(\Lambda_0^k, \Lambda_0'^k, p) \geq G(\Lambda_1^k, \Lambda_1'^k, p)$ for all $p \in [0, 1]$.*

Proof. See appendix 3.A.6. ■

In other words, there is unanimity within the largest class of utility functions $\mathcal{R}^k$ in saying that the distance between the income prospects of two comparable varieties has decreased by effect of policy intervention if and only if the income *cumulated* gap at the k-th order between the distributions at each percentile is smaller in the treated pair of distributions than it is in the non treated one. This induces a form of dominance at the

k-th order of the gaps distributions, once that gaps have been ordered according to their percentile (and not according to the gap magnitude itself).

One can also note that $G(\Lambda_0^k, \Lambda_0'^k, p) - G(\Lambda_1'^k, \Lambda_1'^k, p) = G(\Lambda_1'^k, \Lambda_0'^k, p) - G(\Lambda_1^k, \Lambda_0^k, p)$. Requiring that $G(\Lambda_1'^k, \Lambda_0'^k, p) - G(\Lambda_1^k, \Lambda_0^k, p) \geq 0$ for all p is requiring that the gains by reduction of inequality at order k when moving from $\pi = 0$ to $\pi = 1$ should be greater at all percentiles for the more disadvantaged type.

3.5.2 Identification of ezOP under the expected utility model

As we have seen in the previous section, when considering rank-dependent utility functions, the Shorrocks's measure of economic distance can be easily reformulated as a function of the income gap at different quantiles of the income distributions. The obvious reason is that in the class of rank dependent utility functions, social welfare is a linear function of the income level at each quantile.

We now consider the class $\mathcal{U}$ of additive utility functions, and the related distance measure. Integrating by parts and assuming that the support for the outcomes distributions is bounded by $y \in [\underline{y}, \overline{y}]$, than the additive utility function $W(.) \in \mathcal{U}$ admits the following representation:

$$W(F(y|c,e,\pi)) = \int_y u(y) dF(y|c,e,\pi) = \big(u(\overline{y}) - u(\underline{y})\big) - \int_y u'(y) F(y|c,e,\pi) dy \quad (3.2)$$

We use the higher order derivatives of u with alternating signs to define the restrictions on the set $\mathcal{U}$. This sequence of restrictions characterizes preferences for risk apportionment (Eeckhoudt and Schlesinger 2006). For instance, $\mathcal{U}^2$ represents the class of risk averse preferences. The implementation of the O-ezOP criterion in this case relies on *direct* stochastic dominance comparisons of distributions at order one, two or above. However, direct stochastic dominance analysis presents two main drawbacks with respect to the analysis underlying ezOP identification. First, an infinite sequence of restrictions does not identify EOP-S, because $\mathcal{U}^\infty$ coincides with a well defined class of preferences (see Fishburn 1976). Secondly, distance comparisons are problematic in this setting.

It turns out that, in the general case, the gap in utility levels cannot be expressed as a function of the income gap at the different levels of the underlying distributions. Results can only be obtained in special cases. The reason for this problem originates in the fact that the distance measure is no longer invariant to a translation of all incomes in the compared distributions. As a consequence, the welfare gap cannot only be expressed as a function of the income gaps, independently of the income levels. More specifically, for $W \in \mathcal{U}$, we have:

$$CE_W(F) = u^{-1}\left(\int_0^{\infty} u(x)dF(x)\right).$$

The problem for expressing the distance condition as a restriction on the cdf is that contrary to the rank-dependent case, $\Delta_W(F_0, F_0')$ is not longer linear in $W(F) - W(F')$. In the general case, $u(.)$ is simply an increasing function and so is u^{-1}. Hence the economic distance condition implies no specific restrictions regarding the gap in aggregate welfare between the two distributions.

Some conditions can be expressed in the special case where $u(.)$ is a concave function. In this case, u^{-1} is a convex function. Next assume stochastic dominance of order two of F_i over F_i', for $i = 0, 1$. As a result of dominance, we can get rid of the absolute values in the expression of $\Delta_W(F_\pi, F_\pi')$. Under these assumptions, we have:

$$\Delta_W(F_0, F_0') \geq \Delta_W(F_1, F_1') \quad \Leftrightarrow \quad u^{-1}(W(F_0)) - u^{-1}(W(F_0')) \geq u^{-1}(W(F_1)) - u^{-1}(W(F_1')).$$

The last dominance relation is a condition on the gap in certain equivalent incomes. This condition does not imply any restriction on the sign of $[W(F_0) - W(F_0')] - [W(F_1) - W(F_1')]$. In other words, it does not imply any condition on the gap in social welfare between types before or after the policy. Indeed, because of the convexity of u^{-1}, the gap in social welfare depends not only on the gap in certain equivalents, but also on their magnitude.

Imposing additional conditions on the level of welfare of both types before and after the policy, helps remove this indeterminacy. A relatively clear case occurs when the situation of the worst type is improved as a result of the policy: $W(F_1') \geq W(F_0')$. In this case, using the convexity of the u^{-1} function, one can show that the economic distance between the distributions decreases for all utilities in $\mathcal{U}^2$ if the gap in expected utility falls as a result of the policy. This is summarized by the following proposal:

Proposition 3.7 *Under the following assumptions: (i) $F_\pi \succ_2 F'_\pi$, $\forall\pi$ and (ii) $W(F'_0) \leq W(F'_1)$, $\forall W \in \mathcal{U}^2$, we have:*

$$\forall W \in \mathcal{U}^2 \Delta_W(F_0, F'_0) \geq \Delta_W(F_1, F'_1) \Rightarrow \forall W \in \mathcal{U}^2, W(F_0) - W(F'_0) \geq W(F_1) - W(F'_1).$$

Proof. See appendix 3.A.7. ■

The proposition shows that the right direction in the variation in welfare differential associated to policy intervention is only a necessary condition for decreasing economic distance between outcome prospects. The concavity of the utility function plays the central role in showing that this result depends upon the the validity of $F_1 \succ_{SD2} F'_1 \succ_{SD2} F'_0$. That is, by effect of the policy the situation of the worse off is ameliorated by effect of policy intervention.

Consider firstly a very restrictive case of second order stochastic dominance, the one in which $F_\pi \succ_{SD1} F'_\pi$. Using (3.2), one can equivalently check that the necessary condition in proposition 3.7 is satisfied by looking at the gap between cdfs, $(F_0(y) - F'_0(y)) - (F_1(y) - F'_1(y))$, is less than zero at any quantile $y \in [\underline{y}, \overline{y}]$. Another way to see the relation, is that $(F_0(y) - F_1(y)) - (F'_0(y) - F'_1(y)) \leq 0$, that is the change in the percentage of risk of receiving an income lower than y due to policy change is larger for the more disadvantaged variety (c', e) compared to the more advantaged variety (c, e).

A similar argument can be used to determine a dominance condition for the case where only second order stochastic dominance is verified. The equivalence between $\succ_{SD2}$ and $\succ_{ISD2}$, or alternatively Generalized Lorenz Dominance (Shorrocks 1983), gives the following proposition, which can be easily demonstrated making use of integration by part of (3.2):

Proposition 3.8 *Under the following assumptions: (i) $F_\pi \succ_{SD2} F'_\pi$, $\forall\pi$ and (ii) $W(F'_0) \leq$*

$W(F_1'), \forall W \in \mathcal{U}^2, W(F_0) - W(F_0') \geq W(F_1) - W(F_1')$ *iff*

$$\int_y \left[(F_0(y) - F_0'(y)) - (F_1(y) - F_1'(y)) \right] dy \leq 0 \; \forall y \in [\underline{y}, \overline{y}].$$

The condition provides an alternative, equivalent but empirically attractive necessary condition for the distance dominance in Proposition 3.7. Nevertheless, if one is not willing to go beyond in restricting the class of preferences $\mathcal{U}^2$, it is not possible to restore the equivalence in Proposition 3.7. A meaningful restriction would consists in selecting families of additive utility functions according to the properties of the risk aversion coefficient. Such restrictions allows to keep the risk attitude as constant within a class of utility functions, thus providing more intuitions on the distortion in the certain equivalent caused by the evaluation $u^{-1}(.)$.

3.5.3 Implementation algorithm

We propose a procedure for testing equalization of opportunity based on the ezOP criterion identified by the rank dependent utility model. We first assume that individual outcome, circumstances and effort are observed for a representative sample of the population. We construct an algorithm for comparing the outcome prospects for all pairs of circumstances in C, evaluated at all possible effort levels in E. Admittedly, assuming that all circumstances and effort are observable is a strong requirement that may be easily violated empirically. We leave for the next section the discussion of implementation when effort is not observable.

We construct the algorithm based on D-ezOP and O-ezOP criteria. For each criterion we state the null hypothesis that have to be tested, we propose the testing procedure for ISD and distance comparisons and we also provide the corresponding empirical analog test statistics. For inverse stochastic dominance, we use integrals of the generalized Lorenz curve, for distance dominance we use the cumulative gaps distribution (also called gap curve).

Algorithm 3.1 (Implementable ezOP) *The following sequence of estimations and tests implement ezOP:*

F$(.|\mathbf{c},\mathbf{e},\pi)$ *For any pair* (c,c')*, any policy level* π*, and any effort level* e*, estimate the distributions* $F(y|c,e,\pi)$ *and* $F(y|c,e,\pi)$*.*

$\kappa(\mathbf{c},\mathbf{c}',\mathbf{e},\pi)$ *For each* (c,c',e,π) *compute* $\kappa(c,c',e,\pi)$ *as follows:*

- *For* $k=1,...$ *define the following pair of hypothesis:*
  $\{H_0 : F(y|c,e,\pi) \succ_{ISDk} F(y|c',e,\pi)$ *vs.* $H_a : F(y|c,e,\pi) \not\succ_{ISDk} F(y|c',e,\pi)\}$
  and
  $\{H_0 : F(y|c',e,\pi) \succ_{ISDk} F(y|c,e,\pi)$ *vs.* $H_a : F(y|c',e,\pi) \not\succ_{ISDk} F(y|c,e,\pi)\}$
  Define $I_k=(u,v)$ *the result of this pair of tests, where* u*,* v *are equal to 1 if the null hypothesis is rejected and 0 otherwise, respectively for both null hypothesis.*
- *Compute* I_k *for* $k=1,\ldots$*:*
  - *if* $I_k=(0,0)$*:* $\kappa(c,c',e,\pi)=\infty$ *- stop*
  - *if* $I_k=(0,1)$ *or if* $I_k=(1,0)$*:* $\kappa(c,c',e,\pi)=k$ *- stop*
  - *if* $I_k=(1,1)$*: let* $k=k+1$ *and iterate.*

O-ezOP *Verify that* $\kappa(c,c',e,1)\geq\kappa(c,c',e,0)$*, for all* (c,c') *and all* e*.*

D-ezOP *Verify that distance in outcome prospects is reduced by effect of the policy within the class* $\mathcal{R}^{\kappa(c,c',e,1)}$*. For this class, compute the gaps distribution at order* $\kappa(c,c',e,1)$ $(=\kappa)$ *and test*

$$G\left(\Lambda^{\kappa}(p|c,e,0),\Lambda^{\kappa}(p|c,e,0)\right)\geq G\left(\Lambda^{\kappa}(p|c,e,1),\Lambda^{\kappa}(p|c,e,1)\right)\quad\forall p\in[0,1].$$

If steps O-ezOP and D-ezOP are verified for all (c,c',e)*, ezOP is verified.*

The same algorithm can be used to test Weak ezOP, by testing the minimal degree of restrictions for which *ISD* can be verified, and by comparing if this value has increased by effect of policy implementation. Moreover, the algorithm can be adapted to the study of ex ante equalization criteria, provided that the appropriate conditional distributions are used. The algorithm does not feature a statistical test on the *order* of dominance.

The definitions of the null hypothesis, the test statistics and their asymptotic behavior are provided in the appendix A.5.

3.5.4 Unobservable effort

Assume that circumstances are observable but effort is not. Hence, we only observe the distribution of outcome conditional on circumstances under both policy regimes. This distribution is obtained by averaging quantiles of the outcomes distribution according to the effort distribution:

$$F^{-1}(p|c,\pi) \quad = \quad \int_e F^{-1}(p|c,e,\pi)dG(e|c,\pi). \tag{3.3}$$

We now turn to the following question: is it possible to assess ezOP under observability constraints?

Ex ante perspective

Under observability constraints, it is possible to implement only ex ante criteria of equality of opportunity. In a nutshell, the ex ante approach amounts to assume that individuals do not know their final level of effort when making equality of opportunity judgments. Hence, these judgements are only based on the distribution of outcome conditional on circumstances. All effort levels are aggregated and this is formally equivalent to assuming that there exists only a single effort level. This leads to the *Opportunity Dominance* concept (Fleurbaey 2008, Ch. 9). The criterion requires that the distance between types in the outcome space be reduced as a result of the policy, regardless of their responsibility.

It is straightforward to adapt the ezOP test to the ex ante perspective of equality of opportunity. It suffices to replace the distributions $F(.|c,e,\pi)$ with the ex ante counterpart $F(.|c,\pi)$ for all pairs of circumstances $c \neq c'$ and policy levels in the definitions of O-ezOP, D-ezOP and ezOP. Moreover, if ezOP holds for all pairs of types, than ex ante ezOP is verified. This is summarized in the following definition:

Definition 3.9 (ex ante ezOP) *Define $F(y|c,\pi)$ and $F(y|c',\pi)$ the outcome conditional distributions that can be estimated when effort is not observable. For $c \neq c'$, define $\kappa(c,c',\pi)$ as the minimum order for which $F(y|c,\pi)$ and $F(y|c',\pi)$ can be ordered according to inverse stochastic dominance. Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities in the ex ante perspective if for all pairs of types $c, c' \in C$ holds that $\kappa(c,c',1) \geq \kappa(c,c',0)$ and $\Delta_W(F(y|c,0), F(y|c',0)) - \Delta_W(F(y|c,1), F(y|c',1)) \geq 0$ for all $W \in \mathcal{R}^{\kappa(c,c',1)}$.*

The ex ante ezOP criterion can be empirically tested by resorting to the sequence proposed in Algorithm 3.1 while replacing the estimates of the outcome conditional distribution functions $F(.|c,e,\pi)$ with $F(.|c,\pi)$ at any effort level in the first step of the algorithm (step $\mathbf{F(.|c,e,\pi)}$).

The ex ante ezOP procedure defines the implementation criterion for the ezOP test, based upon an ex post perspective. We discuss in the following section the relation between the two criteria.

Ex post perspective

In this section we study the relation between ex post and ex ante criteria of opportunity equalization. These relations can be studied, however, at the cost of imposing some (often non-testable) restriction on our model. As discussed in previous contributions (e.g. Lefranc et al. 2009), one natural first restriction can be set on the effort distribution, by assuming that $G(e|c,\pi)$ is independent on circumstances and the policy. For a given policy, this corresponds to two distinct situations. Either the analyst takes an absolutistic concept of effort, and in this case nothing guarantees the independence which, nevertheless, may be empirically satisfied. Or the analyst takes a relativistic perspective of effort (as in Roemer). In the latter case one gets that the independence is satisfied by construction.[15]

Assumption 3.1 (Effort Independence) *For all c and π, $G(e|c,\pi) = G(e)$.*

[15] For a more detailed discussion on the implications derived by the assumptions on the effort distribution, see Lefranc et al. (2009).

This assumption, along with a monotonicity requirement on the direction of dominance between pairs of circumstances across policy changes, allows to state at least a necessary condition for ezOP implementation.

Proposition 3.9 *Under Assumption 1, if* $\forall c \neq c'$ $F(.|c,e,\pi) \succcurlyeq_{\kappa(c,c',e,\pi)} F(.|c',e,\pi)$ *holds for all* $e \in E$*, then*

$$ezOP \quad \Rightarrow \quad ex\ ante\ exOP.$$

Proof. See appendix 3.A.8. ∎

A direct result of the proposition is that the ex ante ezOP criterion can identify, at most, a weaker notion of opportunity equalization than the one embodied by the ezOP criterion. We use similar justifications as for LPT's EOP-W2 criterion to define a new notion of equality of opportunity. The criterion entails a type of judgement on the changes in direction and size of the economic advantage attributed to some circumstances, as if it is evaluated by an agent whose preferences have not been attributed and effort is not chosen yet. However, the criterion remains ex-post in the sense that it is always possible to construct comparisons of distributions at similar effort levels.

We define a new concept of equality of opportunity, denoted $\overline{\text{EOP-W}k}$. For a given class of preferences $\mathcal{R}^k$, $\overline{\text{EOP-W}k}$ holds if and only if one of the two following cases occurs: either there exists no effort level for which one circumstance provides an unambiguous advantage (assessed at unanimity within $\mathcal{R}^k$) with respect to another one, or if there is one effort level for which one circumstance provides an unambiguous advantage, then the same circumstance should provide an unambiguous disadvantage for at least one other effort level. In both cases, no type can be unambiguously preferred ex ante, i.e. if effort has not been chosen yet.

Definition 3.10 ($\overline{\textbf{EOP-W}k}$) *For a given policy* π*,* $\overline{\text{EOP-W}k}$ *is satisfied iff*

(a) $\nexists c \neq c'$ *such that* $\forall e$ $F(.|c,e,\pi) \succ_{ISDk} F(.|c',e,\pi)$*,*

(b) EOP-S does not hold.

The definition of $\overline{\text{EOP-W}k}$ is used to construct a new dominance criterion for equalization of opportunity. The ordinal criterion $\overline{\text{O-ezOP}}$ is satisfied, and opportunities equalized, if $\overline{\text{EOP-W}k}$ holds under policy $\pi = 1$ at least in all those cases where $\overline{\text{EOP-W}k}$ holds under $\pi = 0$. Indeed, the reverse implication has to be false.

Criterion 3.5 ($\overline{\textbf{O-ezOP}}$) *Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities according to the dominance-order criterion iff*

$$\overline{\textit{EOP-W}k} \textit{ under } \pi = 0 \quad \Rightarrow \quad \overline{\textit{EOP-W}k} \textit{ under } \pi = 1, \qquad \forall k \in \mathbb{N}_+.$$

In general, it is not said that a k exists for which $\overline{\text{EOP-W}k}$ cannot be verified. Nevertheless, if $\overline{\text{EOP-W}k}$ holds in $\pi = 0$, then it should hold also in $\pi = 1$ for validating $\overline{\text{O-ezOP}}$. On the contrary, if $\overline{\text{EOP-W}k}$ always holds under $\pi = 1$, then it is the case that a k finite exists such that $\overline{\text{EOP-W}k}$ is rejected under $\pi = 0$. Finally, suppose that a k and k' exist such that $\overline{\text{EOP-W}l}$ holds for all $l < k$ under $\pi = 0$ and for all $l < k'$ in $\pi = 1$. Then $\overline{\text{O-ezOP}}$ is equivalent to having:

$$\kappa(0) := k \ \leq \ k' =: \kappa(1),$$

where

$$\kappa(\pi) \ = \min_{c \neq c'} \max_{e} \{\kappa(c, c', e, \pi)\}. \tag{3.4}$$

Along with the ordinal criterion, we also introduce a weak distance criterion for equalization of opportunity, $\overline{\text{D-ezOP}}$. Contrary to the D-ezOP criterion, requiring that economic distance between any pair of comparable circumstances consistently decreases by effect of the policy implementation, the $\overline{\text{D-ezOP}}$ criterion is satisfied whenever, after policy implementation, for any class of preferences $\mathcal{R}^k$ there is no unambiguous increase in the economic distance between any pair of circumstances at all effort levels. This seems to be a rather natural requirement when effort has not been chosen yet.

Criterion 3.6 ($\overline{\textbf{D-ezOP}}$) *Consider a case where the comparison $\pi = 0$ and $\pi = 1$ satisfies O-ezOP, with $\kappa(0) \leq \kappa(1)$. Moving from $\pi = 0$ to $\pi = 1$ equalizes opportunities according* *to the dominance-order criterion iff for all* $k \geq \kappa(1)$*:*

$$\nexists c \neq c' \text{ such that } \forall e: \ \Delta_W(F(.|c,e,0),F(.|c',e,0)) - \Delta_W(F(.|c,e,1),F(.|c',e,1)) < 0 \ \forall W \in \mathcal{R}^k.$$

If $\overline{\text{D-ezOP}}$ is satisfied, two situations occur. Either it is not possible to make an unambiguous assessment on the direction of changes in economic distance, or, if for a pair of circumstances there is an effort level for which distance unambiguously increases, then there must exist another effort level at which distance unambiguously decreases after policy implementation.

By combining the two rather weak criteria, we obtain a third alternative ezOP criterion, denoted $\overline{\text{ezOP}}$.

Definition 3.11 ($\overline{\textbf{ezOP}}$) *Moving from* $\pi = 0$ *to* $\pi = 1$ *equalizes opportunities in the sense of* $\overline{ezOP}$ *iff* $\overline{O\text{-}ezOP}$ *and* $\overline{D\text{-}ezOP}$ *are satisfied.*

Identification from ex ante testing

Assumption 1 is not sufficient to identify the $\overline{\text{ezOP}}$ criterion by having the ex ante ezOP condition satisfied. A key assumption for identification weakens the existence conditions used in Proposition 3.9 by requiring that there exists at least one pair of circumstances whose associated luck distribution can be raked unanimously within $\mathcal{R}^k$ at all effort levels, at least under policy $\pi = 0$.[16]

Assumption 3.2 (Existence) *There exists* $c \neq c'$ *such that there exists a* k *for which* $F(.|c,e,0) \succ_k F(.|c',e,0) \ \forall e$.

The second additional assumption that we introduce is consistency: it requires that if two circumstances can be unanimously ranked at the same order at all effort level, such that $\overline{\text{EOP-W}k}$ fails under $\pi = 0$, than the same two circumstances give that $\overline{\text{EOP-W}k'}$ (for

[16] The assumption requires that in the society there are at lest a pair c and c' such that, whatever effort choice is made, type c' can never face better outcomes that type c, where “better outcome” stems from unanimity in evaluation according to the preferences in $\mathcal{R}^k$.

$k' \geq k$) fails under policy $\pi = 1$.[17]

Assumption 3.3 (Consistency) *If there exists* $(c_\pi, c'_\pi) = arg\min_{c,c'} \max_e\{\kappa(c, c, e, \pi)\}$ *with* $c_\pi \neq c'_\pi$ *and* $F(.|c_\pi, e, \pi) \succ_{\kappa(c_\pi, c'_\pi, e, \pi)} F(.|c'_\pi, e, \pi)$ *holding for all* e *and for all* π*, then* $(c_0, c'_0) = (c_1, c'_1)$.

The existence and consistency assumptions, along with the independence of effort distributions, are necessary (although non testable) assumptions to achieve identification of $\overline{\text{ezOP}}$ through the ex ante ezOP test. Hence, only a weak notion of EOP changes can be identified by implementing the ex ante test when effort is not observable. Moreover, by implementing the Weak ex ante ezOP criterion, one needs only to assume that the distribution of effort is independent on policy and circumstances to achieve identification of the $\overline{\text{ezOP}}$ criterion.

Proposition 3.10 *Under the assumptions of Effort Independence, Existence and Consistency,*

$$\textit{ex ante exOP} \quad \Rightarrow \quad \overline{\textit{ezOP}}.$$

Proof. See appendix 3.A.9. ■

3.6 Application: assessing the opportunity equalization impact of two educational policies in France

In this section we implement the opportunity equalization criterion for the empirical evaluation of public policy. We consider two separate educational policies, taking place in different stages of the educational career of students, and we evaluate whether or not their implementation fosters equalization of opportunity.

In many empirical assessments, economists have manifested a particular interest in estimating the average treatment effect of educational policies (e.g. Angrist and Krueger 1991,

[17] This assumption simply requires that the two type remains the same before and after the policy. To take and extreme example, one should consider the children of homeless fathers compared to the children of super-millionaire fathers. If before the policy the formers are always worse off compared to the latter independently on their effort choices, than policy implementation does not alter this relation.

Card 1993), since it is commonly considered the relevant measure for evaluating policy intervention. However, educational policies may have a sizable redistributive impact on students future earnings distributions that average treatment effects may miss. In this section, we go beyond average returns from policy treatment by illustrating the effectiveness of policy intervention in equalizing earning opportunities of students coming from heterogeneous family backgrounds.

The ezOP criterion can be used to evaluate a large spectrum of policies in optimal taxation (Roemer et al. 2003), health (Garcia-Gomez et al. 2012) or policies designed to alleviate poverty (Van de gaer et al. 2011). Our focus is on educational policies, which have been considered by economists and policy makers as one of the major mechanisms to equalize achievement opportunities for children coming from diverse backgrounds (Meghir and Palme 2005, Björklund and Salvanes 2011) and to promote intergenerational mobility (Hanushek and Woessmann 2011). We focus in particular on two policies that point at expanding access to secondary and higher education though a compulsory schooling requirement, while affecting only a well defined subset of the students population. We use French data to illustrate the role of such policies. For a comprehensive survey and a comparative analysis of the educational policies that took place in Europe in the last 70 years, see Braga, Checchi and Meschi (2011).

It is difficult to obtain reliable and comparable data to assert the impact of the two policy considered. The French case serves at this scope. We use two important institutional changes that took place in France, that potentially covered all the interested population of students and that involved a well defined set of cohorts. The first event is the introduction of the *Loi Berthoin* (see Grenet 2012), which shifted from 14 to 16 the minimal age requirement for leaving school. The second institutional change corresponds to the events of *May 1968* (Maurin and McNally 2008). With the available data, we cannot identify the counterfactual distribution of earnings of the treated students, that is te earning expected by this group if the institutional background would not have changed. However, we can exploit the institutional changes in an instrumental variable approach to identify the impact of educational attainment at high school or university level on earnings quantiles, and then simulate policy intervention by assigning the quantile treatment effects to a target group of students.

The first subsection hereafter describes the problem of identification and simulation in detail, while the following sections illustrate the application of the equalization criterion to observed and simulated data.

3.6.1 Identification and simulation

Let us now state more formally the estimation issues involved in the process of simulating the earnings profiles after policy implementation. We use earnings as the reference outcome variable, and the quantile treatment effects (QTE hereafter) will be measured therefore in Euros. Simulation is grounded on a two steps procedure: in the first step, we estimate the QTE of education on earnings for the group of compliers; in the second step we simulate the policy change by assigning the QTE to the quantiles of a target group of students.

The first step consists in identifying the causal impact of education on earnings quantiles. Both of the applications discussed here require to identify the QTE associated to changes in educational attainment, captured by an indicator D, on conditional earning distributions. In the first application, D indicates being treated with at least some years in the secondary education system, compared to those who do not. In the second application, D represents being treated by spending at least one year in the higher education system, compared to the case of not being treated.

The simple differences of income distributions conditional on $D = 1$ versus $D = 0$ does not measure the causal effect of being treated with D, since individuals may sort into treatment and non treatment group according to unobservables, notably ability or family background characteristics. We resort to an IV strategy to estimate causal QTE in both applications. Our identification information rests on an exogenous change in the underlying *institutional background* inducing discontinuities in the schooling age profiles in the first case (the introduction of the Loi Berthoin), or providing a quasi-natural experimental design in the second case (the May 1968 events).

The two events considered in the two applications allow to separate the population in two groups according to an instrumental variable, represented by the indicator Z. In both applications, the IV takes value one for all observations belonging to some well specified cohorts, and zero for the remaining observations. For instance, in the first application we use $Z = 1$ for those born in cohorts affected by the implementation of the Loi Berthoin, and zero otherwise. In the second application, $Z = 1$ stands for being born in the cohorts who mostly benefitted from the May 1968 events.

The intuition for Z being an instrument for treatment is that conditional on the treatment D received, the distribution of potential earnings profiles is independent on the IV (Card 2001). This is the conditional exogeneity assumption for the instrument. Our intuition is that, conditional on the chosen schooling level D, the shift in Z (which identifies different cohorts) does not have a causal impact on earnings quantiles. This is coherent with the type of unobserved heterogeneity that we would like to control for: ability, family background effects, "hard" and "soft" skills and parent investments are likely to be similarly distributed across adjacent cohorts, while these factors are likely to differ substantially for people selecting into different schooling attainment levels. Provided that our IV is exogenous, identification is achieved if the IV has a causal impact on educational choices. This is the case for the chosen IV, because both the Loi Berthoin and the May 1968 events have increased the educational attainment for some well defined generations at a precise point of the educational careers of the students. An exogenous variation of the IV induces variations in educational attainment that are orthogonal to unobserved heterogeneity, and thus allows to identify the impact of schooling on earnings *as if* heterogeneity is kept as fixed.

The QTE identified by the IV strategy do not apply to the whole population, but only to the *marginal students* (or compliers). This is the group of students that are at the margin of the education system and that are compelled, or are given the possibility, to change their schooling choices by effect of changes in the underlying institutional background. The marginal students are the high school dropouts in the case of the Loi Berthoin application, and the students that completed (or almost completed) the secondary education system requirements for the May 1968 application. However, these groups coincide exactly with the *target students* that we would like to treat with the QTE, to simulate policy intervention.

That is, the students who dropped out from the educational system at some point, and that we would induce to remain some additional years with the simulation. The expected returns for these students coincide with the returns of the group of marginal students, which are identified by the IV model.

We use the conditional IV model in Abadie, Angrist and Imbens (2002). We condition the model for a trend in the survey year (which captures time fixed effects) and other selected covariates.[18]

In this setting, the DiD strategy, a potential alternative to the IV method, which is discussed in appendix B.3, is not appropriate. Identification in the DiD setting rests upon a very strong exogeneity requirement: the distribution of the unobservable heterogeneity is assumed to be fixed within a given cohort, while it can vary across cohorts. In our case, D is an indicator of the schooling achievement, while G (to use the notation in the appendix) is an indicator for cohorts exposed to the events of May 1968 or to the changes introduced by the Loi Berthoin. The exogeneity assumption of the DiD model excludes the exogeneity assumption behind the IV model, since it imposes that unobserved heterogeneity is constant across educational levels, an assumption that is hardly acceptable.

The second step of our procedure consists in combining together the estimated QTE and the observed distribution of earnings of the group of marginal students, also referred to as the target group. In our applications, the target group can be exactly identified in the data.

The simulation is developed in three stages. In the first stage, we detrend the observed earnings by the impact of the cohort of birth and the year of survey, so that all earnings in our database are made comparable. Using the detrended data, we isolate the empirical distribution of earnings of the target group, and we associate the observations in each quantile interval of the target group to the quantile interval of earnings distribution where their income falls into.

[18] These covariates are represented by a set of indicators for the background of origin, and an effort indicator for hours worked. In the second model, we also control for a polynomial function of age left full time schooling above/below the threshold imposed by the law. This is a natural requirement in a regression discontinuity design approach. The model is estimated by using the Stata `ivqte` command by Frölich and Melly (2010).

In a second stage, we treat all observations in the same quantile of the target group with the QTE associated to that quantile. Hence, we simulate policy treatment by attributing its expected impact on the earnings of a selected target group in the population.

In the third and final stage we construct the earning distributions conditional on circumstances by using the *whole sample*. When the observed values of the target group are used, along with the rest of the observed earnings, we obtain a nonparametric estimator of the earning distributions before policy implementation ($\pi = 0$). When the simulated earnings of the target groups are used, along with the observed earnings of the remaining population, we obtain a nonparametric estimator of the earning distributions after policy implementation ($\pi = 1$). The equalization test is performed using these distributions. Further details on the data and the institutional variations exploited in this section are described for each application separately.

3.6.2 Application I: the impact of widening access to secondary education

The objective of this section is to simulate the effect of policies that provide incentives for students to pursue their educational careers in secondary school, up to completion. These policies should, for instance, provide means to reintegrate high school dropouts in the secondary education system. Using actual and simulated distributions we can assess whether widening accessibility to secondary education equalizes opportunities in the sense of the ezOP criterion. Our intuition is that, if a relevant majority of the treated students has grown in a disadvantaged background, then after policy implementation their earnings profiles should look more similar to the earnings profiles associated to students coming from more advantaged backgrounds. In fact, this policy would have no impact on earnings for those who would have achieved at least an high school degree even without the policy.

We simulate this policy by treating all the individuals in a target group, corresponding to all the students that do not complete secondary education, but who have already completed the primary education cycle. This can be done only if returns from high school participation are identified.

The *Loi Berthoin*

To identify the effect of additional years spent in high school on earnings, we exploit the rise of the minimum school leaving age from 14 to 16 years introduced by the *Loi Berthoin* in 1959 in France. The Loi Berthoin interested all the students born from January 1, 1953 onwards. Moreover, the Loi Berthoin organized the French school system as we know it now. A description of the changes after the introduction of the law can be found in Grenet (2012). We use the impact of the reform in a *regression discontinuity design*, where the discontinuity is given by an exogenous shock that affects the minimal mandatory school leaving age. In particular, the students in their 14th year of life born after 1953 who would have dropped out of the school are now compelled to take two additional years of education, presumably during the high school period. We exploit this exogenous shift to identify the effect of staying longer in high school, up to completion.

Grenet (2012) exploits the Berthoin's reform to evaluate the causal impact of years of education (a continuous variable) on earnings. He uses an indicator function taking value one for those born after the cutoff date 1953 as an instrument (IV) for age left full time schooling. Grenet makes use of an IV strategy within a regression discontinuity design analysis, and he identifies the average returns from education only at the discontinuity. In fact, the Loi Berthoin induces an exogenous shift in schooling age that is likely to affect the most (if not exclusively) the group of marginal students that would have dropped out from the system in the absence of the policy, thus allowing to identify, for this group, the returns from spending few years into the secondary education system on their earnings. Estimates of the average returns from education reveal that age left full time schooling has a very narrow and statistically insignificant impact on earnings.

We use a similar strategy as in Grenet (2012) to estimate the causal effect of age left full time schooling on earnings quantiles. Contrary to Grenet, however, we do not look for the impact of one additional year spent in school on earnings, but rather at the effect of staying in school until high school completion. This is actually the educational stage that may be interested by implementation of the type of policy that we aim at simulating. It coincides with the French *lycée* period.

In this application, the *policy treatment* indicator takes value one for the observations who spent at least one year in the secondary education system starting from the age of 15 and above. The indicator takes value zero if the observation dropped out from the educational system just at (or before) the limit minimum age, before policy implementation (that is when he was 14 years old). With this indicator, we capture only the effect of spending some years in high school, while leaving aside concerns on the indirect additional effect that high school completion may have on earning. For instance, high school completion is a necessary condition for entering university or obtaining a certificate for leave (the *baccalauréat*), and the probability of attaining this level of education (and possibly higher earnings) may be modified by policy implementation.

To retrieve a measure comparable to the effects in Grenet (2012), we trim the estimating sample to individuals with an observed educational level lower than high school diploma. With this operation, we are sure to capture the sole effect of high school participation by differentiating out the earnings profiles associated to individuals that completed the secondary education versus those who did not. Trimming is not problematic in this case, because it allows to preserve the group of compliers. In fact, the policy obliges those of cohort 1953 who are 14 years old to spend two additional years and entering the secondary education system (from the age of 15 to 16 at least). Those that go beyond this education level are most likely to be unaffected by the changes in the IV, and therefore do not contribute to the identification of QTE.[19]

We use the cutoff date of January 1, 1953 to identify the IV variable. The IV defines a *treatment* group, denoted by all the individuals born in the cohorts between 1953 to 1955, and a *comparison* group given by individuals born in cohorts 1950 to 1952.[20] The usual conditional independence assumption for the IV states that the unobservable factors affecting the earnings of those that stay longer in high school may differ from the distribution

[19]We do not consider the problem of grade repetition, probably not a relevant phenomenon for those in their fifteens (grade ten).

[20]Cohorts born in 1949 and before may be affected by the impact of the May 1968 events, while we use cohorts up to the year 1955 to construct the comparison groups with a comparable sample size to the treatment group.

of the same factors in the group of the dropouts, although the unobserved heterogeneity is required to be identically distributed in cohorts born immediately after or immediately before the 1953 cutoff date.

Changes in mandatory minimal age requirements fulfill by construction the monotonicity assumption (discussed in appendix B.3), because the policy obliges students that otherwise would have dropped out the system to receive more years of education. This last observation, along with the conditional IV independence permit to identify causal estimates of the impact of spending some years in high school on the distribution of earnings, which can be estimated by resorting on the IV model for quantile treatment effects in Abadie et al. (2002), discussed in detail in appendix B.3.

Data

Educational and labor market outcomes for the cohorts considered in this study can be illustrated using the French Labor Force Survey (LFS, *Enquete Emploi* distributed by INSEE) for the years 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010. The sample is a rotating panel, therefore we select only particular years of the survey to preserve exclusively the cross sectional information.[21] The LFS is a large representative sample of the French population of age 15 and above. There are on average 15,000 respondent per cohort in our pooled sample.

To estimate quantile treatment effects around the discontinuity in age left full time education, we use the cutoff date of 1953 as predicted by the Loi Berthoin implementation. Our sample is restricted to French male workers born between 1950 and 1955, for a total of 26,421 observations, equally distributed across cohorts.

