# Comparing inequality of distributions on the basis of mixed Lorenz curves

Tommaso Lando[1], Lucio Bertoli-Barsotti[2]

**Abstract**

The Lorenz dominance (LD) is generally used to rank Lorenz curves (LCs) or, equivalently, the corresponding distributions, in terms of inequality. When LCs intersect, the LD is not verified, but we can rely on weaker orders such as the upward or downward second-degree Lorenz dominance (2-LD), which emphasize the effect of the left or the right tail, respectively. The main idea of this paper it to propose a dominance relation that lies between the LD and the 2-LD, i.e. weaker than the former and stronger than the latter. For this purpose, we introduce a mixed Lorenz curve, that is, a mix of the original LC and a symmetric trasformation of it. By so doing, our approach is also intended to emphasize both tails of the distribution, rather than one. We provide an exemplification with regard to distributions of income.

**Key words**

Lorenz ordering, majorization, inequality, disparity, stochastic dominance

**JEL Classification:** D31, D63, I32

## 1. Introduction

The Lorenz curve (LC) is a primary tool for comparison of distributions of non-negative variables (e.g. random variables or statistical variables) in terms of inequality. Its use in the field of economics, for the measurement of income inequality is well known. The LC gives rise to a preorder, that is, the Lorenz dominance (LD), that is generally used to rank distributions based on their degree of inequality. In particular, according to the LD, the higher of two non-intersecting Lorenz curves shows less inequality compared to the lower one, therefore we can establish a preference relation between the two corresponding distributions. In an economic framework, the LD is coherent with the Pigou-Dalton condition, i.e. the so called "principle of transfers". According to this principle, the higher of two non-intersecting Lorenz curves can be obtained from the lower one by an iteration of income transfers from "richer" to "poorer" individuals (the so called *elementary* transfers or *T-transforms,* Marshall et al., 2009, p. 32, also called *progressive* transfers, Shorrocks and Foster, 1987). For this reason, the "coherence" with the LD represents a fundamental property for all inequality measures.

However, many empirical studies revealed that Lorenz curves often intersect in the practice, therefore it is not rare to find couples of distributions that cannot be ranked based on the LD. In such cases, we can compare the intersecting distributions by relying on weaker orders of inequality, such as the *second-degree Lorenz dominance, 2-LD* (Aaberge, 2009). The basic idea of the 2-LD is to move from the (first-degree) LD to a dominance relation of higher degree by cumulating LCs i) from the bottom or, ii), from the top. In the first case (i), this idea has been analysed in several works, related to the concept of *third-order stochastic dominance* (see e.g. Atkinson, 2008) or *third-degree inverse stochastic dominance* (Muliere and Scarsini, 1989). These orderings emphasize the left tail of the distribution. Indeed, many authors (see e.g.

[1] Tommaso Lando, University of Bergamo and VŠB-TU Ostrava, tommaso.lando@unibg.it.
[2] Lucio Bertoli-Barsotti, University of Bergamo, lucio.bertoli-barsotti@unibg.it.

Shorrocks and Foster, 1987; Dardanoni and Lambert, 1988; Atkinson, 2008) agree that an elementary transfer should be more equalizing the "lower" it occurs in the distribution, that is, the principle of *aversion to downside inequality* (Davies and Hoy, 1995). On the other hand, the second idea (ii), introduced by Muliere and Scarsini (1989) and more recently developed by Aaberge (2009), emphasizes the right tail of the distribution. This can be interesting, for instance, in an economic context, where a lot of attention is recently given to those variations occurring at the top of the income distribution (Makdissi and Yazbeck, 2014).

