J. Math. Anal. Appl. 514 (2022) 126335

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Regular Articles

Extreme points of Lorenz and ROC curves with applications to inequality analysis

Amparo Baíllo [a], Javier Cárcamo [b,*], Carlos Mora-Corral [a]

[a] Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain
[b] Departamento de Matemáticas, Universidad del País Vasco, Aptdo. 644, 48080 Bilbao, Spain

ARTICLE INFO

Article history:
Received 12 November 2021
Available online 16 May 2022
Submitted by S. Geiss

Keywords:
Extreme points
Gini index
Lorenz curve
Lorenz ordering
Inequality
ROC curve

ABSTRACT

We find the extreme points of the set of convex functions $\ell : [0, 1] \to [0, 1]$ with a fixed area and $\ell(0) = 0$, $\ell(1) = 1$. This collection is formed by Lorenz curves with a given value of their Gini index. The analogous set of concave functions can be viewed as Receiver Operating Characteristic (ROC) curves. These functions are extensively used in economics (inequality and risk analysis) and machine learning (evaluation of the performance of binary classifiers). We also compute the maximal L^1-distance between two Lorenz (or ROC) curves with specified Gini coefficients. This result allows us to introduce a bidimensional index to compare two of such curves, in a more informative and insightful manner than with the usual unidimensional measures considered in the literature (Gini index or area under the ROC curve). The analysis of real income microdata illustrates the practical use of this proposed index in statistical inference.

1. Introduction

Given a non-empty, compact and convex set of a locally convex space, the collection of its extreme points plays a prominent role in optimization and mathematical programming. By Bauer's maximum principle (see, e.g., Phelps [25, Proposition 16.6] or Aliprantis and Border [1, 7.69]) any convex, upper-semicontinuous functional defined on this set attains its maximum at an extreme point. In other words, we can search for maximizers of such functionals within the set of extreme points. The relevance of extreme points can be also comprehended through the Krein–Milman theorem (see, e.g., Simon [30, Theorem 8.14]), which is a central result in convex analysis. This theorem affirms that the original set is actually the closed convex hull of its extreme points. Therefore, we can retrieve the entire convex set by knowing the (usually much smaller) subset of extreme points. The practical application of these powerful results goes through the

* Corresponding author.
E-mail address: javier.carcamo@ehu.eus (J. Cárcamo).

https://doi.org/10.1016/j.jmaa.2022.126335

explicit computation of the extreme points of the convex set under study, which is usually a difficult task in infinite-dimensional spaces.

As a matter of fact, one of the aims of this work is to find the extremal points of some infinite-dimensional collections of functions. Specifically, we define

$$\mathcal{L} = \{\ell : [0,1] \to [0,1] : \ell \text{ convex and } \ell(0) = 0,\ \ell(1) = 1\}, \tag{1}$$

$$\mathcal{R} = \{r : [0,1] \to [0,1] : r \text{ concave and } r(0) = 0,\ r(1) = 1\}, \tag{2}$$

and we consider

$$\mathcal{L}_a = \{\ell \in \mathcal{L} : \|\ell\| = (1-a)/2\}, \quad a \in [0,1], \tag{3}$$

$$\mathcal{R}_a = \{r \in \mathcal{R} : \|r\| = a\}, \quad a \in [1/2,1], \tag{4}$$

where $\|\cdot\|$ is the usual norm in $L^1 = L^1([0,1])$. We will provide a detailed description of the main properties of $\mathcal{L}_a$ and $\mathcal{R}_a$, which are actually compact and convex sets in the space L^1; see Proposition 1.

Our main motivation lies in the fact that the curves in $\mathcal{L}_a$ and $\mathcal{R}_a$ appear repeatedly in many applied disciplines. On the one hand, an important interpretation of the functions in $\mathcal{L}$ is as Lorenz curves of positive and integrable random variables. In addition, $\mathcal{L}_a$ is the collection of Lorenz curves with Gini index a. In economics, Lorenz curves are extensively used to provide a graphical representation of the distributions of income or wealth in populations, while the Gini index is perhaps the most prominent inequality measure. On the other hand, $\mathcal{R}$ can be viewed as the set of Receiver Operating Characteristic (ROC) curves. ROC curves are employed to evaluate the quality of classifiers in probability forecasts (see for example Fawcett [15]) and appear in many scientific disciplines in which classification of binary outcomes is relevant (medical diagnostic, credit scoring, classification of financial transactions, and so on). The Gini coefficient of a ROC curve is defined as twice its area minus 1. However, the area under the ROC curve (AUC scoring) is used more frequently as a measure of the global accuracy of the underlying classifier; see Section 2 for details.

A second goal of this work is to quantify how "far" two of the curves in (1) (or (2)) can be from one another. Specifically, given two Lorenz (or ROC) curves with fixed Gini indices, we want to compute the maximal L^1-distance between them. In other words, for $a, b \in [0,1]$, we are interested in computing the distance between the sets $\mathcal{L}_a$ and $\mathcal{L}_b$ (or $\mathcal{R}_a$ and $\mathcal{R}_b$, for $a, b \in [0,1/2]$). This question turns out to be an infinite-dimensional convex maximization problem with two linear constraints. We will solve this optimization problem by a careful analysis of the distance between the extreme points of the considered sets. This approach also allows us to identify the maximizers of the underlying functional.

This paper is structured as follows: Section 2 reviews the main interpretations of the elements in $\mathcal{L}_a$ and $\mathcal{R}_a$. We recall the definition of the Lorenz curve and the Gini index of an integrable variable, as well as the main elements to describe ROC curves. In Section 3, we determine the set of extreme points of $\mathcal{L}_a$ and $\mathcal{R}_a$. Section 4 is devoted to the computation of the maximal L^1-distance between two of these sets. We also identify *extremal curves*, that is, functions for which this maximal distance is attained, and show their connection with stochastic orderings. As an application, in Section 5 we introduce a bidimensional inequality index to compare different characteristics of two Lorenz or ROC curves and enumerate its main properties. In the case of Lorenz curves, this new index measures simultaneously inequality and dissimilarity, while for ROC curves, it indicates accuracy and dissimilarity as well. We consider the empirical version of this inequality index to be used in practice. We also propose to combine a bootstrap approach with a non-parametric set estimation technique to obtain a confidence region for the population index. The procedure can be further applied to test simple hypotheses related to the index. These techniques have been implemented in the software R and are illustrated via the analysis of some real income microdata samples from Spain. Finally, the proofs of the main results are collected in Section 6.

2. Lorenz and ROC curves and the Gini index

In this section we recall the definitions of Lorenz and ROC curves and the Gini index, as well as their connections with the sets $\mathcal{L}_a$ (in (3)) and $\mathcal{R}_a$ (in (4)). Also, we briefly describe the scientific fields in which these concepts are widely used.

2.1. Lorenz curves and the Gini index in economics

Let X be a positive random variable with finite mean $\mu > 0$ and cumulative distribution function $F(x) = \mathrm{P}(X \leq x)$, for $x \geq 0$. The *Lorenz curve* of the variable X (or of the distribution F) is

$$\ell(t) = \frac{1}{\mu} \int_0^t F^{-1}(x)\,\mathrm{d}x, \quad 0 \leq t \leq 1, \tag{5}$$

where

$$F^{-1}(x) = \inf\{y \geq 0 : F(y) \geq x\} \tag{6}$$

$(0 < x < 1)$ is the *quantile function* of X, that is, the generalized inverse of F.