Within this sample, we consider two distinct groups defined by the IV. We use individuals born in 1953 to 1955 as the treatment group (and, as a consequence, the instrumental

[21] The panel rotation frequency was of three years before 2003 and earnings information are available only after 1990. This explains the choice of the years 1990, 1993, 1996 and 1999. Moreover, the rotation frequency after 2003 changed to one year and a half (that is, one-sixth of the sample is replaced every trimester). Picking up information every two years allows to deal with a renewed sample, as in years 2004, 2006, 2008 and 2010. The year 2002 is not exploited due to imperfections in the data collected.

variable IV). The three cohorts correspond to students that entered the French education system after the introduction of the Loi Berthoin reform. The treatment group amounts to roughly less than half of the whole sample.

As motivated by Grenet (2012), the policy induced a significant increase of roughly one year in age left full time schooling for cohorts born after 1953, with respect to older cohorts. This result is also illustrated in table 3.12, where differences in education and age of leaving school are significantly different between the treatment and the comparison groups. The proportion of students who received the policy treatment (longer staying in secondary education) is also significantly higher in the treatment group, thus explaining the reduction in the size of the target group (which shifts from 27% to 16% of the students population). Treatment and control groups are otherwise similar according to a variety to characteristics reported in table 3.12.

To estimate the quantile treatment effects, we make use of a trimmed subsample of 17,779 observations. Those are the observed individuals in our sample who at most received an high school diploma. The repartition of this subsample into the treatment/control groups is reported in table 3.4.

The outcome used to measure opportunities is monthly earnings after taxes. We partition the distribution of earnings in the sample of interest into twenty groups of 5% population mass, thus defining twenty quantiles, and for each of these quantiles we estimate the treatment effects from policy treatment. Selected quantiles of the overall earnings distributions, as well as for distributions made conditional upon treatment groups status (IV) and policy treatment (High/Low education) are reported in table 3.4 for the sample of cohort 1950 to 1955.

Estimation results

As shown in table 3.4, the differences between earnings quantiles of treated and non treated observed individuals with higher education are sizable. However, these differences are similar across treatment and control groups. This indicates that the treatment effect that can be identified is probably low or statistically not significant.

Table 3.4: Descriptive statistics: earnings by treatment (IV) and policy treatment

Earnings (Monthly)	Overall	Treatment (IV=1)		Control (IV=0)	
		H-educ	L-educ	H-educ	L-educ
	(1)	(2)	(3)	(4)	(5)
Q5%	426.9	387.2	416.0	450.0	472.6
Q10%	762.2	762.2	686.0	800.4	731.8
Q25%	985.6	990.9	911.1	1,067.1	914.7
Q50%	1,219.6	1,250.0	1,092.8	1,311.1	1,112.9
Q75%	1,550.0	1,585.0	1,402.5	1,676.9	1,402.5
Q90%	2,058.1	2,000.0	2,200.0	2,134.3	1,900.0
Q95%	2,500.0	2,400.0	3,000.0	2,500.0	2,591.6
Mean	1,395.0	1,383.9	1,436.6	1,458.5	1,285.4
	[2,160.3]	[1,977.2]	[3,771.0]	[2,185.2]	[886.9]
Overall sample size	26,421	7,357	6,276	5,513	7,275
Estimation sample size	17,779	7,357	1,785	5,513	3,124

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Trimmed sample reduced to French male earners where circumstances have been recorded, cohorts 1950 to 1955. Income quantiles are measured in Euro. IV is a dummy for cohorts 1953 and 1955. Treatment and comparison are defined upon the IV. *Policy treatment* is determined by access to at least one year of secondary education, with age left at school (H-educ) versus no high school (L-educ). Standard deviations reported between brackets. Sample size for the trimmed sample are reported in the final row (*Estimation*). This sample is used to implement the IV QTE estimator.

This conjecture is demonstrated by the results in table 3.5. For a selected number of quantiles we report the IV estimates of the quantile treatment effects for the overall sample and for the sub-samples conditional on circumstances. Despite the important share of compliers, it is not possible to identify a significant effect of the educational indicator for population percentiles that range out of the 40% to the 80% quantiles intervals. Moreover, the effects estimated in this interval are quite small and are significant only for the subsample of Circumstance 2 (those coming from the most disadvantaged background).

The estimated QTE is identified for the group of marginal students, and the full list of QTE estimates is reported in figure 3.2(a). We use these marginal effects, conditional on the individual position in the target group, to simulate a policy change. This is done by assigning the estimated treatment effects to the earnings quantiles intervals of the target group. We use only significant effects computed in the overall population (model (1) in

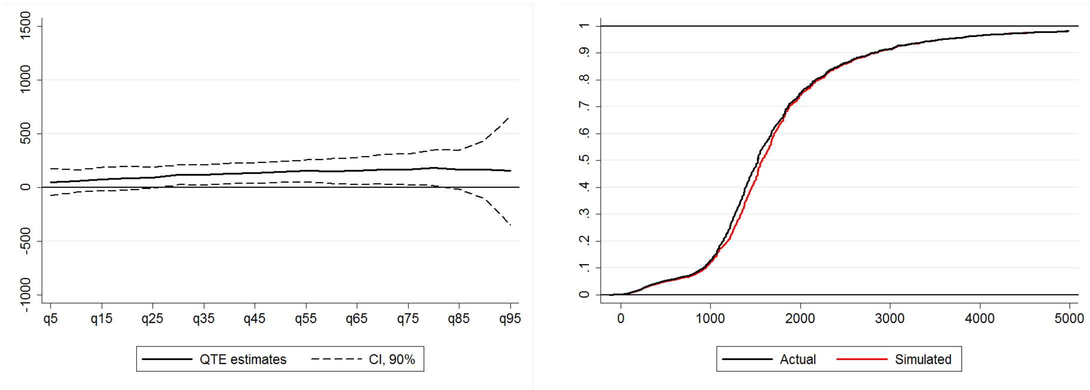
Table 3.5: Quantile treatment effects, IV estimator

Independent variable: Earnings	Overall	Conditional			
		Circ. 1	Circ. 2	Circ. 3	Circ. 4
	(1)	(2)	(3)	(4)	(5)
Treatment Q5%	49.5	-0.8	48.7	24.7	53.7
	(76.9)	(248.9)	(86.5)	(133.6)	(254.3)
Treatment Q10%	57.5	-47.2	53.4	76.2	1.5
	(61.7)	(245.3)	(63.4)	(139.0)	(243.1)
Treatment Q25%	90.6	-93.8	76.2	104.2	135.1
	(59.1)	(674.5)	(59.0)	(147.8)	(437.1)
Treatment Q50%	142.3**	45.7	126.4**	157.2	7.7
	(58.8)	(720.8)	(62.4)	(162.9)	(333.0)
Treatment Q75%	167.7*	-187.3	179.9	155.8*	-152.4
	(88.3)	(697.5)	(100.3)	(188.1)	(542.8)
Treatment Q90%	167.7	-759.6	228.7	228.7	-457.3
	(165.5)	(1,978.8)	(174.3)	(321.4)	(1,035.8)
Treatment Q95%	157.4	-1,021.4	167.7	213.4	-643.8
	(306.9)	(1,409.6)	(278.5)	(640.9)	(1,145.9)
Controls (reported at Q50%)					
$(cob - 1953)^2$	11.0	-48.0	4.8	29.4	-18.0
	(29.7)	(319.3)	(30.2)	(81.2)	(243.0)
$(cob - 1953)^4$	-0.8	2.2	-0.4	-2.7	0.2
	(2.8)	(29.2)	(2.8)	(8.1)	(25.3)
Circumstance 1	-0.0				
	(179.0)				
Circumstance 3	52.6				
	(45.0)				
Circumstance 4	116.9				
	(144.4)				
Survey year (FE)	yes	yes	yes	yes	yes
Sample size	17,779	981	11,351	3,720	1,727
Compliers (%)	18.1	12.7	21.2	15.8	11.4

* $p < .10$, ** $p < .05$, *** $p < .01$ (one-tailed).

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Trimmed sample reduced to French male earners where circumstances have been recorded, cohorts 1950 to 1955 and trimmed to the observations with at most high school degree. The dependent variable measures earnings in 1999, once year effect has been eliminated. Robust standard errors are reported in parenthesis.

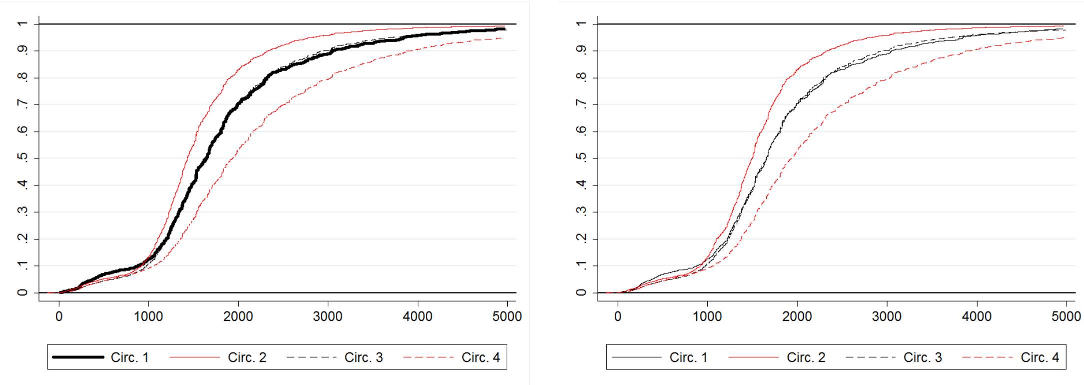


(a) QTE, by 5% quantiles

(b) *cdfs* before/after treatment

Figure 3.2: Quantile treatment effects of the impact of access to secondary education on earnings. Estimates based on the cohorts 1950 to 1955 of French male earners (trimmed sample). Cohorts 1953, 1954 and 1955 define the IV, participation to the higher education system is the policy treatment variable. In panel (a), quantile treatment effects are computed at 5% income intervals (IV estimator), the CI at 90% is computed with robust standard errors. Controls: cohort trends, year of survey, a quartic polynomial of the gap between year 1953 and last year spent in school, and circumstance dummies. Empirical *cdfs* in panel (b) are obtained for detrended earnings data (actual) an by providing policy treatment by quantile of earnings for the *marginal students* (simulated).

table 3.5). The target groups consists in the students in the age interval 11 to 15 years old who did not enter in the secondary education system. This group amounts to 5,585 out of the 26,421 individual observations of the whole sample. Selected quantiles of the earnings distributions before (the status quo, observed in the sample) and after (when target groups are treated with quantile treatment effects) policy implementation are reported in table 3.13 in the appendix. In model (1) of the table are reported the earning quantiles of the whole population. This model may predict overall distributional changes induced by the policy. Models (3) to (6) are more important for our analysis, since they show the differences in the earning distribution of the different subgroups (by family background circumstances) that one can construct from the sample used for model (1). The ezOP criterion is implemented by comparing these distributions before and after policy simulation.



(a) Actual data ($\pi = 0$)   (b) Simulated data ($\pi = 1$)

Figure 3.3: Empirical *cdf*s are obtained for earnings data (a) and by treating the target group of 11 to 15 years old students (b).

Equalization of Opportunity: results

A policy whose objective is to grant to drop out students the possibility to spend at least some additional years in the secondary education system has a very narrow and often non significant impact on these students' earnings profiles. The overall impact of such policy on the population distribution of earnings is shown in figure 3.2(b). When policy impact is redistributed across circumstances, there is no evidence of changes in the earnings patterns associated to different circumstances. Figure 3.3 clearly illustrates this point.

Both before and after policy simulation it is almost always the case that circumstances pairs can be ranked according to first order stochastic dominance. Only Circumstance 1 and Circumstance 3 seem not to be affected by schooling expansion. As shown in figures 3.10 and figure 3.11, the patterns of the differences in earnings across pairs of circumstances, the differences in the *GL* curve ordinates and their integrals are always positive along the earnings percentiles domain. The signs of the differences of these curves, which serves at identifying the gap dominance, are however less conclusive because the curves fluctuates around the zero effect threshold, as shown in figure 3.12.

The observations made by this short overview of the results based on graphical devices are confirmed by the dominance tests listed in table 3.6. We test six comparisons between distinct pairs of circumstances, reported by row, both before policy and after policy simulation. In columns (1) and (2) we report, for each policy and for each pair of circumstances, the direction and the minimal degree of ISD that cannot be rejected by the data at a 5% confidence level. The sequential methodology for testing dominance is coherent with the requirements of the ex ante ezOP Algorithm. For instance, one has to read the first dominance relation in (1) as $Circ.1 \succ_{ISD1} Circ.2$ (but not the inverse) under $\pi = 0$.

For any pair of circumstances, the direction of the advantage as measured by ISD is unaffected by policy implementation. Circumstance 1 provides an unambiguously higher advantage compared to Circumstance 2, according to ISD1. This result reflects possibly a substantial heterogeneity in family background for the group of people with non French fathers. In fact figure 3.10 clearly shows that the bottom decile of the distribution of the group with Circumstance 1 is disadvantaged compared to the group with Circumstance 2, while the ranking reverses for the remaining deciles. The comparison between Circumstance 1 and Circumstance 3 cannot be verified according to ISD1. For the two circumstances, it is necessary to test dominance up to the order ISD3, which is verified both before and after policy intervention. It is nevertheless possible to rank unambiguously the Circumstances 2, 3 and 4 (French father, different socioeconomic classes) according to ISD1 both before and after policy simulation. This result shows that the policy has no impact in reducing agreement over the direction of the disadvantage, nor on changing the direction of disadvantage itself. Notice that it is always possible to rank the distributions or their integrals, since equality tests across pairs of circumstances are always rejected under both policy regimes.

In column (3) of table 3.6 we report the result for the distance comparison. Gap curve dominance relations are tested at 5% significance level. The results for the gap curve dominance are reported in column (4) of table 3.11. For a given model, we report the minimal order at which it is not possible to reject, with a confidence of 5%, that the gap curve generated by that model is either statistically equal to zero, or it always lie above the zero line for all the considered quantiles. The tested model, reported in brackets, gives the order of differentiation of circumstances' earnings distributions under each policy regime, which allows to conclude in favor of gap dominance. For instance, the model associated

Table 3.6: Equalization of Opportunity test: Ordinal and Distance criteria for high school expansion policies.

Circ. c vs Circ. c'	Before policy	After policy	Difference-in-differences	
	$(\pi = 0)$	$(\pi = 1)$	$\Delta_W(F_0^c, F_0^{c'}) - \Delta_W(F_1^c, F_1^{c'})$	Gap dominance k (tested model)*
	(1)	(2)	(3)	(4)
Circ. 2 vs Circ. 1	$\prec_{ISD1}$	$\prec_{ISD1}$	$\geq 0\ \forall W \in \mathcal{R}^1$	> 0 for $k = 1$ (12-12)
Circ. 3 vs Circ. 1	$\succ_{ISD3}$	$\succ_{ISD3}$	$= 0\ \forall W \in \mathcal{R}^3$	$= 0$ for $k = 1$ (31-31)
Circ. 4 vs Circ. 1	$\succ_{ISD1}$	$\succ_{ISD1}$	$= 0\ \forall W \in \mathcal{R}^1$	$= 0$ for $k = 1$ (41-41)
Circ. 3 vs Circ. 2	$\succ_{ISD1}$	$\succ_{ISD1}$	$\geq 0\ \forall W \in \mathcal{R}^1$	> 0 for $k = 1$ (32-32)
Circ. 4 vs Circ. 2	$\succ_{ISD1}$	$\succ_{ISD1}$	$\geq 0\ \forall W \in \mathcal{R}^1$	> 0 for $k = 1$ (42-42)
Circ. 4 vs Circ. 3	$\succ_{ISD1}$	$\succ_{ISD1}$	$= 0\ \forall W \in \mathcal{R}^1$	$= 0$ for $k = 1$ (43-43)

	Opportunity amelioration
	Δ Policy impact
Overall	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$
Circumstance 1	After $(\pi = 1) \sim_{ISD1}$ Before $(\pi = 0)$
Circumstance 2	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$
Circumstance 3	After $(\pi = 1) \sim_{ISD1}$ Before $(\pi = 0)$
Circumstance 4	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$

Source: Estimates from Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Earnings distribution corrected by the age effect. Sample reduced to male French earners where circumstances have been recorded, cohorts 1950 to 1955. Circumstances defined by father status: *Circ. 1*: foreign born; *Circ. 2*: French farmer or manual worker; *Circ. 3*: French artisan or non-manual worker; *Circ. 4*: French professional.

EzOP tested at 5% significance level on a selected sample of twenty quantiles. Both inverse stochastic equality and dominance null hypothesis have been separately tested for any degree $k = 1$ up to 5. Only the minimal degree of dominance $\kappa(c, c', \pi)$ is reported. The notation $c \succ_{ISD\kappa} c'$ means that the earnings distribution of circumstance c ISD at order κ the earnings distribution of circumstance c'. The distance test is defined over the class $\kappa(c, c', \pi = 1)$. Direction of the gaps dominance is reported, along with information on the direction of distance for the model verifying gap dominance. ISDk for $k = 1, 2$ is estimated as in Beach and Davidson (1983), while for $k \geq 3$ tests are constructed by using the asymptotic estimators of the vector of conditional Gini SWF.

*: For model $(ij - kh)$ we tested the gaps curve of circumstances i vs j in $\pi = 0$ minus the gaps curve of k vs h in $\pi = 1$, exclusively for configurations $k = i$ and $h = j$ or $k = j$ and $h = i$.

to circumstances Circ. 2 and Circ. 1 is $(12 - 12)$, which means that to find dominance in gap curves at order one it is necessary to take the difference of the earnings distribution of Circumstance 1 minus the earnings distribution of Circumstance 2 both under policy $\pi = 0$ and policy $\pi = 1$. Otherwise, alternative models for gap dominance always reject the null hypothesis of equality or dominance even at orders of inverse stochastic dominance higher than one.

Despite the inconclusive result of the ordinal criterion, we find evidence that the gap curve dominance at the first order cannot be rejected at the 5% confidence level for all the pairs of circumstances. The gap curve dominance tests are coherent with the direction of advantage measured by ISD under both policy regimes, although for many comparisons the change in distance is statistically zero (that is, the gap curve coincides with the zero line). This result is coherent with the fact that the simulated policy has no sizable impact on the earnings distribution of Circumstances 1 and 3. The distance between Circumstance 2 and the Circumstances 3 and 4 is reduced by effect of policy simulation, while the distance between Circumstances 1 and 4 remains unaffected. This result is consistent with the fact that an expansion of the secondary education system provides benefits for students coming from more disadvantaged backgrounds, who can catch up the differences with the rest. The policy does not have a statistical impact on the distribution associated to Circumstance 1. As a result, the gap between Circumstance 1 and Circumstance 2 decreases at order one by effect of the policy.

We conclude that under the assumption of the rank dependent model for preferences, the ex ante O-ezOP criterion is validated by the data, although there is no apparent change in consensus due to policy simulation. Moreover, the ex ante D-ezOP criterion is also satisfied by the data. This is so because the gaps dominance cannot be rejected at order one for all pairs of circumstances using an estimation model that is consistent with the direction of the advantage underlying the ordinal criterion. For Circumstances 3 and 1, gaps dominance is verified at order one, and therefore it is also at order 3, which gives the result.

Globally, the ex ante ezOP criterion is met by our data, so we conclude that the policy aimed at increasing participation in the high school system equalizes opportunities in the sense of $\overline{\text{ezOP}}$.

3.6.3 Application II: the impact of widening access to higher education

The objective of this second application is to study the impact of a widening in the access to higher education, and to assess if this policy equalizes earnings patterns across circumstances. Our intuition is that only the students at the margin of the secondary education system would benefit from granting larger accessibility to the higher education system. Moreover, if a relevant majority of these students has grown in a (relatively) disadvantaged background, then this type of policy would make their earnings profiles more similar to the profiles associated to students coming from more advantaged families. In fact, this policy is not supposed to have an impact on earnings for those who would have achieved a university diploma or degree even without the policy being implemented.

To simulate the impact of this policy we need to treat marginal students with the true causal returns from spending some years in the higher education system for those who are at the margin. We resort on the events that took place in May 1968 in France as a *natural experiment* that allows to identify, for a well established group of cohorts, these effects.

One of the most relevant consequences of the famous events of May 1968 in France is that the normal examination procedure to obtain high school diplomas (a necessary condition to access higher education) were abandoned, so that the pass rate for various qualifications increased substantially. Moreover, there are numerous examples of delays and modifications to university examinations taking place in that year which favored in pursuing the academic career of students in the early stages of higher education. Maurin (2007, 2008) provides an exhaustive treatment of the events taking place in that period, and a discussion on the nature of the marginal students.

We use the shift in educational outcomes induced by the May 1968 events to identify the impact of the participation into higher education on quantiles of the earnings distributions of students as adults, made conditional upon the background of origin. To do so, we resort to quantile treatment effects methods and we use the IV procedure in Abadie et al. (2002).

The *May 1968* events

Maurin and McNally (2008) motivate that the May 1968 events define a quasi-natural setting for estimating treatment effects. In fact, the events took place in a very precise moment in history, and did not leave long term effects on the educational system. Hence, those who had to enter the higher education system in 1965 faced the same rules as those in 1971. Only the cohorts born in 1949 and 1948 were subject to the events of May 1968. Hence, we use cohorts 1948 and 1949 as an IV for identifying and isolate the *causal* impact of variation in accessibility to higher education on earnings. The IV defines the *treatment* group. This group has to be associated to a *comparison* group given by the adjacent cohorts of 1946 and 1952. By comparing these two groups, the impact of other potential factors is in large part neutralized.[22]

The *policy treatment* is represented in our study by the possibility of spending at least some periods in the higher education system. Selection into policy treatment is highly endogenous. Unobservable factors such as cognitive and non cognitive abilities, family background or individual motivation are not only important determinants of the schooling choices of the marginal students, but also explicative for individual earnings. Hence, conditional or unconditional QTE estimates exploiting the exogeneity assumption (see appendix B) are inconsistent. We resort on an IV strategy to recover the causal impact of higher education participation on earnings.

The key assumption for identification is that for the group of marginal students, the distributions of potential outcome with and without the treatment (higher education) are jointly independent from the assignment to the cohort of birth, which is used to construct the IV. As in Maurin and McNally (2008), it is hard to argue that cohorts 1948 and 1949 differ in many respects from cohorts 1946 and 1952, which are indeed similar also in terms of a large set of demographic and socioeconomic characteristics (as shown by the data). The main difference across groups generated by the IV is instead given by the educational

[22] Cohorts born in 1945 and before may suffer from the impact of WWII, while people born in 1953 and after have been exposed to the effect of the *Loi Berthoin*, that raised by two years (up to 16) the legal age to drop out from the schooling system.

chances that these cohorts have been given. This exogenous variation identifies the QTE.

Data

Educational and labor market outcomes for the cohorts considered in this study can be illustrated using the French Labor Force Survey (LFS, *Enquete Emploi* distributed by INSEE) for the years 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010. To estimate quantile treatment effects exploiting the quasi-natural experiment of May 1968, we restrict our sample to include only French male workers born between 1946 and 1952: the sample is therefore reduced to nearly 27,500 units, equally distributed across cohorts.

Within this sample, we consider two distinct groups. We use individuals born in 1948 and 1949 to define the treatment group (and, as a consequence, the instrumental variable IV). The two cohorts correspond to students that were 19 in 1968, and so were called to repeat the *baccalauréat* examination. These students were likely the ones more affected by the relaxation of the examination rules. Table 3.7 allows to identify the cohorts and the degrees who were more exposed to the events of May 1968. In that period, cohorts 1948 and 1949 were expected to be in the early stages of the higher education process, and a significant share of them benefitted from staying at the university at least some additional years (a formal test is reported in table 3.14). The rest of the sample seems not to show particular deviations with respect to the treatment group.

The *comparison* group is given by individuals born in 1946 and in 1952. In this way, it is possible to obtain treatment and control groups of approximatively 8,500 units each (table 3.9), for a total of nearly 17,500 observations that are used to estimate the treatment effects at various quantiles.

The *policy treatment* variable is given by access to higher education. We generate an indicator function taking value one if the individual spent at least some years in the higher education education, got a diploma (two years after *baccalauréat*) or a degree (three or more years after *baccalauréat*), both in public or private (*Grandes Ecoles*) institutions. The contributions of each of these educational groups in defining the indicator distribution by cohort of birth are illustrated in Table 3.7.

Table 3.7: Distribution of education across male workers, by cohort of birth

	Baccalauréat		Higher education system									
			All		Some years		Diploma		Degree		Grand Ecole	
Cohort:	(1)		(2)		(3)		(4)		(5)		(6)	
1946	0.119	[0.3]	0.218	[0.4]	0.029	[0.2]	0.073	[0.3]	0.116	[0.3]	0.048	[0.2]
1947	0.110	[0.3]	0.224	[0.4]	0.024	[0.2]	0.076	[0.3]	0.124	[0.3]	0.047	[0.2]
1948	0.114	[0.3]	0.213	[0.4]	0.023	[0.2]	0.075	[0.3]	0.115	[0.3]	0.045	[0.2]
1949	0.109	[0.3]	0.237	[0.4]	0.033	[0.2]	0.091	[0.3]	0.113	[0.3]	0.046	[0.2]
1950	0.114	[0.3]	0.220	[0.4]	0.033	[0.2]	0.089	[0.3]	0.098	[0.3]	0.038	[0.2]
1951	0.121	[0.3]	0.205	[0.4]	0.027	[0.2]	0.084	[0.3]	0.094	[0.3]	0.039	[0.2]
1952	0.109	[0.3]	0.205	[0.4]	0.033	[0.2]	0.079	[0.3]	0.093	[0.3]	0.042	[0.2]
Sample size	27,536		27,536		27,536		27,536		27,536		27,536	

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Sample reduced to French male earners where circumstances have been recorded, cohorts 1946 to 1952. Scores for different degrees of education (and a dummy for *Grand Ecole* in model (6)) by cohort of birth. Standard deviations reported between brackets.

The LFS database reports, for each observed individual, the information on the socioeconomic characteristics of the father when the observed individual was a children. We construct four indicators for *family background circumstances*: *Circumstance 1* collects the individuals whose father is non French, nearly 9% of the overall sample. The remaining circumstances are obtained by partitioning the sample of those with a French father according to her socioeconomic background, thus giving: *Circumstance 2* if father was a farmer or a manual worker; *Circumstance 3* if the father was an artisan or non-manual worker; *Circumstance 4* if the father was involved in a professional activity. The distribution of the four circumstances is reported in table 3.8. The same table shows that the patterns of variation across cohorts of the policy treatment variable are substantially similar among circumstances, although scaled according to the social prestige of the occupation of the father.

The outcome used to measure opportunities is monthly earnings after taxes. We partition the distribution of earnings in the sample of interest into twenty groups of 5% population mass, thus defining twenty quantiles, and for each of these quantiles we estimate the treatment effects. Selected quantiles of the overall earnings distributions, as well as for distributions made conditional upon treatment groups status (IV) and policy treatment (High/Low education) are reported in table 3.9 for the sample ranging from cohort 1946 to 1952. Figures for the estimation sample are similar.

Along with the outcomes distribution, we report in table 3.14 in the appendix the average value of other covariates, computed separately for treatment/control groups (defined by the IV). As illustrated in the table, there is convincing evidence that the two groups are statistically not distinguishable on a variety of attributes regarding individual characteristics or family background. Most importantly, the two groups differ in terms of policy treatment (higher education participation), which is statistically higher (at 5%) in the treatment group. These two facts provide sufficient evidence in favor of the identifying information behind the IV QTE estimator.

Table 3.8: Distribution of male workers with at least some years spent in higher education across birth cohorts, by circumstances of origin.

Cohort of birth	Circ. 1 (1)		Circ. 2 (2)		Circ. 3 (3)		Circ. 4 (4)	
1946	0.227	[0.4]	0.113	[0.3]	0.230	[0.4]	0.503	[0.5]
1947	0.324	[0.5]	0.108	[0.3]	0.246	[0.4]	0.536	[0.5]
1948	0.259	[0.4]	0.111	[0.3]	0.256	[0.4]	0.459	[0.5]
1949	0.286	[0.5]	0.120	[0.3]	0.291	[0.5]	0.510	[0.5]
1950	0.286	[0.5]	0.111	[0.3]	0.258	[0.4]	0.470	[0.5]
1951	0.272	[0.4]	0.097	[0.3]	0.225	[0.4]	0.506	[0.5]
1952	0.294	[0.5]	0.103	[0.3]	0.242	[0.4]	0.459	[0.5]
Sample size by cohorts								
1946-1952	1,637		14,566		6,725		4,608	
1946, 1948, 1949 and 1952	931		8,267		3,795		2,626	

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Sample reduced to French male earners where circumstances have been recorded, cohorts 1946 to 1952. The policy treatment variable is an indicator equal to one if the sample unit has spent at least one year in the higher education system, owns a diploma or a degree. Circumstances are defined by father socioeconomic status: *Circ. 1* if father does not hold French nationality, *Circ. 2* if father is French and worked as a farmer or manual worker, *Circ. 4* if father is French and worked as an artisan or non manual worker, *Circ. 4* if father is French and worked in a professional activity. Standard errors are reported in brackets. Sample's size under two alternative sample designs (depending on the cohorts analyzed) is also reported.

Estimation results

The IV estimates of the quantile treatment effects for selected quantiles are reported in table 3.10. The model (1) is estimated on the restricted sample of male French workers in cohorts 1946, 1948, 1949 and 1952. QTE estimates are obtained by using cohorts 1948 and 1949 as an IV. The model is based on a propensity score matching, estimated by controlling for age and cohort trends as in Maurin and McNally (2008). Moreover, we control for an indicator for survey years (taking value one if the unit is surveyed after 1999) to capture age and cohort trends on earnings. The whole set of quantile treatment effects in model (1) are reported in panel (a) of figure 3.4. Treatment effects are positive, significant and increasing with wage for the largest part of the sample.

Table 3.9: Descriptive statistics: earnings by treatment (IV) and policy treatment

Earnings (Monthly)	Overall	Treatment (IV=1)		Control (IV=0)	
		H-educ	L-educ	H-educ	L-educ
	(1)	(2)	(3)	(4)	(5)
Q5%	592.9	990.9	579.3	869.0	502.8
Q10%	838.5	1,295.8	821.1	1,217.0	792.7
Q25%	1,067.1	1,676.9	1,028.3	1,600.7	1,006.2
Q50%	1,417.8	2,234.8	1,300.0	2,167.9	1,295.8
Q75%	1,981.8	3,049.0	1,699.9	3,049.0	1,676.9
Q90%	2,913.4	4,500.0	2,289.4	4,408.0	2,286.7
Q95%	3,811.2	5,793.0	2,999.3	5,495.5	2,800.0
Mean	1,801.9	2,756.4	1,624.2	2,721.2	1,511.9
	[3,812.6]	[4,978.0]	[4,432.1]	[4,955.6]	[2,856.3]
Overall sample	27,536	1,788	6,140	4,193	15,415
Estimation sample	15,619	1,788	6,140	1,620	6,071

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Sample reduced to French male earners where circumstances have been recorded, cohorts 1946 to 1952. IV is a dummy for cohorts 1948 and 1949. Treatment and comparison are defined upon the IV. *Policy treatment* is determined by access to at least one year of higher education (H-educ) versus secondary or lower schooling achievement (L-educ). Standard deviations reported between brackets. Sample size for Cohorts 1946, 1948, 1949 and 1952 reported in the final row (*Estimation sample*). This sample is used to implement the IV QTE estimator.

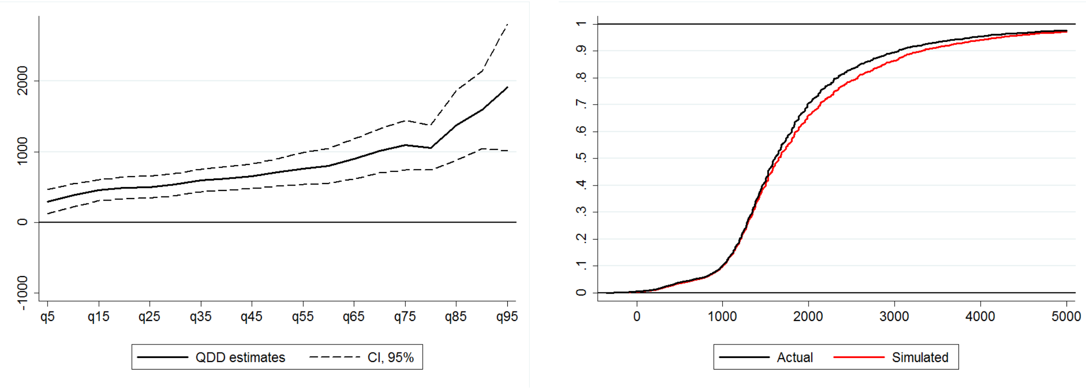
Table 3.10: Quantile treatment effects, IV estimator

Independent variable: Earnings	Overall	Conditional			
		Circ. 1	Circ. 2	Circ. 3	Circ. 4
	(1)	(2)	(3)	(4)	(5)
Treatment Q5%	292.1***	26.2	-689.1	365.9*	284.2
	(86.7)	(420.0)	(172,419.0)	(164.3)	(310.2)
Treatment Q10%	381.1***	127.0	-672.2	431.9*	333.5
	(83.9)	(603.9)	(.)	(190.9)	(249.7)
Treatment Q25%	499.7***	455.2	583.0***	487.8	503.1
	(79.0)	(438.0)	(75.5)	(256.2)	(269.9)
Treatment Q50%	706.0***	701.3	661.6***	749.2**	617.8*
	(98.9)	(553.9)	(106.9)	(263.2)	(301.8)
Treatment Q75%	1,092.6***	1,216.3*	1,145.0***	1,076.2**	1,049.4**
	(179.8)	(560.6)	(227.3)	(398.6)	(365.2)
Treatment Q90%	1,591.2***	1,566.5	1,720.4***	1,676.9**	-412.3
	(280.4)	(1,292.3)	(396.7)	(624.5)	(492.9)
Treatment Q95%	1,910.7***	1,901.4	2,396.1**	2,082.3*	1,829.4*
	(457.2)	(2,887.0)	(798.8)	(866.6)	(833.4)
Controls (reported at Q50%)					
Age trend	36.4***	27.9	34.0**	33.9	73.5
	(10.0)	(45.8)	(11.8)	(26.2)	(45.0)
Cohort trend	19.5	73.7	504.1	-0.0	-77.3
	(65.8)	(218.7)	(3,694.7)	(157.8)	(483.5)
Circumstance 1	109.1				
	(131.5)				
Circumstance 3	139.6				
	(77.5)				
Circumstance 4	286.7*				
	(128.2)				
Survey year (FE)	yes	yes	yes	yes	yes
Sample size	15,619	931	8,267	3,795	2,626
Compliers (%)	3.9	4.3	4.4	6.1	.5

* $p < .10$, ** $p < .05$, *** $p < .01$ (one-tailed).

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Sample reduced to French male earners where circumstances have been recorded, cohorts 1946, 1948, 1949, 1952. The dependent variable measures earnings in 1999 Euro, once year effect has been eliminated. Robust standard errors are reported in parenthesis.



(a) QTE, by 5% quantiles

(b) *cdf*s before/after treatment

Figure 3.4: Quantile treatment effects of the impact of access to the higher education system on earnings. Sample constructed on cohorts 1946, 1948, 1948 and 1952 of male earners. Cohorts 1948 and 1949 define the IV, participation to the higher education system is the policy treatment variable. In panel (a), quantile treatment effects are computed at 5% population tranches, IV estimator. The CI at 95% is computed using robust standard errors. Controls: age and cohort trends, year of survey and circumstances dummies. Empirical *cdf*s in panel (b) are obtained for row earnings data (actual) an by providing policy treatment by quantile of earnings for the *marginal students* (simulated).

This is a conditional quantile model. This means that the estimated treatment effects only have a meaning if attributed to the respective quantiles of the earning distributions made conditional upon the same covariates. Our covariates are either age or cohorts trends, or years of survey fixed effects, that we use to eliminate the impact of common trends across cohorts. This is not a big issue for retrieving the unconditional QTE, since we will consider outcome distributions that are detrended from cohorts and year effects in our study. Therefore, the conditional quantiles estimated by this model can be correctly associated to the appropriate quantiles of the detrended earnings distributions.

In order to check the robustness of our estimates, we also compute estimates of the treatment effects separately for each circumstance group, using similar controls as in model (1). Results are reported in table 3.10, models (2) to (5), for selected quantiles, while the whole set of estimates is represented in figure 3.14 in the appendix. The patterns identified by the circumstances sub-samples are consistent with the estimates in the whole sample, with the notable exception of Circumstance 1, for which the QTE are all statistically zero.