The basic idea of this paper is to overcome some possible (related) issues of the 2-LD. First, the 2-LD may be a too weak dominance relation for comparing distributions. In this regard, we recall that, distributions with equal values of the Gini index can be easily ranked with the 2-LD, especially if they are single-crossing (see e.g. Zoli, 1999 or Lando and Bertoli-Barsotti, 2016). Thus, 2-LD may be too restrictive, because the weight that it attaches to one of the two tails may be excessive so that it may reverse an extreme situation such as the equality of the Gini. Moreover, by choosing to emphasize the left (right) tail, the 2-LD logically downsizes the effect of those variations occurring in the right (left) one. Thus, it might be preferable to combine those two approaches into a single dominance relation that emphasizes both of the tails at the same time, as recently proposed by Lando and Bertoli-Barsotti (2016).

## 2. Preliminaries

In this section we define the LD and analyze its relation with the majorization preorder. We recall that a preorder is a binary relation $\leq$ over a set $S$ that is reflexive and transitive. In particular, observe that a preorder $\leq$ does not generally satisfy the antisymmetry property (that is, $a \leq b$ and $b \leq a$ does not necessarily imply $a = b$) and it is generally not total (that is, each pair $a, b$ in $S$ is not necessarily related by $\leq$).

Let $\mathcal{F}$ be the space of non-negative distributions $F$ with finite expectation $\mu_F$, $\mathcal{F} = \{F: F(z) = 0 \; \forall \, z < 0 \; \wedge \; \int_0^\infty z dF = \mu_F < \infty\}$. The (generalized) inverse or *quantile distribution* of a distribution function $F \in \mathcal{F}$ is given by

$$F^{-1}(p) = \inf\{z: F(z) \geq p\}, p \in (0,1) \tag{1}$$

If $F$ has finite expectation, $\mu_F$, then the *Lorenz curve* $L_F: [0,1] \to [0,1]$ is defined as follows [16]:

$$L_F(p) = \frac{1}{\mu_F} \int_0^p F^{-1}(t) dt \,, p \in (0,1). \tag{2}$$

We recall that the *Gini index* is given by twice the area between the Lorenz curve and the 45° line:

$$\Gamma(F) = 1 - 2 \int_0^1 L_F(t) dt. \tag{3}$$

Let us also define the complementary curve [13], $\bar{L}_F: [0,1] \to [0,1]$, given by

$$\bar{L}_F(p) = \frac{1}{\mu_F} \int_0^p F^{-1}(1-t) dt = 1 - L_F(1-p), p \in (0,1) \tag{4}$$

Actually, for a given percentage $p$, $L_F(p)$ represents the percentage of "total" possessed by the low $100p\%$ part of the distribution, while $\bar{L}_F(p)$ represents the percentage of "total" corresponding to the top $100p\%$ part of the distribution. From a geometrical point of view,

$\bar{L}_F(p)$, increasing and concave, is the 180° rotation of $L_F(p)$, increasing and convex, with respect to the point $(0.5,0.5)$.
In the sequel we shall also need the following curve:

$$\bar{L}_F^{-1}(p) = 1 - L_F^{-1}(1-p), \tag{5}$$

where $\bar{L}_F^{-1}$ and $L_X^{-1}$ denote, respectively, the inverse functions of $\bar{L}_F$ and $L_F$ (that are clearly invertible).

$\bar{L}_F^{-1}(p)$ represents the percentage of population that holds the top $100p\%$ part of the "total". We observe that $\bar{L}_F^{-1}$ is also a Lorenz curve, in that it is non-decreasing, convex, differentiable almost everywhere and defined on the set $[0,1]$ ($\bar{L}_F^{-1}(0) = 0, \bar{L}_F^{-1}(1) = 1$), see Sordo et al. (2014). From a geometrical point of view, the LC $\bar{L}_F^{-1}$ is symmetric with respect to the LC $L_F$, where the axis of symmetry is the line $t = 1 - p$.

# 3. Dominance relations

The Lorenz dominance (or ordering) $\leq_L$ is a pre-order defined over the space $\mathcal{F}$. It can be defined as follows.

**Definition 1**. Let $F, G \in \mathcal{F}$: we write $F \leq_L G$ if and only if $L_F(p) \geq L_G(p), \forall\, p \in (0,1)$.