If X measures income in a population, for each value $t \in [0, 1]$, the function in (5) gives us the (normalized) total income accumulated by the proportion t of the poorest in that population. Note that F^{-1} is non-decreasing, $\mu = \int_0^1 F^{-1}(x)\,\mathrm{d}x$ and $\ell'(t) = F^{-1}(t)/\mu$ a.e. $t \in (0, 1)$. In particular, ℓ is continuous except perhaps at the point 1 and has positive second derivative ℓ'' a.e. Moreover, as the quantile function in (6) characterizes the probability distribution, ℓ determines the distribution of the underlying variable up to a (positive) scale transformation. Explicit analytic expressions for the Lorenz curves of the usual parametric distributions can be found in Kleiber and Kotz [20, Section 2.1.2].

It is easy to check that the set $\mathcal{L}$ in (1) is the closure (with respect to the pointwise convergence) of the set of Lorenz curves of positive and integrable random variables with strictly positive expectation. For simplicity, we will refer to $\mathcal{L}$ as the class of Lorenz curves. By convexity, for every $\ell \in \mathcal{L}$ it holds that

$$\ell_{\mathrm{pi}} \leq \ell \leq \ell_{\mathrm{pe}}, \tag{7}$$

where

$$\ell_{\mathrm{pe}}(t) = t \quad (0 \leq t \leq 1) \quad \text{and} \quad \ell_{\mathrm{pi}}(t) = \begin{cases} 0, & \text{if } 0 \leq t < 1, \\ 1, & \text{if } t = 1. \end{cases} \tag{8}$$

Fig. 1 shows a graphical representation of the inequalities in (7). The function ℓ_{pe} is called the *perfect equality curve* as it corresponds to the Lorenz curve of a Dirac delta measure, i.e., the probability measure corresponding to a population in which all individuals have equal (and positive) incomes. Additionally, ℓ_{pi} is the *perfect inequality curve* because it can be viewed as the limit (when the total number of individuals tends to infinity) of Lorenz curves in finite populations where only one person accumulates all the wealth. Note that the function ℓ_{pi} defined in (8) (see also Fig. 1), which is *not* a proper Lorenz curve, belongs to $\mathcal{L}$.

In practice, it is very common to synthesize the information of the Lorenz curve in a single numerical value that quantifies income inequality. Different characteristics, functionals and values of the Lorenz curve are employed to construct those inequality indices; see Arnold and Sarabia [4]. The most popular inequality measure derived from the Lorenz curve is the *Gini index*. This index has a vast number of interesting

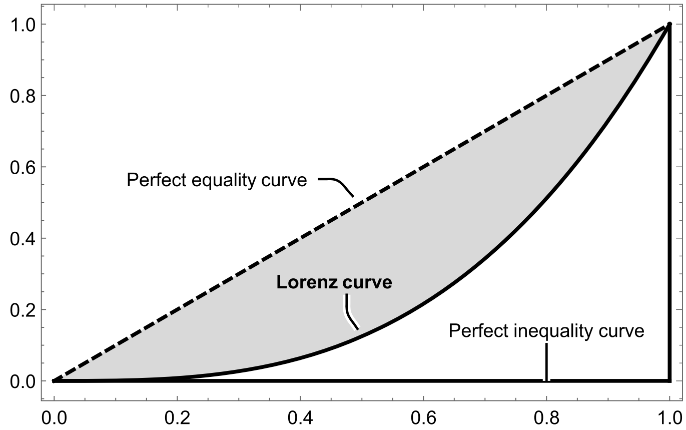

Fig. 1. A Lorenz curve together with the perfect equality and inequality curves.

interpretations and representations; see Yitzhaki and Schechtman [35, Chapter 2]. One possible way to define it is the following:

$$G(\ell) = 2\int_0^1 (t - \ell(t))\,\mathrm{d}t = 1 - 2\|\ell\|. \tag{9}$$

Therefore, we have that for $a \in [0, 1]$, the set $\mathcal{L}_a$ defined in (3) is precisely

$$\mathcal{L}_a = \{\ell \in \mathcal{L} : G(\ell) = a\}, \tag{10}$$

the collection of Lorenz curves with Gini index a.

Observe that

$$G(\ell) = \frac{\|\ell - \ell_{\mathrm{pe}}\|}{\|\ell_{\mathrm{pe}} - \ell_{\mathrm{pi}}\|}. \tag{11}$$

The denominator in (11) equals 1/2 (the maximum L^1-distance between Lorenz curves) and acts as a normalizing constant so that $0 \le G(X) \le 1$. Graphically, $G(\ell)$ is twice the shaded area in Fig. 1.

The Gini index has many convenient properties: it is scale-free (because the Lorenz curve is itself invariant under positive scaling); it can be computed whenever the considered random variable is integrable (so finite second moment is not necessary); it is normalized so that it takes values between 0 (perfect equality) and 1 (perfect inequality); it has a simple and effective interpretation (small values of this index amount to fair income distributions, whereas high values indicate unequal distributions).

Another important instrument to compare distributions according to inequality is the so-called Lorenz ordering. Let X_1 and X_2 be two variables with Lorenz curves ℓ_1 and ℓ_2, respectively. It is said that X_1 is less than or equal to X_2 in the *Lorenz order*, written $X_1 \le_{\mathrm{L}} X_2$, if $\ell_1(t) \ge \ell_2(t)$, for all $t \in [0, 1]$. In this case, we have that $\ell_{\mathrm{pe}} \ge \ell_1 \ge \ell_2$, where ℓ_{pe} is the perfect equality curve defined in (8). In other words, income is distributed in a more equitable manner in X_1 than in X_2.

2.2. ROC curves in machine learning

Binary supervised classification is one of the main statistical techniques in machine learning; see Hastie et al. [17]. In this context, we want to classify an object in one of two groups labelled 0 and 1. We observe a random vector (X, Y), where X is the predictor, usually a multidimensional (or functional) variable, and

$Y \in \{0, 1\}$ indicates the group membership. Usual procedures in machine learning combine the information in X to construct a score or marker S to predict Y. This score is typically an estimate of the posterior probability $\mathrm{P}(Y = 1|X = x)$ or some increasing function of this quantity. We can assume that members of the class $\{Y = 0\}$ have often smaller values of the score; if not, we can interchange the labels. Then, higher values of S provide stronger evidence in favour of the event $\{Y = 1\}$.

We denote by $F_i(t) = \mathrm{P}(S \le t|Y = i)$, $i = 0, 1$ and $t \in \mathbb{R}$, the conditional distribution functions of the score in the groups. Any reasonable $x \in \mathbb{R}$ can be used as a cut-off to obtain a classification rule by assigning $\{Y = 1\}$ whenever $\{S > x\}$, and $\{Y = 0\}$ if $\{S \le x\}$. In practice, the selected value for the cut-off point x typically depends on the prior probability of the groups and misclassification costs. The *probability of detection* of the classifier generated by the threshold x is

$$\mathrm{PD}(x) = \mathrm{P}(S > x|Y = 1) = 1 - F_1(x)$$

and the *probability of false detection* is

$$\mathrm{PFD}(x) = \mathrm{P}(S > x|Y = 0) = 1 - F_0(x).$$

The functions $\mathrm{PD}(x)$ and $\mathrm{PFD}(x)$ are also known as *hit rate* and *false alarm rate*, respectively.