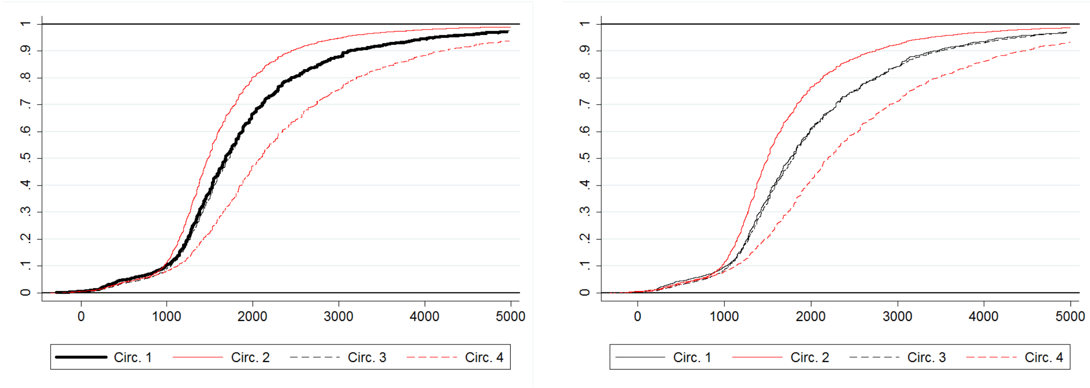
Using the sample of individuals born between 1946 and 1952, we derive a consistent estimator of the earnings cumulative distribution function given before policy implementation. To obtain the simulated distributions, we proceeds in two steps. First, we target only the students at the margin of the higher educational system, who own a *baccalauréat* or another secondary education diploma, but who did not proceed further to spend some years at the university. Second, we assign quantile treatment effects to the corresponding quantiles of the *target group*, to simulate the expected distribution of earning after widening the accessibility to the higher education system. Treatment effects vary accordingly to the position of the target group in the overall earnings distribution.

We stress here that it is not possible to interpret QTE at quantile p as a measure of the impact of the May 1968 events on on individual earnings, but rather on the same quantile of the earning distribution of the target group. These people are treated with quantiles estimated for the entire population, rather than repeating the simulation exercise for each circumstance separately.

The distributions of earnings before and after simulating the policy are summarized in figure 3.4. Given that treatment effects are significantly positive and increasing with income quantile, there is an unambiguous positive effect of policy treatment on earnings, concentrated in the central and top quantiles of the initial distribution. These results are mainly driven by the interaction between the distribution of the target group across earnings quantiles, more concentrated in quantiles above the median (see figure 3.13(a) in the appendix), and the distribution of treatment effects, which are particularly high above the median. The positive effect of the treatment is, however, not evenly distributed across circumstances.

The earnings for sample subgroups defined by individual circumstances are reported in figure 3.5, for the actual situation (panel (a)) and after implementing the widening in higher education access (panel (b)). Earnings prospects associated to all circumstances are unambiguously improved by policy implementation. Nevertheless, these gains are unequally distributed across circumstances and quantiles.

The earnings profiles associated to Circumstances 1 and 3 are the most affected by



(a) Actual data ($\pi = 0$) (b) Simulated data ($\pi = 1$)

Figure 3.5: Empirical *cdfs* are obtained for row earnings data (a) an by simulating policy treatment by quantile of earnings for the *marginal students* (b).

the policy, because these two circumstances and the target group are both concentrated above the earnings' median. The earnings profile of Circumstance 4 before and after policy simulation are, on the contrary, very stable, because the type defined by this circumstance is more concentrated among top 75% earners, where the target group is less represented.

Figure 3.5 shows that the distance between the patterns of the earnings profiles associated to Circumstances 2, 3 and 4 seems to be reduced by effect of the policy. Meanwhile, the three earnings cumulative distribution functions diverge substantially from the profile of Circumstance 1. The ezOP criterion will verify if policy simulation has in fact equalized opportunities with respect to the observed situation.

Equalization of Opportunity: results

A policy whose objective is to extend accessibility to the higher educational system has sizable, positive and significant effects on the earnings percentiles of the group of marginal students. However, a large part of the simulated gains from this policy are concentrated in the top quantiles, as shown in figure 3.4(b). It is not possible to associate the target group's members only to a particular subset of background circumstances. As a consequence the positive impact of the policy simulation is mapped by an amelioration of opportunities for all circumstances. In fact, it is not possible to reject at conventional significance levels that the simulated earnings distributions associated to different circumstances first order stochastically dominates the respective distributions constructed from the observed data.

Figures 3.15 and 3.16 show the distribution across quantiles of the differences (and their integrals) in the earning distributions for pairs of circumstances across population percentiles. It is possible to identify a clear pattern of dominance across pairs of circumstances, at leat looking at GL curves (the dashed curve marked with $D2$ in the pictures). What emerges is that both before and after policy implementation the pairs of circumstances can be ranked according to ISD criteria, with Circumstance 4 the most advantaged (those having French fathers working as professionals) and Circumstance 2 the most disadvantaged (those having French fathers employed as manual workers). In figure 3.17 in the appendix we differentiate the curves in figures 3.15 and 3.16 across policy regimes, to obtain the difference between gap curves for each pair of circumstances. The patterns identified in figure 3.17 deserve more attention. In all the six pairs of comparisons obtained out of the four circumstances (marked with letters (a) to (f)), the differences in gap curves for the quantile function (marked with letters $DD1$) always cross the zero line at some point. The integrals of these curves (marked with $DD2$ and $DD3$) follow a similar pattern, making gap curves dominance not convincing.

The inference results for inverse stochastic dominance are reported in the upper panel of table 3.11. The interpretation of the table is similar to the one given in the previous application. In this application we find that the order of dominance is preserved by policy implementation for almost all pairs of circumstances. We find that equality of the conditional cumulative distribution functions associated to Circumstances 1 and 3 cannot be rejected after policy simulation.

In addition to this inconclusive result, we find evidence that gaps dominance is in most cases not satisfied. We report the result of the gap curve dominance in column (4) of table 3.11. The interpretation of the table is as before. For instance, the model associated to circumstances Circ. 3 and Circ. 2 is $(32-23)$, which means that to find gap curve dominance at order one it is necessary to take the difference of the earnings distribution of

Table 3.11: Equalization of Opportunity test: Ordinal and Distance criteria for university accessibility expansion.

Circ. c vs Circ. c'	Before policy	After policy	Difference-in-differences	
	$(\pi = 0)$	$(\pi = 1)$	$\Delta_W(F_0^c, F_0^{c'}) - \Delta_W(F_1^c, F_1^{c'})$	Gap dominance k (tested model)*
	(1)	(2)	(3)	(4)
Circ. 2 vs Circ. 1	$\prec_{ISD1}$	$\prec_{ISD1}$	$= 0\ \forall W \in \mathcal{R}^1$	$= 0$ for $k = 1$ $(12 - 12)$
Circ. 3 vs Circ. 1	$\succ_{ISD1}$	$\sim_{ISD1}$	$= 0\ \forall W \in \mathcal{R}^1$	$= 0$ for $k = 1$ $(31 - 31)$
Circ. 4 vs Circ. 1	$\succ_{ISD1}$	$\succ_{ISD1}$	$= 0\ \forall W \in \mathcal{R}^1$	$= 0$ for $k = 1$ $(41 - 41)$
Circ. 3 vs Circ. 2	$\succ_{ISD1}$	$\succ_{ISD1}$	Not verified $\forall W \in \mathcal{R}^1$	> 0 for $k = 1$ $(32 - 23)$
Circ. 4 vs Circ. 2	$\succ_{ISD1}$	$\succ_{ISD1}$	Not verified $\forall W \in \mathcal{R}^1$	> 0 for $k = 1$ $(42 - 24)$
Circ. 4 vs Circ. 3	$\succ_{ISD1}$	$\succ_{ISD1}$	$\geq 0\ \forall W \in \mathcal{R}^1$	> 0 for $k = 1$ $(43 - 43)$

	Opportunity amelioration
	Δ Policy impact
Overall	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$
Circumstance 1	After $(\pi = 1) \sim_{ISD1}$ Before $(\pi = 0)$
Circumstance 2	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$
Circumstance 3	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$
Circumstance 4	After $(\pi = 1) \succ_{ISD1}$ Before $(\pi = 0)$

Source: Estimates from Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Earnings distribution corrected by the age effect. Sample reduced to male earners where circumstances have been recorded, cohorts 1950 to 1955. Circumstances defined by father status: *Circ. 1*: foreign born; *Circ. 2*: French farmer or manual worker; *Circ. 3*: French artisan or non-manual worker; *Circ. 4*: French professional.

EzOP tested at 5% significance level on a selected sample of twenty quantiles. Both inverse stochastic equality and dominance null hypothesis have been separately tested for any degree $k = 1$ up to 5. Only the minimal degree of dominance $\kappa(c, c', \pi)$ is reported. The notation $c \succ_{ISD\kappa} c'$ means that the earnings distribution of circumstance c ISD at order κ the earnings distribution of circumstance c'. The distance test is defined over the class $\kappa(c, c', \pi = 1)$. The direction of the gaps dominance is estimated according to the direction provided by the ISD dominance relation. The ISDk for $k = 1, 2$ is estimated as in Beach and Davidson (1983), while for $k \geq 3$ tests are constructed by using the asymptotic estimators of the vector of conditional Gini SWF.

*: For model $(ij - kh)$ we tested the gap curve of circumstances i vs j in $\pi = 0$ minus the gap curve of k vs h in $\pi = 1$, only for configurations $k = i$ and $h = j$ or $k = j$ and $h = i$.

Circumstance 3 minus the earning distribution of Circumstance 2 under policy $\pi = 0$, and the inverse under policy $\pi = 1$. Otherwise, the alternative gap curves always reject the null hypothesis of equality or dominance even at higher orders.

For cases in which the model that does not allow to reject gaps dominance is coherent with the order of circumstances reported in columns (1) and (2), it is possible to identify changes in economic distance. The results are reported in column (3) of table 3.11. The economic distance between Circumstance 1 and the remaining circumstances is unaffected by policy simulation. On the contrary, the distance between Circumstance 3 and Circumstance 4 is reduced by effect of the policy. This result is consistent with the findings of Maurin and McNally (2008), who assert that a large part of the benefits from the opening of universities are redistributed to the type defined by Circumstance 3. In fact, the distribution associated to this circumstance approaches now the one of Circumstance 4, representing the most affluent background. For the other pairs of circumstances, it is not possible to verify the effect of the policy in terms of economic distance.

We conclude that under the assumption of the rank dependent model for preferences, the ex ante O-ezOP criterion is validated by the data, although there is no apparent change in consensus due to policy simulation. However, the ex ante D-ezOP criterion cannot be verified by the data. Globally, the ex ante ezOP criterion is rejected and we cannot conclude with certitude that the simulated policy equalizes opportunities. The result is mostly driven by the fact that the gains from policy treatment are distributed in a way that the gap between more advantaged circumstances is reduced after policy intervention, but this comes at the cost of widening the distance with the other, less advantaged circumstances.

3.6.4 A comparison of the two applications

This result provides new evidence in favor of the opportunity equalizing impact of schooling policies that promote inclusion into the educational system and that take place early in the educational career of individuals. In fact, although both policies analyzed here provide unambiguous amelioration of opportunities for all circumstances (see tables 3.6 and 3.11), and despite the weak impact on earnings of policies promoting inclusion in the secondary education system, this last type of policies have a robust (and statistically significant) impact in terms of equalizing earnings prospects across circumstances. This depends on the interplay between the distribution of treatment effects (where positive treatments are assigned only to the quantiles around the median of the treated group) and the distribution of the target group across circumstances and earnings quantiles. In fact the policy assigns a positive significant treatment only to target students who mainly come from disadvantaged backgrounds and that do not score among top earners.

Policies facilitating access to the higher education system assign treatment effects that are increasing with the target group's earnings quantiles. Since top earners in the target group are expected to come from more advantaged circumstances (see figures 3.8 and 3.13), it is possible that this type of policies redistribute higher gains to more advantaged groups, thus explaining the rejection of gap dominance for many pairs of circumstances, thus ruling out ex ante equalization.

The results in this section may also be seen in terms of efficiency. There is growing evidence (Cunha, Heckman and Lochner 2006, Cunha and Heckman 2007) that it is cheaper and more efficient for the society to invest in policies meant to compensate disadvantaged individuals/groups early in their educational career rather than to provide late intervention measures. Our results show that the gains from early intervention mostly affect those in the center of the distribution, while leaving (statistically) unchanged the tails of the earnings distribution. However, we find that this allocation of gains from policy along the earnings curve promotes opportunity equalization. We leave for future investigations the assessment of the opportunity equalizing impact of policies which allow to compensate disadvantaged students at the beginning of their educational career, such as kindergarten expansion policies, *vis à vis* a more traditional cost-benefit analysis or opportunity amelioration comparisons.

3.7 Conclusions

In this chapter we propose an innovative criterion for evaluation of public polices, that builds on the notion of equality of opportunity in Lefranc et al. (2009). The ezOP criterion entails a difference-in-differences type of comparison between distribution functions conditional on effort levels.

In a first stage, differences are taken across distributions *within* each policy regime separately, in order to exploit the direction and distribution of the economic advantage among pairs of types outcomes distributions. This is done by imposing sequential restrictions on a class of evaluation preferences until agreement is reached in assessing the disadvantaged circumstance. We can therefore define a new notion of equality of opportunity that is based on lack of consensus over the direction of disadvantage.

In a second stage, we compare differences across circumstances *between* policy regimes. We do so by mean of two criteria that are meaningfully combined to achieve robustness in opportunity equalization assessments. The first criterion is ordinal, and it requires that the degree of lack of consensus over the direction of disadvantage (as measured by the sequence of restriction on a class of preferences) among every comparison of pairs of distributions conditional on type, increases by effect of policy changes. This amounts to verify that the "degree" of equality of opportunity is increased by effect of policy implementation. The second is a distance criterion, asking that the extent of the disadvantage (an economic measure of the distance between pairs of distributions) for a given pair of circumstances measured both before and after policy implementation, falls by effect of the policy.

The ezOP criterion is implemented if and only if the ordinal and cardinal criteria are verified for all pairs of circumstances at all effort levels. We discuss possible and meaningful alternatives to weaken this very demanding criterion for equalization. We propose sequential methods of elimination of dominated circumstances based on agreement on a well defined class of preferences.

Finally, we discuss implementation issues for the ezOP criterion. Identification of the ezOP criterion relies on the choice of a specific class of preferences and of a sequence of restrictions. Our identification strategy relies on the class of Yaari's rank dependent utility functions and the VNM class of preferences. Restrictions are respectively placed on the derivative of the weighting function and on VNM utility functions. Within these two models, one can find tests for ezOP based on inverse or, respectively, direct stochastic dominance.

However, we are able to obtain tractable empirical orders for distance comparisons only in the class of Yaari functions.

Within the rank dependent preferences model, we also discuss identification when effort is not observable. In that case, we show that we are only able to identify a very special equalization criterion that aggregates comparisons across circumstances and effort levels. An implementation algorithm to test the ezOp criterion is proposed. Finally, we provide an empirical illustration of the validity of our ezOP criterion in evaluating the opportunity equalization impact of educational policies in France.

This work contributes in four directions. First, we study a new definition for equality of opportunity that encompasses other models proposed in the literature. The advantage of our model is that it allows to assign a degree of equality of opportunity to each configuration. Moreover, the model is enough general to take into considerations other components of outcome achievement such as luck. We use preferences under risk as a tool to evaluate the distributional risk of the outcome prospect associated to different circumstances. Since we construct ex-post criteria, every comparison is conditional in the same effort level. Our main normative principle is based on robustness: the smaller is the class of preferences that unanimously agree upon ordering the prospects associated to two circumstances compared at the same level of effort choices, the weaker the contribution of these two varieties to the overall level of inequality of opportunity. Hence, the higher is the equality of opportunity granted.

Second, we make use of the rank comparison, along with a distance comparison, to gain robustness in the evaluation of equalization of opportunity. We cross the two criteria to derive an ex post criterion of equalization of opportunity, which allows to compare, for instance, the distributional impact of a policy *vis à vis* the status quo. We derive ex ante criteria to cope with cases in which effort is not observable, and we show that only a very weak criterion of opportunity equalization in the ex post setting can be identified by the ex ante version of our criteria. In this respect, we let for future research the definition of alternative criteria of opportunity equalization that are stronger than $\overline{\text{ezOP}}$ but that can still be identified by the ex ante criteria. The ezOP criterion induces a partial order of policy regimes. We intend to study also alternative criteria for opportunity equalization that are related to the ezOP criterion but that produce less partial orderings of distributions. The Weak ezOP criterion is only one possibility.

Third, we attempt to derive inference procedure for inverse stochastic dominance at orders higher than the second (i.e. Generalized Lorenz dominance). We adopt a threshold procedure to test ISD that exploits the conditional Gini social welfare functions, a method which is dual to Davidson and Duclos (2000). Asymptotic normality of the conditional Gini SWF is illustrated making use of the influence function (Barrett and Donald 2009). We intend to develop the direct testing procedure, as an extension of the methods introduced by Beach and Davidson (1983). We are able to derive the asymptotic distribution of the integrals up to order k of the quantile function. The difficulty rests on deriving an empirically tractable formulation of the asymptotic covariance matrix above order three. We leave this task for future research.

Finally, our fourth contribution is in the empirical literature. We evaluate two alternative simulated policies. Both policies are supposed to widen access to the educational system, although they take place in different periods of the students educational career. A policy that widens access to the educational system early in life seems to have a very mild impact on future students' earnings, although these effects are distributed in such a way that opportunities are equalized in the sense of the ezOP criterion. We let for further investigations the impact of other types of policies, such as kindergarten expansion, that take place very early in the educational career of individuals, if not before. Research in this field would provide additional information on hidden benefits of such policies that are often overlooked by traditional cost-benefit methods for policy evaluation.

3.A Appendix: Definitions and proofs

3.A.1 Notions of stochastic dominance and the rank dependent utility model

Following Gastwirth (1971), the integral function

$$GL(p|c,e,\pi) \;=\; \int_0^p F^{-1}(t|c,e,\pi)\,dt$$

defines the *Generalized Lorenz curve* (GL) of the distribution $F(y|c,e,\pi)$. The integral condition of order k constructed from the GL curve, $\Lambda^k(p|c,e,\pi)$, is defined in a recursive way by the following relations:

$$\begin{aligned}\Lambda^k(p|c,e,\pi) &= \int_0^p \Lambda^{k-1}(t|c,e,\pi)\,dt, \quad p\in[0,1]\\ \Lambda^2(p|c,e,\pi) &= GL(p|c,e,\pi).\end{aligned}$$

Muliere and Scarsini (1989) introduced the *inverse stochastic dominance* partial order $\succcurlyeq_{ISDk}$ as a criterion to rank distributions. The partial order is defined implicitly by the following inequalities. Let $F(y|c,e,\pi)$ and $F(y|c',e,\pi)$ be two distribution functions:

$$F(y|c,e,\pi) \;\succcurlyeq_{ISDk}\; F(y|c',e,\pi) \quad\Leftrightarrow\quad \Lambda^k(p|c,e,\pi)\geq\Lambda^k(p|c',e,\pi) \;\;\forall p\in[0,1].$$

Note that $\succcurlyeq_{ISD1}$ denotes rank dominance and $\succcurlyeq_{ISD2}$ denotes GL dominance. Moreover, by Proposition 1 in Maccheroni et al. (2005), if $F(y|c,e,\pi) \succcurlyeq_{ISDk} F(y|c',e,\pi)$ then $F(y|c,e,\pi) \succcurlyeq_{ISDl} F(y|c',e,\pi)$, for all $l>k$. It follows that GL dominance is sufficient for any other inverse dominance comparison.[23]

The inverse stochastic dominance at order k identifies a sequence of restrictions on the set of preferences determined by the Yaari's (1987) rank dependent model. Let this set of preferences be denoted by $\mathcal{R}$. Every utility function in $W\in\mathcal{R}$ can be represented by a

[23] It is well known (e.g. Muliere and Scarsini 1989) that first and second order inverse stochastic dominance are equivalent to direct first and second order stochastic dominance, which is in turn equivalent to generalized Lorenz dominance for incomes distributions with different means (Shorrocks 1983). Atkinson (1970) showed the logical relation between GL dominance with fixed means and an the utilitarian social welfare function, later generalized to all S-concave social welfare functions and to income distributions with different means.

weighted average of the quantiles of the distribution function F:

$$W(F) = \int_0^1 w(p)F^{-1}(p)dp,$$

where $w(p)$ is called the *distortion function.* Properties of the distortion function have been discussed, among others, by Maccheroni et al. (2005) and Aaberge (2009). The distortion function is such that $w(p) \geq 0$ for all $p \in [0,1]$ and $\int_0^1 w(p)dp = 1$ is insured by the fact that $w(0) = w(1) = 0$. Depending on the restrictions on the distortion function, one obtains a weighting scheme and therefore a utility function. For instance, by imposing $dw(p)/dp \leq 0$ one obtain a class of preferences that give larger weight to the bottom of the distribution of possible realization of lottery F. This restriction introduces risk aversion, and it is the dual condition with respect to requiring concave Von Neumann Morgerstern utility functions in the expected utility setting.

The nature of comparisons of lotteries allows to restrict attention to the set $\mathcal{R} := \mathcal{R}^1$ of distortion functions that are non negative. We focus on a sequence of restrictions on the derivatives of the weighting function, which introduces restrictions on the risk attitude of the preferences, by giving larger weight to the realizations occurring at the bottom of the distribution. Let use $k \in \mathbb{N}_{++}$ to map these restrictions into a scalar indicator. Let $\mathcal{R}^k \subseteq \mathcal{R}$ define the set of rank dependent utilities $W(.)$ restricted by introducing k restrictions on the sign of the high order derivatives of the weighting function $w(.)$. The resulting set of preferences $\mathcal{R}^k$ is characterized as follows, for $k \geq 2$:

$$\mathcal{R}^k = \left\{ W \in \mathcal{R} : \ (-1)^{i-1} \cdot \frac{d^{i-1}w(p)}{dp^{i-1}} \geq 0, \ \ \frac{d^{i-1}w(1)}{dp^{i-1}} = 0 \ \forall p \in [0,1] \text{ and } i = 2,3,\ldots,k \right\}.$$

The restrictions on the sign of the derivatives of the weighting function define a sequence of nested sets of preferences such that for $l < k$: $\mathcal{R}^k \subseteq \mathcal{R}^l \subseteq, \ldots, \subseteq \mathcal{R}^1 = \mathcal{R}$.

This sequence of restrictions on the class of preferences admitting the rank dependent representation is associated to the ranking of pairs of distributions produced by the inverse stochastic dominance, as shown by the following proposition:

Proposition 3.11 *For any $F(y|c,e,\pi)$ and $F(y|c',e,\pi)$:*

$$F(y|c,e,\pi) \succcurlyeq_{ISDk} F(y|c',e,\pi) \quad \Leftrightarrow \quad W(F(y|c,e,\pi)) \geq W(F(y|c',e,\pi)), \quad \forall W \in \mathcal{R}^k.$$

Proof. See, among others, Aaberge (2009). ■

As shown in Proposition 3.3 it is always possible to rank two distributions according to inverse stochastic dominance, if the two distributions do not coincide. Equivalently, it is always possible to determine a finite sequence of restrictions k on $\mathcal{R}$ so that all preferences in $\mathcal{R}^k$ rank one distribution as preferred to another.

Similar restrictions can be constructed for the set of preferences $\mathcal{U}$ satisfying the expected utility representation. In this context, the restrictions have a clear interpretation in terms of preferences towards risk apportionment (Eeckhoudt and Schlesinger 2006), and can be verified by resorting on direct stochastic dominance partial orders (Fishburn 1976).

3.A.2 Proof of Proposition 3.2

Proof. The first implication (O-ezOP $\Rightarrow$ O-ezOP2) is verified whenever, for a pair of circumstances i and h such that $\kappa_{ih}(0) < 0$, it also holds that $\kappa_{ih}(1) < 0$ for the same pair. Consider $h = \arg\min_j\{|\kappa_{ij}|\, s.t.\, \kappa_{ij}(1) < 0\}$. For $h \neq i$, then by construction $\kappa_i(0) < \kappa_{ih}(0) < 0$. Moreover, by O-ezOP, $\kappa_{ih}(0) \leq \kappa_{ih}(1)$, which gives the desired implication.

The second implication (O-ezOP2 $\Rightarrow$ O-ezOP3) can be demonstrated by contradiction. Let abuse notation by writing $\sigma_\pi(i)$ as the function assigning a position $\sigma_\pi(i)$ into the ordered sequence $\widehat{\kappa}_{\sigma_\pi(i)}(\pi)$ to the degree of dominance associated to circumstance i. It follows that $\widehat{\kappa}_{\sigma_\pi(i)}(\pi) = \kappa_i(\pi)$. The two sequences obtained for $\pi = 0$ and $\pi = 1$ has to be compared. To simplify, we consider the case $\sigma_0(i) = i$. The set of circumstances has cardinality $n(= |C|)$.

Suppose that the desired implication does not hold, that is when O-ezOP2 is true, than there exists one case in which O-ezOP3 is false. This occurs if there exists one circumstance i such that

$$\widehat{\kappa}_{\sigma_1(i)}(1) = \kappa_i(1) < \kappa_{\sigma_1(i)}(0). \tag{3.5}$$

As long as O-ezOP2 holds, then $\kappa_i(1) = \kappa_i(0) + \Delta_i$, with $\Delta_i \in \mathbb{N}_{++}$. There are two possible cases, each of them leading to a contradiction.

If $\sigma_1(i) \leq i$ then $\kappa_{\sigma_1(i)}(0) \leq \kappa_i(0) \leq \kappa_i(1)$, where the first inequality comes from the fact that the sequence $\kappa_i(0)$ is already ordered in increasing magnitude, while the second inequality is a consequence of O-ezOP2. This is a clear violation of (3.5).

On the contrary, if $\sigma_1(i) > i$ then $\kappa_{\sigma_1(i)}(0) \leq \kappa_h(0) \leq \kappa_h(1)$, for all $h = \sigma_1(i) < \ldots < n$.

Assume that (3.5) holds, then one gets that $\kappa_i(1) = \widehat{\kappa}_{\sigma_1(i)}(1) < \kappa_{\sigma_1(i)}(0) \leq \kappa_h(1)$ for all $h = \sigma_1(i) < \ldots < n$, which is always true since O-ezOP2 holds and the circumstances under $\pi = 0$ have been already ordered. Then one gets a vector $(\widehat{\kappa}_{\sigma_1(i)}(1), \ldots, \widehat{\kappa}_n(1))$ of $n - \sigma_1(i) + 1$ elements that are all larger or, at most, equal to $\widehat{\kappa}_{\sigma_1(i)}(1)$. This elements has to be paired with the entries of the remaining vector $(\kappa_{\sigma_1(i)+1}(0), \ldots, \kappa_n(0))$ of size $n - \sigma_1(1)$, which elements are higher than $\kappa_{\sigma_1(i)}(0)$. A clear contradiction of the fact that $\kappa_i(1) = \widehat{\kappa}_{\sigma_1(i)}(1)$, because the element in the position $\sigma_1(i)$ cannot coincide with $\kappa_i(1)$. Hence it must hold that O-ezOP2 is violated which gives the desired result.

Finally, O-ezOP3 implies O-ezOP4 by construction. ■

3.A.3 Proof of Proposition 3.3

Proof. The proof consists in showing that if $F(y|c, e, \pi)$ inverse stochastically dominates $F(y|c', e, \pi)$ at the first order for some positive percentiles between 0 and $p_\beta > 0$, then we have a sufficient condition for the two distribution to be comparable at a finite degree of integration k^*.

Define $\Delta F^{-1}(p) := F^{-1}(p|c, e, \pi) - F^{-1}(p|c', e, \pi)$ and $\Delta\Lambda^k(p) := \Lambda^k(p|c, e, \pi) - \Lambda^k(p|c', e, \pi)$ at any $p \in [0, 1]$. Integrate by part up to $k - 2$ times the function $\Delta\Lambda^k(p)$ to obtain the following:

$$
\begin{aligned}
\Delta\Lambda^k(p) &= \int_0^p \Delta\Lambda^{k-1}(t)dt = -\int_0^p t \cdot \Delta\Lambda^{k-2}(t)dt + \Big[t\Delta\Lambda^{k-1}(t)\Big]_0^p \\
&= \int_0^p (p - t)\Delta\Lambda^{k-2}(t)dt \\
&= \int_0^p \frac{1}{2}(p - t)^2\Delta\Lambda^{k-3}(t)dt + \left[\frac{1}{2}(p - t)^2\Lambda^{k-2}(t)\right]_0^p \\
&= \int_0^p \frac{1}{(k-2)!}(p - t)^{k-2}\Delta F^{-1}(t)dt \qquad (3.6)
\end{aligned}
$$

To see the result in (3.6) it is sufficient to note that $\Lambda^k(0) = 0$ and therefore $\Delta\Lambda^k(0) = 0$ for any k, and that $\Delta\Lambda^2(p) = \int_0^p \Delta F^{-1}(t)dt$.

The sufficient conditions of the proposition states that $\Delta F^{-1}(p) \geq 0$ for all $p \in [0, p_\beta)$ and there exists a p such that the strong inequality holds. As long as we use continuous or at most left inverse cumulative distribution functions, we make sure that the function $\Delta F^{-1}(p)$ is well behaved on the whole percentile domain. Moreover, the function takes only finite values even in $p = 1$ or $p = 0$. As a consequence the value p_β exists.

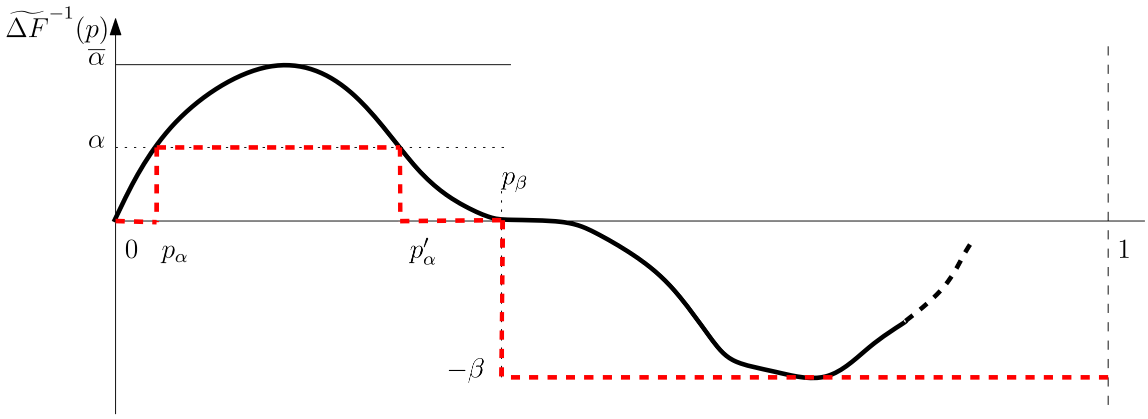


Figure 3.6: Proof of Proposition 3.3. The curves $\Delta F^{-1}(p)$ (solid black) and $\widetilde{\Delta F}^{-1}(p)$ (dashed red)

Moreover, consider the two bounds values $\overline{\alpha} := \sup\{\Delta F^{-1}(p) : p \in [0, p_\beta)\} > 0$ and $-\beta := \inf\{\Delta F^{-1}(p) : p \in [p_\beta, 1]\} < 0$, that corresponds respectively to the largest positive and negative horizontal distance between two distributions. They both exist finite, provided that the sufficient conditions given above are satisfied. [24] The curve of ΔF^{-1} is marked with a solid lines on the graph in figure 3.6, along with the corresponding values of α and $-\beta$.

Let $0 < \alpha \leq \overline{\alpha}$ such that it is possible to define *at least* two points $p_\alpha, p'_\alpha \in [0, p_\beta)$, such that for $p_\alpha \leq p \leq p'_\alpha$, $\Delta F^{-1}(p) > 0$ holds. Consequently, we define the new differences curve $\widetilde{\Delta F}^{-1}(p)$ in the following way:

$$\widetilde{\Delta F}^{-1}(p) := \begin{cases} 0 & \text{if } p \in [0, p_\alpha) \\ \alpha & \text{if } p \in [p_\alpha, p'_\alpha] \\ 0 & \text{if } p \in (p'_\alpha, p_\beta) \\ -\beta & \text{if } p \in [p_\beta, 1] \end{cases}$$

The curve is represented by the dashed line in Figure 3.6. It is not difficult to see that α and $-\beta$ are defined by the distribution functions, while it always hold that $\widetilde{\Delta F}^{-1}(p) \leq \Delta F^{-1}(p)$ for all p. As a consequence, also the value of the integrals of $\widetilde{\Delta F}^{-1}(p)$ lie always below the value of the integral of $\Delta F^{-1}(p)$ calculated in p. The function reduces the positive domain of the difference $\Delta F^{-1}(p)$ for percentiles in the lower side of the domain, while it

[24]If the conditions do not hold we have either that type c' dominates type c or type c dominates on the first order type c'

magnify the negative effect of the difference for the percentiles in the remaining side of the domain. Therefore, making use of (3.6), if it is possible to find a value of $\widetilde{k}^*$ such that $\forall k > \widetilde{k}^*$:

$$\int_0^p \frac{1}{(k-2)!}(p-t)^{k-2}\widetilde{\Delta F}^{-1}(t)dt \geq 0 \quad \forall p \in [0,1], \tag{3.7}$$

then there must exists also a value k^* satisfying our proposition (that is inverse stochastic dominance at a finite order is always granted).

Not that in the interval $[0, p_\alpha)$ and (p'_α, p_β) the expression (3.7) is always zero. Moreover, (3.7) is always strictly positive on the interval domain $[p_\alpha, p'_\alpha]$. It remains to check the condition for any $p \geq p_\beta$.

$$\begin{aligned}\int_0^p \frac{(p-t)^{k-2}}{(k-2)!}\widetilde{\Delta F}^{-1}(t)dt &= \frac{1}{(k-2)!}\left[\int_{p_\alpha}^{p'_\alpha}(p-t)^{k-2}\alpha dt + \int_{p'_\alpha}^{p_\beta}(p-t)^{k-2}(-\beta)dt\right] \\ &= \frac{\left\{\alpha\left[(p-p_\alpha)^{k-1} - (p-p'_\alpha)^{k-1}\right] - \beta(p-p_\beta)^{k-1}\right\}}{(k-2)!} \geq 0 \;\forall p \geq p_\beta.\end{aligned}$$

To check the solution it suffice that there exists a $\widetilde{k}^*$ such that:

$$\frac{(p-p_\alpha)^{k-1} - (p-p'_\alpha)^{k-1}}{(p-p_\beta)^{k-1}} \geq \frac{\beta}{\alpha}, \quad \forall p \geq p_\beta. \tag{3.8}$$

By construction of $\widetilde{\Delta F}^{-1}(p)$, if the condition holds for $p = 1$, then it must hold for all $p < 1$, because the differential takes only negative values for $p \geq p_\beta$. Note that the numerator and denominator of the left hand side of (3.8) are positive, but the ratio is not said to be greater than one. Nevertheless, one can always pick up a value of $\alpha < \overline{\alpha}$ such that $(p - p'_\alpha) \approx (p - p_\beta)$ and (3.8) is therefore satisfied if and only if the following holds:

$$\left(\frac{1-p_\alpha}{1-p_\beta}\right)^{k-1} \geq 1 + \frac{\beta}{\alpha}. \tag{3.9}$$

Both sides of (3.9) are positives and greater than one. Thus, by taking logs on the left and right side, it is easy to show that the integral condition in (3.6) is satisfied if and only if the integration order $\widetilde{k}^*$ is large enough to verify:

$$\widetilde{k}^* \geq 1 + \frac{\ln(1+\beta/\alpha)}{\ln(1-p_\alpha) - \ln(1-p_\beta)}.$$

Note that $\widetilde{k}^*$ is positive and greater than one and it always exists *finite* for any $0 < p_\alpha < p_\beta < 1$ and for $\alpha, \beta > 0$. Therefore the value k^* exists as well, which concludes the proof. ■

3.A.4 Proof of Proposition 3.4

Proof. As a consequence of the dominance hypothesis, we have:

$$\forall W \in \mathcal{R}, \forall \pi \int_0^1 w(p) F_\pi^{-1}(p) dp > \int_0^1 w(p) F_\pi'^{-1}(p) dp$$

Consequently, for all $W \in \mathcal{R}$, we can write:

$$\Delta_W(F_\pi, F'_\pi) = \int_0^1 w(p) G(F_\pi, F'_\pi, p) dp$$

Hence, we have:

$$\Delta_W(F_0, F'_0) - \Delta_W(F_1, F'_1) = \int_0^1 w(p)[G(F_0, F'_0, p) - G(F_1, F'_1, p)] dp \tag{3.10}$$

If $[G(F_0, F'_0, p) - G(F_1, F'_1, p)] \geq 0$ for all p, since the weights $w(p)$ are non-negative, the integrand in equation (3.10) is positive for all p and the integral is positive.