When the LD is not fulfilled, i.e. when LCs intersect, we need to introduce some weaker criteria in order to obtain unambiguous rankings. Muliere and Scarsini (1989) suggest using the third-degree inverse stochastic dominance to rank intersecting Lorenz curves. While the LD compares the percentages of total (wealth) corresponding to the low $100t\%$ parts of the distributions, by using the 3-ISD an integration is performed. Hence the comparison concerns the cumulated percentages of total corresponding to the low $100t\%$ parts of the distributions. In other words, by so doing we emphasize the left tail of the distribution (i.e. lower incomes). A parallel approach consists in cumulating LCs from the right: that is, attaching more weighting to top incomes. In this paper we find it more convenient to adopt the normalized (i.e. based on the LC) version of the 3-ISD: that is, the second-degree Lorenz dominance (Aaberge, 2009), defined as follows. Note that our definition slightly differs from Aaberge's definition in that, coherently with the literature (see e.g. Marshall et al., 2009) and our definition of LD (Def. 1), we consider "dominant" the distribution that presents greater inequality.

**Definition 2**. We say that $G$ second-degree upward Lorenz dominates $F$, and write $F \leq_L^2 G$ iff:

$$\int_0^t L_F(p)dp \geq \int_0^t L_G(p)dp, \forall t \in [0,1].$$

We say that $G$ second-degree downward Lorenz dominates $F$, and write $F \leq_{\bar{L}}^2 G$ iff any of the following equivalent conditions is true:

i) $\int_0^t \bar{L}_F(p)dp \leq \int_0^t \bar{L}_G(p)dp, \forall t \in [0,1]$

ii) $\int_t^1 1 - L_F(p)dp \leq \int_t^1 1 - L_G(p)dp, \forall t \in [0,1]$

iii) $\int_0^t \bar{L}_F^{-1}(p)dp \geq \int_0^t \bar{L}_G^{-1}(p)dp, \forall t \in [0,1]$

Observe that $F \leq_L G$ implies $F \leq_L^2 G$ and $F \leq_{\bar{L}}^2 G$, but the converse is not necessarily true, i.e. $F \leq_L^2 G$ and $F \leq_{\bar{L}}^2 G$ do not imply the LD. Moreover, note that $F \leq_L^2 G$ implies that $L_F$ starts above $L_G$ and presents a larger (underlying) area: that is, a lower (or equal) value of the Gini index. Differently, $F \leq_L^2 G$ implies that $L_F$ starts below $L_G$ but still has a lower (or equal) Gini. Hence it is apparent that in both cases the condition $\int_0^1 L_F(p)dp \geq \int_0^1 L_G(p)dp$ (i.e. $\Gamma(F) \leq \Gamma(G)$) is necessary to establish a dominance. In particular, Zoli (1999) argues that if Lorenz curves cross only once, the value of the Gini index is crucial for determining the (upward) second-degree Lorenz dominance (this result can be trivially extended to the downward 2-LD).

## 4. A mixed approach

As stated in the introduction, the objective of this paper is to propose a dominance relation that emphasizes both the tails of the distributions. Moreover, we aim at defining a preorder that is finer than the LD (i.e. it can rank a major number of couples of distributions, such as those that intersect) but less restrictive, i.e. stronger, then the 2-LD. In particular, we focus on the issue of ranking single-crossing LCs. Indeed, it should be stressed that, in an economic context, single-crossing LCs occur in most practical cases, as highlighted by the empirical analyses of Atkinson (1970) and Davies and Hoy (1985), whilst multiple-crossing LCs are very rare. The formal definition of single-crossing LCs is as follows.

**Definition 3.** $L_F$ single-crosses $L_G$ from above if there exist a point $t_0$ (where $0 < t_0 < 1$) such that $L_F(t) \leq L_G(t)$ for $0 \leq t < t_0$, where the inequality is strict for some for some $0 < t' < t_0$, and $L_F(t) \geq L_G(t)$ for $t_0 \leq t < 1$, where the inequality is strict for some $t_0 \leq t'' < 1$. Similarly, $L_F$ single-crosses $L_G$ from below iff $L_G$ single-crosses $L_F$ from above.