In this context, ROC curves are commonly used to evaluate the predictive ability of binary classifiers. Formally, the ROC curve is the parametric curve in $[0, 1]^2$ given by $\{(\mathrm{PFD}(x), \mathrm{PD}(x)) : x \in \mathbb{R}\}$. If F_i, $i = 0, 1$ are continuous and strictly increasing, the ROC curve is the graph of the function

$$r(t) = 1 - F_1(F_0^{-1}(1 - t)), \quad t \in (0, 1),$$

with $r(0) = 0$ and $r(1) = 1$.

We observe that a good classifier should have a high probability of detection and low probability of false detection. Therefore, classifiers with ROC curves close to the constant 1 are preferible. A perfect classifier has the ROC curve $r(0) = 0$ and $r(x) = 1$, $x \in (0, 1]$, while a random classifier has a ROC curve on the diagonal of $[0, 1]^2$. Therefore, the ROC curve gives information about the precision of a binary classifier. Further, ROC curves are concave (see Lloyd [22]) and the area under the ROC curve (AUC) is used to evaluate the performance of the classifier. In particular, $\mathcal{R}_a$ in (4) is the set of ROC curves with AUC scoring a. At this point it should be commented that within the machine learning community ROC curves are considered convex, as they are viewed from the line $\{(x, 1) : x \in (0, 1]\}$. We follow here the usual terminology in mathematics. Finally, if two classifiers have ordered ROC curves, the above curve corresponds to the classifier that is uniformly better than the other one; that is, it gives better results for each cut-off point x that is selected to carry out the classification procedure.

3. Extreme points of $\mathcal{L}_a$ and $\mathcal{R}_a$

The sets $\mathcal{L}_a$ defined in (3) (see also (10)) and $\mathcal{R}_a$ in (4) are clearly convex. The following result asserts that they are also compact in L^1.

Proposition 1. *For each $a \in [0, 1]$ and $b \in [1/2, 1]$, the sets $\mathcal{L}_a$ and $\mathcal{R}_b$ are compact in L^1.*

As $\mathcal{L}_a$ and $\mathcal{R}_b$ are convex and compact, their extreme points acquire special relevance. We recall that an extreme of a convex set is a point that cannot be expressed as a proper convex combination of other points within the set. Formally, given a convex set C, $x \in C$ is an *extreme point* of C if $x = tx_1 + (1 - t)x_2$, for some $t \in (0, 1)$ and $x_1, x_2 \in C$, implies that $x_1 = x_2$. In the following we denote by $\mathrm{Ext}(C)$ the set of extreme points of C.

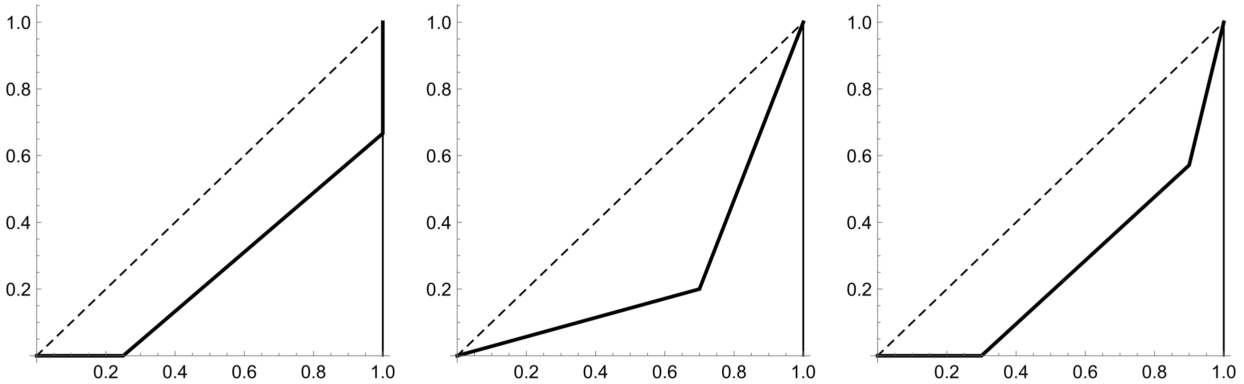

Fig. 2. The functions $\ell^a_{x_1}$, with $a = 0.5$ and $x_1 = 0.25$ (left panel), $m^a_{x_2}$, with $a = 0.5$ and $x_2 = 0.7$ (central panel) and $n^a_{x_1,x_2}$, with $a = 0.5$ and $x_1 = 0.3$, $x_2 = 0.9$ (right panel).

The next theorem determines the set of extreme points of $\mathcal{L}_a$.

Theorem 1. *For $a \in [0, 1]$, we have that*

$$\operatorname{Ext}(\mathcal{L}_a) = \left\{\ell^a_{x_1} : x_1 \in [0, a]\right\} \cup \left\{m^a_{x_2} : x_2 \in (a, 1)\right\} \cup \left\{n^a_{x_1,x_2} : x_1 \in (0, a),\ x_2 \in (a, 1)\right\},$$

where $\ell^a_{x_1}$, $m^a_{x_2}$ and $n^a_{x_1,x_2}$ are the piecewise affine functions of $\mathcal{L}_a$ such that

$$\ell^a_{x_1} : \begin{cases} 0 \mapsto 0 \\ x_1 \mapsto 0 \\ 1^- \mapsto \frac{1-a}{1-x_1}, \end{cases} \qquad m^a_{x_2} : \begin{cases} 0 \mapsto 0 \\ x_2 \mapsto x_2 - a \\ 1 \mapsto 1, \end{cases} \qquad n^a_{x_1,x_2} : \begin{cases} 0 \mapsto 0 \\ x_1 \mapsto 0 \\ x_2 \mapsto \frac{x_2-a}{1-x_1} \\ 1 \mapsto 1 \end{cases} \tag{12}$$

(with the notation $\ell^a_{x_1}(1^-) = \lim_{t\uparrow 1} \ell^a_{x_1}(t)$ and the convention $\ell^1_1 = \ell_{\text{pi}}$ in (8)).

Theorem 1 summarizes the information of $\mathcal{L}_a$ (an infinite-dimensional collection) in the set of its extreme points, which has only dimension 2. Furthermore, as $\mathcal{L}_a$ is compact in L^1, it identifies all possible maximizers of convex and continuous functionals. To prove Theorem 1 (Section 6) we first show that twice differentiation determines an affine isomorphism between $\mathcal{L}_a$ and the set of non-negative measures on $(0, 1)$ with some restrictions. Afterwards, we identify those combinations of delta measures that are extreme points.

In Fig. 2 we have depicted various extreme points of $\mathcal{L}_a$, with $a = 0.5$. The probabilistic and economic meaning of some of these Lorenz curves is described in Section 4.