If on the contrary $[G(F_0, F'_0, p) - G(F_1, F'_1, p)]$ is negative in the neighborhood of a quantile p_0, we can find a weight profile $w(p)$ that is arbitrarily small outside this neighborhood and it makes the integral negative. ■

3.A.5 Proof of Proposition 3.5

Proof. We use the same type of proof argument as in Aaberge (2009). As a consequence of the dominance hypothesis, we have:

$$\forall W \in \mathcal{R}^2, \forall \pi \int_0^1 w(p) F_\pi^{-1}(p) dp > \int_0^1 w(p) F_\pi'^{-1}(p) dp$$

Consequently, for all $W \in \mathcal{R}^2$, we can write:

$$\Delta_W(F_\pi, F'_\pi) = \int_0^1 w(p) G(F_\pi, F'_\pi, p) dp$$

Hence, $\forall W \in \mathcal{R}^2$ we have:

$$\Delta_W(F_0, F_0') - \Delta_W(F_1, F_1') = \int_0^1 w(p)[G(F_0, F_0', p) - G(F_1, F_1', p)]dp \tag{3.11}$$

It is possible to integrate (3.11) by parts once,

$$\begin{aligned}\Delta_W(F_0, F_0') - \Delta_W(F_1, F_1') &= w(1)\int_0^1 \left[G(F_0, F_0', p) - G(F_1, F_1', p)\right] \\ &+ \int_0^1 (-1)w'(p)\int_0^p \left[G(F_0, F_0', t) - G(F_1, F_1', t)\right] dtdp\end{aligned}$$

By $W \in \mathcal{R}^2$ then $w(1) = 0$ and the first term disappears. By $w'(p) \leq 0$ for all p makes $\int_0^p \left[G(F_0, F_0', t) - G(F_1, F_1', t)\right] dt$ sufficient for (3.11). Moreover, Lemma 1 in Aaberge (2009) gives the necessary part. ■

3.A.6 Proof of Proposition 3.6

Proof. We use the same type of proof argument as in Aaberge (2009). As a consequence of the dominance hypothesis, we have:

$$\forall W \in \mathcal{R}^k, \forall \pi \int_0^1 w(p)F_\pi^{-1}(p)dp > \int_0^1 w(p)F_\pi'^{-1}(p)dp$$

Consequently, for all $W \in \mathcal{R}^k$, we can write by proposition 3.11:

$$\Delta_W(F_\pi, F_\pi') = \int_0^1 w(p)G(F_\pi, F_\pi', p)dp$$

Hence, $\forall W \in \mathcal{R}^k$ we have:

$$\Delta_W(F_0, F_0') - \Delta_W(F_1, F_1') = \int_0^1 w(p)[G(F_0, F_0', p) - G(F_1, F_1', p)]dp \tag{3.12}$$

It is possible to integrate (3.12) by parts k times,

$$\begin{aligned}
\Delta_W(F_0,F_0')-\Delta_W(F_1,F_1') &= w(1)\int_0^1 \left[G(F_0,F_0',p)-G(F_1,F_1',p)\right] \\
&+ \sum_{j=1}^{i}(-1)^j \frac{d^j w(1)}{dp^j}\left[G(\Lambda_0^k,\Lambda_0'^k,1)-G(\Lambda_1^k,\Lambda_1'^k,1)\right] \\
&+ (-1)^i \int_0^1 \frac{d^i w(p)}{dp^i}\left[G(\Lambda_0^k,\Lambda_0'^k,p)-G(\Lambda_1^k,\Lambda_1'^k,p)\right]dp
\end{aligned}$$

By $W \in \mathcal{R}^k$ then $w(1)=0$ and $\frac{d^j w(1)}{dp^j}=0$ for all $j \leq i$ and the first term disappears. Thus follows that the conditions for $W \in \mathcal{R}^k$ makes $\left[G(\Lambda_0^k,\Lambda_0'^k,1)-G(\Lambda_1^k,\Lambda_1'^k,1)\right] \geq 0$ sufficient for (3.12). Moreover, Lemma 1 in Aaberge (2009) gives the necessary part. ■

3.A.7 Proof of Proposition 3.7

Proof. As already noted:

$$\Delta_W(F_0,F_0') \geq \Delta_W(F_1,F_1') \quad \Leftrightarrow \quad u^{-1}(W(F_0))-u^{-1}(W(F_0')) \geq u^{-1}(W(F_1))-u^{-1}(W(F_1'))$$

We first prove that the condition cannot be satisfied if $W(F_1)-W(F_1') \geq W(F_0)-W(F_0')$.

For any function $\phi(x)$ defined on $\mathbb{R}$ with $\phi'>0$ and $\phi''>0$, and for all $d>0$, the function $\psi(x)$ defined by $\psi(x)=\phi(x+d)-\phi(x)$ is an increasing function of x. Hence u^{-1} is increasing and convex. We have $W(F_1') \geq W(F_0')$. Let $dW=W(F_0)-W(F_0')$. We have:

$$\begin{aligned}
u^{-1}(W(F_1')+dW)-u^{-1}(W(F_1')) &> u^{-1}(W(F_0')+dW)-u^{-1}(W(F_0')) \\
&> u^{-1}(W(F_0))-u^{-1}(W(F_0')).
\end{aligned}$$

Hence, $W(F_1)-W(F_1') \geq W(F_0)-W(F_0') \Rightarrow u^{-1}(W(F_1')+dW)-u^{-1}(W(F_1')) > u^{-1}(W(F_0))-u^{-1}(W(F_0'))$.

The reciprocal is however not true as illustrated by the graphical argument in figure 3.7. ■

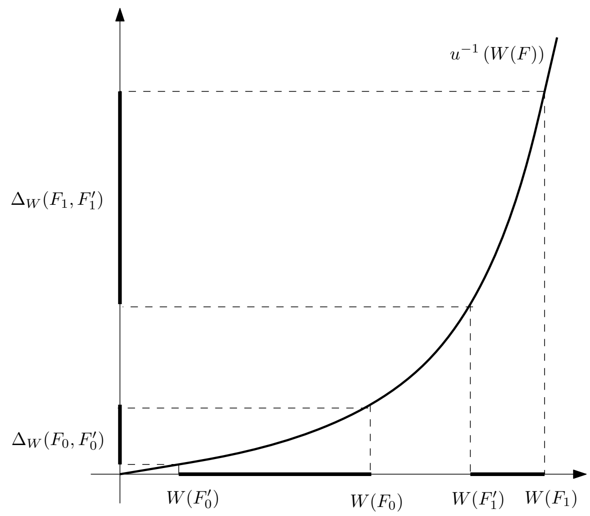


Figure 3.7: Proof of Proposition 3.7. $[W(F_0) - W(F_0')] - [W(F_1) - W(F_1')] \geq 0 \not\Rightarrow \Delta_W(F_0, F_0') - \Delta_W(F_1, F_1') \geq 0$

3.A.8 Proof of Proposition 3.9

Proof. The proof uses Fubini's Theorem to integrate the appropriate dominance conditions across effort levels. Under Assumption 1, effort is independently distributed from circumstances and from the policy. Assume additionally that the direction of dominance of any pair of different circumstances does not reverse between effort level within the same policy (this assumption simplifies the proof but it does not invalidate the result). Suppose that O-ezOP is satisfied. Then it follows that:

$$\forall c \neq c', \ \forall e: \ \Lambda^{\kappa}(p|c, e, \pi) \geq \Lambda^{\kappa}(p|c', e, \pi) \ \forall p \in [0, 1],$$

where $\kappa = \kappa(c, c', e, \pi)$ and for at least one p the inequality holds strict.

Hence, by integrating $\Lambda^{\kappa}(p|c, \pi) = \int_e \Lambda^{\kappa}(p|c', e, \pi) dG(e)$ for both policies π, under Assumption 1, one gets:

$$\forall c \neq c': \ \Lambda^{\kappa}(p|c, \pi) \geq \Lambda^{\kappa}(p|c', \pi) \ \forall p \in [0, 1], \tag{3.13}$$

where $\kappa = \max_e\{\kappa(c, c', e, \pi)\}$ and for at least one p the inequality holds strict.

If O-ezOP criterion is satisfied, then $\max_e\{\kappa(c, c', e, 0)\} \leq \max_e\{\kappa(c, c', e, 1)\}$ for all pairs $c \neq c'$. This is the ex ante ordinal criterion. Moreover, if D-ezOP is satisfied one gets that:

$$\forall c \neq c',\ \forall e:\ \Lambda^{\kappa}(p|c,e,0) - \Lambda^{\kappa}(p|c',e,0) \geq \Lambda^{\kappa}(p|c,e,1) - \Lambda^{\kappa}(p|c',e,1)\ \forall p \in [0,1],$$

where $\kappa \leq \kappa(c,c',e,1)$ and for at least one p the inequality holds strict.

By integrating the distance function (exploiting the linearity of the function for the class of preferences marked by κ for which the economic advantage associated to circumstance c is unambiguously higher than the economic advantage associated to c'), one gets

$$\forall c \neq c':\ \Lambda^{\kappa}(p|c,0) - \Lambda^{\kappa}(p|c',0) \geq \Lambda^{\kappa}(p|c,1) - \Lambda^{\kappa}(p|c',1)\ \forall p \in [0,1], \tag{3.14}$$

where $\kappa \leq \max_e\{\kappa(c,c',e,1)\}$ and for at least one p the inequality holds strict. This is the ex ante distance criterion. When the conditions in (3.13) and (3.14) are jointly satisfied, then the ex ante ezOP test is also satisfied. ■

3.A.9 Proof of Proposition 3.10

Proof. Assume that $\overline{\text{ezOP}}$ is not satisfied, we want to show that in this case the ex ante ezOP test fails, which proofs the implication in the proposition. The $\overline{\text{ezOP}}$ test fails in two cases.

The first case occurs when $\overline{\text{O-ezOP}}$ fails. As a consequence, there should exist a class of preference $\mathcal{R}^k$ for which $\overline{\text{EOP-W}k}$ holds under $\pi = 0$ but not under $\pi = 1$. Assumption 2 guarantees that a pair $c \neq c'$ exists under both policy regimes for which $F(.|c,e,\pi) \succ_{ISDk} F(.|c',e,\pi)$, $\forall e$. Therefore, $\overline{\text{O-ezOP}}$ fails if and only if $\kappa(0) > \kappa(1)$, where $\kappa(\pi)$ is defined in the text, equation (3.4). By using Assumption 1 and Fubini's theorem, the dominance condition leads to the following integral condition (where c_π, c'_π indicated the pair of circumstances that are unambiguously ranked according to the class $\mathcal{R}^{\kappa(\pi)}$ at all effort levels):

$$\int_e \Lambda^{\kappa(\pi)}(p|c_\pi,e,\pi)dG(e) \geq \int_e \Lambda^{\kappa(\pi)}(p|c'_\pi,e,\pi)dG(e)\ \forall p \in [0,1],$$

where the inequality holds strict for at least one percentile p. This condition is implied by the fact that $\overline{\text{EOP-W}k}$ is not satisfied for all effort levels. As a consequence, one gets that $F(.|c_\pi,\pi) \succ_{ISD\kappa(\pi)} F(.|c'_\pi,\pi)$ for all π, with $\kappa(0) > \kappa(1)$ (since the order of dominance is preserved under integration). By Assumption 3, $(c_0,c'_0) = (c_1,c'_1)$, which implies that the ex ante ordinal criterion of equalization fails.

The second case determining $\overline{\text{ezOP}}$ failure occurs when $\overline{\text{D-ezOP}}$ is not satisfied. We assume that $\overline{\text{O-ezOP}}$ is indeed satisfied. Otherwise, one does not need to check the distance criterion.

Under Assumption 2 and 3, the pair of circumstances (c, c') for which $\overline{\text{D-ezOP}}$ fails is the same under both policy levels. Moreover $\overline{\text{O-ezOP}}$ holds, hence $\kappa(0) \leq \kappa(1)$. If $\overline{\text{D-ezOP}}$ is violated, then the economic advantage associated to circumstance c compared to c' is unambiguously increased by effect of policy intervention at all effort levels, as evaluated by all preference relations in $\mathcal{R}^{\kappa(1)}$. It follows that the difference between the integral at order $\kappa(1)$ of the conditional quantile function associated to circumstances c and c' is always positive along its domain. Under Assumption 1, the distance condition can be integrated across effort levels, such that $\forall p \in [0, 1]$:

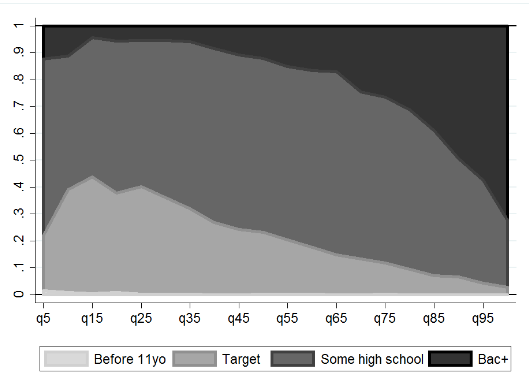
$$\int_e \Lambda^{\kappa(1)}(p|c, e, 0) - \Lambda^{\kappa(1)}(p|c', e, 0)dG(e) < \int_e \Lambda^{\kappa(1)}(p|c, e, 0) - \Lambda^{\kappa(1)}(p|c', e, 0)dG(e). \quad (3.15)$$

Integrating (3.15), one obtains that $\forall p \in [0, 1]$:

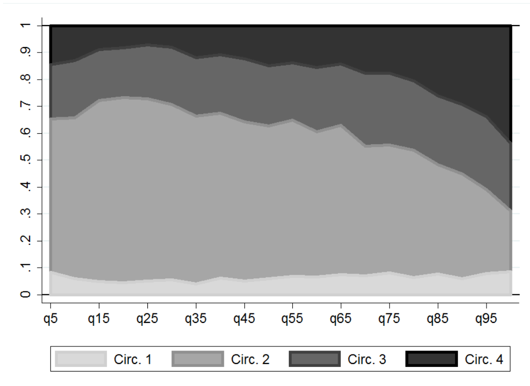
$$\Lambda^{\kappa}(p|c, 0) - \Lambda^{\kappa}(p|c', 0) < \Lambda^{\kappa}(p|c, 1) - \Lambda^{\kappa}(p|c', 1).$$

This is a clear violation of the distance criterion for the ex ante equalization test. ■

3.B Appendix: Additional material for the empirical analysis

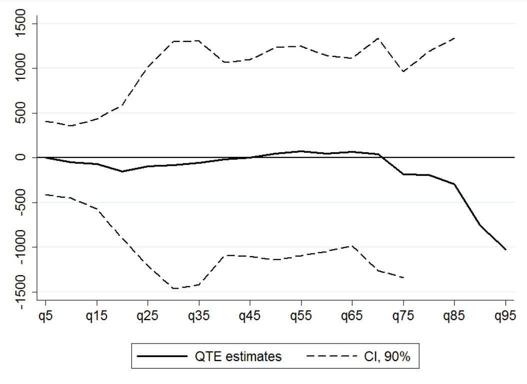


(a) By educational achievement (treatment)

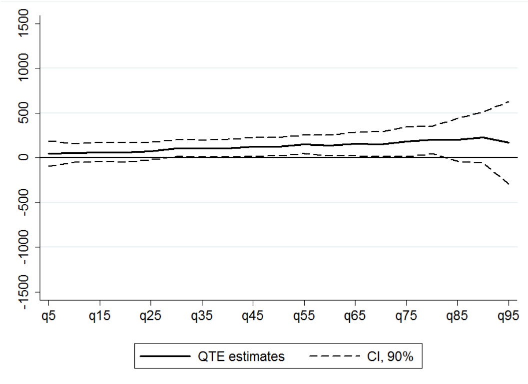


(b) By circumstances

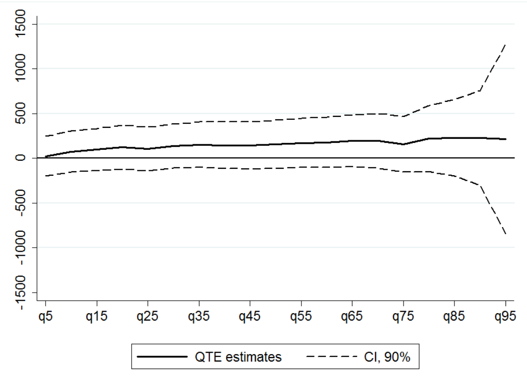
Figure 3.8: Distribution of population across educational achievement levels (a) and circumstances (b), by earnings quantiles, 5% intervals, expressed in cumulative shares. Scores have been calculated from a multinomial logit model. In panel (a), the *target* group refers to students between age 11 to 15 who are in junior-high school (*College*). Circumstances are defined according to the father socioeconomic status. qX represent a 5% share of the population between quantile QX% and QX%-5% in the overall earnings distribution.



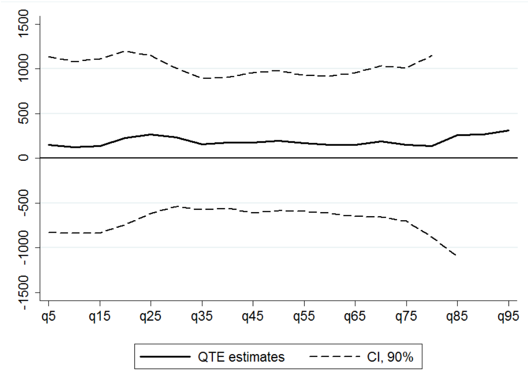
(a) Circumstance 1



(b) Circumstance 2

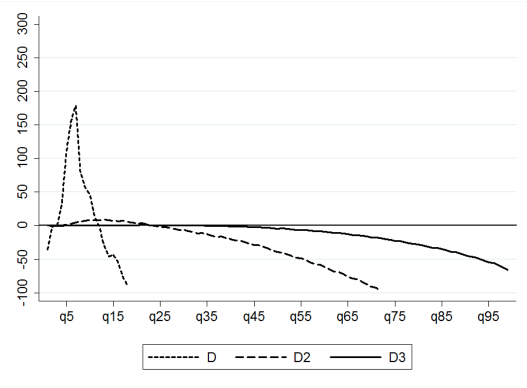


(c) Circumstance 3

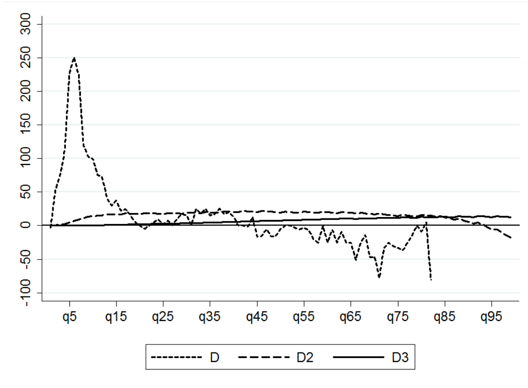


(d) Circumstance 4

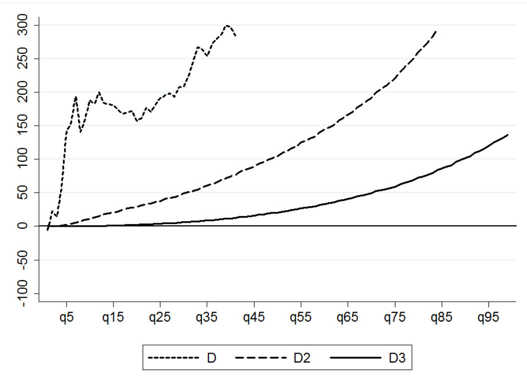
Figure 3.9: QTE of access to the secondary education system on earnings, by circumstances. Sample of cohorts 1950 to 1955, French male earners. Cohorts 1953 to 1955 define the IV (introduction of the Berthoin's law). IV estimates at 5% intervals, CI at 90% is computed with robust standard errors. Controls: cohort trends, year of the survey, age left full time schooling (polynomial).



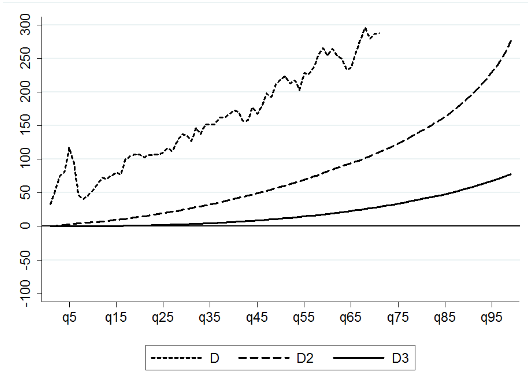
(a) Circumstance 2 - Circumstance 1



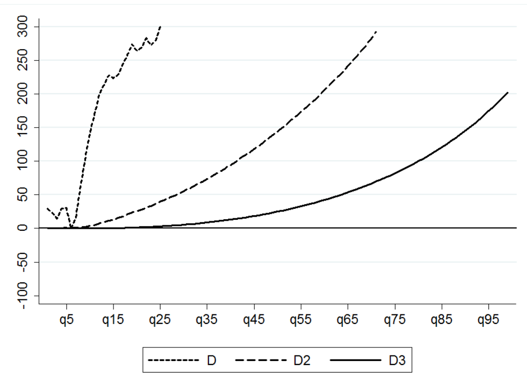
(b) Circumstance 3 - Circumstance 1



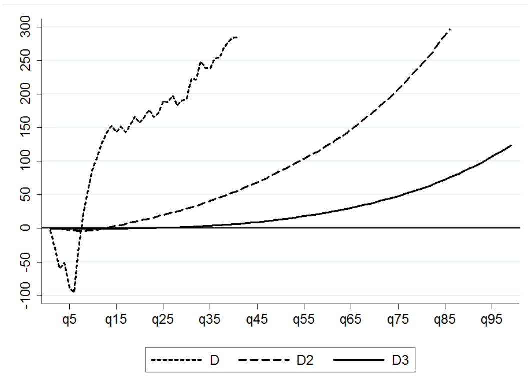
(c) Circumstance 4 - Circumstance 1



(d) Circumstance 3 - Circumstance 2

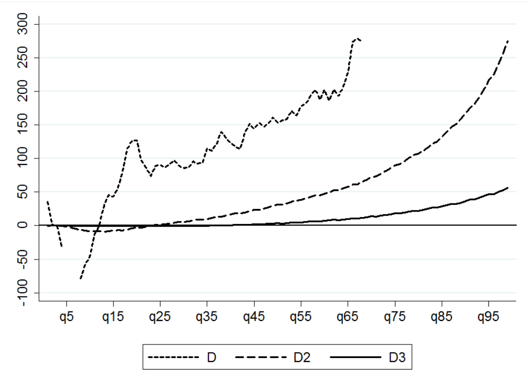


(e) Circumstance 4 - Circumstance 2

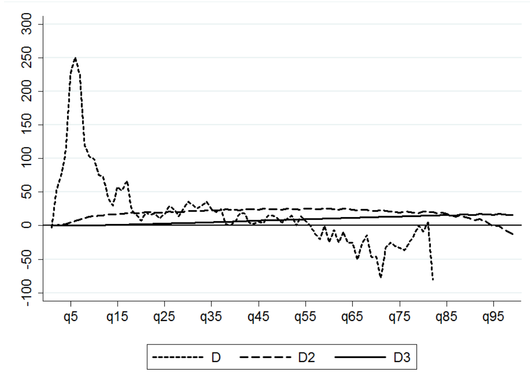


(f) Circumstance 4 - Circumstance 3

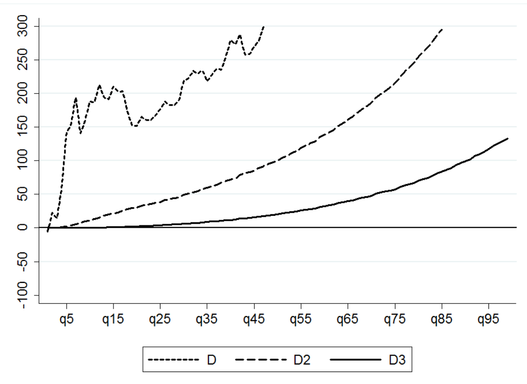
Figure 3.10: Differences in quantile functions (D), *GL* curves (D2) and integrals of the *GL* curves (D3) computed at each percentile of the actual earnings distribution without policy treatment. Values on the horizontal axis refer to percentiles of the actual earnings distribution. Values on the vertical axes express the difference between curves, in Euros. The curves represent the differences between the outcomes prospect associated to two distinct circumstances, for a total of six comparisons. Earnings differences are trimmed at 300 and -100 Euro.



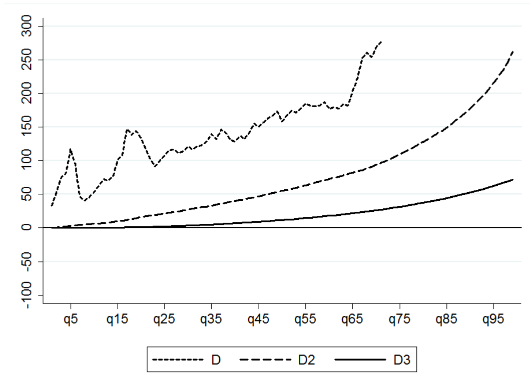
(a) Circumstance 2 - Circumstance 1



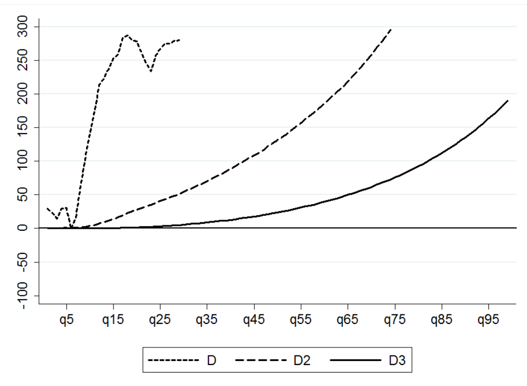
(b) Circumstance 3 - Circumstance 1



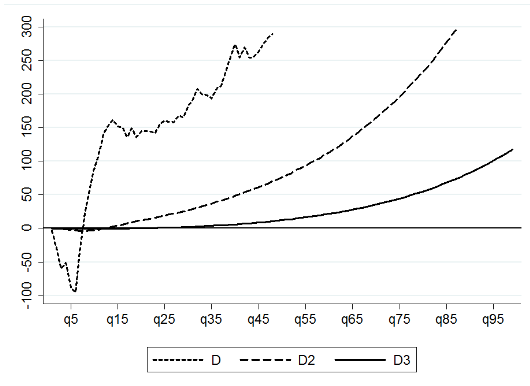
(c) Circumstance 4 - Circumstance 1



(d) Circumstance 3 - Circumstance 2

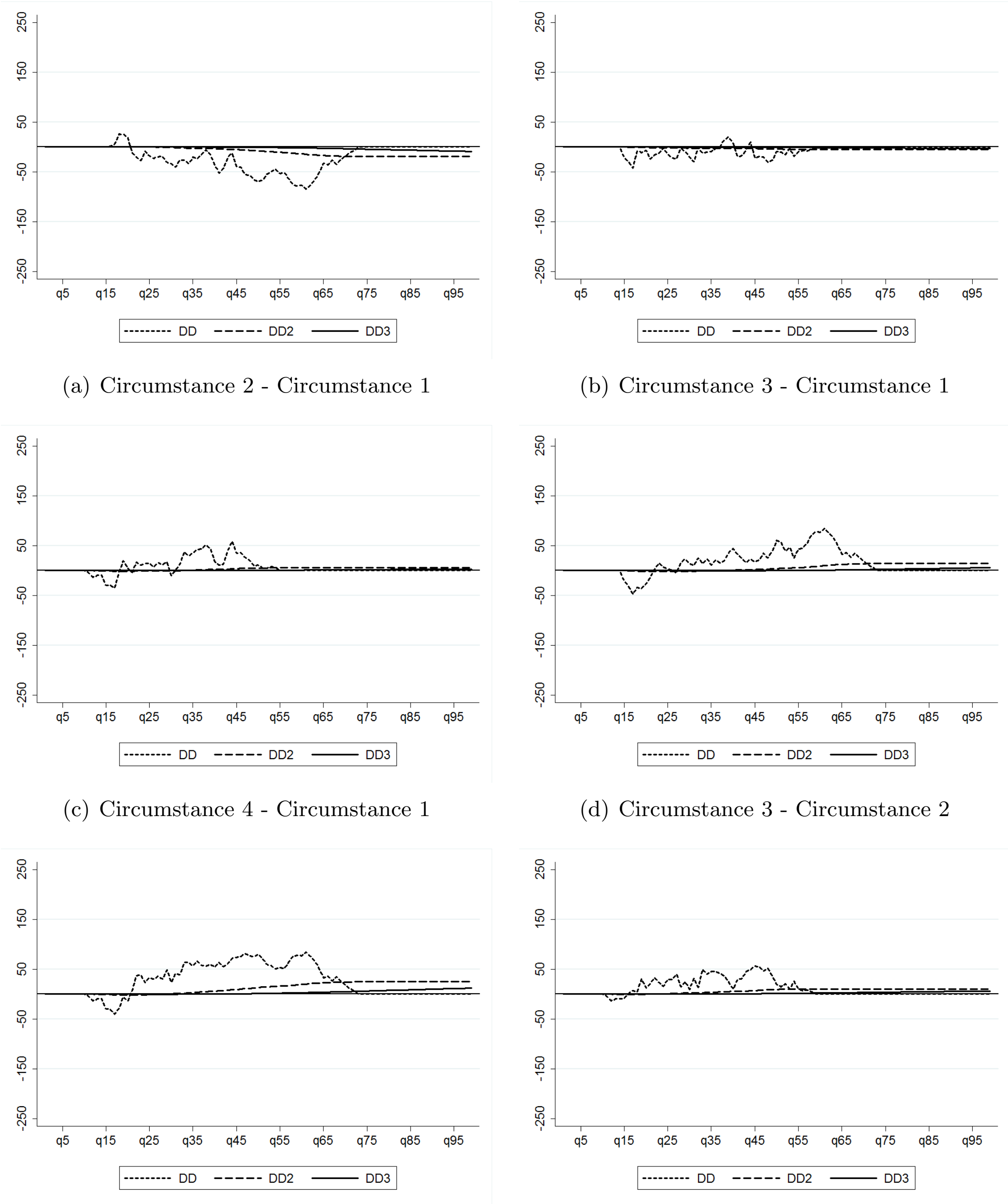


(e) Circumstance 4 - Circumstance 2



(f) Circumstance 4 - Circumstance 3

Figure 3.11: Differences in quantile functions (D), *GL* curves (D2) and integrals of the *GL* curves (D3) computed at each percentile of the simulated earnings distribution with policy treatment. Values on the horizontal axis refer to percentiles of the simulated earnings distribution. Values on the vertical axes express the difference between curves, in Euros. The curves represent the differences between the outcomes prospect associated to two distinct circumstances, for a total of six comparisons. Earnings differences are trimmed at 300 and -100 Euro.



(a) Circumstance 2 - Circumstance 1

(b) Circumstance 3 - Circumstance 1

(c) Circumstance 4 - Circumstance 1

(d) Circumstance 3 - Circumstance 2

(e) Circumstance 4 - Circumstance 2

(f) Circumstance 4 - Circumstance 3

Figure 3.12: Difference in differences in quantile functions (D), GL curves (D2) and integrals of the GL curves (D3) computed at each percentile of the actual and simulated earnings distributions. Values on the horizontal axis refer to percentiles of the earnings distribution. Values on the vertical axes express the difference across policies in the differences between earning gaps, GL curves gaps and gaps in the integrals of GL curves associated to pairs of circumstances, in Euros. Earnings differences in differences trimmed at 250 and -250 Euro.

Table 3.12: Descriptive statistics: covariates, by treatment group (IV)

	Treatment (IV=1)		Comparison (IV=0)	
	(1)		(2)	
Individual characteristics:				
Wage, monthly, in Euro	1,676.578	[2,876.4]	1,737.303	[3,246.8]
Prizes	0.511	[0.5]	0.525	[0.5]
Weekly working hours	40.120	[9.4]	40.338	[9.7]
Self employed	0.022	[0.1]	0.026	[0.2]
Employed in the public sector	0.244	[0.4]	0.251	[0.4]
Education, years	12.116	[3.3]	11.903	[3.6]
Age, in years (above 15)	43.984	[6.5]	46.165	[6.0]
Marriage status	0.758	[0.4]	0.790	[0.4]
Number of children below 18	1.034	[1.1]	0.907	[1.1]
Socioeconomic conditions of the father:				
Father without french nationality	0.066	[0.2]	0.060	[0.2]
Circumstance 2	0.539	[0.5]	0.533	[0.5]
Farmers	0.113	[0.3]	0.119	[0.3]
Manual worker	0.456	[0.5]	0.443	[0.5]
Circumstance 3	0.220	[0.4]	0.242	[0.4]
Artisans	0.101	[0.3]	0.109	[0.3]
Non manual workers	0.140	[0.3]	0.151	[0.4]
Circumstance 4	0.174	[0.4]	0.165	[0.4]
H-grade prof.	0.075	[0.3]	0.075	[0.3]
L-grade prof.	0.115	[0.3]	0.104	[0.3]
Age of leaving education	18.116	[3.3]	17.903	[3.6]
$(cob - 1953)^2$	1.667	[1.7]	4.559	[3.3]
$(cob - 1953)^3$	3.002	[3.6]	-11.658	[10.9]
$(cob - 1953)^4$	5.672	[7.3]	31.634	[34.5]
Trimmed proportion of sample size	0.672	[0.5]	0.676	[0.5]
Groups interested by policy intervention:				
Receives policy treatment	0.540	[0.5]	0.432	[0.5]
Δ *policy treatment*		0.108*** (.006)		
Marginal students (target)	0.160	[0.4]	0.268	[0.4]
Sample size	13,364		12,516	

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

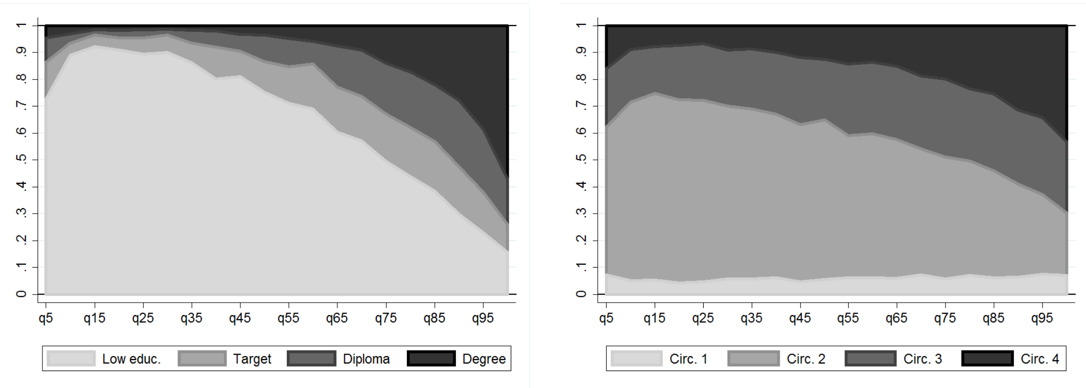
Notes: Sample reduced to French male earners where circumstances have been recorded, cohorts 1950 to 1955. IV is a dummy for cohorts 1953 to 1955. Treatment and comparison groups are defined upon the IV. Standard deviations of all the covariates are reported in brackets. Difference between control and treatment groups are not statistically different form zero at standard significance levels. *cob* identifies the cohort of birth. Trimmed sample size refers to the sub-sample of those who at most have an high school diploma. The group receiving policy treatment is given by those who completed primary education but did not qualify above this level. Marginal students are defined as the target group used to simulate policy intervention. *** indicates significance at 1%

Table 3.13: Earnings distributions, by cohorts before and after introduction of the policy for selected quantiles.