Let us introduce a new LC, that is a compound between the LCs $L_F$ and $\bar{L}_F^{-1}$:

$$L_F^a(p) = aL_F(p) + (1-a)\bar{L}_F^{-1}(p),\ 0 \leq a \leq 1. \tag{6}$$

We denote $L_F^a$ as the *mixed Lorenz curve (MLC) of order* $a$. By construction, $L_F^a(p)$ represents a weighted average between the percentage of "total" possessed by the low $100p\%$ part of the distribution and the percentage of population that holds the top $100p\%$ part of the "total". Clearly, inequality decreases if both these percentage increase, in that inequality aversion generally corresponds to the principle that i) the low parts of the population should not be too "poor", and, symmetrically, ii) the top parts of the total should not be possessed by a too small number of "rich" people. Therefore, $L_F^a$ combines two complementary approaches into a single representation of inequality, where the number $a$ may be used as a "mixing" coefficient. For $a = 1$ we obtain $L_F^1 = L_F$, whilst, for $a = 0$, the right tail is moved to the left and vice-versa, so that $L_F^0 = \bar{L}_F^{-1}$. Hence, generally, $a$ represent the proportion that we want to use in order to mix the left and right parts of the distribution. We also observe the following properties.

1) $L_F^a$ is a Lorenz curve.
2) The LCs $L_F, \bar{L}_F^{-1}, L_F^a$ yield equal values of the Gini coefficient.
3) The LCs $L_F, \bar{L}_F^{-1}, L_F^a$ cross the line $t = 1 - p$ in the same point.
4) For $a = 0.5$, $L_F^{0.5}$ is symmetric w.r.t. the line $t = 1 - p$.

Clearly, if $L_F(p) \geq L_G(p), \forall\, p \in (0,1)$ then $L_F^a(p) \geq L_F^a(p), \forall a, p \in (0,1)$. But when the original LCs intersect it may be useful to introduce a dominance relation that is based on the integral of $L_F^a$, rather than $L_F$ or $\bar{L}_F^{-1}$. By cumulating such a "mix", we emphasize both tails at the same time (with different weights, according to $a$) and obtain a preorder that generalizes the upward and downward 2-LD. We define this preorder as the *second-degree mixed Lorenz dominance of order* $a$ (2-MLD of order $a$).

**Definition 4.** We say that $G$ second-degree Lorenz dominates $F$ with mixing order $a$ ($a \in [0,1]$), and write $F \leq_L^{2,a} G$, iff $\int_0^t L_F^a(p)dp \geq \int_0^t L_G^a(p)dp$, $\forall t \in [0,1]$

Because the cumulation of the MLC is performed from the left to the right, we may easily understand that the coefficient $a$ determines the disproportion of weights that we want to attach to the right tail compared to the left one. Moreover, we observe that the 2-MLD has the following interesting properties.

1) $F \leq_L^{2,1} G$ is equivalent to $F \leq_L^2 G$.
2) $F \leq_L^{2,0} G$ is equivalent to $F \leq_{\bar{L}}^2 G$.
3) $F \leq_L G$ implies $F \leq_L^{2,a} G, \forall a \in [0,1]$.
4) If $L_F, L_G$ cross once from above and $F \leq_L^{2,a_0} G$, then $F \leq_L^{2,a} G$ $\forall a \in [a_0, 1]$.
5) If $L_F, L_G$ cross once from below and $F \leq_L^{2,a_0} G$, then $F \leq_L^{2,a} G$ $\forall a \in [0, a_0]$.