Remark 1. Let us consider the map

$$T\ell(x) = 1 - \ell(1 - x), \quad \ell \in \mathcal{L}, \quad x \in [0, 1]. \tag{13}$$

Observe that, for $a \in [0, 1]$, T defines a natural bijection between $\mathcal{L}_a$ and $\mathcal{R}_{(1+a)/2}$ whose inverse is itself. Further, for $t \in [0, 1]$ and $\ell_1, \ell_2 \in \mathcal{L}$, it holds that

$$T(t\ell_1 + (1 - t)\ell_2) = tT(\ell_1) + (1 - t)T(\ell_2).$$

Therefore, T preserves extreme points and the set $\operatorname{Ext}(\mathcal{R}_a)$ is directly obtained from Theorem 1:

$$\operatorname{Ext}(\mathcal{R}_a) = \{T(\ell) : \ell \in \operatorname{Ext}(\mathcal{L}_{2a-1})\}, \quad \text{for } a \in [1/2, 1].$$

4. Maximum distance between Lorenz or ROC curves

Given two Lorenz (or ROC) curves with fixed Gini indices, in this section we quantify how "far" they can be from one another. We only consider the problem for Lorenz curves, as the corresponding question for ROC curves is analogous (see Remark 3). We also show another property of those curves for which the maximal distance is attained, which is connected with stochastic orders.

4.1. Computation of the maximum distance

We are interested in computing the value $d(\mathcal{L}_a, \mathcal{L}_b)$ $(a, b \in [0, 1])$, for a suitable metric d on $\mathcal{L} \times \mathcal{L}$, where $\mathcal{L}$ is defined in (1), and $\mathcal{L}_a$ and $\mathcal{L}_b$ as in (3) (see also (10)). Theorem 1 is extremely useful for this purpose. If d is defined through a norm, d is a convex and continuous functional on (the convex set) $\mathcal{L}_a \times \mathcal{L}_b$. Therefore, as long as $\mathcal{L}_a$ and $\mathcal{L}_b$ are compact, by Bauer's maximum principle, the supremum of d on $\mathcal{L}_a \times \mathcal{L}_b$ is attained in $\mathrm{Ext}(\mathcal{L}_a \times \mathcal{L}_b) = \mathrm{Ext}(\mathcal{L}_a) \times \mathrm{Ext}(\mathcal{L}_b)$. Thus, thanks to Theorem 1, we reduce the calculation of $d(\mathcal{L}_a, \mathcal{L}_b)$ to a finite-dimensional problem.

The exact computation of $d(\mathcal{L}_a, \mathcal{L}_b)$ will eventually depend on the particular choice of the metric d. We note that the Gini coefficient itself is defined in terms of a (normalized) L^1-distance between Lorenz curves; see formula (11). This is indeed a sensible and convenient choice to measure dissimilarities between Lorenz curves (and their associated probability distributions). The L^1 distance between Lorenz curves has also been used in Zheng [36] related to *almost stochastic dominance* of Leshno and Levy [21]. Explicitly, endow the set $\mathcal{L}$ in (1) (or, analogously, the set $\mathcal{R}$ in (2)) with the *Lorenz distance*

$$d_{\mathrm{L}}(\ell_1, \ell_2) = \frac{\|\ell_1 - \ell_2\|}{\|\ell_{\mathrm{pe}} - \ell_{\mathrm{pi}}\|} = 2\|\ell_1 - \ell_2\| = 2\int_0^1 |\ell_1 - \ell_2|, \quad \ell_1, \ell_2 \in \mathcal{L}. \tag{14}$$

Remark 2. Let us consider X_1, X_2 two positive and integrable random variables with positive expectation and Lorenz curves ℓ_1 and ℓ_2, respectively. We can define $d(X_1, X_2) = d_{\mathrm{L}}(\ell_1, \ell_2)$. We have that d is actually a *pseudo-metric* (on the space of positive and integrable random variables) because $d(X_1, X_2) = 0$ holds if and only if $X_1 =_{\mathrm{st}} cX_2$, where $c > 0$ is a constant and '$=_{\mathrm{st}}$' stands for stochastic equality.

By (7), the diameter of $\mathcal{L}$ with respect to the metric d_{L} is

$$\mathrm{diam}(\mathcal{L}) = \sup\{d_{\mathrm{L}}(\ell_1, \ell_2) : \ell_1,\, \ell_2 \in \mathcal{L}\} = d_{\mathrm{L}}(\ell_{\mathrm{pe}}, \ell_{\mathrm{pi}}) = 1.$$

We further observe that $\mathcal{L}_a$ in (10) is the set of $\ell \in \mathcal{L}$ such that $d_{\mathrm{L}}(\ell, \ell_{\mathrm{pe}}) = a$. For any fixed $a, b \in [0, 1]$, $\mathcal{L}_a$ and $\mathcal{L}_b$ are compact sets in L^1 (see Proposition 1). Therefore, from Theorem 1, the maximum

$$M(a, b) = \max\{d_{\mathrm{L}}(\ell_1, \ell_2) : \ell_1 \in \mathcal{L}_a \text{ and } \ell_2 \in \mathcal{L}_b\} \tag{15}$$

is attained at $\mathrm{Ext}(\mathcal{L}_a) \times \mathrm{Ext}(\mathcal{L}_b)$.

Note that

$$M(a, b) = \max_{\ell_1, \ell_2 \in \mathcal{L}} 2\,\|\ell_1 - \ell_2\| \quad \text{subject to} \int_0^1 \ell_1 = \frac{1-a}{2} \text{ and } \int_0^1 \ell_2 = \frac{1-b}{2}.$$

Therefore, the computation of (15), which in principle is an infinite-dimensional convex maximization problem with two linear constraints, is reduced to a finite-dimensional optimization problem.

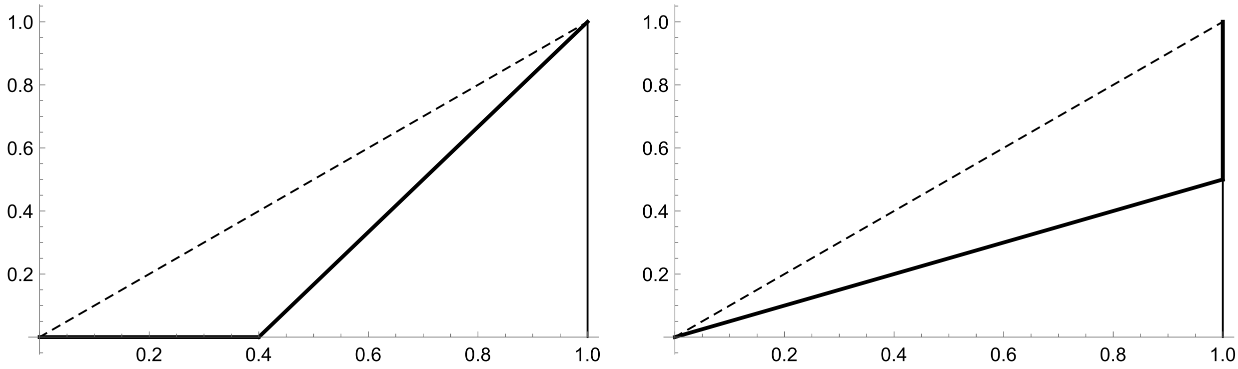

Fig. 3. The functions $\ell_{0.4}^{-}$ (left panel) and $\ell_{0.5}^{+}$ (right panel).