Quantiles	Overall (1)	Target (2)	Circ. 1 (3)	Circ. 2 (4)	Circ. 3 (5)	Circ. 4 (6)
	Before policy implementation					
Q5%	499.1	618.2	394.0	474.8	606.3	569.1
Q10%	944.6	883.6	883.6	914.7	975.1	1,066.5
Q25%	1,226.7	1,097.0	1,269.4	1,173.3	1,275.2	1,448.3
Q50%	1,534.3	1,305.6	1,638.4	1,427.6	1,620.6	1,934.4
Q75%	2,011.7	1,529.1	2,164.2	1,808.7	2,134.3	2,748.7
Q90%	2,825.0	1,840.1	3,049.0	2,316.6	2,935.4	3,876.1
Q95%	3,535.4	2,147.0	3,841.1	2,779.7	3,665.4	4,976.7
Mean	1,825.7	1,378.5	1,940.3	1,597.6	1,875.4	2,431.4
	[3,026.9]	[2,102.1]	[3,868.8]	[2,378.3]	[2,270.0]	[4,785.9]
Gini	0.303	0.204	0.330	0.256	0.287	0.352
	(0.006)	(0.015)	(0.022)	(0.008)	(0.010)	(0.018)
Sample size	26,421	5,585	1,682	14,134	6,103	4,502
	After policy implementation					
Q5%	499.1	618.2	394.0	474.8	606.3	569.1
Q10%	944.6	883.6	883.6	914.7	975.1	1,066.5
Q25%	1,264.3	1,097.0	1,290.2	1,219.6	1,310.5	1,473.3
Q50%	1,574.9	1,447.9	1,656.3	1,493.4	1,656.3	1,934.4
Q75%	2,011.7	1,676.1	2,164.2	1,808.7	2,134.3	2,748.7
Q90%	2,825.0	1,840.1	3,049.0	2,316.6	2,935.4	3,876.1
Q95%	3,535.4	2,147.0	3,841.1	2,779.7	3,665.4	4,976.7
Mean	1,842.5	1,458.3	1,950.0	1,621.0	1,888.0	2,436.3
	[3,024.7]	[2,102.7]	[3,867.5]	[2,376.5]	[2,267.6]	[4,784.9]
Gini	0.299	0.197	0.326	0.251	0.284	0.351
	(0.006)	(0.014)	(0.022)	(0.008)	(0.010)	(0.018)
Sample size	26,421	5,585	1,682	14,134	6,103	4,502

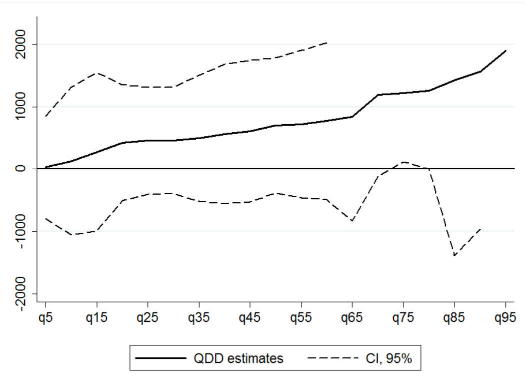
Source: Estimates from Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Earnings quantiles for earnings distribution detrended by the age effect. Sample reduced to Frech male earners where circumstances have been recorded, cohorts 1950 to 1955. Earnings after policy implementation are obtained by assigning quantile treatment effects estimated by model (1) in table 3.5 to the target group. Standard deviations reported in brackets. Gini index are reported for each subgroup's earnings distribution. Standard errors in parentheses are calculated by bootstrapping 100 replications of the Gini index.

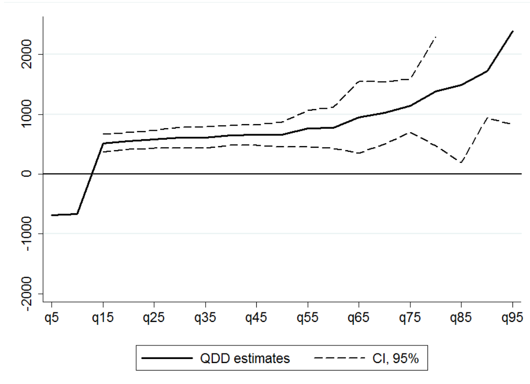


(a) By educational achievement (treatment) (b) By circumstances

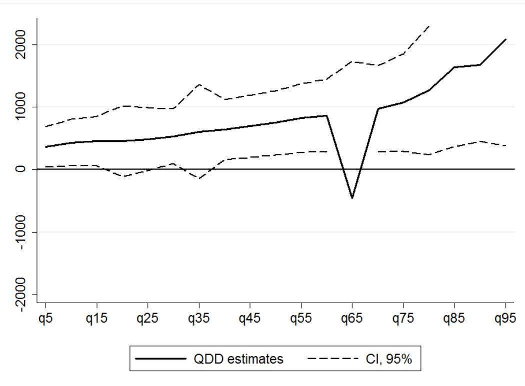
Figure 3.13: Distribution of population across educational achievement levels (a) and circumstances (b), by earnings quantiles, 5% intervals, expressed in cumulative shares. Scores have been calculated from a multinomial logit model. In panel (a), the *target* group refers to having a technical degree or *baccalauréat*. Circumstances are defined according to the father socioeconomic status. qX represent a 5% share of the population between quantiles QX% and QX%-5% in the overall earnings distribution.



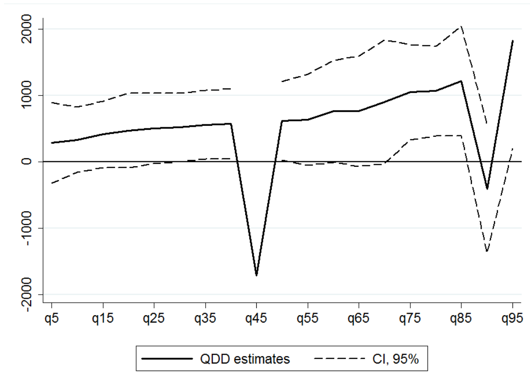
(a) Circumstance 1



(b) Circumstance 2

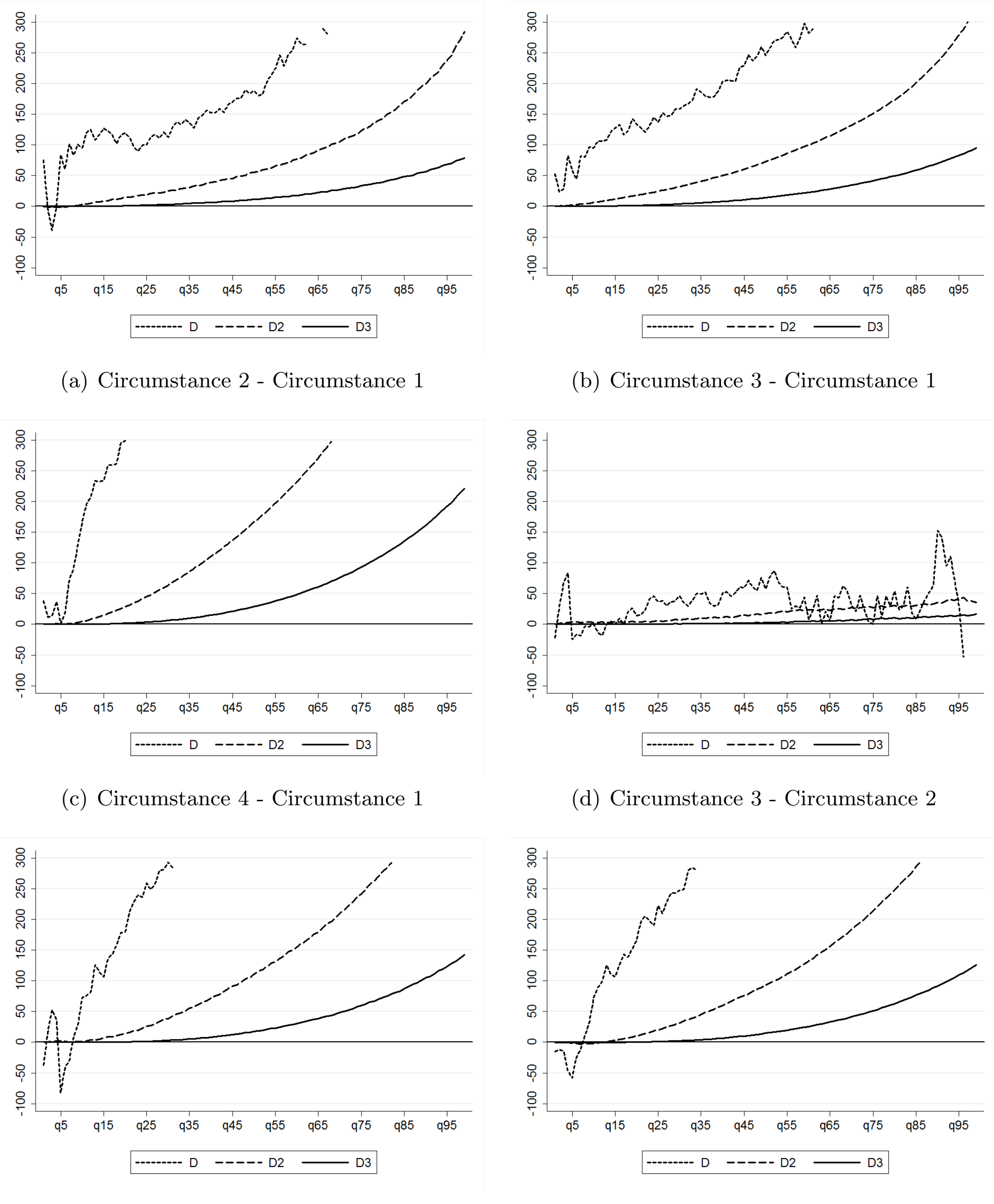


(c) Circumstance 3



(d) Circumstance 4

Figure 3.14: QTE of the impact of access to the higher education on earnings, by circumstances. Sample of cohorts 1946, 1948, 1948 and 1952, male earners. Quantile treatment effects estimated via IV are computed at 5% intervals, the CI at 95% is computed with robust standard errors. Controls: age and cohort trends and year of survey. Data omitted if above (below) 2000 (-2000) Euro



(a) Circumstance 2 - Circumstance 1

(b) Circumstance 3 - Circumstance 1

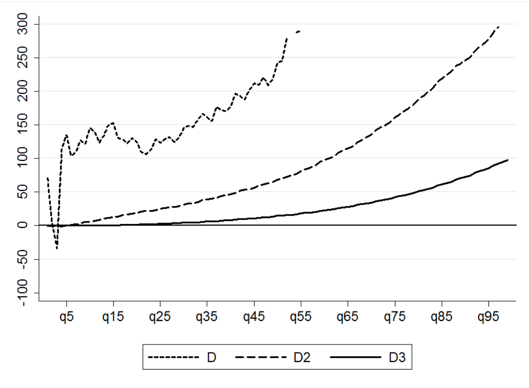
(c) Circumstance 4 - Circumstance 1

(d) Circumstance 3 - Circumstance 2

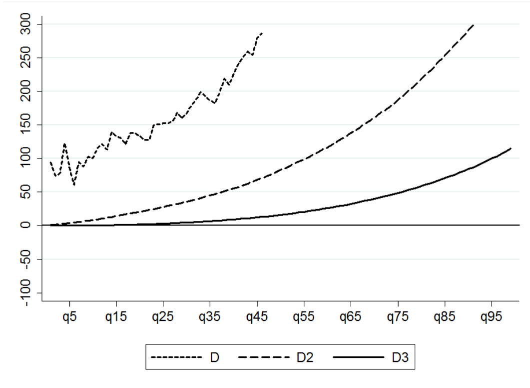
(e) Circumstance 4 - Circumstance 2

(f) Circumstance 4 - Circumstance 3

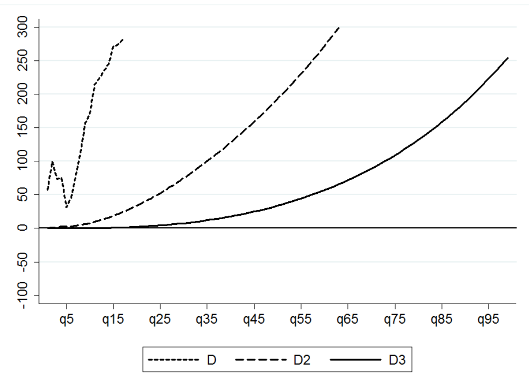
Figure 3.15: Differences in quantile functions (D), *GL* curves (D2) and integrals of the *GL* curves (D3) computed at each percentile of the actual earnings distribution without policy treatment. Values on the horizontal axis refer to percentiles of the actual earnings. Values on the vertical axes express the differences between curves, in Euros. The curves represent the differences between the outcomes prospect associated to two distinct circumstances, for a total of six comparisons. Earnings differences are trimmed at 300 and -100 Euro.



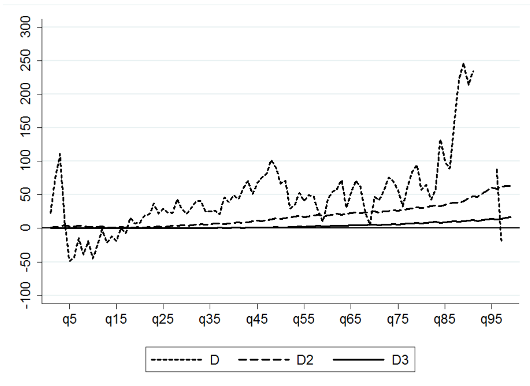
(a) Circumstance 2 - Circumstance 1



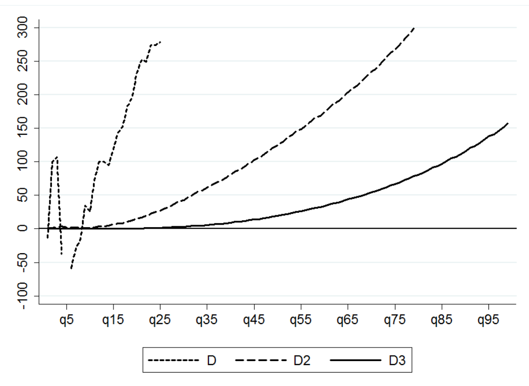
(b) Circumstance 3 - Circumstance 1



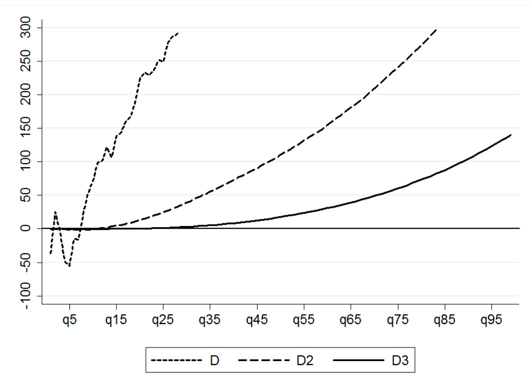
(c) Circumstance 4 - Circumstance 1



(d) Circumstance 3 - Circumstance 2

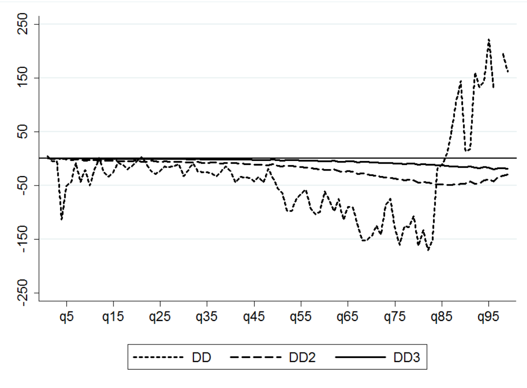


(e) Circumstance 4 - Circumstance 2

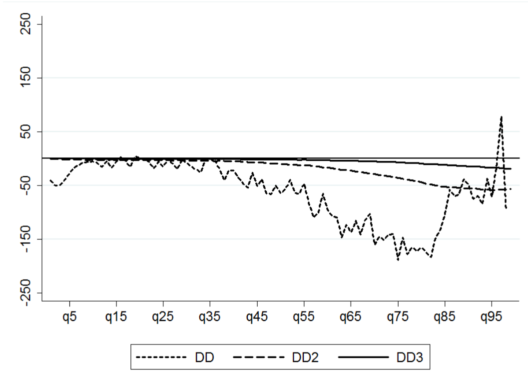


(f) Circumstance 4 - Circumstance 3

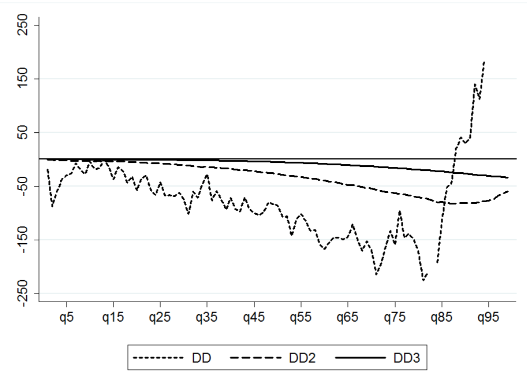
Figure 3.16: Differences in quantile functions (D), GL curves (D2) and integrals of the GL curves (D3) computed at each percentile of the simulated earnings distribution with policy treatment. Values on the horizontal axis refer to percentiles of the simulated earnings distributions. Values on the vertical axes express the difference between curves, in Euros. The curves represent the differences between the outcomes prospect associated to two distinct circumstances, for a total of six comparisons. Earnings differences are trimmed at 300 and -100 Euro.



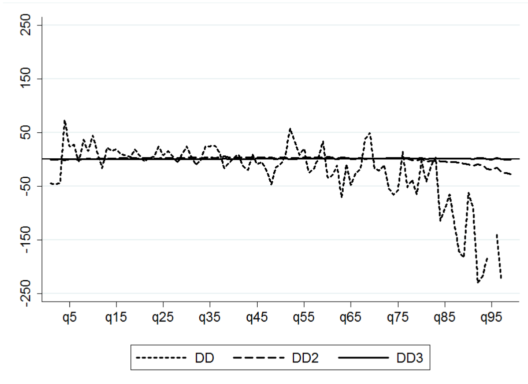
(a) Circumstance 2 - Circumstance 1



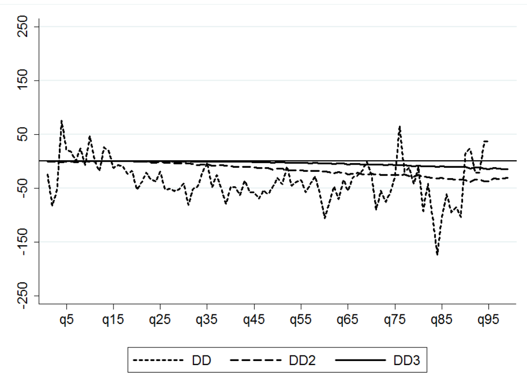
(b) Circumstance 3 - Circumstance 1



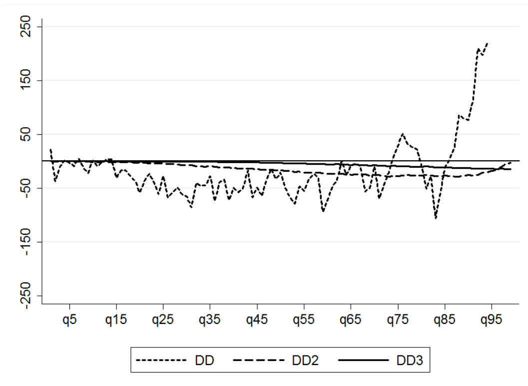
(c) Circumstance 4 - Circumstance 1



(d) Circumstance 3 - Circumstance 2



(e) Circumstance 4 - Circumstance 2



(f) Circumstance 4 - Circumstance 3

Figure 3.17: Difference in differences in quantile functions (D), GL curves (D2) and integrals of the GL curves (D3) computed at each percentile of the actual and simulated earnings distributions. Values on the horizontal axis refer to percentiles of the earnings distribution. Values on the vertical axes express the difference across policies in the differences between earning gaps, GL curves gaps and gaps in the integrals of GL curves associated to pairs of circumstances, in Euros. Earnings differences in differences are trimmed at 250 and -250 Euro.

Table 3.14: Descriptive statistics: covariates, by treatment group (IV)

	Treatment (IV=1)		Comparison (IV=0)	
	(1)		(2)	
Individual characteristics				
Monthly earnings (Euro)	1,890.137	[4,622.8]	1,784.303	[3,482.9]
Prizes	0.526	[0.5]	0.530	[0.5]
Hours worked (week)	40.446	[10.4]	40.300	[10.0]
Self employed	0.029	[0.2]	0.028	[0.2]
Employed (public)	0.269	[0.4]	0.258	[0.4]
Education (years)	11.881	[3.8]	11.852	[3.7]
Age (above 15)	47.799	[5.4]	47.174	[5.8]
Marriage status	0.822	[0.4]	0.802	[0.4]
Nember of children below 18	0.752	[1.0]	0.796	[1.0]
Socioeconomic conditions of the father:				
Father without French nationality	0.056	[0.2]	0.061	[0.2]
Circumstance 2	0.534	[0.5]	0.528	[0.5]
Farmers	0.132	[0.3]	0.127	[0.3]
Manual worker	0.428	[0.5]	0.429	[0.5]
Circumstance 3	0.239	[0.4]	0.246	[0.4]
Artisans	0.110	[0.3]	0.111	[0.3]
Non manual workers	0.146	[0.4]	0.153	[0.4]
Circumstance 4	0.170	[0.4]	0.165	[0.4]
H-grade prof.	0.072	[0.3]	0.075	[0.3]
L-grade prof.	0.113	[0.3]	0.103	[0.3]
Cohort 1946	-		0.165	[0.4]
Cohort 1952	-		0.227	[0.4]
Groups interested by policy intervention				
Receives policy treatment	0.225	[0.4]	0.213	[0.4]
Δ *policy treatment*		0.014** (.006)		
Marginal students (target)	0.121	[0.3]	0.113	[0.3]
Sample size	7,786		19,207	

Source: Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Sample reduced to French male earners where circumstances have been recorded, cohorts 1946 to 1952. IV is a dummy for cohorts 1948 and 1949. Treatment and comparison groups are defined upon the IV. Standard deviations of all the covariates are reported between brackets. No difference between control and treatment groups is significant at standard significance levels. The group receiving policy treatment is given by those who completed the higher education system with no qualification. Marginal students are defined as the target group where policy has to be implemented.

Table 3.15: Earnings distributions, by cohorts before and after introduction of the policy for selected quantiles.

Quantiles	Overall (1)	Target (2)	Circ. 1 (3)	Circ. 2 (4)	Circ. 3 (5)	Circ. 4 (6)
	Before policy implementation					
Q5%	666.7	540.2	543.5	655.2	731.8	666.7
Q10%	990.9	1,118.4	990.9	972.1	1,043.9	1,124.6
Q25%	1,270.8	1,505.7	1,306.5	1,202.4	1,346.3	1,544.6
Q50%	1,605.5	1,839.9	1,680.3	1,477.6	1,719.8	2,083.5
Q75%	2,139.0	2,342.9	2,282.1	1,886.8	2,263.6	2,954.0
Q90%	3,042.9	3,049.0	3,106.4	2,442.9	3,137.6	4,183.5
Q95%	3,868.7	3,806.6	4,091.1	3,042.9	4,021.1	5,383.5
Mean	1,963.1	2,076.2	1,965.7	1,710.5	2,052.5	2,630.1
	[3,807.1]	[2,958.3]	[1,321.2]	[3,211.4]	[4,154.7]	[5,248.7]
Gini	.319	.267	.332	.273	.310	.357
	(0.007)	(0.010)	(0.019)	(0.011)	(0.015)	(0.019)
Sample size	27,536	3,185	1,637	14,566	6,725	4,608
	After policy implementation					
Q5%	704.1	815.3	597.0	681.3	765.1	723.1
Q10%	1,002.6	1,590.3	1,021.4	972.1	1,048.4	1,133.6
Q25%	1,277.0	2,135.5	1,339.9	1,216.1	1,353.3	1,605.5
Q50%	1,651.9	2,718.3	1,757.9	1,499.5	1,794.9	2,191.7
Q75%	2,291.5	3,395.0	2,496.6	1,956.8	2,490.4	3,131.8
Q90%	3,308.6	4,423.8	3,411.3	2,741.2	3,487.0	4,433.5
Q95%	4,186.1	5,017.1	4,268.6	3,353.9	4,383.5	5,424.6
Mean	2,047.1	2,831.0	2,069.6	1,779.0	2,173.3	2,707.6
	[3,709.6]	[1,145.2]	[1,350.7]	[3,219.4]	[4,179.9]	[4,765.1]
Gini	.296	.215	.326	.264	.285	.318
	(0.005)	(0.001)	(0.013)	(0.006)	(0.011)	(0.015)
Sample size	27,420	3,069	1,627	14,533	6,695	4,565

Source: Estimates from Labor Force Survey 1990, 1993, 1996, 1999, 2004, 2006, 2008 and 2010.

Notes: Earnings quantiles for earnings distribution detrended by the age effect. Sample reduced to French male earners where circumstances have been recorded, cohorts 1946 to 1952. Earnings after policy implementation are obtained by assigning quantile treatment effects estimated by model (1) in table 3.10 to the target group. Standard deviations reported in brackets. Gini index are reported for each subgroups earnings distribution. Standard errors in parentheses are calculated by bootstrapping 100 replications of the Gini index.

Conclusions

This thesis studies the notion of dissimilarity, and shows that many relevant economic and sociological problems involve dissimilarity comparisons. Problems such as segregation, discrimination, multidimensional inequality, intergenerational mobility across economic or social classes are studied by representing data as distribution matrices, and then by ordering these matrices according to the dissimilarity order. In this setting, social groups are rows, and discrete outcome achievements are represented by columns of the matrices. The entries of a distribution matrix are the relative population frequencies associated to each outcome level, made conditional upon groups. Chapter 1 contributes by showing that a minimal number of operations on the rows and columns of a distribution matrix allows to obtain another matrix, where the groups distributions represented in the latter are less dissimilar than the groups distributions represented in the former.

Dissimilarity is a multidimensional phenomenon that involves the comparison of more than two groups at a time. In fact, when two distribution matrices involving more than two groups can be ranked according to the dissimilarity criterion, one obtains that also any distribution matrix obtained from the original one by considering only pairs of groups can also be ranked in the same way. Of course, the inverse is not true, thus showing the relevance of the dissimilarity order for studying multi-group social phenomena.

Chapter 1 illustrates that there are different and meaningful transformations of the data that can be used to characterize the dissimilarity order. However, not all these transformation are allowed in different frameworks. Two of them, the merge and the exchange of population masses, may generate counterintuitive results when both are applied (or combined) with ordered or non-ordered classes indistinctly. One important contribution of chapter 1 is to show that depending on the choice of the merge or exchange operations, one obtains two separated majorization criteria to compare distribution matrices coherently with the dissimilarity partial order. This innovative result provides the necessary background to exploit implementable criteria for the dissimilarity comparisons.

In the case of two groups, dissimilarity can be exploited by looking at the ranking produced by segregation curves (if the outcome levels are non-ordered) or discrimination curves (if ordered). In the multi-group setting, the same comparisons applied to every possible pair of groups may lead to the wrong conclusion that dissimilarity, as measured by these two criteria, is decreased, when actually it is not. This is well illustrated, for instance, in the short empirical application on the changes in multi-group immigrants segregation registered in the an Italian city. Chapter 1 shows that dissimilarity comparisons can be implemented by studying the inclusion properties of well defined geometric objects: the Zonotopes inclusion order (if outcome levels are non-ordered) or the Path Polytope inclusion order (if outcomes are ordered). Both criteria can be implemented by comparing two matrices of finite size.

The algorithm illustrated in the appendix of chapter 1 proves that the Zonotopes inclusion is implementable. The Path Polytope inclusion is more problematic, since the central symmetry characterizing the Zonotopes cannot be exploited to construct feasible procedures to verify the inclusion. However, chapter 1 shows that the Path Polytope partial order can be equivalently tested by a finite sequence of Lorenz dominance comparisons between cumulative groups proportions at fixed overall population proportion. This last criterion is implementable, and the only difficulty consists in defining the relevant population shares at which the Lorenz comparisons has to be made. A definition of the rules that such division of the original data matrix has to satisfy is provided. The implementation of this criterion is left for future research, along with the study of the sample properties of the Zonotope and the Path Polytope inclusion criteria. In fact, when groups frequencies distributions are obtained from survey data, one can at most *estimate* the probability masses and associate to them their standard errors. Hence, Zonotope or Path Polytope inclusion criteria can only be statistically verified, for a given level of confidence.

The Zonotope and the Path Polytope inclusion criteria define partial orders. When inclusion is not verified, two configurations cannot be ranked in the sense of the dissimilarity order. However, additional structure may provide a solution: an index of dissimilarity. This index should be of course consistent with the axiomatic approach underlying the dissimilarity order. Some possible extensions in this direction are provided. For instance, one can use the volume of the Zonotope as an index of dissimilarity in the non-ordered setting. This is done, for instance, in chapter 2. The idea is, however, not new (Gini 1912, Koshevoy and Mosler 1997). In the ordered case, one can instead focus on distances between distributions. In the two groups case, a distance measure coherent with this setting can be represented as an average of the differences between the values of the two cumulative distribution functions associated to the groups, calculated in correspondence of the same quantile, and weighted by the overall population share at that quantile. It is shown that, in the two groups case, this distance measures the area of the two-dimensional Path Polytope. A direct extension of this distance indicator to the multi-group case may be constructed by looking at the multi-group Path Polytope volume.[25]

Chapter 2 exploits the dissimilarity partial order in the context of measuring multi-group exposure segregation from the perspective of the individual. Hence, segregation can be studied by mean of Zonotopes inclusion, since in this setting individuals are now interpreted as the "non-ordered outcome classes" in a distribution matrix. The innovative result of the paper consists in defining the class of segregation indicators that are coherent with the dissimilarity partial order. A new index of segregation from this class is also studied. It is the the Gini Exposure index, which coincides with the volume of the Zonotope. It is not therefore by chance that the empirical rank correlation between this index and the multi-group dissimilarity index (Reardon and Firebaugh 2002) is sizable and, as shown by the data, also robust with respect to the variability of the main factors used to compute the segregation measures, such as population shares, ratios of shares, group proportions and so on.

[25] The Path Polytope volume, differently from the Zonotope volume, is always defined, since the dimension of the Path Poytope can never reduce below the dimension of the space in which the Path Polytope is defined.

It is not possible to show if the multi-group dissimilarity index belongs to the family of indicators studied in this chapter, because the multi-group dissimilarity indicator has been designed to deal with issues of segregation from an organizational unit perspective (where interaction profiles cannot be meaningfully measured, unless everybody in the same unit is assumed to be endowed with the same interaction profile).

The analysis in chapter 2 stems from the comparison, at individual level, of the interaction profiles, listing the probabilities of interaction with the different groups in which a population is partitioned. The analysis does not build directly on the dissimilarity between the distributions of the groups across organizational units. This setting allows to tackle an important issue in segregation measurement that is represented by the interplay between the anonymity of the groups and of the individuals. The analysis in chapter 2 exploits this interplay, which is a result of the axiomatization of the exposure segregation ordering. In general, however, one may motivate that two symmetric patterns that give the same measured segregation level (according to virtually all indices of segregation) may not be ranked as equally segregated (as in the introductory example involving the Greens and the Reds). A possible extension of the segregation ordering, that is left for future research, consists in replacing the anonymity requirements with weaker conditions. This can be done, for instance, by imposing that the label of the group of the individuals must be also taken into account in the segregation measurement, and especially in assessing how segregation changes according to which interaction profiles are merged.

The detailed analysis of dissimilarity performed in chapter 1 illustrates three fundamentals of any dissimilarity comparison: First, one configuration is ranked less dissimilar than another if it is closer to the configuration where perfect similarity is achieved. How far or close a set of distributions is from the configuration displaying similarity rests on the normative approach undertaken in the dissimilarity analysis. Second, the degree of dissimilarity does not (in general) depend on a reference distribution that is exogenously imposed. Third, dissimilarity is a multi-group phenomenon. The analysis of pairwise comparisons of groups may not be sufficient to assess the degree of dissimilarity, and it may even lead to counterintuitive result. Chapter 3 proposes an equalization of opportunity criterion that is grounded on these three fundamentals.

The equalization of opportunity criterion applies the notion of dissimilarity to the evaluation of changes in the degree of equality of opportunity due to policy intervention, and builds on the three fundamentals listed above, in a context where the ordinal and cardinal information associated to an outcome distribution across groups are both taken into consideration. The interpretation of the three fundamentals in this context is as follows. First, the reference configuration displaying perfect similarity is the situation where Roemer's equality of opportunity is reached, that is the outcome distributions of the different groups coincide. How far or close a configuration is from satisfying equality of opportunity is modeled by exploiting ordinal and distance criteria.

Second, implementation of the ordinal and distance comparisons only rests upon the information on the outcome distributions made conditional on circumstances. The equalization of opportunity criterion builds on these two fundamentals to evaluate equalization from the perspective of pairwise comparisons of distributions.

Third, only very weak criteria of equalization of opportunity, discussed in the chapter, can be implemented by looking at the whole distribution of circumstances. Two alternatives are proposed: the first alternative exploits an ordinal criteria based on the set of non-dominated circumstances; the second alternative defines the sufficient number of economic distance comparisons that insure opportunity equalization, when the outcomes prospects associated to different circumstances can be ordered according to some well defined class of preferences.

However, when satisfied, these weak multi-group criteria are not sufficient to guarantee equalization between all pairs of distributions. Although this may seem in clear contrast with the third fundamental of dissimilarity comparisons, there is no contradiction at all. In fact, the generalizations of the equalization of opportunity criterion are still grounded on pairwise comparisons of distributions, although it is motivated that some of these comparisons are not useful from a normative perspective, and therefore overlooked when implementing weaker criteria. This problem is however common to all the alternative evaluation methods proposed in the literature, which define either partial or complete rankings of distributions on the ground of pairwise comparisons.

This thesis concludes that equalization of opportunity has to be stated on the ground of a dissimilarity comparison between conditional distribution functions. To build this comparison, one has to achieve the three fundamentals at the basis of the dissimilarity order. Two of them has been incorporated in the equalization of opportunity criterion proposed in chapter 3. However, this *paretian* equalization of opportunities criterion is defined on pairwise comparisons, and it does not completely fulfill the third fundamental behind a dissimilarity order. There is, however, a tradeoff between normative requirements and empirical implementation possibilities, and the opportunity equalization criterion exposed here seems already a very demanding one. The experiment conducted with actual and simulated data seems to validate this concern. In fact, even the reforms with sizable simulated effects, such as the opening of the higher education system analyzed in chapter 3, may fail to equalize opportunities because of the role of few comparisons of circumstances pairs.

The empirical application in chapter 3 is not only meant to show that the equalization criterion is empirically testable, but also it shows that the equalization evaluation that one obtains does not go at odd with the expected effects of the simulated policies. Two possible extensions are probably of interest. The first extension, is to apply the equalization criterion to educational policies that take place early in the life of the students. These policies, such as the opening of kindergarten accessibility, have a strong tendency to "equalize the playing field" at the very beginning of the students lives, notwithstanding the impact on household reallocation of time into labor, leisure and child nurturing. This motivates the second extension, that should point in the direction of calculating the economic value of equality of opportunity, both from a social planner and individual perspective, thus fulfilling the Atkinson's (1998) view that equality of opportunity has to do also with "ex ante expectations." This measure can be for instance compared with the economic returns from policy implementation. The Opportunity Index defined in chapter 3 is a (very partial) attempt to construct such a measure, and it can be used to attenuate or inflate the actual average impact of policy treatment in a way that is consistent with equality of opportunity concerns.

Bibliography

Aaberge, R. (2009). Ranking intersecting lorenz curves, *Social Choice and Welfare* **33**: 235–259.

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models, *Journal of the American Statistical Association* **97**(457): 284–292.

Abadie, A., Angrist, J. and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings, *Econometrica* **70**(1): 91–117.

Alonso-Villar, O. and del Rio, C. (2010). Local versus overall segregation measures, *Mathematical Social Sciences* **60**: 30–38.

Andreoli, F. (2012). Multigroup segregation patterns and determinants: The case of immigrants in an Italian city, *in* J. A. Bishop and R. Salas (eds), *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber*, Vol. 20 of *Research on Economic Inequality*, Emerald Group Publishing Limited, chapter 4, pp. 91–116.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?, *The Quarterly Journal of Economics* **106**(4): pp. 979–1014.

Asplund, E. and Bungart, L. (1966). *A First Course in Integration*, Holt Rinehart & Winston, New York.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models, *Econometrica* **74**(2): pp. 431–497.

Atkinson, A. B. (1998). Social exclusion, poverty and unemployment, *in* A. B. Atkinson and J. Hills (eds), *Exclusion, Employment and Opportunity*, Centre for Analysis of Social Exclusion, LSE, London.

Atkinson, A. B. and Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status, *The Review of Economic Studies* **49**(2): 183–201.

Atkinson, A. B. and Marlier, E. (2010a). Income and living conditions in Europe, *Technical report*, Eurostat, European Union.

Atkinson, A. and Marlier, E. (2010b). Analysing and measuring social inclusion in a global context, *Technical report*, United Nations.

Atkinson, T. B. (1970). On the measurement of inequality, *Journal of Economic Theory* **2**: 244–263.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance, *Econometrica* **71**(1): pp. 71–104.

Barrett, G. F. and Donald, S. G. (2009). Statistical inference with generalized Gini indices of inequality, poverty, and welfare, *Journal of Business & Economic Statistics* **27**(1): 1–17.

Beach, C. M. and Davidson, R. (1983). Distribution-free statistical inference with Lorenz curves and income shares, *The Review of Economic Studies* **50**(4): 723–735.

Bertino, S., Drago, E., Landenna, G., Leti, G. and Marasini, D. (1987). Dissimilarity and connection, *in* A. Naddeo (ed.), *Italian Contributions to the Methodology of Statistics*, Cleup Padova, chapter 3, pp. 83 – 109.

Björklund, A. and Salvanes, K. G. (2011). *Education and Family Background: Mechanisms and Policies*, Vol. 3 of *Handbook of the Economics of Education*, Elsevier, chapter 3, pp. 201–247.