From the properties above we may observe that the 2-MLD is a preorder between the first and the second-degree LD. Indeed, the 2-MLD is weaker than the LD but, when LCs are single crossing, from above or from below, it is stronger than the upward 2-LD or the downward LD, respectively. Moreover, the 2-MLD is progressively stronger if $a$ gets closer to 0 (if $F, G$ cross once from above) or to 1 (if $F, G$ cross once, from below). Clearly, if $F, G$ are single-crossing, it is not possible to have $F \leq_L^{2,a} G$ $\forall a \in [0,1]$ because this would imply that the LCs do not intersect. Thus, it would be interesting to find the extreme value of $a$ (that is, the minimum if the LCs cross from above, the maximum if they cross form below) because this number provides information about the "strenght" of the dominance. Indeed, $F, G$ cross once from above 1) $a$ is close to 1 means that the dominance of the second-degree is very weak because it mainly depend on the disproportion between the left and right tails, whilst 2) $a$ is close to 0 means that the dominance is stronger and almost independent from such disproportion, as it still holds even if we almost invert the weights of the tails. The concept is very similar, but inverse, if , $F, G$ cross once from below.

### 4.1 A practical exemplification

In order to better understand the basic logic of the 2-MLD, in this section we present a case study. As well known, the LD is commonly used to compare income distributions. We dowloaded the data from the *WIID*, that is, the *World Income Inequality Database*, (UNU-WIDER, 2015) which collects and stores information on income inequality for developed, developing, and transition countries. The WIID provides a comprehensive set of income inequality statistics available and can be downloaded for free. In particular, the WIID provides the deciles of income distributions, that can be used to determine the LCs, at least approximately (i.e. consisting of 10 segments). The last version of WIID, that is version 3.3, was released in 2015 and the income distributions correspond to different years and different sources (we refer to the user guide for further information). The most recent data, for most countries, are generally referred to years 2009-2010 and, in the whole database, there is no information available after year 2013. Let us focus on a particular couple of european countries, that is, France and Spain. The deciles (expressed in percentage) of the income distribution of France, for the year 2011 (source=Eurostat) are

3.5, 5.1, 6.2, 7, 7.9, 8.8, 10, 11.5, 14, 25.9;

whilst, for Spain (year=2011, source=Eurostat), we have

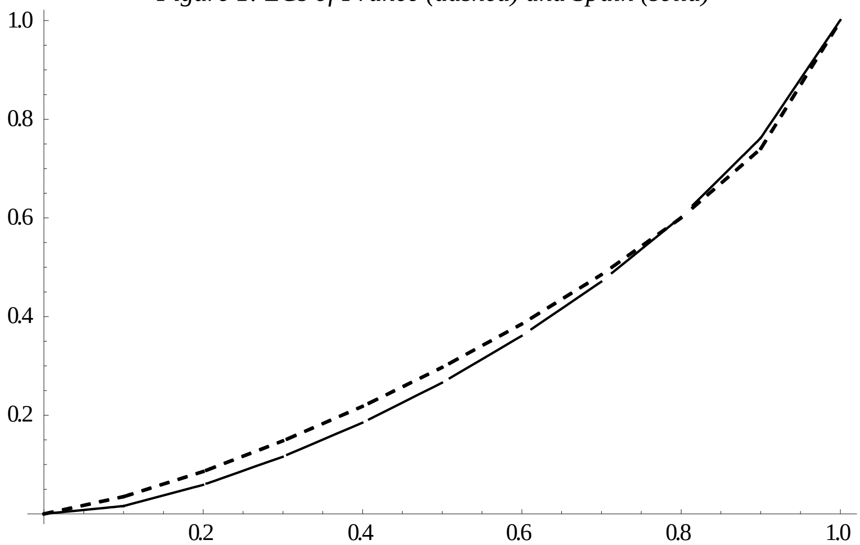
1.6, 4.3, 5.7, 6.9, 8.1, 9.5, 11, 13.1, 16, 23.9.

Observe that the LCs can be derived very easily by cumulating the deciles.