Definition 1. We say that the pair $(\ell_1, \ell_2) \in \mathcal{L} \times \mathcal{L}$ is *extremal* if

$$d_{\mathrm{L}}(\ell_1, \ell_2) = M(G(\ell_1), G(\ell_2)).$$

The pair of probability distributions associated to an extremal pair of Lorenz curves will also be called *extremal distributions.*

For notational convenience, we rename the functions ℓ_a^a and ℓ_0^a in (12) as ℓ_a^- and ℓ_a^+, respectively. In other words, for $0 \le a \le 1$, $\ell_a^-, \ell_a^+ \in \mathcal{L}_a$ are defined as

$$\ell_a^-(t) = \max\left\{0, \frac{t-a}{1-a}\right\} \quad \text{and} \quad \ell_a^+(t) = \begin{cases} (1-a)t, & \text{if } 0 \le t < 1, \\ 1, & \text{if } t = 1 \end{cases} \tag{16}$$

(with the agreement that $\ell_1^- \equiv \ell_{\mathrm{pi}}$ defined in (8)). These two functions will play an essential role in the rest of the section. In Fig. 3 we display two of these functions.

The following theorem, which is the main theoretical result of this section, provides an explicit expression for $M(a, b)$ and shows that this maximum distance is precisely attained at functions of the form (16). The computation of $M(a, b)$, which begins at Theorem 1 and is collected in Section 6, reveals that this issue is more delicate and complex than expected.

Theorem 2. *For $0 \le a, b \le 1$, let $M(a, b)$ be as in (15). We have that*

$$M(a, b) = \frac{(1-a)b^2 + (1-b)a^2}{a + b - ab} \tag{17}$$

(the value $M(0, 0) = 0$ is taken by continuity). Moreover, (ℓ_a^-, ℓ_b^+) and (ℓ_a^+, ℓ_b^-) are pairs of extremal Lorenz curves within the set $\mathcal{L}_a \times \mathcal{L}_b$.

Theorem 2 asserts that the maximum distance in (15) is attained at the pairs (ℓ_a^-, ℓ_b^+) and (ℓ_a^+, ℓ_b^-). Hence, the associated probability distributions (unique up to positive scale transformations) are extremal. The function ℓ_a^- is the Lorenz curve of a population in which a proportion a of the people have 0 income and the rest, a proportion $1 - a$, have equal and positive income. In other words, (up to positive scale transformations) ℓ_a^- is the Lorenz curve of a variable X_a^- with Bernoulli $(1 - a)$ distribution, that is, $\mathrm{P}(X_a^- = 0) = a$ and $\mathrm{P}(X_a^- = 1) = 1 - a$. On the other hand, ℓ_b^+ is not a proper Lorenz curve, but it can be expressed as the limit (as n goes to infinity) of Lorenz curves of populations with n individuals where $n - 1$ of them fairly share a proportion $(1 - b)$ of the wealth and there is only one “lucky person” who accumulates the rest of the total wealth (the proportion b). Hence, ℓ_b^+ can be obtained as the limit (as n goes to infinity) of the Lorenz curves of a sequence of random variables $X_b^+(n)$ taking the values $(1-b)n/(n-1)$ (with probability $1-1/n$) and bn (with probability $1/n$).

The minimum value of $d_{\mathrm{L}}(\ell_1, \ell_2)$ for $\ell_1 \in \mathcal{L}_a$ and $\ell_2 \in \mathcal{L}_b$ is easily seen to be $|b-a|$ (see Lemma 1 in Section 6). With this and Theorem 2 we can see the range of values of the distance $d_{\mathrm{L}}(\ell_1, \ell_2)$.

Corollary 1. *For $0 \leq a, b \leq 1$, we have that*

$$\{d_{\mathrm{L}}(\ell_1, \ell_2) : \ell_1 \in \mathcal{L}_a \text{ and } \ell_2 \in \mathcal{L}_b\} = [|b-a|, M(a,b)],$$

where $M(a,b)$ is given in (17).

Proof. The minimum and maximum of $d_{\mathrm{L}}(\ell, m)$ among $\ell \in \mathcal{L}_a$ and $m \in \mathcal{L}_b$ are $|b-a|$ and $M(a,b)$, as calculated in Lemma 1 and Theorem 2, respectively. Now, the set $\{d_{\mathrm{L}}(\ell, m) : \ell \in \mathcal{L}_a$ and $m \in \mathcal{L}_b\}$ is compact and connected as it is the image by the continuous function d_{L} of the compact and connected set $\mathcal{L}_a \times \mathcal{L}_b$ (in the space L^1). The result follows. □

Theorem 2 also allows us to compute the maximum distance between Lorenz curves with a given difference of their Gini indices.

Corollary 2. *For $-1 \leq c \leq 1$, let us consider*

$$M^*(c) = \max\{M(a,b) : a, b \in [0,1] \text{ and } b-a=c\}.$$

We have that

$$M^*(c) = M(a_c, a_c + c) \quad \text{with} \quad a_c = (4 - c - \sqrt{8+c^2})/2 \tag{18}$$

and

$$M^*(c) = 8 - \frac{8 + (c^2+8)^{3/2}}{c^2+4}. \tag{19}$$

Definition 2. We say that the pair $(\ell_1, \ell_2) \in \mathcal{L}^2$ is *super-extremal* if

$$d_{\mathrm{L}}(\ell_1, \ell_2) = M^*(G(\ell_1) - G(\ell_2)).$$

The associated pairs of probability distributions will be also called *super-extremal distributions*.

Obviously, each super-extremal pair is extremal because it always holds that

$$M(a,b) \leq M^*(a-b), \quad \text{for } 0 \leq a, b \leq 1.$$

However, from Theorem 2 and for any $0 \leq c \leq 1$, among all the pairs (ℓ_a^-, ℓ_{a+c}^+) and (ℓ_{a+c}^-, ℓ_a^+) (with $a \in [0, 1-c]$) of extreme Lorenz curves with a value c for the difference of their Gini indices there are only two super-extremal curves. Namely, the pairs corresponding to $a = a_c$ in (18).

Observe that $M^*(0)$ is the maximum possible distance between Lorenz curves with equal Gini indices. By Theorem 2 and Corollary 2, we have that the maximum distance between two income distributions both with Gini indices equal to a is

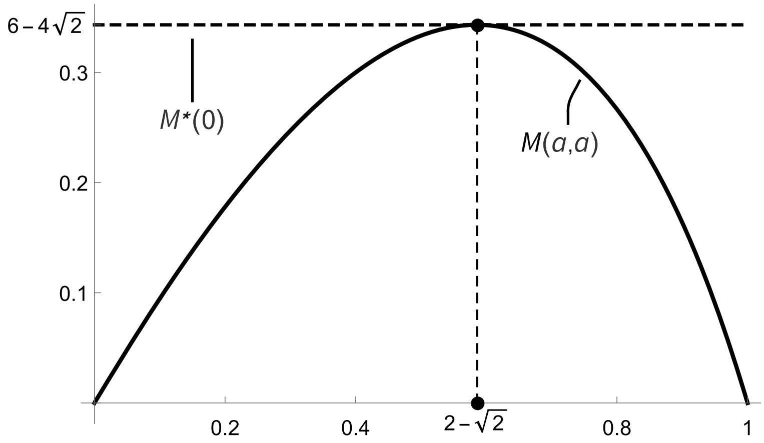

Fig. 4. The function $M(a,a)$ $(0 \leq a \leq 1)$ in (20) and its maximum value $M^*(0)$.