Blackwell, D. (1953). Equivalent comparisons of experiments, *The Annals of Mathematical Statistics* **24**(2): 265–272.

Borjas, G. J. (1992). Ethnic capital and intergenerational mobility, *The Quarterly Journal of Economics* **107**(1): pp. 123–150.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities, *The American Economic Review* **85**(3): 365–390.

Bourguignon, F., Ferreira, F. and Walton, M. (2007). Equity, efficiency and inequality traps: A research agenda, *Journal of Economic Inequality* **5**: 235–256.

Braga, M., Checchi, D. and Meschi, E. (2011). Institutional reforms and educational attainment in Europe: A long run perspective, *IZA Discussion Papers 6190*, Institute for the Study of Labor (IZA).

Butler, R. J. and McDonald, J. B. (1987). Interdistributional income inequality, *Journal of Business & Economic Statistics* **5**(1): pp. 13–18.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling, *NBER Working Papers 4483*, National Bureau of Economic Research, Inc.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems, *Econometrica* **69**(5): 1127–60.

Carrington, W. J. and Troske, K. R. (1997). On measuring segregation in samples with small units, *Journal of Business & Economic Statistics* **15**(4): 402–409.

Chakravarty, S. R. and Dutta, B. (1987). A note on measures of distance between imcome distributions, *Journal of Economic Theory* **41**(1): 185 – 188.

Chakravarty, S. R. and Silber, J. (2007). A generalized index of employment segregation, *Mathematical Social Sciences* **53**(2): 185 – 195.

Checchi, D. and Peragine, V. (2010). Inequality of opportunity in Italy, *Journal of Economic Inequality* **8**: 429–450.

Cunha, F. and Heckman, J. (2007). The technology of skill formation, *American Economic Review* **97**(2): 31–47.

Cunha, F., Heckman, J. J. and Lochner, L. (2006). *Interpreting the Evidence on Life Cycle Skill Formation*, Vol. 1 of *Handbook of the Economics of Education*, Elsevier, chapter 12, pp. 697–812.

Dahl, G. (1999). Matrix majorization, *Linear Algebra and its Applications* **288**: 53 – 73.

Dardanoni, V. and Forcina, A. (1999). Inference for Lorenz curve orderings, *Econometrics Journal* **2**: 49–75.

Davidson, R. and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality, *Econometrica* **68**(6): 1435–1464.

Donaldson, D. and Weymark, J. A. (1983). Ethically flexible Gini indices for income distributions in the continuum, *Journal of Economic Theory* **29**(2): 353 – 358.

Donaldson, D. and Weymark, J. A. (1998). A quasiordering is the intersection of orderings, *Journal of Economic Theory* **78**(2): 382 – 387.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes, *American Sociological Review* **20**(2): 210–217.

Dworkin, R. (1981a). What is equality? Part 1: Equality of welfare, *Philosophy & Public Affairs* **10**(3): pp. 185–246.

Dworkin, R. (1981b). What is equality? Part 2: Equality of resources, *Philosophy & Public Affairs* **10**(4): pp. 283–345.

Ebert, U. (1984). Measures of distance between income distributions, *Journal of Economic Theory* **32**(2): 266 – 274.

Echenique, F. and Fryer, R. G. (2007). A measure of segregation based on social interactions, *Quarterly Journal of Economics* **122**(2): 441–485.

Echenique, F., Fryer, Roland G., J. and Kaufman, A. (2006). Is school segregation good or bad?, *American Economic Review* **96**(2): 265–269.

Eeckhoudt, L. and Schlesinger, H. (2006). Putting risk in its proper place, *The American Economic Review* **96**(1): pp. 280–289.

Ferreira, F. H. G. and Gignoux, J. (2011). The measurement of educational inequality: Achievement and opportunity, *Policy Research Working Paper Series 5873*, The World Bank.

Fields, G. S. and Fei, J. C. H. (1978). On inequality comparisons, *Econometrica* **46**(2): pp. 303–316.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects, *Econometrica* **75**(1): pp. 259–276.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009). Unconditional quantile regressions, *Econometrica* **77**(3): 953–973.

Fishburn, P. C. (1976). Continua of stochastic dominance relations for bounded probability distributions, *Journal of Mathematical Economics* **3**(3): 295 – 311.

Fleurbaey, M. (2008). *Fairness, Responsability and Welfare*, Orxford University Press.

Flückiger, Y. and Silber, J. (1999). *The Measurement of Segregation in the Labour Force*, Physica-Verlag, Heidelberg, Germany.

Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures, *Econometrica* **52**(3): 761–66.

Frankel, D. M. and Volij, O. (2011). Measuring school segregation, *Journal of Economic Theory* **146**(1): 1 – 38.

Frölich, M. and Melly, B. (2008). Unconditional quantile treatment effects under endogeneity, *IZA Discussion Papers 3288*, Institute for the Study of Labor (IZA).

Frölich, M. and Melly, B. (2010). Estimation of quantile treatment effects with Stata, *Stata Journal* **10**(3): 423–457.

Fusco, A. and Silber, J. (2011). Ordinal variables and the measurement of polarization, *CEPS/INSTEAD Working Paper Series 2011-33*, CEPS/INSTEAD.

Garcia-Gomez, P., Schokkaert, E., Ourti, T. V. and d'Uva, T. B. (2012). Inequity in the face of death, *Tinbergen Institute Discussion Papers 12-084/3*, Tinbergen Institute.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve, *Econometrica* **39**(6): 1037–1039.

Gini, C. (1912). Variabilità e mutabilità, *in* E. Pizetti and T. Salvemini (eds), *Reprinted in Memorie di metodologica statistica*, Rome: Libreria Eredi Virgilio Veschi.

Gini, C. (1914). Di una misura di dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche, *Atti del R. Istitituto Veneto di Scienze Lettere e Arti* **LXXIII**.

Gini, C. (1965). La dissomiglianza, *Metron* **24**: 85–215.

Giovagnoli, A., Marzialetti, J. and Wynn, H. (2009). Bivariate dependence orderings for unordered categorical variables, *in* L. Pronzato and A. Zhigljavsky (eds), *Optimal Design and Related Areas in Optimization and Statistics*, Vol. 28 of *Optimization and its Applications*, Springer, New York., chapter 4.

Grant, S., Kajii, A. and Polak, B. (1998). Intrinsic preference for information, *Journal of Economic Theory* **83**(2): 233 – 259.

Grenet, J. (2012). Is it enough to increase compulsory schooling to raise earnings? Evidence from French and British compulsory schooling laws, *Scandinavian Journal or Economics* **forthcoming**.

Hanushek, E. A. and Woessmann, L. (2011). *The Economics of International Differences in Educational Achievement*, Vol. 3 of *Handbook of the Economics of Education*, Elsevier, chapter 2, pp. 89–200.

Hardy, G. H., Littlewood, J. E. and Polya, G. (1934). *Inequalities*, London: Cambridge University Press.

Hasani, A. and Radjabalipour, M. (2007). On linear preservers of (right) matrix majorization, *Linear Algebra and its Applications* **423**(2-3): 255 – 261.

Havnes, T. and Mogstad, M. (2010). Is universal child care leveling the playing field? Evidence from non-linear difference-in-differences, *Technical Report 4978*, IZA.

Hutchens, R. M. (1991). Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations, *Mathematical Social Sciences* **21**(1): 31 – 51.

Hutchens, R. M. (2001). Numerical measures of segregation: Desirable properties and their implications, *Mathematical Social Sciences* **42**(1): 13 – 29.

Jenkins, S. J. (1994). Earnings discrimination measurement: A distributional approach, *Journal of Econometrics* **61**: 81–102.

Kodde, D. A. and Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions, *Econometrica* **54**(5): pp. 1243–1248.

Koenker, R. and Bassett, Gilbert, J. (1978). Regression quantiles, *Econometrica* **46**(1): pp. 33–50.

Kolm, S.-C. (1969). *The Optimal Production of Social Justice*, Public Economics, London: McMillan, pp. 145–200.

Kolm, S.-C. (1977). Multidimensional egalitarianisms, *The Quarterly Journal of Economics* **91**(1): 1–13.

Koshevoy, G. (1995). Multivariate Lorenz majorization, *Social Choice and Welfare* **12**: 93–102.

Koshevoy, G. (1997). The Lorenz zonotope and multivariate majorizations, *Social Choice and Welfare* **15**(1): 1–14.

Koshevoy, G. A. and Mosler, K. (1997). Multivariate Gini indices, *Journal of Multivariate Analysis* **60**(2): 252–276.

Koshevoy, G. and Mosler, K. (1996). The Lorenz zonoid of a multivariate distribution, *Journal of the American Statistical Association* **91**(434): 873–882.

Le Breton, M., Michelangeli, A. and Peluso, E. (2012). A stochastic dominance approach to the measurement of discrimination, *Journal of Economic Theory* **147**(4): 1342 – 1350.

Lefranc, A., Pistolesi, N. and Trannoy, A. (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies all alike?, *Review of Income and Wealth* **54**(4): 513–546.

Lefranc, A., Pistolesi, N. and Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France, *Journal of Public Economics* **93**(11-12): 1189 – 1207.

Maccheroni, F., Muliere, P. and Zoli, C. (2005). Inverse stochastic orders and generalized Gini functionals, *METRON* **63**(3): 529–559.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Springer.

Marshall, A. W., Olkin, I. and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, Springer.

Martínez Pería, F. D., Massey, P. G. and Silvestre, L. E. (2005). Weak matrix majorization, *Linear Algebra and its Applications* **403**: 343 – 368.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation, *Social Forces* **67**(2): 281–315.

Maurin, E. (2007). *La Nouvelle Question Scolaire*, Seuil, Paris.

Maurin, E. and McNally, S. (2008). Vive la révolution! Long-term educational returns of 1968 to the Angry Students, *Journal of Labor Economics* **26**(1): pp. 1–33.

McMullen, P. (1971). On zonotopes, *Transactions of the American Mathematical Society* **159**: 91–109.

Meghir, C. and Palme, M. (2005). Educational reform, ability, and family background, *American Economic Review* **95**(1): 414–424.

Mosler, K. (1994). Majorization in economic disparity measures, *Linear Algebra and its Applications* **199**(Supplement 1): 91 – 114. Special Issue Honoring Ingram Olkin.

Muliere, P. and Scarsini, M. (1989). A note on stochastic dominance and inequality measures, *Journal of Economic Theory* **49**(2): 314 – 323.

Ooghe, E., Schokkaert, E. and Van de gaer, D. (2007). Equality of opportunity versus equality of opportunity sets, *Social Choice and Welfare* **28**: 209–230.

Peragine, V., Palmisano, F. and Brunori, P. (2011). Economic growth and equality of opportunity, *Working Papers 232*, ECINEQ, Society for the Study of Economic Inequality.

Ramos, X. and Van de gaer, D. (2012). Empirical approaches to inequality of opportunity: Principles, measures, and evidence, *Technical report*, ECINEQ WP 2012 – 259.

Rao, C. R. (1982). Diversity and dissimilarity coefficients: A unified approach, *Theoretical Population Biology* **21**(1): 24 – 43.

Reardon, S. F. (2009). Measures of ordinal segregation, *in* J. S. Yves Flückiger, Sean F. Reardon (ed.), *Occupational and Residential Segregation (Research on Economic Inequality)*, Vol. 17, Emerald Group Publishing Limited, pp. 129–155.

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation, *Sociological Methodology* **32**: 33–67.

Reardon, S. F. and O'Sullivan, D. (2004). Measures of spatial segregation, *Sociological Methodology* **34**(1): 121–162.

Roemer, J. (1998). *Equality of Opportunity*, Harvard University Press, Cambridge.

Roemer, J. E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S. P., Lefranc, A., Marx, I., Page, M., Pommer, E., Ruiz-Castillo, J., Segundo, M. J. S., Tranaes, T., Trannoy, A., Wagner, G. G. and Zubiri, I. (2003). To what extent do fiscal regimes equalize opportunities for income acquisition among citizens?, *Journal of Public Economics* **87**(3–4): 539 – 565.

Shaked, M. and Shanthikumar, G. (2006). *Stochastic Orders*, Springer, New York.

Shephard, G. C. (1974). Combinatorial properties of associated zonotopes, *Canadian Journal of Mathematics* **26**: 302–321.

Shorrocks, A. F. (1982). On the distance between income distributions, *Econometrica* **50**(5): pp. 1337–1339.

Shorrocks, A. F. (1983). Ranking income distributions, *Economica* **50**(197): pp. 3–17.

Silber, J. (2012). Measuring segregation: Basic concepts and extensions to other domains, *in* J. A. Bishop and R. Salas (eds), *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber*, Vol. 20 of *Research on Economic Inequality*, Emerald Group Publishing Limited, chapter 1, pp. 1–35.

Silber, J. G. (1989). On the measurement of employment segregation, *Economics Letters* **30**(3): 237 – 243.

Torgersen, E. (1992). *Comparison of Statistical Experiments*, Vol. 36 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press.

Van de gaer, D. (1993). *Equality of Opportunity and Investment in Human Capital*, PhD thesis, Katholieke Universiteit Leuven.

Van de gaer, D., Vandenbossche, J. and Figueroa, J. L. (2011). Children's health opportunities and project evaluation: Mexico's oportunidades program, *Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 11/749*.

Weymark, J. A. (2003). Generalized gini indices of equality of opportunity, *Journal of Economic Inequality* **1**: 5–24. 10.1023/A:1023923807503.

White, M. J. (1983). The measurement of spatial segregation, *The American Journal of Sociology* **88**(5): pp. 1008–1018.

Yaari, M. E. (1987). The dual theory of choice under risk, *Econometrica* **55**(1): pp. 95–115.

Yaari, M. E. (1988). A controversial proposal concerning inequality measurement, *Journal of Economic Theory* **44**(2): 381 – 397.

Ziegler, G. (1995). *Lectures on Polytopes*, number 152 in *Graduate Texts in Mathematix*, Springer-Verlag, New York.

Zoli, C. (1999). Intersecting generalized Lorenz curves and the Gini index, *Social Choice and Welfare* **16**: 183–196.

Zoli, C. (2002). Inverse stochastic dominance, inequality measurement and Gini indices, *Journal of Economics* **9**: 119–161.

Appendix A

Hypothesis, test statistics and limiting distributions for the inverse stochastic dominance

Contents

A.1 Objective

This section develops the distribution theory which is used to test for inverse stochastic dominance and gap dominance, as prescribed by the ezOP algorithm. We firstly derive the asymptotic distributions, as well as the asymptotic variance covariance matrices, for the quantiles of the empirical cumulative distribution functions, empirical generalized Lorenz curves and their integrals up to order k. We exploit these results to derive asymptotic distributions and standard errors of Wald-type test statistics, which allow to test stochastic dominance for a selected vector of population shares, in the form of equality and inequality constraints.

In this appendix, we also propose an innovative result that is dual to the approach by Davidson and Duclos (2000), who test direct stochastic dominance by comparing the values of the poverty index by Foster, Greer and Thorbecke (1984) at different poverty lines. We propose instead to test inverse stochastic dominance at order k by comparing the values of the single parameter Gini indix (Donaldson and Weymark 1983) calculated at different quantiles of the distributions. We build on the results in Barrett and Donald (2009) to obtain the asymptotic normality of the indicator's distribution. We also survey bootstrapping methods to construct the inference for inverse stochastic dominance.

A.2 Setting

Let Y define a random variable with cumulative distribution function F and inverse F^{-1}. Let $\{Y_i\}_{i=1}^n$ be a sequence of independent random variables with common distribution F. The distribution F can, for instance, represent the distribution of income conditional to a given circumstance under a given policy regime. We use instead $Y_{(i)}$ to indicate that the random variables have been ranked according to the index i, from the lowest $Y_{(1)}$ to the highest $Y_{(n)}$ rank. We assume that random samples are available, so that each conditional distribution function can be estimated from the data. A sample $\mathcal{Y}$ is a collection of n realizations from i.i.d. variables. Let denote these realizations by small letters y such that $y_1 \leq y_2 \leq \ldots \leq y_n$. We use y_i to index the observation in position i in the ranking. For simplicity, we assume that all the observed values in our sample are distinct (thus n values are observed). We use i also to indicate observations. Moreover, let $[x]$ denote the integer part of the real number x.

The empirical distribution for the sample is $\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \leq y)$ while the empirical quantile function is $\widehat{F}^{-1}(p) = \inf\{y : \widehat{F}(y) \geq p\}$.

Let us simplify the notation by using ${}_{\pi}F_1$ and ${}_{\pi}F_2$ to represent the distribution functions corresponding to $F(.|c = 1, \pi)$ and $F(.|c = 2, \pi)$ respectively, where $c = 1$ and $c = 2$ are taken to be two different circumstances in the set C. The empirical counterparts can be estimated through ${}_{\pi}\widehat{F}_1$ and ${}_{\pi}\widehat{F}_2$ respectively, where in general ${}_{\pi}n_1 \neq {}_{\pi}n_2$. Using a similar notation, we define ${}_{\pi}\Lambda_1^k(p)$ and ${}_{\pi}\Lambda_2^k(p)$ as the $(k-1)$-th integral of ${}_{\pi}F_1^{-1}$ and ${}_{\pi}F_2^{-1}$ respectively, evaluated at percentile $p \in [0, 1]$. The empirical counterparts of the two processes are ${}_{\pi}\widehat{\Lambda}_1^k(p)$ and ${}_{\pi}\widehat{\Lambda}_2^k(p)$, and they can be estimated from samples with different sizes.

We omit the subscript π when it is clear that comparisons are made under the same policy regime π. We also omit the subscript indicating the circumstance when dealing with the asymptotic properties of a single distribution function. Through this section we focus on comparisons of either two outcomes distributions conditional on two different circumstances under the same policy regime or, alternatively, we extend the same comparisons to two different policy regimes. The style of the presentation of our arguments is closely related to Dardanoni and Forcina (1999).

A.3 Convergence results for testing ISDk

A.3.1 Tools and methods

We exploit non-parametric methods to estimate and testing assumptions on F, whose parametric form is in general not known by the econometrician. Following Muliere and Scarsini (1989), one gets the following equivalent representation of the $(k-1)$-th integral of $\widehat{F}^{-1}$:

$$\widehat{\Lambda}^k(p) = \frac{1}{(k-2)!}\int_0^p (p-t)^{k-2}\widehat{F}^{-1}(t)dt, \quad p \in [0,1],\ k = 2, 3, \ldots \tag{A.1}$$

Since $\widehat{F}$ is a consistent estimator of F, $\widehat{\Lambda}^k(p)$ is a consistent estimator of $\Lambda^k(p)$. The literature on inequality offers alternative tools for measuring the information embedded in a distribution of a continuous variable. We make extensive use of a parametric form of the *Generalized Gini Social Welfare Function* $W^k(F)$ (Donaldson and Weymark 1983),[1] defined

[1] One can derive from this SWF the whole family of absolute or relative single parameter *S-Gini* inequality measures.

as:

$$W^k(F) \; = \; k\int_0^1 (1-p)^{k-1}F^{-1}(p)dp.$$

It is also possible to specify $W^k(F)$ at a given population share. This is the *conditional* SWF at population percentile q, denoted $W^k(q,F)$ (see for instance Zoli 2002):

$$W^k(q,F) \; := \; k\int_0^1 (q-p)^{k-1}F^{-1}(p)\mathbf{1}(p \le q)dp, \tag{A.2}$$

where $W^k(1,F) = W^k(F)$. The empirical counterpart calculated at population percentile q, that is $\widehat{W}^k(q,F)$, can be obtained by replacing $\widehat{F}^{-1}$ in (A.2). Thus $\widehat{W}^k(q,F)$ is a consistent estimator of $W^k(q,F)$. We study two approaches to estimate inverse stochastic dominance.

In the first approach, we propose a *direct estimator* of the inverse stochastic dominance for a finite number m of population percentiles levels $0 < p_1 \le \ldots \le p_m \le 1$. This approach is direct in the sense that, as in Beach and Davidson (1983), we aim at exploiting the distributional features of the quantiles of F and F's integrals. Hence, the method is based upon the direct comparison of the quantiles associated to the population shares $p_1,\ldots,p_m$.

The direct approach involves the calculation of a finite sequence of values $\widehat{\Lambda}^k(p_1),\ldots,\widehat{\Lambda}^k(p_m)$ corresponding to the relevant empirical process associated to $\Lambda^k(p)$ for all $p \in \{p_1,\ldots,p_m\}$. It can be represented in compact vector notation as:

$$\widehat{\mathbf{\Lambda}}^k = \left(\widehat{\Lambda}^k(p_1),\ldots,\widehat{\Lambda}^k(p_m)\right)^t \in \mathbb{R}^m,$$

where $\widehat{\mathbf{\Lambda}}^k$ is a $m \times 1$ vector of the integral of the quantile function ordinates for the sample under analysis, with $\mathbf{\Lambda}^k$ being the corresponding vector in the population.

In the second approach, we propose a threshold estimator for ISDk. It is based on a comparison of conditional Gini SWF. This dominance is tested by comparing the values of the conditional Gini SWF $W^k(q,F)$ associated to different distributions at a finite number of population percentiles $q \in \{q_1,\ldots,q_m\}$. The values of the conditional Gini SWF can be written in compact vector notation with a $m \times 1$ vector $\mathbf{W}^k$ whose empirical counterpart is:

$$\widehat{\mathbf{W}}^k = \left(\widehat{W}^k(q_1,\widehat{F}),\ldots,\widehat{W}^k(q_m,\widehat{F})\right)^t \in \mathbb{R}^m.$$

The objective of the rest of the section is to determine the asymptotic distribution of the random variables $\sqrt{n}\left(\widehat{\mathbf{\Lambda}}^k - \mathbf{\Lambda}^k\right)$ and $\sqrt{n}\left(\widehat{\mathbf{W}}^k - \mathbf{W}^k\right)$ for $k = 3, 4, \ldots$. For $k = 1, 2$, the asymptotic distribution of the first estimator has been already established in Lemma 1 and Theorem 1 by Beach and Davidson (1983). We build on their contribution to obtain tractable estimators of the asymptotic covariances of $\widehat{\mathbf{\Lambda}}^k$ and $\widehat{\mathbf{W}}^k$, defined over the m population percentiles.

A.3.2 Direct ISDk testing

We make use of two different approaches to test ISDk. For two empirical distributions $\widehat{F}$ and $\widehat{F'}$, the *direct approach* is based on the direct comparison of the estimators $\widehat{\mathbf{\Lambda}}^k$ and $\widehat{\mathbf{\Lambda}'}^k$ at a finite number of population percentiles. The ISDk can be statistically tested by resorting on a joint test of hypothesis. To derive the test distribution, one needs the asymptotic distribution, the asymptotic covariance matrix and its estimator associated to $\widehat{\mathbf{\Lambda}}^k$ and to $\widehat{\mathbf{\Lambda}'}^k$.

For a given distribution F, the asymptotic distribution and the respective covariance matrix of the quantile functions and the GL curve (that are $\widehat{\mathbf{\Lambda}}^k$ for $k = 1, 2$ respectively) have been derived by Beach and Davidson (1983). We summarize their results in the following Lemma:

Lemma A.1 *Suppose that for a set of proportions $\{p_j | j = 1, \ldots, m\}$ such that $0 < p_1 < \ldots < p_m < 1$, $\widehat{\mathbf{\Lambda}} = \left(\widehat{F}^{-1}(p_1), \ldots, \widehat{F}^{-1}(p_m)\right)^t$ is a vector of m sample quantiles and $\widehat{\mathbf{\Lambda}}^2 = \left(\widehat{GL}(p_1), \ldots, \widehat{GL}(p_m)\right)^t$ is a vector of m ordinates of the GL curve estimator (where $\widehat{GL}(1) = \widehat{\mu}$, the sample mean) obtained from a sample of size n drawn from a continuous population density $f(y)$ with cdf $F(y)$ which is strictly monotonic with quantile function $F^{-1}(p)$. Then:*

i) *the vector $\sqrt{n}\left(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\right)$ converges in distribution to a m variate normal distribution with mean zero and asymptotic covariance matrix $\mathbf{\Sigma}^1$, where the element j, j' corresponding to population proportions p_j and $p_{j'}$ is:*

$$\sigma^1(j, j') = \frac{p_j(1 - p_{j'})}{f(F^{-1}(p_j))f(F^{-1}(p_{j'}))}.$$

ii) *the vector $\sqrt{n}\left(\widehat{\mathbf{\Lambda}}^2 - \mathbf{\Lambda}^2\right)$ converges in distribution to a m variate normal distribution* *with mean zero and asymptotic covariance matrix* $\mathbf{\Sigma}^2$*, where the element* j, j' *corresponding to population proportions* p_j *and* $p_{j'}$ *is:*

$$\begin{aligned}\sigma^2(j,j') &= p_j v_{p_j}^2 + p_j(1-p_{j'})(F^{-1}(p_j)-\mu_{p_j})(F^{-1}(p_{j'})-\mu_{p_{j'}}) + \\ &\quad + p_j(F^{-1}(p_j)-\mu_{p_j})(\mu_{p_{j'}}-\mu_{p_j}) \qquad for\ p_j \le p_{j'},\end{aligned}$$

where $v_{p_j}^2$ *and* μ_{p_j} *are respectively the variance and expected value of a random variable* Y *distributed as* F *conditional on* $Y \le F^{-1}(p_j)$*.*

The sample counterparts of $\sigma^1(j,j')$ and $\sigma^2(j,j')$ can be obtained by replacing the population moments (quantiles, population shares and the conditional means and variances) with the respective sample estimators, while one can use a kernel estimator to obtain a consistent estimator of f.

We contribute by deriving a similar result for higher orders of integration, that is for $\widehat{\mathbf{\Lambda}}^k$ with $k \ge 3$. However, the resulting covariance matrix is hardly tractable and it is in practice difficult to obtain its empirical counterpart.[2]

Proposition A.1 *Suppose that for a set of proportions* $\{p_j | j = 1, \ldots, m\}$ *such that* $0 < p_1 < \ldots < p_m < 1$, $\widehat{\mathbf{\Lambda}}^k = \left(\widehat{\Lambda}^k(p_1), \ldots, \widehat{\Lambda}^k(p_m)\right)^t$ *for* $k = 3, 4, \ldots$ *is a vector of* m *ordinates of the estimator of* $\Lambda^k(.)$*, the* $(k-1)$*-th integral of the quantile function, obtained from a sample of size* n *drawn from a continuous population density* $f(y)$ *with cdf* $F(y)$ *which is strictly monotonic with quantile function* $F^{-1}(p)$*. Then the vector* $\sqrt{n}\left(\widehat{\mathbf{\Lambda}}^k - \mathbf{\Lambda}^k\right)$ *converges in distribution to a* m *variate normal distribution with mean zero and asymptotic covariance matrix* $\mathbf{\Sigma}^k$ *for* $k = 3, 4, \ldots$*. The element* j, j' *of* $\mathbf{\Sigma}^k$ *corresponding to population proportions* p_j *and* $p_{j'}$ *is:*

$$\begin{aligned}\sigma^k(j,j') &= \frac{1}{(k-2)!}\left[p_j\Lambda^{k-1}(p_j) + (k-1)\Lambda^k(p_j)\right]\left[(1-p_{j'})\Lambda^{k-1}(p_{j'}) + (k-1)\Lambda^k(p_{j'})\right] + \\ &\quad + \int_0^{F^{-1}(p_j)} (p_j - F(x))^{k-2}F(x)dx \int_x^{F^{-1}(p_j)} (p_{j'} - F(y))^{k-2}dy \ - \\ &\quad - \int_0^{F^{-1}(p_j)} (p_j - F(x))^{k-2}dx \int_x^{F^{-1}(p_j)} (p_{j'} - F(y))^{k-2}F(y)dy.\end{aligned}$$

[2] In a recent paper, Aaberge, Havnes and Mogstad (2011) derive a similar expression for the covariance matrix by treating F as a continuous process and they show that it converges in distribution to a Gaussian process.

Proof. We use the fact that $\widehat{\Lambda}^k(p) = \frac{1}{n}\Lambda^k(p) + O(n^{-1})$, which comes from the fact that the empirical quantile function estimates only a lower bound for the real population quantile at population share p. Using the fact that $E[F^{-1}(p)] = F^{-1}(p) + o(n^{-1/2})$ from Lemma A.1 (result i)), by linearity of the operator in (A.1) one obtains that:

$$E\left[\widehat{\Lambda}^k(p)\right] \;=\; \Lambda^k(p) + o(n^{-1/2}).$$

The central limit theorem applies and therefore the vector $\sqrt{n}\left(\widehat{\boldsymbol{\Lambda}}^k - \boldsymbol{\Lambda}^k\right)$ is a multivariate normal with zero means and finite covariance. We consider:

$$\text{cov}\left[n^{-1/2}\widehat{\Lambda}^k(p_j),\ n^{-1/2}\widehat{\Lambda}^k(p_{j'})\right] \quad \text{where } p_j \leq p_{j'}. \tag{A.3}$$

We use the empirical estimator of $\widehat{\Lambda}^k(p)$ given by:

$$\widehat{\Lambda}^k(p_j) = \frac{1}{(k-2)!}\sum_{i=1}^{[p_j n]}\left(\frac{[p_j n]-i}{n}\right)^{k-2} Y_{(i)} + o(n^{-1}),$$

so that the covariance in (A.3) can be rewritten, for $p_j \leq p_{j'}$ as:

$$\frac{1}{n}\frac{1}{(k-2)!}\sum_{i=1}^{[p_j n]}\sum_{h=1}^{[p_{j'} n]}\left(\frac{[p_j n]-i}{n}\right)^{k-2}\left(\frac{[p_{j'} n]-h}{n}\right)^{k-2}\text{cov}(Y_{(i)},Y_{(h)}) + o(n^{-1}). \tag{A.4}$$

Making use of the consistent estimator of the covariance $\text{cov}\left[Y_{(i)},Y_{(h)}\right]$ in Lemma A.1, it is possible to write (A.4) as:

$$\frac{1}{n}\frac{1}{[(k-2)!]^2}\sum_{i=1}^{[p_j n]}\sum_{h=1}^{i}\left(\frac{[p_j n]-i}{n}\right)^{k-2}\left(\frac{[p_{j'} n]-h}{n}\right)^{k-2}\frac{\left(\frac{h}{n}\right)\left(1-\frac{i}{n}\right)}{n\widehat{F}'(\widehat{F}^{-1}(h/n))\widehat{F}'(\widehat{F}^{-1}(i/n))} +$$
$$+\frac{1}{n}\frac{1}{[(k-2)!]^2}\sum_{i=1}^{[p_j n]}\sum_{h=i+1}^{[p_{j'} n]}\left(\frac{[p_j n]-i}{n}\right)^{k-2}\left(\frac{[p_{j'} n]-h}{n}\right)^{k-2}\frac{\left(1-\frac{i}{n}\right)\left(\frac{h}{n}\right)}{n\widehat{F}'(\widehat{F}^{-1}(h/n))\widehat{F}'(\widehat{F}^{-1}(i/n))} + o(n^{-1}) \tag{A.5}$$

The estimator $n\,\text{cov}\left[n^{-1/2}\widehat{\Lambda}^k(p_j),\ n^{-1/2}\widehat{\Lambda}^k(p_{j'})\right]$ can be estimated with a asymptotic precision equal to $o(1)$. As a consequence, summations can be replaced by integrals of population to obtain the following formulation of (A.5):

$$\int_0^{p_j} (p_j - p)^{k-2}(1-p)\frac{dF^{-1}(p)}{dp}\int_0^{p} (p_{j'} - q)^{k-2} q\frac{dF^{-1}(q)}{dq}dqdp +$$
$$+\int_0^{p_j} (p_j - p)^{k-2} p\frac{dF^{-1}(p)}{dp}\int_p^{p_{j'}} (p_{j'} - q)^{k-2}(1-q)\frac{dF^{-1}(q)}{dq}dqdp.$$

After a change in variables, this integral can be written as:

$$\int_0^{F^{-1}(p_j)} (p_j - F(x))^{k-2}(1-F(x))dx\int_0^{x} (p_{j'} - F(y))^{k-2}F(y)dy +$$
$$+\int_0^{F^{-1}(p_j)} (p_j - F(x))^{k-2}F(x)dx\int_x^{F^{-1}(p_{j'})} (p_{j'} - F(y))^{k-2}(1-F(y))dy.$$

Integrating by parts, and appropriate substitutions of the integration terms give the desired result. ■

Unfortunately, we are able to derive an empirically tractable estimator for $\sigma^k(j,j')$ (i.e. that depends only on a sum or product of population moments that can be consistently estimated by their empirical counterparts as the sample size grows) only for $k = 3$ and not for higher orders of integration.[3]

A direct consequence of the previous Lemma A.1 and Proposition A.1 is that:

$$\widehat{\mathbf{\Lambda}}^k \text{ is asymptotically distributed as } \mathcal{N}\left(\mathbf{\Lambda}^k,\ \frac{\mathbf{\Sigma}^k}{n}\right). \tag{A.6}$$

The implementation problems related to orders higher than $k = 3$ forced us to define an alternative methodology to test ISDk which turns out to be implementable in the data. We describe this method in the following section.

A.3.3 Threshold ISDk testing

We propose an innovative method for testing ISDk which exploits a result in Muliere and Scarsini (1989) and Maccheroni et al. (2005), Remark 3.5, reported below:

[3] The estimator is available upon demand from the author.

Remark A.1 Given two distributions F and F'

$$W^{k-1}(q,F) \geq W^{k-1}(q,F') \text{ for all } q \in [0,1] \;\Leftrightarrow\; F \succcurlyeq_{ISDk} F'.$$

Suppose that the conditional Gini SWF at order k can be empirically measured for a finite number m of quantiles, thus giving vectors $\widehat{\mathbf{W}}^k$ and $\widehat{\mathbf{W}'}^k$ for F and F' respectively. Hence, one can use the Remark A.1 to derive the inference behind the test for ISDk. To derive the test statistic distribution, one need the asymptotic distribution and asymptotic covariance matrix associated to the estimators of the conditional Gini SWF. This innovative method is dual to the procedure based on poverty lines studied by Davidson and Duclos (2000) (Theorem 1) to test direct stochastic dominance at any order.

To derive the asymptotic distribution and the covariance matrix of $\widehat{\mathbf{W}}^k$ we rely on the results in Barrett and Donald (2009), although the procedure for testing ISDk was not considered by the authors. They provide the asymptotic standard errors of the Gini SWF and provide its sample estimator making use of *influence functions.* As in Beach and Davidson (1983), the approach is fully non parametric since it imposes no structure on the underlying distribution beyond weak regularity conditions.

Integrating by parts one can show that the conditional Gini SWF $\widehat{W}^k(p,F)$ can be written as:

$$\widehat{W}^k(p,\widehat{F}) \;=\; k(k-1)\int_0^1 (q-p)^{k-2}\widehat{GL}(p)\mathbf{1}(p \leq q)dp \;=\; T(\widehat{H}), \tag{A.7}$$

where $T(\widehat{H})$ is a scalar valued functional of some process $\widehat{H}$ that is defined on $[0,1]$. In (A.7), $\widehat{H}$ is the GL curve estimator. Using Hadamard differentiability of the functional T (see Barrett and Donald 2009, for references), one obtains that:

$$\sqrt{n}\left(T(\widehat{H}) - T(H)\right) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} T'_H\left(\phi_i(.,\widehat{H})\right) + o(1),$$

where the random variables $\phi_i(.,\widehat{H})$ are often referred to as the *influence functions*, and give the effect of an observation i on the estimator. In many estimation problems, the variables $\phi(.,\widehat{H})$ are iid. The value of the function at p is $\phi(p,\widehat{H})$.

In our case, we use $\widehat{H} = \widehat{GL}$, so that the estimator $\sqrt{n}\left(\widehat{H} - H\right)$ can be decomposed into influence functions $\phi_i(p,\widehat{GL})$. These functions are used to construct the conditional

Gini SWF influence function in the following way:

$$\begin{aligned}
\sqrt{n}\left(\widehat{W}^k(q,\widehat{F}) - W^k(q,F)\right) &= k(k-1)\int_0^1 (q-p)^{k-2}\sqrt{n}\left(\widehat{GL}(p) - GL(p)\right)\mathbf{1}(p \leq q)dp \\
&= k(k-1)\int_0^1 (q-p)^{k-2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_i(p,\widehat{GL})\right)\mathbf{1}(p \leq q)dp + o(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n k(k-1)\int_0^1 (q-p)^{k-2}\phi_i(p,\widehat{GL})\mathbf{1}(p \leq q)dp + o(1) \quad (A.8) \\
&:= \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_i(\widehat{W}^k(q,\widehat{F})) + o(1). \quad (A.9)
\end{aligned}$$

We can extend this result to a $m \times 1$ *vector* of conditional Gini SWF computed at different shares q. This vector can be represented using (A.9) ad:

$$\sqrt{n}\left(\widehat{\boldsymbol{W}}^k - \boldsymbol{W}^k\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\phi_i(\widehat{W}^k(q_1,\widehat{F})),\ldots,\phi_i(\widehat{W}^k(q_m,\widehat{F}))\right)^t + o(1)(1,\ldots,1)^t.$$

The following proposition gives the asymptotic distribution of $\widehat{W}^k(p,\widehat{F})$.