Let us denote by $F, G$, respectively, the income distributions of France and Spain. The LCs of France and Spain, represented in Figure 1, are single-crossing from above, thus the LD is not verified. In particular, Spain exhibits a greater degree of inequality in the low part of the distribution, whilst France in exhibits a greater degree of inequality in the high part. The main questions at this point are: 1) is the right tail of $F$ "heavier", or comparable to the left tail of $G$? 2) Are we "close" to the LD or the difference (in terms of inequality) between the two distributions is weak?

As the Gini index of France is higher than that of Spain (this is evident from the ares under the LCs in Fig.1), a result of Zoli (1999) guarantees that France dominates Spain w.r.t. the upward 2-LD, so the left tail of $G$ is "heavier" than the right tail of $F$. Moreover, by using the MLD, we can analyze and understand the strenght of the 2-LD dominance. After some iterations, we obtain that $F \leq_L^{2,a} G, \forall a \in [0.04,1]$, so the dominance relation is quite strong because, in order to compensate the disproportion between the inequality in the left tail of $G$ and that in the right tail of $F$, we need a mixing coefficient that is very close to 0. Put otherwise, the dominance is strong and quite independent from the tails, because it holds even if we move the 96% of the left tail to the right and vice-versa.

*Figure 1: LCs of France (dashed) and Spain (solid)*




# 5. Conclusion

We proposed a new ordering for comparing distributions in terms of inequality. Our objective was 1) to obtain an dominance relation that can be placed between the LD and the 2-LD; 2) to emphasize both tails of the distribution rather than one. Future work will be aimed at further analysing the mathematical relations beween the 2-MLD and other orders and at performing larger empirical analysis in order to verify the usefulness of our approach when dealing with real data.

## Acknowledgement

The research was supported through the Czech Science Foundation (GACR) under project 15-23699S and through SP2015/15, an SGS research project of VŠB-TU Ostrava, and furthermore by the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296.

## References

[1] Aaberge, R. (2009). Ranking intersecting Lorenz curves. *Soc. Choice Welf*. 33, pp. 235-259.

[2] Atkinson B. (1970). On the measurement of inequality. *Journal of economic theory*, 2 (3), pp. 244-263.

[3] Atkinson, A.B. (2008). More on the measurement of inequality. *J. Econ. Inequal*. 6, pp. 277-283.

[4] Dardanoni, V. and Lambert, P.J. (1988). Welfare rankings of income distributions: A role for the variance and some insights for tax reforms. Soc. Choice Welf. 5, pp. 1-17.

[5] Davies J., Hoy M.: *Comparing income distributions under aversion to downside inequality*. London: Department of Economics, University of Western Ontario (1985).

[6] Davies, J.B. and Hoy, M. (1995). Making inequality comparisons when Lorenz curves intersect. *Am. Econ. Rev*. 85, pp. 980-986.

[7] Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Econometrica* 39: pp. 1037-1039.

[8] Lando, T. and Bertoli-Barsotti, L. (2016). Weak orderings for intersecting Lorenz curves. *Metron* 7, DOI 10.1007/s40300-016-0087-6, pp.1-16.

[9] Madkissi, P. and Yakbeck, M. (2015). On the measurement of plutonomy. *Soc. Choice Welf*. 44, pp.703-717.

[10] Marshall, A.W., Olkin, I. and Arnold, B.C. (2009). *Inequalities: theory of majorization and its applications*. (second edition), Springer.

[11] Muliere P., Scarsini M. (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory*, 49 (2), pp. 314-323.

[12] Sordo, M. A., Navarro, J., & Sarabia, J. M. (2014). Distorted Lorenz curves: models and comparisons. *Social Choice and Welfare*, 42(4), pp. 761-780.

[13] Shorrocks, A.F. and Foster J.E. (1987). Transfer sensitive inequality measures. *Rev. Econ. Stud*., 14, pp. 485-497.

[14] UNU-WIDER, World Income Inequality Database (WIID3c), September 2015, <http://www.wider.unu.edu/research/Database/en_GB/wiid/>

[15] Zoli C. (1999). Intersecting generalized Lorenz curves and the Gini index. i, 16 (2), pp. 183-196.