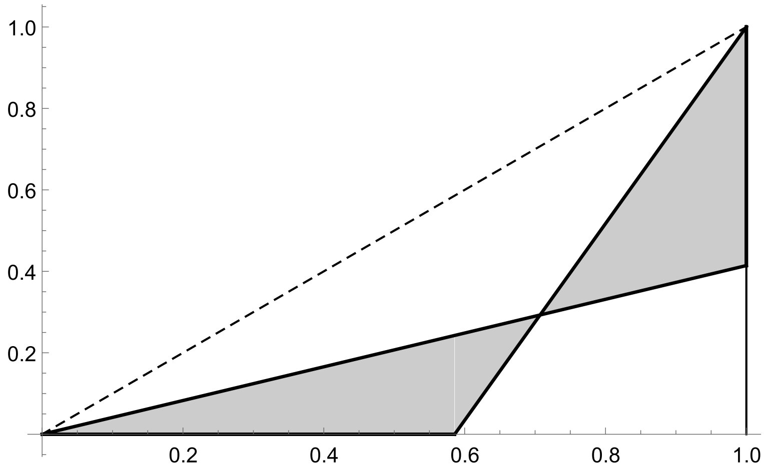

Fig. 5. Lorenz curves with equal Gini indices and maximal distance.

$$M(a,a) = \frac{2a(1-a)}{2-a}, \quad \text{for } a \in [0,1], \tag{20}$$

which attains its maximum at the point $a_0 = 2 - \sqrt{2} \approx 0.59$. Therefore,

$$M^*(0) = \max_{0 \leq a \leq 1} M(a,a) = M(a_0, a_0) = 6 - 4\sqrt{2} \approx 0.34.$$

Additionally, $M(a,a)$ is the d_{L}-diameter of $\mathcal{L}_a$. A graphical representation of $M(a,a)$ and $M^*(0)$ is presented in Fig. 4. The Lorenz curves $\ell^-_{a_0}$ and $\ell^+_{a_0}$ are hence super-extremal Lorenz curves with equal Gini indices; see Fig. 5.

Remark 3. The mapping T given in (13) is an isometry, that is, for ℓ_1, $\ell_2 \in \mathcal{L}$, it holds that

$$\|T(\ell_1) - T(\ell_2)\| = \|\ell_1 - \ell_2\|.$$

Therefore, all the results obtained in this section for Lorenz curves can be immediately translated to ROC curves.

4.2. An extremal property related to stochastic orders

The curves ℓ_a^- and ℓ_a^+ satisfy another extremal property within the class $\mathcal{L}_a$ which is related to inequality and stochastic orderings. Its proof is also postponed to Section 6.

Proposition 2. *Let $a \in [0,1]$ be fixed. For all $\ell \in \mathcal{L}_a$ and all $t \in [0,1]$, it holds that*

$$\int_0^t \ell_a^+(x)\,\mathrm{d}x \geq \int_0^t \ell(x)\,\mathrm{d}x \geq \int_0^t \ell_a^-(x)\,\mathrm{d}x. \tag{21}$$

Proposition 2 leads naturally to introduce new stochastic orders to compare random variables in terms of inequality. First, for $r \in \mathbb{N}$ and $0 \leq t \leq 1$, we define

$$\ell^{[1]}(t) = \ell(t) \quad \text{and} \quad \ell^{[r]}(t) = \int_0^t \ell^{[r-1]}(x)\,\mathrm{d}x, \quad r = 2, 3, \ldots$$

Upon integration by parts, this function can be expressed as

$$\ell^{[r]}(t) = \frac{1}{\mu\Gamma(r)} \int_0^t (t-x)^{r-1} F^{-1}(x)\,\mathrm{d}x, \quad 0 \leq t \leq 1,\ r \geq 1, \tag{22}$$

where Γ is the Euler gamma function. We note that this alternative representation of the function $\ell^{[r]}$ is valid for real values of $r \geq 1$. If the underlying variable measures income, the quantity (22) is a weighted average of the income accumulated by the proportion t of the poorest in that population. The weight function, $w_r(x) = (t-x)^{r-1}$, for $0 \leq x \leq t$ and $r > 1$, places more weight on the poorest. The parameter $r \geq 1$ modifies the form of the weight function; the higher the value of r, the greater importance is given to the lower part of the distribution.

Given two random variables X_1 and X_2 with Lorenz curves ℓ_1 and ℓ_2, respectively, we can define a stochastic order by comparing the curves $\ell_1^{[r]}$ and $\ell_2^{[r]}$. We say that X_1 is less than (or equal to) X_2 in the *Lorenz order of degree* r if $\ell_1^{[r]} \geq \ell_2^{[r]}$, and we write $X_1 \leq_{\mathrm{L}(r)} X_2$. Obviously, the case $r = 1$ corresponds to the usual Lorenz ordering. Further, if $X_1 \leq_{\mathrm{L}(r)} X_2$, then $X_1 \leq_{\mathrm{L}(r+1)} X_2$. Hence, '$\leq_{\mathrm{L}(r)}$' is a nested family of stochastic orders.

The orders '$\leq_{\mathrm{L}(r)}$' are closely related to *inverse stochastic dominance rules* in the econometric literature. Inverse stochastic orders were introduced in Muliere and Scarsini [23] to analyze social inequality and they have also been considered in different contexts; see Zoli [37], Dentcheva and Ruszczyński [9], de la Cal and Cárcamo [7] and Andreoli [3], among others. Inverse dominance is defined through the comparison of successive integrals of the quantile functions of the variables and it is easy to check that

$$X_1 \leq_{\mathrm{L}(r)} X_2 \quad \text{if and only if} \quad \frac{X_2}{\mu_2} \leq_{\mathrm{ISD}(r+1)} \frac{X_1}{\mu_1},$$

where '$\leq_{\mathrm{ISD}(r+1)}$' stands for the inverse dominance of degree $r+1$. Therefore, Lorenz orders are equivalent to inverse rules after a normalization of the variables to ensure that they have equal mean.

We observe that inequalities in (21) provide the extreme elements with respect to the Lorenz ordering of degree 2 within the set of variables with equal Gini index a. Recall the interpretation of ℓ_a^- and ℓ_a^+ explained after Theorem 2: ℓ_a^- is the Lorenz curve of the Bernoulli $(1-a)$ variable X_a^-, while ℓ_a^+ is the limit (as n goes to infinity) of the Lorenz curves of the random variables $X_a^+(n)$. In particular, from Proposition 2 it can be shown that for $n \geq 1/a$

$$X_a^+(n) \leq_{\mathrm{L}(2)} X_a \leq_{\mathrm{L}(2)} X_a^-,$$

for each random variable X_a with Gini index a. In other words, the Lorenz curve ℓ_a^+ (respectively, ℓ_a^-) is the most (respectively, least) equitable within the class $\mathcal{L}_a$ with respect to the Lorenz order of degree 2.

5. A bidimensional inequality index

In the literature there is a great amount of statistics to assess economic inequality. In addition to the ubiquitous Gini index, we can also mention the well-known *Theil*, *Hoover*, *Amato*, *Atkinson* and *generalized entropy* indices. These are only a few examples among many others, and even new measures are introduced from time to time; see Prendergast and Staudte [27] for a recent proposal. The interested reader might consult the book by Cowell [8] or Eliazar and Sokolov [13] for a panoramic overview on equality measures, their relevance and applications.