Proposition A.2 *Suppose that for a set of proportions* $\{q_j | j = 1,\ldots,m\}$ *such that* $0 < q_1 < \ldots < q_m < 1$, $\widehat{\mathbf{W}}^k = \left(\widehat{W}^k(q_1,\widehat{F}),\ldots,\widehat{W}^k(q_m,\widehat{F})\right)^t$ *for* $k = 2,3,4,\ldots$ *is a vector of* m *ordinates of the estimator of* $W^k(q,F)$*, which is the conditional Gini SWF conditional on population share* q *parametrized at* k*, obtained from a sample of size* n *drawn from a continuous population density* $f(y)$ *with cdf* $F(y)$ *which is strictly monotonic with quantile function* $F^{-1}(p)$*. Then, the vector* $\sqrt{n}\left(\widehat{\mathbf{W}}^k - \mathbf{W}^k\right)$ *converges in distribution to a* m *variate normal distribution with mean zero and asymptotic covariance matrix* $\mathbf{\Sigma}^k$ *for* $k = 2,3,\ldots,$ *The element* j,j' *of* $\mathbf{\Sigma}^k$ *corresponding to population proportions* q_j *and* $q_{j'}$ *is:*

$$\varsigma^k(q_j,q_{j'}) = E\left[\phi_i(\widehat{W}^k(q_j,\widehat{F}))\,\phi_i(\widehat{W}^k(q_{j'},\widehat{F}))\right]. \quad (A.10)$$

Proof. By (A.8), for a given q the random variable $\sqrt{n}\left(\widehat{W}^k(p,\widehat{F}) - W^k(p,F)\right)$ satisfies the central limit theorem, and it is therefore asymptotically normal with mean zero and finite variance. Using (A.9), we have that $\sqrt{n}\left(\widehat{\mathbf{W}}^k - \mathbf{W}^k\right) = \sum_{i=1}^n \boldsymbol{\phi}_i + o(1)\mathbf{e}_m$ where $\boldsymbol{\phi}_i = \left(\phi_i(\widehat{W}^k(q_1,\widehat{F})),\ldots,\phi_i(\widehat{W}^k(q_m,\widehat{F}))\right)^t$ and $\mathbf{e}_m$ is a $m \times 1$ vectors of ones. This collection of variables is multivariate normal with zero mean and finite covariance equal to

$$\mathrm{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\phi}_i\cdot\frac{1}{\sqrt{n}}\sum_{i'=1}^{n}\boldsymbol{\phi}_{i'}^{t}\right], \tag{A.11}$$

which is equal to:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathrm{E}\left[\boldsymbol{\phi}_i\cdot\boldsymbol{\phi}_{i'}^{t}\right].$$

In fact, since the random vectors $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_{i'}$ are independent by construction of the influence function, one gets that $\mathrm{E}\left[\boldsymbol{\phi}_i\cdot\boldsymbol{\phi}_{i'}^{t}\right]=0$ for $i\neq i'$ and that $\mathrm{E}\left[\boldsymbol{\phi}_i\cdot\boldsymbol{\phi}_{i}^{t}\right]=\mathrm{E}\left[\boldsymbol{\phi}_{i'}\cdot\boldsymbol{\phi}_{i'}^{t}\right]$ for all i,i'. This implies that the covariance in (A.11) is equal to $\mathrm{E}\left[\boldsymbol{\phi}_i\cdot\boldsymbol{\phi}_{i}^{t}\right]$, which defines the covariance matrix parameters. The term $\varsigma^k(p_j,p_{j'})$ in (A.10) corresponds to one of the cells of this matrix. ■

A direct consequence of the previous proposition is that:

$$\widehat{\mathbf{W}}^k \text{ is asymptotically distributed as } \mathcal{N}\left(\mathbf{W}^k,\ \frac{\boldsymbol{\Sigma}^k}{n}\right). \tag{A.12}$$

The implementation problems and empirical solutions are discussed in the following section.

A.3.4 Bootstrapping techniques

The direct approach does not allow to conclude and construct the sample counterparts of the covariance in Proposition A.1 that can be also shown to be a consistent estimators of the population covariance matrix.

This problem is solved by constructing the test for ISDk by exploiting conditional Gini SWF relations calculated at different population shares. However, the calculation of the sample counterpart of the asymptotic covariance matrix in Proposition A.2 is computationally demanding: for each observation one has to calculate the empirical influence function. This has to be done for each of the variable under analysis and, on the top of that, for each of the m population shares considered.

An alternative procedure, often exploited in the literature of inequality measurement, consists in bootstrapping the conditional Gini SWF for each of the $\{q_1,\ldots,q_m\}$ population shares, separately for a sufficiently large number of sub-samples. Then, one has to construct the covariance matrix of the conditional Gini SWF calculated for each bootstrapped sample, and then use these estimates to test ISDk.

Let $\mathcal{Y}$ be the original sample of size n drawn from the distribution F. Bootstrap computations are constructed conditional on $\mathcal{Y}$. Let a random sample of size n^* drawn with replacement from $\mathcal{Y}$ be denoted as $\mathcal{Y}^b = \{Y_i^b\}_{i=1}^{n^*}$, and with empirical distribution $\widehat{F}^b$. One can construct all the empirical measures discussed in the following section by using the data of sub-sample $\mathcal{Y}^b$. With this random sample we can calculate the conditional Gini SWF at order k for each of the the m conditioning population shares, denoted by $\widehat{W}^k(q, \widehat{F}^b)$. By repeatedly drawing random samples from $\mathcal{Y}$, say B times, and calculating for each of the sub-samples the vector of Gini SWF conditional on the m population shares, one obtains a $B \times m$ matrix of data. The $m \times m$ covariance matrix derived from these data is the bootstrap estimator of $\mathbf{\Sigma}^k/n$, the asymptotic covariance matrix of $\widehat{\mathbf{W}}^k$ in (A.12). Hence:

$$\frac{\widehat{\varsigma}_B^k(q_j, q_{j'})}{n} = \frac{1}{B-1}\left(\sum_{b=1}^{B} \widehat{W}^k(q_j, \widehat{F}_b)\widehat{W}^k(q_{j'}, \widehat{F}_b) - B\overline{\widehat{W}}^k(q_j, \widehat{F}_b)\,\overline{\widehat{W}}^k(q_{j'}, \widehat{F}_b)\right) \quad (A.13)$$

$$\overline{\widehat{W}}^k(q, \widehat{F}_b) = \frac{1}{B}\sum_{b=1}^{B} \widehat{W}^k(q, \widehat{F}_b).$$

This application of the bootstrap only requires the calculation of a vector of m conditional Gini SWF estimators within each re-sampling stage, although in general it does not offer a refinement of the asymptotic approximation.

A.4 Sample implementation

Consider a sample of size n and the associated (observed or estimated) outcomes $y_1, \ldots, y_n$ ordered such as $y_1 \leq \ldots \leq y_i \leq \ldots \leq y_n$. The estimator of the empirical *cdf* $\widehat{F}$ at any point y is:

$$\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(y_i \leq y), \quad (A.14)$$

where $\mathbf{1}(.)$ is the indicator function returning one if the argument is true. The *inverse cdf* estimator at population share p, $\widehat{F}^{-1}(p)$ is:

$$\widehat{F}^{-1}(p) = y_i \quad \text{where} \quad i-1 < pn \leq i. \quad (A.15)$$

For all quantiles $(j-1)/n < p \leq j/n$, the estimator of (A.1) is given by:

$$\widehat{\Lambda}^k(p) \quad := \quad \widehat{\Lambda}^k\left(\frac{j}{n}\right) \;=\; \frac{1}{(k-2)!}\frac{1}{n}\sum_{i=1}^{j}\left(\frac{j-i}{n}\right)^{k-2} y_i, \quad \forall k > 2. \qquad \text{(A.16)}$$

The conditional Gini SWF can be calculated using the following approximation, resulting from our assumption that each realization is associated with only one observation. Let $\widehat{p}_i = \widehat{F}(y_i)$ be the population share associated to observed value y_i, which amounts to $\widehat{p}_i = i/n$ in our case. We use the fact that $\int_0^1 = \sum_{j=1}^{n}\int_{p_{i-1}}^{p_i}$ with $p_0 = 0$ to derive the sample estimator of the conditional Gini SWF as follows:

$$\begin{aligned}
\widehat{W}^k(q,\widehat{F}) &= k\int_0^1 (q-p)^{k-1}\widehat{F}^{-1}(p)\mathbf{1}(p \leq q) \\
&= \sum_{i=1}^{n} y_i \int_{\widehat{p}_{i-1}}^{\widehat{p}_i} k(q-p)^{k-1}\mathbf{1}(p \leq q)dp \\
&= \sum_{i=1}^{n} y_j \left\{(q-\widehat{p}_{i-1})^k - (q-\widehat{p}_i)^k\right\}\mathbf{1}(i \leq qn) + y_{[qn]}(q-\widehat{p}_{[qn]})\mathbf{1}(qn \neq [qn]) \\
&\approx \sum_{i=1}^{n} y_i \left\{\left(\frac{n-i+1}{n}\right)^k - \left(\frac{n-i}{n}\right)^k\right\}\mathbf{1}(i \leq [qn]), \qquad \text{(A.17)}
\end{aligned}$$

where the last line is a consequence of the fact that, unless $p_i = q$ there remain a residual proportional to $[qn] - i$ which vanishes as soon as the sample grows to infinity. Given the large size of the sample used in our applications, we take the empirical estimators (A.17) as the consistent estimator of $W^k(q, F)$.

The consistent estimator of the asymptotic covariance matrix of the conditional Gini SWF, $\mathbf{\Sigma}^k/n = \frac{1}{n}E\left[\boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_i^t\right]$ can be obtained by replacing the population moment with the sample moments, such that the empirical counterpart becomes:

$$\frac{\widehat{\mathbf{\Sigma}}^k}{n} = \frac{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_i^t}{n}.$$

For a given pair of population shares q_j, $q_{j'}$ the estimator of the asymptotic covariance is:

$$\frac{\widehat{\varsigma}^k(q_j, q_{j'})}{n} = \frac{\frac{1}{n}\sum_{i=1}^{n}\phi_i\left(\widehat{W}^k(q_j,\widehat{F})\right)\phi_i\left(\widehat{W}^k(q_{j'},\widehat{F})\right)}{n}.$$

The influence function $\widehat{\phi}_i\left(\widehat{W}^k(q_j,\widehat{F})\right)$ is the estimator of $\phi_i\left(\widehat{W}^k(q_j,\widehat{F})\right)$ given in (A.8). For each conditioning population share q, this estimator has to be calculated for each observation, once that the data have been sorted by increasing magnitude of the variable y. The values of $\widehat{\phi}_i\left(\widehat{W}^k(q_j,\widehat{F})\right)$ for those observations ranked higher than $[qn]$ are replaced by a zero. One obtains therefore a $m \times n$ matrix of new observations which are used to construct the estimator of the asymptotic covariance of $\widehat{\mathbf{W}}^k$, for a k given. We adapt the estimator of the influence function proposed by Barrett and Donald (2009) to our case to compute $\widehat{\phi}_i\left(\widehat{W}^k(q,\widehat{F})\right)$. To do so we make use of (A.8), where $\widehat{\phi}_i(q,\widehat{GL})$ can be estimated by using:

$$\widehat{\phi}_i(q,\widehat{GL}) \;=\; (q\widehat{F}^{-1}(q)-\widehat{GL}(q))-\mathbf{1}(y_i\leq \widehat{F}^{-1}(q))(\widehat{F}^{-1}(q)-y_i). \tag{A.18}$$

Substituting (A.18) into (A.8) we obtain a form of the influence function of $\widehat{W}^k(q,\widehat{F})$ that can be decomposed into three components:

$$\widehat{\phi}_i\left(\widehat{W}^k(q,\widehat{F})\right) \;=\; \sum_{h=1}^{3} I_h.$$

Depending on the value of q and on the sample size, we can derive either exact or approximate (by a number $qn-[qn]$) empirical estimators. Since the correction term vanishes when the sample size is large, we only provide the formulation of the approximate estimators. These terms are:

i) $I_1 := \sum_{j=1}^{n} y_j\left[-k\left(\frac{j}{n}\left(\frac{[qn]-j}{n}\right)^{k-1}-\frac{j-1}{n}\left(\frac{[qn]-j+1}{n}\right)^{k-1}\right)-d_j(k)\right]\mathbf{1}(j\leq[qn])$,

ii) $I_2 := \sum_{j=1}^{n}\left[k\left(\left(\frac{[qn]-j}{n}\right)^{k-1}\widehat{GL}(j/n)-\left(\frac{[qn]-j+1}{n}\right)^{k-1}\widehat{GL}((j-1)/n)\right)+d_j(k)\right]\mathbf{1}(j\leq[qn])$,

iii) $I_3 := \sum_{j=1}^{n} -k\,(y_i-y_j)\,\mathbf{1}(y_i\leq y_j)\,[d_j(k-1)]\,\mathbf{1}(j\leq[qn])$,

where $d_j(\alpha)=\left(\frac{[qn]-j}{n}\right)^{\alpha}-\left(\frac{[qn]-j+1}{n}\right)^{\alpha}$ with α a positive natural number.

A.5 Testing for inverse stochastic dominance

The objective of the section is to propose an empirical test and its limiting distribution for inverse stochastic dominance at order k, constructed for any pair of distributions $F(.|c=$

$1, \pi)$ and $F(.|c = 2, \pi)$ under the same policy regime. The test allows to implement step $\kappa(\mathbf{c}, \mathbf{c}', \mathbf{e}, \pi)$ in the implementation algorithm for ezOP.

The general test for inverse stochastic dominance would require a null and alternative hypothesis formulated as in Barrett and Donald (2003). We take therefore inverse stochastic dominance at order k ($F_1 \succ_{ISDk} F_2$) as the null hypothesis, and intersection of the k-th order integrals of the quantile function as the alternative. Hence:

$$\begin{array}{lll} H_0^k: & \Lambda_1^k(p) \geq \Lambda_2^k(p) & \text{for all } p \in [0,1]; \\ H_1^k: & \Lambda_1^k(p) < \Lambda_2'^k(p) & \text{for some } p \in [0,1] \end{array}.$$

One can easily test for equality by reversing the role of $c = 1$ and $c = 2$ and testing if dominance is accepted also in this case.

We use a similar definition of the null and alternative hypothesis to test dominance in the conditional Gini SWF. To test ISDk it is sufficient to express the null hypothesis and the alternative as follows:

$$\begin{array}{lll} H_0^k(W): & W^{k-1}(q, F_1) \geq W^{k-1}(q, F_2) & \text{for all } q \in [0,1]; \\ H_1^k(W): & W^{k-1}(q, F_1) < W^{k-1}(q, F_2) & \text{for some } q \in [0,1] \end{array}.$$

In practice, dominance can be tested only for a finite number of percentiles $\{p_1, \ldots, p_m\}$. We take a similar stance as in Dardanoni and Forcina (1999) and Lefranc et al. (2009), among others, by constructing direct tests of dominance for a finite number m of linear constraints. We write $\Lambda_c^k(p_j)$ for the ordinate of the k-th integral of the quintile function, corresponding to the j-th fraction of the population with circumstance c. We also write $\mathbf{\Lambda}_c^k$ for the $m \times 1$ vector of the integral quantile function ordinates for population with circumstances c, with $\widehat{\mathbf{\Lambda}}_c^k$ being the corresponding vector of sample estimates.

Alternatively, one can test ISDk by using the conditional Gini SWF for a finite number of conditioning population shares $\{q_1, \ldots, q_m\}$. We take a similar stance as in Davidson and Duclos (2000) by constructing conditional tests of dominance for a finite number m of linear constraints (thresholds) on the conditional Gini SWF. In this case, we write $\mathbf{W}_c^{k-1}$ for the $m \times 1$ vector of conditional Gini SWF coordinates corresponding to conditioning at population shares q_1 to q_m, associated to the population with circumstances c. The vector $\widehat{\mathbf{W}}_c^{k-1}$ is the corresponding sample estimator.

A.5.1 Setting

The results in (A.6) and (A.12) state the asymptotic normality of the estimators $\widehat{\mathbf{\Lambda}}_c^k$ and $\widehat{\mathbf{W}}_c^{k-1}$. We build on this fact to retrieve the asymptotic properties of the test statistics. To simplify notation, we use $\widehat{\mathbf{\Theta}}_c^k$ to identify the sample estimator used to test ISDk. According to the direct testing procedure, $\widehat{\mathbf{\Theta}}_c^k = \widehat{\mathbf{\Lambda}}_c^k$, while for the conditional testing we use $\widehat{\mathbf{\Theta}}_c^k = \widehat{\mathbf{W}}_c^{k-1}$.

Let $\mathbf{\Theta}^k$ be the $2m \times 1$ vector obtained by staking the vectors $\mathbf{\Theta}_1^k$ and $\mathbf{\Theta}_2^k$, corresponding to the populations with circumstances $c = 1$ and $c = 2$ respectively. The sample estimates are collected in $\widehat{\mathbf{\Theta}}^k$, and we use $n = n_1 + n_2$ to indicate the overall sample population, while $r_c = n_c/n$ depict the relative size of the sample whose circumstance is c.

The hypothesis of dominance can be reformulated as a sequence of m linear constraints placed on the vector $\mathbf{\Theta}^k$. Let $\mathbf{R} = (\mathbf{I}_m, -\mathbf{I}_m)$ be the $m \times 2m$ *differences* matrix, with $\mathbf{I}_m$ indicating the $m \times m$ identity matrix. Define the parametric vector $\boldsymbol{\delta}_k \in \mathbb{R}^m$ as:

$$\boldsymbol{\delta}_k = \mathbf{R}\mathbf{\Theta}^k.$$

We maintain the (non testable) assumption that F_1 and F_2 are generated by independent processes. The various hypothesis of dominance or equality can be written in terms of linear inequalities involving $\boldsymbol{\delta}_k$. By exploiting the result in (A.6) or (A.12) and the independence assumption, one obtains the following asymptotic result:

$$\sqrt{n}\,\widehat{\boldsymbol{\delta}}_k = \sqrt{n}\,\mathbf{R}\,\widehat{\mathbf{\Theta}}^k \text{ is asymptotically distributed as } \mathcal{N}\left(\sqrt{n}\,\mathbf{R}\,\mathbf{\Theta}^k,\ \mathbf{\Omega}\right), \tag{A.19}$$

where $\widehat{\boldsymbol{\delta}}_k$ denotes the sample estimator of $\boldsymbol{\delta}_k$, and

$$\mathbf{\Omega} = \mathbf{R}\,\mathrm{diag}\left(\frac{\mathbf{\Sigma}_1^k}{r_1}, \frac{\mathbf{\Sigma}_2^k}{r_2}\right)\mathbf{R}^t.$$

For the direct method for testing ISDk, the empirical implementation is possible by using an estimator for the asymptotic variance $\widehat{\mathbf{\Omega}}$, which is obtained by plugging $\widehat{\mathbf{\Sigma}}_c^k$, the estimator of the asymptotic covariance of $\widehat{\mathbf{\Lambda}}_c^k$, in the previous formula. In the conditional (threshold) method for testing, $\widehat{\mathbf{\Omega}}$ is obtained by plugging the estimator of the asymptotic covariance of $\widehat{\mathbf{W}}_c^{k-1}$ in (A.10). We exploit the convergence result to test the discretized versions of the dominance or equality hypothesis, which is defined byy the set of m constraints.

A.5.2 Testing equality

In the case of equality testing, the null and alternative hypothesis for direct and threshold ISDk testing can be stated as follows:

$$H_0^k : \boldsymbol{\delta}_k = \mathbf{0} \qquad H_1^k : \boldsymbol{\delta}_k \neq \mathbf{0}.$$

Under the null hypothesis, it is possible to resort to a Wald test static T_1^k:

$$T_1^k \ := \ n \, \widehat{\boldsymbol{\delta}}_k^t \, \widehat{\boldsymbol{\Omega}}^{-1} \, \widehat{\boldsymbol{\delta}}_k.$$

Given the convergence results in (A.19), the asymptotic distribution of the test T_1^k is χ_m^2. The p-value tabulation follows the usual rules.

A.5.3 Testing dominance

In the case of strong dominance testing (i.e. $F_1 \succ_{ISDk} F_2$), the null and alternative hypothesis for direct and threshold ISDk testing can be stated as follows:

$$H_0^k : \boldsymbol{\delta}_k \in \mathbb{R}_+^m \qquad H_1^k : \boldsymbol{\delta}_k \not\in \mathbb{R}_+^m.$$

The Wald test statistic under inequality constraints has been developed by Kodde and Palm (1986). For this set of hypothesis, the test statistics T_2^k is defined as:

$$T_2^k = \min_{\boldsymbol{\delta}_k \in \mathbb{R}_+^m} \left\{ n \, (\widehat{\boldsymbol{\delta}}_k - \boldsymbol{\delta}_k)^t \, \widehat{\boldsymbol{\Omega}}^{-1} \, (\widehat{\boldsymbol{\delta}}_k - \boldsymbol{\delta}_k) \right\}.$$

Kodde and Palm (1986) have shown that the statistic T_2^k is asymptotically distributed as a mixture of χ^2 distributions, provided that the result in (A.19) holds:

$$T_2^k \ \sim \ \overline{\chi}^2 \ = \ \sum_{j=0}^{m} w\left(m, m-j, \widehat{\boldsymbol{\Omega}}\right) \Pr\left(\chi_j^2 \geq c\right),$$

with $w\left(m, m-j, \widehat{\boldsymbol{\Omega}}\right)$ the probability that $m-j$ elements of $\boldsymbol{\delta}_k$ are strictly positive.[4]

[4] To estimate $w\left(m, m-j, \widehat{\boldsymbol{\Omega}}\right)$, we draw 10,000 multivariate normal vectors and covariance matrix $\widehat{\boldsymbol{\Omega}}$, provided it is positive definite. Then we compute the proportion of vectors with $m-j$ positive elements.

To test the reverse dominance order, that is $F_2 \succ_{ISDk} F_1$, it is sufficient to replace $-\widehat{\boldsymbol{\delta}}_k$ and $-\boldsymbol{\delta}_k$ for their positive counterparts.

A.6 Testing dominance in the gap curve

We extend the previous analysis on stochastic dominance by constructing a test for dominance at order k in the differences of the gap curves, defined by $({}_0\Lambda_1^k(p) - {}_0\Lambda_2^k(p)) - ({}_1\Lambda_1^k(p) - {}_1\Lambda_2^k(p))$.[5] It is assumed in this section that there exists a degree of inverse stochastic dominance, κ, for which ${}_\pi F_1 \succ_{ISD\kappa} {}_\pi F_2$ for all π. The ezOP algorithm requires to test the gap dominance only at order $\kappa(c_1, c_2, 1)$. In general, the gap dominance defined over the difference between gap curves can be tested for any pair of gap curves. However, it can be related to changes in certain equivalents only by assuming that each of the gap curves can be related to inverse stochastic dominance at a given order. We focus on testing gap dominance for the cases in which one can assess the direction of dominance within each policy. The null and alternative hypothesis take the form:

$$\begin{aligned} H_0^k: &\quad {}_0\Lambda_1^k(p) - {}_0\Lambda_2^k(p) \geq {}_1\Lambda_1^k(p) - {}_1\Lambda_2^k(p) \quad \text{for all } p \in [0,1]; \\ H_1^k: &\quad {}_0\Lambda_1^k(p) - {}_0\Lambda_2^k(p) < {}_1\Lambda_1^k(p) - {}_1\Lambda_2^k(p) \quad \text{for some } p \in [0,1]. \end{aligned}$$

One can easily test for equality by reversing the role of $c = 1$ and $c = 2$ and testing if gap dominance is accepted also in this case.

Alternatively, the conditional testing procedure for the gap curve can be formulated by looking at the differences in the conditional Gini SWF. In fact, the conditional Gini SWF are expressed as certain equivalents. If there is a reduction of distance for this family, then there is gap dominance. The null and alternative can be expressed in the following terms for gaps at order k:

$$\begin{aligned} H_0^k(W): &\quad W^{k-1}(q,{}_0F_1) - W^{k-1}(q,{}_0F_2) \geq W^{k-1}(q,{}_1F_1) - W^{k-1}(q,{}_1F_2) \quad \text{for all } q \in [0,1]; \\ H_1^k(W): &\quad W^{k-1}(q,{}_0F_1) - W^{k-1}(q,{}_0F_2) < W^{k-1}(q,{}_1F_1) - W^{k-1}(q,{}_1F_2) \quad \text{for some } q \in [0,1]. \end{aligned}$$

In practice, dominance can be tested only for a finite number m of population shares, both in the direct and threshold testing procedures. We follow the previous section in defining $\boldsymbol{\Theta}_G^k$ as the $4m \times 1$ vector obtained by staking the vectors ${}_0\boldsymbol{\Theta}^k$ and ${}_1\boldsymbol{\Theta}^k$ in this precise order. The index π identifies the sample estimator under policy π. The sample

[5] Note that the role of circumstances $c = 1$ and $c = 2$ can be reversed under different policy regimes.

estimates are collected in the $4m \times 1$ vector $\widehat{\mathbf{\Theta}}_G^k$, and we use $n = {}_0n_1 + {}_0n_2 + {}_1n_1 + {}_1n_2$ to depict the overall sample size with circumstances $c = 1$ and $c = 2$ under $\pi = 0$ and $\pi = 1$, while ${}_\pi r_c = {}_\pi n_c/n$ is the relative size of each sub-sample.

The hypothesis of dominance in the differences of the gap curve can be reformulated as a sequence of m linear constraints placed on the vector $\mathbf{\Theta}_G^k$. Let $\mathbf{R}_G = (\mathbf{R}, -\mathbf{R})$ be the $m \times 4m$ *difference-in-differences* matrix, where $\mathbf{R}$ is defined as above. Define the parametric vector $\boldsymbol{\gamma}_k = \mathbf{R}_G \, \mathbf{\Theta}_G^k$.

We maintain the (non-testable) assumption that ${}_\pi F_1$ and ${}_\pi F_2$ can be described by independent processes for all π. Moreover, we introduce the assumption of independence between ${}_0F_c$ and ${}_1F_c$ for all c. This latter assumption is verified when the sampling scheme is based upon randomized assignment to treatment and control groups. The various hypothesis of dominance or equality can be written in terms of linear inequalities involving $\boldsymbol{\gamma}_k$. By exploiting the result in (A.6) and (A.12), the hypothesis of interest specify restrictions of an asymptotic multivariate normal variable:

$$\sqrt{n}\,\widehat{\boldsymbol{\gamma}}_k \;=\; \sqrt{n}\,\mathbf{R}_G\,\widehat{\mathbf{\Theta}}_G^k \;\text{ is asymptotically distributed as }\; \mathcal{N}\left(\sqrt{n}\,\mathbf{R}_G\,\mathbf{\Theta}_G^k,\;\mathbf{\Phi}\right), \quad \text{(A.20)}$$

where $\widehat{\boldsymbol{\gamma}}_k$ denotes the sample estimate of $\boldsymbol{\gamma}_k$, and

$$\mathbf{\Phi} \;=\; \mathbf{R}_G \,\mathrm{diag}\left(\frac{{}_0\mathbf{\Sigma}_1^k}{{}_0r_1}, \frac{{}_0\mathbf{\Sigma}_2^k}{{}_0r_2}, \frac{{}_1\mathbf{\Sigma}_1^k}{{}_1r_1}, \frac{{}_1\mathbf{\Sigma}_2^k}{{}_1r_2}\right)\mathbf{R}_G^t.$$

The empirical estimator of the asymptotic variance, $\widehat{\mathbf{\Phi}}$, is obtained by plugging ${}_\pi\widehat{\mathbf{\Sigma}}_c^k$ in the previous formula. As in the case of ISDk testing, the empirical covariance estimator can be obtained by using the empirical counterpart of the covariance matrices in Lemma A.1 and Proposition A.1 and A.2. We exploit the convergence result to test the discretized versions of the dominance or equality hypothesis, which is defined on the set of proportions $\{q_1, \ldots, q_m\}$.

A.6.1 Testing equality in gap curves difference

In the case of equality testing, the null and alternative hypothesis for dominance in gaps can be stated as follows:

$$H_0^k : \;\boldsymbol{\gamma}_k = \mathbf{0} \qquad H_1^k : \;\boldsymbol{\gamma}_k \neq \mathbf{0}.$$

Under the null hypothesis, it is possible to resort to a Wald test static ${}_{G}T_1^k$:

$${}_{G}T_1^k \;:=\; n\,\widehat{\boldsymbol{\gamma}}_k^t\,\widehat{\boldsymbol{\Phi}}^{-1}\,\widehat{\boldsymbol{\gamma}}_k.$$

Given the convergence results in (A.20), the asymptotic distribution of the test ${}_{G}T_1^k$ is χ^2_m. The p-value tabulation follows the usual rules.

A.6.2 Testing dominance in gap curves difference

In the case of strong dominance testing, such that $G({}_0\Lambda_1^k,\,{}_0\Lambda_2^k,u) \geq G({}_1\Lambda_1^k,\,{}_1\Lambda_2^k,u)$ for all $u \in [0,1]$, the null and alternative hypothesis concerning the dominance in gap curves can be stated as follows:

$$H_0^k:\ \boldsymbol{\gamma}_k \in \mathbb{R}_+^m \qquad H_1^k:\ \boldsymbol{\gamma}_k \not\in \mathbb{R}_+^m$$

The Wald test statistics with inequality constraints has been developed by Kodde and Palm (1986). For this set of hypothesis, the test statistics ${}_{G}T_2^k$ is defined as:

$${}_{G}T_2^k = \min_{\boldsymbol{\gamma}_k \in \mathbb{R}_+^m} \left\{ n\left(\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\right)^t \widehat{\boldsymbol{\Phi}}^{-1} \left(\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\right) \right\}.$$

Kodde and Palm (1986) have shown that the statistic T_2^k is asymptotically distributed as a mixture of χ^2 distributions, provided that (A.20) holds:

$$T_2^k \;\sim\; \overline{\chi}^2 \;=\; \sum_{j=0}^{m} w\left(m, m-j, \widehat{\boldsymbol{\Phi}}\right) \Pr\left(\chi_j^2 \geq c\right),$$

with $w\left(m, m-j, \widehat{\boldsymbol{\Phi}}\right)$ the probability that $m-j$ elements of $\boldsymbol{\gamma}_k$ are strictly positive.[6]

To test the reverse dominance order, that is $G({}_1\Lambda_1^k,\,{}_1\Lambda_2^k,p) \geq G({}_0\Lambda_1^k,\,{}_0\Lambda_2^k,p)$ for all $p \in [0,1]$, it is sufficient to replace $-\widehat{\boldsymbol{\gamma}}_k$ and $-\boldsymbol{\gamma}_k$ for their positive counterparts.

[6] To estimate $w\left(m, m-j, \widehat{\boldsymbol{\Phi}}\right)$, we draw 10,000 multivariate normal vectors and covariance matrix $\widehat{\boldsymbol{\Phi}}$, provided it is positive definite. Then we compute the proportion of vectors with $m-j$ positive elements.

Appendix B

Semi and nonparametric models for identification of quantile treatment effects

Contents

B.1 Objective

Nowadays, 95% of applied econometrics is concerned with averages. Many variables, however, have continuous distributions and they can change in a way not revealed by the examination of (conditional) averages. For instance, the policymaker may be interested in the distributional impact of a policy, when the impact may not be uniformly distributed across population percentiles. A redistributive policy may increase earnings for the poorest quantiles while decreasing earnings for the richest quantiles of the earnings distribution. Changes may compensate on average, while distributional effects may be sizable. Models for estimation of the *quantile treatment effects* have been proposed to analyze the distributional impact of policy implementation. Although quantile regression models are not new, the implementation of these models in a semi- or non-parametric setting is recent, and the issues of identification and estimations of these models under endogeneity are still debated.

This section is a survey of the most recent innovations on identification and empirical implementation of quantile regression models under endogeneity of some of the regressors. We want to emphasize the role of identification, and therefore we have organized this appendix according to the identifying assumptions of the models surveyed. One of these models, the instrumental variables estimator by Abadie et al. (2002), is used in the empirical application in chapter 3.

B.2 Quantile function, quantile regression and quantile treatment effects

Let us denote $Q_Y(p)$ the quantile $p \in [0,1]$ associated to a random variable Y with distribution $F(y)$. It is determined by the following relation:

$$Q_Y(p) \;=\; \inf\{y : F(y) \geq p\}.$$

When F is strictly increasing, $Q_Y(p) = F_Y^{-1}(p)$. The quantile function is increasing, left continuous and satisfies $Q_{\alpha Y+\beta}(p) = \alpha Q_Y(p) + \beta$ for α positive. However, the quantile function is not additive. Given two random variables Y and Y', $Q_{Y+Y'}(p) \neq Q_Y(p) + Q_{Y'}(p)$. This property is crucial for the analysis of changes in quantiles, because it limits the possibility to aggregate quantiles of different (possibly conditional) variables evaluated at similar percentiles p of the population. This is an important drawback compared to the expectations, which are linear in their arguments.

To estimate quantiles, and in general ordered moments of a (conditional) distribution, we resort on a very general optimization problem. Let identify the object of interest by the parameter (or vector of parameters) $\beta(p)$, calculated for quantile p. We restrict attention to the class of moments solving the following program:

$$\beta(p) \quad := \quad \text{argmin}_b \, E\left[W(p)\rho_p(Y-b)\right], \tag{B.1}$$

where $W(p)$ is a weighting scheme and $\rho_p(.)$ is called the *check function*, such that $\rho_p(u) = (p - \mathbf{1}(u<0))u$, with $\mathbf{1}(.)$ the usual indicator function. In general, b is a function of $\beta(p)$.

The empirical estimator of $\beta(p)$, denoted by $\widehat{\beta}(p)$, can be also expressed as the solution of a minimization program. Given a collection of n independent realizations of Y (a random sample), let consider the following program:

$$\widehat{\beta}(p) \quad := \quad \text{argmin}_b \sum_{i=1}^{n} W_i \rho_p(Y_i - b), \tag{B.2}$$

where i identifies an observed value, W_i is the realization of a weighting scheme and $\rho_p(.)$ is the check function.

When b is a scalar number, the program can be used to estimate the empirical quantiles. Therefore the solution of the program (B.2) is $\widehat{\beta}(p) = \widehat{Q}_Y(p)$.[1] In many cases, however, one may be interested in defining a model for b, so that to derive the *conditional quantiles.*

Let assume that Y can be modeled as a function of observables X, with $Y = Y(X)$. By using $b := Y(X)$ in program (B.1) one obtains the *conditional quantile function* $\beta(p) = Q_{Y|X}(p)$. The empirical estimator of the conditional quantile function can be constructed by modeling the check function in program (B.2), provided that a functional form for $Y(.)$ is chosen.

In many instances, one is more interested in appraising the impact of small changes of X (or in just one element X_j of X) on the quantiles of the distribution of Y. For X_j, the effect of interest is the marginal impact $\frac{\partial Q_{Y|X}(p)}{\partial X_j}$ of the population quantiles at p conditional on

[1] For instance, if $W_i = 1$ for all i and the check function is $\rho_{.5}(Y_i - b) = |Y_i - b|$, program (B.2) minimizes the Least Absolute Deviation and the solution $\widehat{\beta}_{.5}$ identifies the *median* value of the distribution $Y_1, \ldots, Y_n$. Other quantiles can be estimated by choosing the appropriate functional form of $\rho_p(.)$. In particular situations, program (B.2) provides the same solution for the empirical quantiles as if they are calculated by using $\widehat{Q}_Y(p) = Y_{[np]}$, where Y_i indicates the i-th observed realization of Y when observations are arranged by increasing magnitude, and $[np]$ is the natural part of the real number np.

covariates X. If one assumes that $Y(X) = X\beta(p)$ is linear, that is the marginal impact of X_j is constant across levels of X, than the solution of program (B.1) gives the parameters of interest $\beta(p) = \beta_j(p)$, evaluated at quantile p. The parameter can be estimated by resorting on the *quantile regression* model of Koenker and Bassett (1978).

In general, one may want to provide estimates of the marginal effect of small changes in X irrespectively of the functional form of $Y(.)$ (which has not to be necessarily linear), or to asses the impact of non marginal changes in X, for instance when X_j is an indicator for policy treatment. These comparisons are possible only at the cost of conditioning the analysis to the covariates vector X. In the rest of the appendix, we discuss the fundamentals for the analysis of quantile treatment effects, the objet of our analysis (section B.2.1) and the identification issues in a variety of cases (section B.3).