In this section we introduce a two-dimensional index for pairs of Lorenz or ROC curves that combines the Gini coefficients of two curves with the Lorenz distance defined in (14). Since the results and the underlying ideas are equivalent for Lorenz or ROC curves, we will only illustrate those concepts with Lorenz curves. The proposed index simultaneously measures relative inequality and dissimilarity between two populations. It is well known (see, e.g., De Maio [10]) that two very different income or wealth distributions can have the same Gini indices, the most popular summary of income inequality. The advantage of the new bidimensional index introduced in this section is that we can distinguish dissimilar variables yet with similar Gini indices. In addition, this index allows us to detect in a simple way the possible Lorenz ordering between two variables (see Section 2.1). In Section 5.1 we define the population version of the bidimensional inequality index and we sketch out its main properties. In Section 5.2 we propose a plug-in estimator of the new index to use in practice. In Section 5.3 we apply the estimated bidimensional index to compare income inequality in Spain from 2009 onwards with the year 2008, the onset of the biggest economic crisis of the last decades in that country. In Section 5.4 we propose to blend a standard bootstrap scheme with a set-estimation technique to construct a confidence region for the unknown population index. These ideas allow testing statistical hypotheses related to the values of the index. We illustrate the performance of these procedures with Spanish income data from 2008 and 2019.

5.1. Definition of the population index $\mathcal{I}$

Let ℓ_1 and ℓ_2 be two Lorenz curves in $\mathcal{L}$. As a measure of relative inequality we simply consider the difference of the Gini indices, that is, $G(\ell_2) - G(\ell_1)$. To quantify dissimilarity we employ the distance $d_{\mathrm{L}}(\ell_1, \ell_2)$ in (14). Therefore, a natural proposal for a new two-dimensional inequality index is the following:

$$\mathcal{I}(\ell_1, \ell_2) = (G(\ell_2) - G(\ell_1), d_{\mathrm{L}}(\ell_1, \ell_2)). \tag{23}$$

From the results in Section 4 we obtain the region of $\mathbb{R}^2$ where $\mathcal{I}$ takes values.

Proposition 3. *Let $\mathcal{I}$ be defined in (23) and let us consider the region of $\mathbb{R}^2$ defined by*

$$\Delta = \mathcal{I}(\mathcal{L} \times \mathcal{L}) \equiv \{\mathcal{I}(\ell_1, \ell_2) : \ell_1, \ell_2 \in \mathcal{L}\}.$$

We have that

$$\Delta = \{(x, y) \in [-1, 1] \times [0, 1] : |x| \leq y \leq M^*(x)\}, \tag{24}$$

where the function M^* is defined in (19).

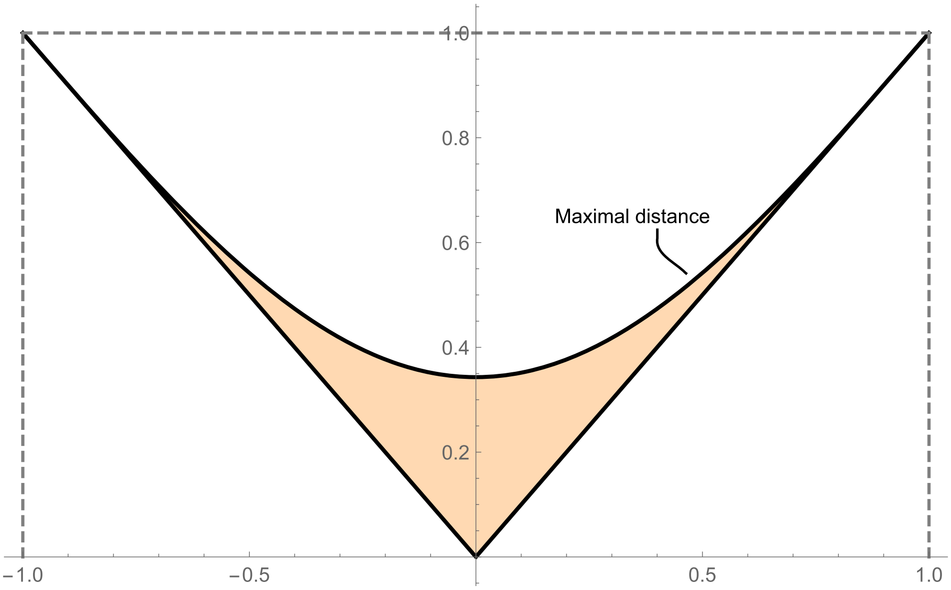

Fig. 6. The region Δ in (24).

Proof. This result follows from Corollaries 1 and 2. □

In Fig. 6 we have plotted the region Δ specified in (24). This graphical representation is very informative because we see how different the Gini indices of two variables can be in accordance with the distance between their Lorenz curves. Fig. 6 also reveals that if two distributions are very different, that is, $d_{\mathrm{L}}(\ell_1, \ell_2)$ is large, then the difference of their Gini indices cannot vary too much.

Lorenz curves and Gini indices satisfy many desirable properties that can be translated to the index $\mathcal{I}(\ell_1, \ell_2)$ in (23). The Gini index satisfies the *population principle* and the Pigou-Dalton *principle of transfers* (see Shorrocks and Foster [29]). The population principle states that if we compare a distribution that takes n values with another one taking the $2n$ values obtained by repeating twice the values of the first distribution, there should be no difference in inequality between the two distributions. This is still valid for $\mathcal{I}(\ell_1, \ell_2)$ as the Lorenz curves do no change under this replication. The next proposition provides some other properties of $\mathcal{I}(\ell_1, \ell_2)$ that are related to its bidimensional character.

Proposition 4. *Let (X_1, X_2) be a pair of random variables with Lorenz curves ℓ_1 and ℓ_2. We consider the index $\mathcal{I} = (\mathcal{I}_1, \mathcal{I}_2)$ defined in (23). The following properties hold:*

(i) $\mathcal{I}(\ell_1, \ell_2) \in \Delta$. Moreover, $\mathcal{I}_1(\ell_1, \ell_2) \geq 0$ if and only if $G(\ell_1) \leq G(\ell_2)$, whereas $\mathcal{I}_1(\ell_1, \ell_2) \leq 0$ if and only if $G(\ell_2) \leq G(\ell_1)$. In other words, positive (respectively, negatives) values of the first component of $\mathcal{I}$ indicates that ℓ_1 (respectively, ℓ_2) is fairer than ℓ_2 (respectively, ℓ_1) according to the Gini index.
(ii) Symmetry: We have that $\mathcal{I}(\ell_2, \ell_1) = (-\mathcal{I}_1(\ell_1, \ell_2), \mathcal{I}_2(\ell_1, \ell_2))$.
(iii) Extreme values (frontier of Δ):
  (1) $\mathcal{I}(\ell_1, \ell_2) \in L_1 = \{(x, x) : x \in [0, 1]\}$ if and only if $\ell_1 \geq \ell_2$, i.e., $X_1 \leq_{\mathrm{L}} X_2$.
  (2) $\mathcal{I}(\ell_1, \ell_2) \in L_2 = \{(-x, x) : x \in [0, 1]\}$ if and only if $\ell_1 \leq \ell_2$, i.e., $X_2 \leq_{\mathrm{L}} X_1$.
  (3) $\mathcal{I}(\ell_1, \ell_2) \in L_3 = \{(x, M^*(x)) : x \in [-1, 1]\}$ if and only if (ℓ_1, ℓ_2) is super-extremal, i.e., $d_{\mathrm{L}}(\ell_1, \ell_2) = M^*(G(\ell_1) - G(\ell_2))$, where M^* is given in (19).
(iv) Value at extreme points of Δ:
  (1) $\mathcal{I}(\ell_1, \ell_2) = (0, 0)$ if and only if $\ell_1 = \ell_2$. Therefore, the value at the origin means that the associated distributions satisfy that $X_1 =_{\mathrm{st}} cX_2$, where $c > 0$ is a constant.