B.2.1 QTE and ATE

The object of interest for a policymaker concerned by the distributional impact of a policy is the *quantile treatment effect* (QTE hereafter) $\alpha(p)$ at quantile p. Following Athey and Imbens (2006) and Firpo, Fortin and Lemieux (2009), let assume that individual i can be assigned either to a group treated by a given policy or to a group that is not treated, according to an indicator $I_i \in \{0, 1\}$. There are two potential outcomes associated to individual i, denoted by Y_i^I. We use Y_i^1 for the potential outcome of i when treated with the policy, and outcome Y_i^0 for i's potential outcome when not treated with the policy. The policy impact for this individuals is defined as $\Delta_i = Y_i^1 - Y_i^0$, which gives the magnitude of the difference in outcomes if i is treated compared to the outcome that i receives if she is not treated, keeping everything else as constant.

Economists are often interested in the *average treatment effect* (ATE, or *mean policy effect*) associated to a "policy change" or "intervention", defined as the expected gains from treatment for a randomly drawn person from the population:

$$ATE \; := \; E[Y^1 - Y^0].$$

In a similar manner, we can define the *quantile treatment effects* (QTE) for a given quantile p. Let F_{Y^I} be the distribution function of potential outcomes Y^I, while we note with $Q_{Y^I}(p)$ the quantile p of this distribution. The QTE $\alpha(p)$ coincides with the expected gains from treatment associated to quantile p of outcomes distributions:

$$\alpha(p) \ := \ Q_{Y^1}(p) - Q_{Y^0}(p).$$

It is now clear that the QTE cannot be assigned to any individual in particular in the distribution, but rather the QTE consists in a measure of the change in incomes at a given quantile of the distribution due to policy shifts. If, for instance, $\alpha(.01) > 0$, one cannot infer that the group of individuals who were relatively "poor" compared to the others when not treated becomes relatively "less poor" with the treatment (a comparison in absolute terms). Rather, it is only possible to say that those who are "poor" in the regime without treatment are "less poor" than those who are "poor" in the regime with the treatment. Hence, the QTE on distributions does not coincide with the distribution of QTE, unless in the cases where intervention is rank-preserving.

B.2.2 Conditional versus unconditional QTE

The model presented above can be easily expanded to include covariates that can be observed (X) or not observed (ε). Consider the case in which the potential outcome is indeed a function of the observables and unobservables, such that $Y^I = Y^I(X, \varepsilon)$. In this setting, one can compare policy changes that are either *conditional* or *unconditional* with respect to observable covariates. Let define the unconditional ATE (UATE) as:

$$UATE \ = \ E[Y^1(X,\varepsilon)] - E[Y^0(X,\varepsilon)] = E[E[Y^1(X,\varepsilon) - Y^0(X,\varepsilon)|X]], \qquad \text{(B.3)}$$

where the second identity follows by the law of iterated expectations and an orthogonality condition for the unobserved component. Hence, one can derive UATE by estimating the conditional ATE (CATE) with a model for conditional expectations, and then averaging CATE over the values of the conditioning observable X. However, the same relation between UATE and CATE cannot be retrieved when looking at unconditional QTE versus conditional QTE, the latter being defined as:

$$\alpha_X(p) \ = \ Q_{Y^1|X}(p) - Q_{Y^0|X}(p). \qquad \text{(B.4)}$$

In fact, the conditional QTE $\alpha_X(p)$ identifies the policy impact associated to the quantile p of the distribution of outcome Y made conditional on X. Similar concerns do not arise for the identification of the ATE, since one can model the conditional population expectations $E[Y|X]$ and then retrieve, by the law of iterated expectations, information on the unconditional population mean $E[Y]$.[2]

As shown in Firpo et al. (2009), even if one proposes a model for the conditional quantile function which allows to recover a reliable estimate of the *conditional QTE* as a function of the observables, one is not able to reconstruct the *unconditional QTE*, the impact of a policy on the *overall* outcome distribution, by simply averaging the conditional QTE estimates over the values of X, apart from very particular cases where one is allowed to work with linear models (thus achieving parametric identification). However, under the common independence assumption $\varepsilon \perp X$, Firpo et al. (2009) were able to show that the unconditional QTE can be written as a complicated weighted average of conditional QTE, but not as its expectation.

We discuss in the remaining the issues related to the identification and estimation of the QTE under different exogeneity assumptions on ε (Firpo 2007, Koenker and Bassett 1978), and possible solutions for endogenity of the unobservable components stemming from the literature on instrumental variables (Abadie 2002, Abadie et al. 2002, Frölich and Melly 2008) and on difference-in-differences design (Athey and Imbens 2006). Along with these identification issues, we also discuss under which conditions one can estimate unconditional QTE from a conditional model (Firpo et al. 2009).

B.3 Identification of QTE

In randomized control trials, the QTE can be estimated by taking the horizontal distance at a given population percentile p between the outcome distribution of the treated group, $Q_{Y|D=1}(p)$, and the distribution of the non-treated group, $Q_{Y|D=0}(p)$, where $D \in \{0,1\}$ is an indicator function that randomizes observations into treated and non-treated groups independently on pre-assigned (un)observable characteristics. Therefore:

$$\alpha(p) \; = \; Q_{Y|D=1}(p) - Q_{Y|D=0}(p). \tag{B.5}$$

When randomized data are available, one can estimate the unconditional treatment effect by simply looking at the empirical counterparts of the quantile functions in (B.5),

[2]For instance, $E[Y|X] = X\beta$ implies that $E[Y] = E[X]\beta$ and OLS estimates of β give information on the population impacts of changes in X (which may contain the policy variable).

because the treated ($D = 1$) and non treated ($D = 0$) groups will have asymptotically identical distributions of potential outcomes. Hence, (B.5) allows to isolate the causal impact of policy intervention.

In real economic applications, however, randomization does not (almost) always take place because of self-selection across policy treatments based on characteristics that may be observable, unobservable or according to both of them. We survey the solutions offered in literature to cope with this endogeneity issue undermining the QTE identification, along with the appropriate empirical strategies to adopt. There are four possible cases that one may construct by looking at (i) if selection takes place on observables or not, and (ii) if the interest is in conditional or unconditional QTE. In what follows, we discuss each single case in detail.

Before proceeding, let us stress that the realization of the random variable Y is defined by a non specified model $Y = Y(D, X, \varepsilon)$ which depends upon (i) the policy variable D (for which we aim at capturing the QTE), (ii) the observables X and (iii) the unobservable factors in ε. The function $Y(.)$ is a model for the *potential* individual outcomes at D, and it can be represented by very flexible and general functional forms. However, only under some particular restrictions one can achieve identification of such general model.

B.3.1 Conditional exogenous QTE

To estimate conditional models for the QTE under the exogeneity assumption, it is sufficient to use standard quantile regression analysis. The identification rests on two assumptions. The first is linearity of the quantiles of the outcome variable Y, modeled as the following:

$$Y(p) = Y(D, X, \varepsilon) = \alpha(p)D + X\gamma(p) + \varepsilon \quad \text{and} \quad Q_\varepsilon(p) = 0,$$

where the QTE $\alpha(p)$ and the marginal effects of the observables $\gamma(p)$ are calculated at percentile p. The second assumption, exogeneity, consists in requiring that unobserved heterogeneity is independently distributed from the treatment variable *and* the other observable covariates, that is $\varepsilon \perp (D, X)$. The two assumptions together imply that conditional quantiles take the form:

$$Q_{Y|D,X}(p) \;=\; \alpha(p)D + X\gamma(p).$$

It is now clear that $\alpha(p)$ can be identified only by comparing treated versus non treated observations that have been randomized on the observables. Hence, QTE is identified within the group of individuals with similar realizations of X, such that

$$\alpha(p) = \alpha_X(p) = Q_{Y|D=1,X}(p) - Q_{Y|D=0,X}(p).$$

Koenker and Bassett (1978) have shown that the QTE in this model is identified, and its estimator, $\widehat{\alpha}(p)$, coincides with the solution of the program (B.2) under linearity of $b = \alpha D_i + X_i\gamma$ and by using the following weighting scheme:

$$W_i = 1 \quad \forall i.$$

Under the linearity assumption, the unconditional QTE trivially coincides with the conditional QTE. This can be seen by adapting the result of Proposition 1 in Firpo et al. (2009) to non-marginal changes of the treatment variable. However, this tie breaks down as soon as $Y(D, X, \varepsilon)$ is non-linear. This case is treated in the next section. In both cases the identification assumption rests on an independence assumption.

B.3.2 Unconditional exogenous QTE

Consider again the case of selection on observables X. As before, we assume that unobservables and observables are independently distributed, while we make use of a new exogeneity assumption, namely *unconfoundedness*: $(Y^0, Y^1) \perp D|X$. For a given control variable, unconfoundedness states that observable and unobservable factors that explain the response variable are independent, given the control variables X. Firpo (2007) demonstrates how to use the selection on observables assumption, along with a common support assumption, to calculate the unconditional quantiles for the treated and for the non-treated outcomes without computing the corresponding conditional QTE. As a consequence, by comparing the marginal quantiles it is possible to identify unconditional QTE.

Firpo (2007) proposes a semi-parametric procedure to compute QTE. This estimation technique requires two steps. In the first step, a nonparametric estimator $\widehat{p}(X_i)$ of the propensity score of observation i is estimated. The second step estimates parametrically the QTE as the differences between two unconditional quantiles by solving the minimization problem (B.2), where the parametric model for the potential outcome quantile is given by

$b = \gamma + \alpha D_i$ and the weighting scheme is defined as follows:

$$W_i \;=\; \frac{D_i}{\widehat{p}(X_i)} + \frac{1 - D_i}{\widehat{p}(X_i)}. \tag{B.6}$$

In an empirical study, Firpo (2007) makes use of logistic power series approximation to estimate (B.6). The estimator $\widehat{\alpha}(p)$ is consistent for the unconditional QTE in the population.

The QTE is identified provided that the unconfoundedness assumption is not violated. If there is selection on unobservable factors, one may expect that given X the potential outcomes realizations are no more independent on the treatment variable D. Hence identification fails. We look now at alternative solutions for this problem based upon *instrumental variables* and *difference-in-differences* methods.

B.3.3 Endogenous QTE: the IV approach

Exogeneity or unconfoundedness assumptions are violated when unobservable factors affect both the treatment decision (even conditional on X) and the distribution of potential treatment outcomes. Hence, the treatment status is said to be endogenous to the outcome variable.

Instrumental variables (IV) methods provide powerful tools for identifying causal estimates of QTE under endogeneity. We introduce a new indicator, $Z \in \{0, 1\}$, which affects the potential treatment status of an observation. We use D^1 to indicate the potential treatment status when $Z = 1$ and D^0 the potential treatment status when $Z = 0$. We consider the following independence assumption: given X (the observable covariates), potential outcomes and potential treatment status (Y^1, Y^0, D^1, D^0) are jointly independent from Z. Z is therefore an IV for treatment D. This is the traditional instrument-error independence assumption. Variations in the IV identify the causal effect of the treatment status on the outcome quantiles, while potential outcomes should not be directly affected by the IV.

Abadie et al. (2002) showed that the traditional instrument independence assumption, i.e. $(Y^1, Y^0, D^0, D^1) \perp Z|X$, implies that $(Y^1, Y^0) \perp D|X, D^1 > D^0$, which means that in the population of *compliers* (where $D^1 > D^0$, that is those whose potential treatment assignment status changes by effect of changes in the IV), comparisons by D conditional on X have a causal interpretation. Using this assumption, Abadie et al. (2002) show that the QTE are identified for the group of compliers, where the QTE can be written as a solution of program (B.1) where $b = Q_{Y|X,D,D^1>D^0}(p)$ and the weighting function takes the form of:

$$W(p) \; = \; 1 - \frac{D(1-Z)}{(1-p(Z=1|X))^2} - \frac{(1-D)Z}{(p(Z=1|X))^2}, \tag{B.7}$$

where $p(Z=1|X)$ measures the probability that the instrument assigns to the treatment given observable factors X.

A particular measure of interest for policy is the QTE $\alpha(p)$, identified by a linear model of the form

$$Q_{Y|X,D,D_1>D_0}(p) \; = \; \alpha(p)D \; + \; X\gamma(p).$$

The estimator for $\alpha(p)$ under linearity, $\widehat{\alpha}(p)$, is obtained by solving program (B.2) where b is replaced by $\alpha(p) + X_i\gamma(p)$ and where W_i is estimated by the sample analog of (B.7). Abadie et al. (2002) show that this estimator is asymptotically normal and they provide a tractable form of its covariance matrix.

To estimate $\widehat{\alpha}(p)$, one has to proceed in two steps. The fist step consists in estimating W_i, the weighting scheme, at individual level. This can be done by using parametric or non parametric estimators of $p(.)$ in (B.7). Under independence assumption, this propensity score estimator identifies the group of the compliers. Then, the second step consists in finding a solution of program (B.2) by modeling the weighting scheme as the estimated sample counterpart of (B.7) and under the linearity assumption introduced above. This is a convex problems and a solution always exists.

We summarize here few facts related to the IV strategy for estimating QTE. First, IV allows to identify QTE only in the group of compliers. This is so because of the mechanics of the IV: changes in the IV are assumed to provide exogenous shifts on the otherwise potentially endogenous treatment assignment variable D, which in turn identify the causal impact of treatment on $Q(p)$. However, exogenous changes can only be identified for the group of compliers. Compliers cannot be identified within the data because one observation corresponds only to a pair (D, Z). However, one can estimates distributions of Y^1, Y^0 by using the outcome distributions under $Z = 1$ and $Z = 0$ which are, by exogeneity, identical.

Second, the identifying power of the IV estimation grows inversely with the size of compliers in the comparison group defined by $Z = 0$. If nobody in the group with $Z = 0$ is also treated by the policy change, then by using the estimator by Abadie et al. (2002) it is possible to identify the QTE for the entire group of treated individuals.

Third, asymptotic properties of the conditional QTE are identified only for parametric models. However, the parametric restrictions are sometimes difficult to justify theoretically, and in particular the linearity assumption, which implies that QTE are constant across realizations of X. One particular case where linearity does not need to be justified is when the covariates represent discrete variables or fixed effects components.

Frölich and Melly (2008) propose an estimator of unconditional QTE under the instrument conditional exogeneity assumption that relies upon IV identification. They show that that the QTE $\alpha(p)$ is non-parametrically identified by solving a version of the convex program in (B.1) where $b = \gamma(p) + \alpha(p)D$ and $W(p)$ can be written as a function of Z, D and the propensity score $p(X)$. It corresponds therefore to a particular case of the QTE in Abadie et al. (2002). The estimator $\widehat{\alpha}(p)$ of the QTE can be estimated by a two step procedure. In the first step, a non-parametric model is used to estimate the empirical propensity score of observation i, denoted $\widehat{p}(X_i)$. In the second step, $\widehat{\alpha}(p)$ is obtained as the solution of problem (B.2) where $b = \gamma(p) + \alpha(p)D_i$ and where the weighting scheme is defined as follows:

$$W_i = \frac{Z_i - \widehat{p}(X_i)}{\widehat{p}(X_i)(1 - \widehat{p}(X_i))}(2D_i - 1).$$

There are cases where the conditional and unconditional IV QTE estimator coincide (asymptotically). For instance, in the *location shift model*, where the QTE is the same independently on the realization of the covariates and of the value of the quantiles p.[3] Another interesting case occurs when the IV is in fact independent on X, so that one can identify QTE only for exogenous variation in the IV that are not dependent upon the conditioning variables. This is particularly true when IV is randomly assigned across all observations. This is the case, for instance, when the variable Z indicates the cohort, the historic period or the region where a given individual is observed. If one considers two adjacent cohorts, historic periods or regions it is very likely that the IV independence to unobservables holds.

[3] If the second condition is not verified, one may still have that the bottom percentile in a particular group defined by X are treated with QTE that are similarly associated to the bottom percentile in another groups defined by X's realizations, although the bottom percentile in the former group may be in absolute terms quite higher than the bottom percentile in the latter group.

B.3.4 Unconditional QTE: the difference-in-differences approach

The standard model for the difference-in-differences (DiD) design assumes that one individual is assigned in a *treatment* or *control* group according to an indicator $G_i \in \{0,1\}$, and she is either *treated* or *non treated* with a given policy according to the indicator D_i. As it is often the case, D_i represent a time indicator, taking value 1 if the individual is observed after that a given policy is implemented. In the DiD setting, identification relies on randomization across G_i groups (while in the IV setting randomization occurs across Z_i levels).

Potential individual outcomes Y_i^I (where I is defined above) are never observed, but only the realized outcome is observed, which can be written as a function of the potential outcomes:

$$Y_i \;=\; (1-I_i)\,Y_i^0 + I_i\,Y_i^1 = Y_i^0 + (Y_i^1 - Y_i^0)\,I_i = Y_i^0 + (\alpha_i)\,I_i.$$

The parameter α_i gives the impact of the policy at individual level, the parameter of interest. In the DiD design, we consider $I_i = G_i \cdot D_i$, since only individuals in group $G_i = 1$ are effectively treated in period $D_i = 1$. Let assume a linear model for the potential outcome without treatment,

$$Y_i^0 \;=\; \gamma_0 + \gamma_1\,G_i + \gamma_2\,D_i + \varepsilon_i,$$

and assume furthermore that the treatment effect is constant across individuals, i.e. $Y_i^1 - Y_i^0 = \alpha$. This leads to the following model of realized outcomes:

$$Y_i \quad = \quad \gamma_0 + \gamma_1\,G_i + \gamma_2\,D_i + \alpha\,I_i + \varepsilon_i. \tag{B.8}$$

When the error term is assumed to be independent of the group indicator and it has the same distribution over policy treatment statuses, i.e. $\varepsilon \perp (G,D)$, the ATE in model (B.8) is identified for the treated group:

$$\begin{aligned}\alpha \;=\; & E[Y_i|D_i=1,G_i=1] - E[Y_i|D_i=0,G_i=1] - \\ & - \left[E[Y_i|D_i=1,G_i=0] - E[Y_i|D_i=0,G_i=0]\right].\end{aligned}$$

This model can be generalized to the non-linear framework, by modeling the counterfactual outcomes according to what Athey and Imbens (2006) calls a production function $Y(.)$ and a group composition term (unobservable heterogeneity) ε. To analyze the policy impact they assume that in the presence of intervention, $Y_i^I = Y^I(\varepsilon_i, D_i)$. Identification in this setting is granted by the following exogeneity assumption: $\varepsilon \perp D|G$. It requires that the distribution of the heterogeneity within (but not across) the treatment and the control groups is identically distributed both before and after policy treatment. Hence, even in a non-linear framework, ATE for the treated can be identified provided that the production function if fixed and the heterogeneity do not vary, within the treatment or control group, across treatment regimes, although heterogeneity may vary across treatment and comparison groups.

Under independence, differences in expected outcomes within groups defined by G_i eliminate the impact of the heterogeneity, while the difference of these differences across groups allows to identify the treatment effect as differences in the levels of the production function. In the traditional linear model, this amount to assume that:

$$\begin{aligned} \varepsilon &= \gamma_2 G + \nu, \\ Y(\varepsilon, D) &= \gamma_0 + \gamma_1 D + \varepsilon. \end{aligned}$$

The independence assumption introduces a very strong exogeneity requirement: the distribution of the heterogeneity is assumed to be fixed within the control or the treatment group, while it can vary across the two groups. This allows to interpret the differences of the outcomes of treated versus non treated units within the control group as the residual earnings after eliminating the impact of unobserved heterogeneity (although this does not require that heterogeity is similarly distributed across groups). Then, the only difference between residual achievements in the treated and control group is the fact that units within the treatment group have been exposed to policy treatment, while the units in the control group are not. Hence, by differentiating residual incomes one obtains an estimate of the impact of the policy on the earnings of the treated.

In many empirical applications, the indicator D refers to *time*, assigning value one to the observations after that a given policy is implemented. The indicator G usually randomly assigns units across groups of the population, like treated and control (i.e. non-treated) regions (see Havnes and Mogstad 2010, for an application within this setting). ATE in the DiD setting is the difference in average achievement of the treated observations (those in the treated regions after policy implementation) and these observations' counterfactual average outcome. The counterfactual outcome is estimated by considering the average outcome before the treatment takes place in the treated regions, augmented by the time trend estimated as the change across time in average outcome in non-treated regions (that is differentiating the outcome conditional on $D = 1$, $G = 0$ and $D = 0$, $G = 0$).

The quantile DiD design (QDiD) extends the DiD analysis to the comparison of distribution functions. Instead of considering individual potential outcomes, the QDiD relies on conditional potential outcome distributions $F_{Y|D,G}(y)$ and the corresponding quantile functions. In essence, the QDiD strategy assumes that the control group provides information about what would have been the earning distribution of the treatment group in the absence of the treatment. This can be done, under different exogeneity assumption, by estimating the *counterfactual* earning distribution $k(p)$ for the treated in the treatment group. The quantile $k(p)$ is obtained from the distribution of those non treated in the treatment sample, $F_{Y|D=0,G=1}(y)$, and augmented by the impact of changes in D on the counterfactual group. Depending on the exogeneity assumption considered, one is able to identify different, and somehow alternative versions of the QTE. We survey three alternative definitions of the QTE in the DiD setting, focussing in particular on estimators for unconditional QTE.

The first model that we consider is the traditional QDiD estimator discussed, among others, in Athey and Imbens (2006). This model estimates the quantiles of the counterfactual distribution for the treated in the treatment group at the quantile corresponding to the earnings distribution of the non-treated in the treatment group, with $F_{Y|D=0,G=1}(y) = p$:

$$\begin{aligned} \alpha^{QDiD}(p) &= Q_{Y|D=1,G=1}(p) \; - \; k^{QDiD}(p), \\ k^{QDiD}(p) &= y \; + \; [Q_{Y|D=1,G=0}(p) - Q_{Y|D=0,G=0}(p)]. \end{aligned}$$

This method compares individuals both across groups and treatment regimes, according to their quantile. The QTE $\alpha^{QDiD}(p)$ is identified by the QDiD estimator $\widehat{\alpha}^{QDiD}(p)$ under the following assumptions: (i) $\varepsilon \perp (D, G)$, so that the traditional DiD model is a special case of QDiD; (ii) $Y(.)$ is monotone in unobservables; (iii) group effects and treatment effects are additively separable (common trend in levels across treatment and comparison group at the quantile). These assumptions have been criticized in literature because they identify QTE that are not invariant to the scaling of the y. Figure B.1 shows in a graph the two differences (marked with Δ_G) that are used to identify the QTE at p, denoted as $\alpha(p) = \Delta_1 - \Delta_0$.

The second model that we consider identifies a QTE that is invariant with respect to

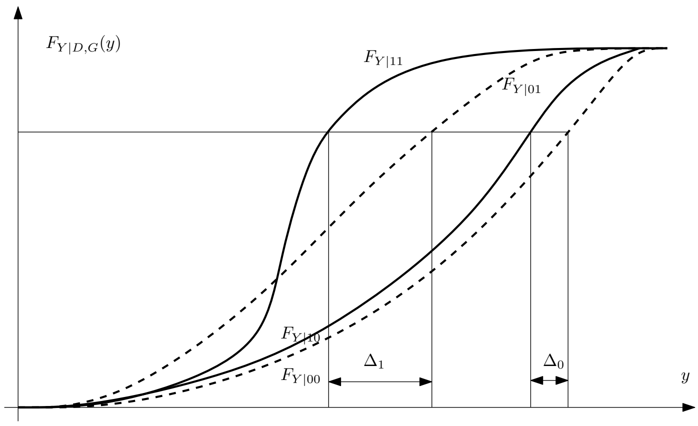


Figure B.1: The QDiD design.

monotonic transformations of the distribution functions (and thus to rescaling). This is the *Changes-in-Changes* design (CiC) proposed by Athey and Imbens (2006). The model relies on less stringent (and more convincing) identifying assumptions than the QDiD model.

Assume that (i) the outcome of the individuals in the absence of intervention does not depend on the group, (ii) potential outcomes vary monotonically with unobservables (a natural assumption when unobservable comprise factors such as individual ability) and (iii) that the heterogeneity generated by unobservables within a group does not vary with policy treatment, that is $\varepsilon \perp D|G$; then the counterfactual distribution of earnings of the treated in the treatment group is identified: $k^{CiC}(y) := F_{Y|D=0,G=1}(Q_{Y|D=0,G=0}(F_{Y|D=1,G=0}(y)))$ (Theorem 3.1 in Athey and Imbens 2006). By a change in variable transformation, it is possible to obtain the CiC QTE $\alpha^{CiC}(p)$ at a given quantile p of the treated, i.e. for $y = Q_{Y|D=0,G=1}(p)$, as follows:

$$\begin{aligned}
\alpha^{CiC}(p) &= Q_{Y|D=1,G=1}(p) - k^{CiC}(p) \\
k^{CiC}(p) &= y + \\
&\quad + Q_{Y|D=1,G=0}\left(F_{Y|D=0,G=0}(Q_{Y|D=0,G=1}(p))\right) - \\
&\quad - Q_{Y|D=0,G=0}\left(F_{Y|D=0,G=0}(Q_{Y|D=0,G=1}(p))\right) \\
&= Q_{Y|D=1,G=0}\left(F_{Y|D=0,G=0}(y)\right).
\end{aligned}$$

The second line follows by definition of the quantile function and by the fact that QTE is, as

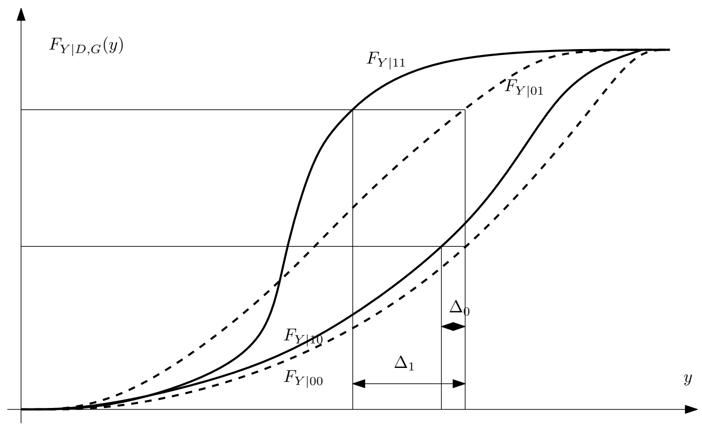


Figure B.2: The CiC design.

usual, associated to the quantile p of the non treated. The rationale behind the CiC model is shown in figure B.2. In this case, the quantile p of the counterfactual outcomes distribution for the treated is obtained by the distribution of the non treated in the treatment group augmented by the the differences in the quantiles of treated versus non treated in the control group, calculated in correspondence of the realization $y = Q_{Y|D=0,G=1}(p)$ of the non treated in the control group.

The CiC estimator does not rely upon linearity (or separability) of the underling model for the counterfactual distribution, and is identified up to a monotone transformation of the outcome variable. However, the CiC estimator is not identified in the model with covariates.

The third and last model that we survey is the *unconditional quantile regression* (UQR in brief) by Firpo et al. (2009). The UQR model consists in running a regression of a transformation of the outcome variable on a set of explicative variables, as in the standard quantile regression. Therefore, this model allows to deal with covariates. However, differently from the quantile regression design (Koenker and Bassett 1978), the UQR allows to retrieve consistent estimates of the *unconditional* QTE. To do so, Firpo et al. (2009) make use of an innovative technique that is based on modeling (by regression analysis) the expectation of the *recentered influence function* (RIF), made conditional on a set of covariates. Then, it is possible to derive the UQTE by differentiating the conditional expectations.

For a given quantile $Q_Y(p)$, the influence function (IF) represent the influence of a given observation on that quantile. For the p-th quantile, the IF is known to be equal to $IF(Y; Q_Y(p)) = (p - \mathbf{1}(Y \leq Q_Y(p)))/f_Y(Q_Y(p))$. The recentered influence function is simply obtained by adding the statistic to the influence function, so that at p:

$$RIF(Y; Q_Y(p)) \;=\; Q_Y(p) \;+\; IF(Y; Q_Y(p)).$$

The unconditional quantile regression is a RIF regression model for the expectations of $RIF(Y; Q_{Y|X}(p))$ at quantile p made conditional upon explanatory variables:

$$E\left[RIF(Y; Q_Y(p)) \mid X\right] \;=\; m_p(X). \tag{B.9}$$

Firpo et al. (2009) showed that the partial effect of small location shifts in the distribution of a covariate X_j on the unconditional quantile corresponds to the average derivative of the unconditional quantile regression in (B.9), that is:

$$\frac{\partial Q_Y(p)}{\partial X_j} \;=\; E\left[\frac{\partial m_p(X)}{\partial X_j}\right].$$

This result can be extended to identify the impact of discrete changes in the policy regime on the unconditional p-th quantile of the outcome distribution (which is measured by the quantity $\Delta_D Q_Y(p)$). Consider a vector of covariates (D, X), where the indicator function D represent policy regime, Firpo et al. (2009) showed that:

$$\Delta_D Q_Y(p) \;=\; m_p(D = 1, X) - m_p(D = 0, X). \tag{B.10}$$

For a general structural model $Y = Y(D, X, \varepsilon)$, these results imply that one can always write the unconditional QTE at p as a weighted average of a family of conditional QTE $\alpha_X(p)$ as in (B.4) (see Proposition 1 in Firpo et al. 2009). However, for linear structural models:

$$Y = Y(D, X, \varepsilon) = \alpha D + X\gamma + \varepsilon,$$

the unconditional and conditional QTE trivially coincide, by the law of iterated expectations on the RIF conditional regression. Hence, under the usual conditional exogeneity assumption, i.e. $\varepsilon \perp D|X$, and the linearity assumption, the QTE is identified as in (B.10). The QTE can be consistently estimated by an OLS regression of the values of the influence function on D and X.

Havnes and Mogstad (2010) propose to use this identification strategy to construct a RIF-based DiD estimator of policy treatment. In their general model, they maintain the

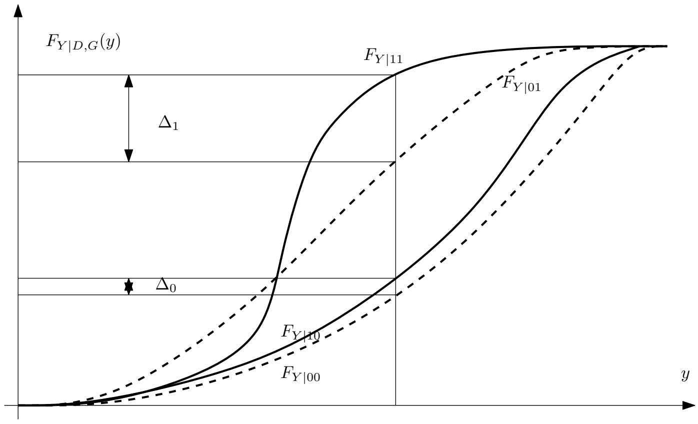


Figure B.3: The RIF-DiD design.

assumption of selection on observable and they consider that the vector of covariates can be split as $X = (G, X')$, where G is the variable defining treatment and control groups as above. To identify DiD, they suggest the following structural model for Y:

$$Y \quad = \quad Y(D, G, X', \varepsilon) = \gamma_0 + \gamma_1 D + \gamma_2 G + \alpha D \cdot G + X'\gamma_3 + \varepsilon. \qquad \text{(B.11)}$$

By assuming (B.11) and the conditional exogeneity $\varepsilon \perp D|G, X'$, the *unconditional* QTE $\alpha(p)$ for quantile p is identified by the RIF-DiD coefficient α at quantile p in (B.11). The estimator of the QTE in the RIF-DiD design (that is, of $\alpha(p)$) is denoted $\widehat{\alpha}^{RIF-DiD}(p)$. It can be obtained by following the two steps procedure explained in Firpo et al. (2009). The first step consists in estimating non parametrically the density of Y at quantile p, $\widehat{f}_Y(Q_Y(p))$. The second step consists in running a probability model $g(.)$ with the DiD structure as in (B.11), such as:

$$\Pr[Y \geq p|D, G, X] = g(\gamma_0(p) + \gamma_1(p)D + \gamma_2(p)G + \alpha(p)D \cdot G + X'\gamma_3(p) + \varepsilon).$$

The parameter $\alpha(p)$ can be visualized as the difference between the quantities Δ_1 and Δ_0 represented in figure B.3. These effects can be inverted using the empirical distribution to find QTE expressed in the same units as the variable Y: $\widehat{\alpha}^{RIF-DiD}(p)/\widehat{f}_Y(Q_Y(p))$.

The UQR has a comparative advantage over alternative models for DiD estimation: it allows to retrieve unconditional QTE even when controlling for potential selection on observables. Similarly to what happens for the other estimators, identification of the QTE in the RIF-based difference-in-differences design rests on the same weak exogeneity assumption in Athey and Imbens (2006). Moreover, asymptotic properties of the RIF-DiD estimator are not known, and in empirical studies the variances has to be obtained by bootstrapping techniques.

Sur la dissemblance et l'égalisation des chances

Résumé : Cette thèse se concentre sur la mesure des dissemblances dans la distribution des attributs économiques, et sur les implications pour l'inégalité des chances. L'égalité des chances a gagné en popularité pour définir l'objectif de la distribution d'une vaste gamme de résultats économiques entre les groupes sociaux. Cette thèse est motivée par le fait que l'évaluation des politiques publiques fondée sur l'égalité des chances s'appuie toujours sur des comparaisons de dissemblance entre des distributions conditionnelles. Nous proposons des critères empiriques pour vérifier ces comparaisons. Dans le premier chapitre, nous caractérisons axiomatiquement le pre-ordre de dissemblance permettant de classer les distributions conditionnelles au groupe d'origine, que sont définies sur des classes de résultats discrètes. Lorsque les classes sont permutables, nous démontrons que la dissemblance est rationalisée par un ordre de majorisation de matrices et mis en œuvre en vérifiant l'inclusion des Zonotopes. Lorsque les classes sont ordonnées, nous fondons le jugement de dissemblance sur un nombre fini de comparaisons au sein de la majorisation au sens de Lorenz entre les proportions des groupes, vérifiées à des étapes différentes de cumulation de la population agrégée. Dans le deuxième chapitre, on examine la pertinence du pre-ordre de dissemblance pour étudier la ségrégation au niveau individuel. On obtient une caractérisation complète d'une famille bien définie d'indicateurs de ségrégation et nous étudions l'un d'eux, l'indice d'exposition de Gini, en utilisant des données italiennes. Le dernier chapitre présente un critère d'égalisation des chances. L'égalité des chances est atteinte lorsqu'il n'y a pas de consensus, selon une classe de préférences donnée, sur l'identité du groupe défavorisé. Nous utilisons les changements de (manque de) consensus sur l'existence et l'étendue du désavantage pour caractériser le critère d'égalisation des chances. Les restrictions nécessaires, autant que des procédures possibles d'agrégations, sont également discutées. Nous démontrons que ce critère est identifié selon la classe de préférences représentées par les fonctions d'utilité dépendantes du rang, et on obtient des résultats innovants d'inférence sur la dominance stochastique inverse qui nous permettent de tester ce critère. Deux applications sur des données françaises illustrent l'impact en termes d'égalisation de chances des politiques éducatives qui ont lieu tôt dans la vie des étudiants.

Mots clés : Dissemblance, égalité de chances, ségrégation, inégalité, politiques publiques, éducation, effet du traitement par quantile, dominance stochastique inverse.

On Dissimilarity and Opportunity Equalization

Abstract: This thesis focuses on the measurement of dissimilarity in the distribution of relevant economic attributes and inequality of opportunity. Equality of opportunity has gained popularity for defining the relevant equalitarian objective for the distribution of a broad range of social and economic outcomes among social groups. I show that equality of opportunity concerns in policy evaluation always rely on *dissimilarity* comparisons between conditional distributions, and I provide empirically testable criteria to implement these comparisons. In the first chapter, I characterize axiomatically the dissimilarity partial order for discrete conditional distributions of groups across outcome classes. I prove that, when classes are permutable, dissimilarity is rationalized by matrix majorization and implemented by checking Zonotopes inclusion, while when classes are ordered the dissimilarity criterion resorts on a finite number of Lorenz majorization comparisons among groups' proportions, performed at different cumulation stages of the overall population. In the second chapter, I discuss the relevance of the dissimilarity partial order for the study of segregation at individual level. I fully characterize a well defined family of segregation indicators and I study one of them, the Gini exposure index, by using Italian data. The final chapter presents the equalization of opportunity criterion for outcome achievements. The guiding principle is that equality of opportunity is reached if there is no consensus, for a given class of preferences, in determining the disadvantaged group out of pairwise comparisons. I use the changes in (lack of) consensus on the existence and on the extent of this type of disadvantage to characterize the equalization of opportunity criterion. Meaningful restrictions and possible aggregation procedures are also discussed. I motivate that this criterion is identified within the rank dependent utility model, and I provide innovative inference results for inverse stochastic dominance to test it. Two applications on French data illustrate the equalizing impact of educational policies taking place early in students life.

Keywords: Dissimilarity, equality of opportunity, segregation, inequality, policy evaluation, education, quantile treatment effects, inverse stochastic dominance.

Discipline : Sciences économiques

Intitule et adresse de l'u.f.r. :
Université de Cergy-Pontoise
33, Boulevard du Port
95011 CERGY CEDEX