(2) $\mathcal{I}(\ell_1, \ell_2) = (1, 1)$ *if and only if* $\ell_1 = \ell_{\text{pe}}$ *and* $\ell_2 = \ell_{\text{pi}}$.

(3) $\mathcal{I}(\ell_1, \ell_2) = (-1, 1)$ *if and only if* $\ell_1 = \ell_{\text{pi}}$ *and* $\ell_2 = \ell_{\text{pe}}$.

(v) Continuity: If $\{\ell_{1,n_1}\}_{n_1 \geq 1} \subset \mathcal{L}$ *and* $\{\ell_{2,n_2}\}_{n_2 \geq 1} \subset \mathcal{L}$ *are sequences such that* $\ell_{1,n_1} \to \ell_1$ *and* $\ell_{2,n_2} \to \ell_2$ *pointwise as* $n_1, n_2 \to \infty$*, then* $\mathcal{I}(\ell_{1,n_1}, \ell_{2,n_2}) \to \mathcal{I}(\ell_1, \ell_2)$.

5.2. Estimation of the bidimensional inequality index $\mathcal{I}$

In practice, inequality measures are computed using a sample of income or wealth microdata from each population. Here we briefly explain how to define the plug-in estimator of the index (23). Let X_1, X_2 be the two random variables of interest with distribution functions F_1 and F_2 and Lorenz curves ℓ_1 and ℓ_2, respectively. For $j = 1, 2$, we consider random samples from X_j, $\{X_{j,i}\}_{i=1}^{n_j}$, $n_j \in \mathbb{N}$. To estimate the inequality index, the starting point is the natural estimator of the distribution function of the sample. Namely, for $j = 1, 2$, we denote by $\hat{F}_j$ the *empirical distribution functions* of the samples, i.e.,

$$\hat{F}_j(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} 1_{\{X_{j,i} \leq x\}}, \quad x \in [0, \infty),$$

where 1_A stands for the indicator function of the set A. The corresponding *empirical quantile functions* are $\hat{F}_j^{-1}(x) = \inf\{y \geq 0 : \hat{F}_j(y) \geq x\}$ $(0 < x < 1)$ and the *empirical Lorenz curves* are

$$\hat{\ell}_j(t) = \frac{1}{\hat{\mu}_j} \int_0^t \hat{F}_j^{-1}(x)\, \mathrm{d}x, \quad t \in [0, 1],$$

where $\hat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{j,i}$ are the sample means. Therefore, the *plug-in estimator of the index $\mathcal{I}$* in (23) is given by

$$\hat{\mathcal{I}}(\ell_1, \ell_2) = \mathcal{I}(\hat{\ell}_1, \hat{\ell}_2). \tag{25}$$

5.3. Application to Spanish income data

To illustrate the practical performance of the bidimensional inequality index introduced in (23), we have computed its estimated value (25) using Spanish cross-sectional income microdata (at the household level) in the period 2008–2020. The data were obtained from the Living Conditions Survey (LCS) carried out by the Instituto Nacional de Estadística (INE), the administrative organism in charge of the official statistics of Spain.

The random variable X under consideration is the *equivalised disposable income*, the total disposable income of a private household divided by the equivalised household size. The total disposable income represents the total income of a household which is available for saving or spending. The equivalised household size is the number of household members converted into "equivalised" adults by the modified OECD (Organisation for Economic Co-operation and Development) equivalence scale. The equivalised disposable income is one of the variables describing income at the household level which is used by Eurostat to compute Gini coefficients and other inequality indicators.

We can use the empirical bidimensional index $\hat{\mathcal{I}}$ in (25) to carry out a cross-temporal comparison within the same country (Spain in this case). The idea is to fix a *reference year* (here 2008) and analyze the evolution of the index in that country (relative to the initial year) over a span of years (here 2009–2020). Thus, in this case, the two random variables of interest X_1 and X_2 are the Spanish income in 2008 and in a posterior year, respectively. This gives insight into the evolution of income inequality in that specific society

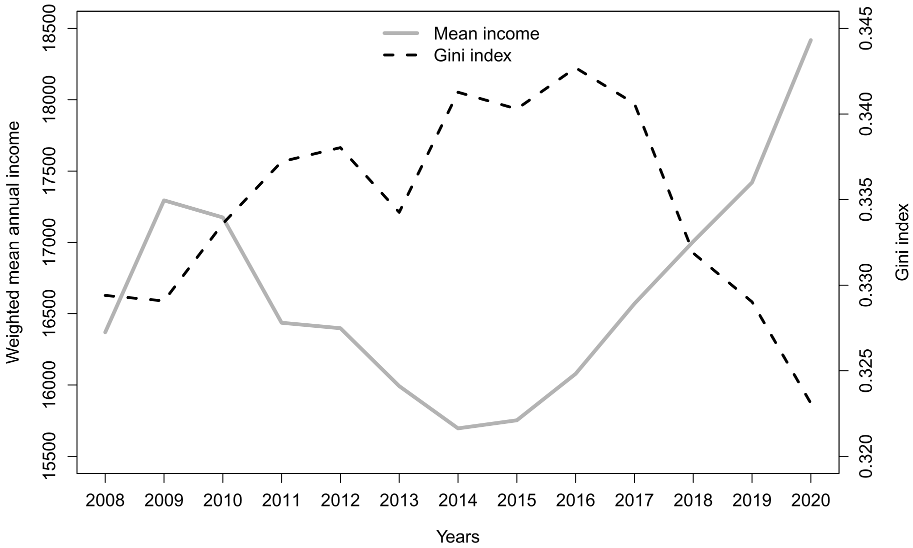

Fig. 7. Evolution of mean income and Gini coefficient in Spain (2008–2020).

along the period under study. Another possible analysis of practical interest is to take X_1 (respectively, X_2) as the equivalised disposable income in country 1 (respectively, country 2) and compute the corresponding index $\hat{\mathcal{I}}$ to compare the relative income inequality between the two countries. This relative inequality index can be computed for several years and the graphical evolution of the bidimensional $\hat{\mathcal{I}}$ along the span of years under study is easy to interpret.

First, to gain better understanding of the analyzed income data, we give some background on the period under study in Spain. Year 2008 was the onset of a severe Spanish financial and economic crisis, officially ending in 2014. There was a first recession period between 2008 and 2010 and a second one between 2011 and 2013. Fig. 7 plots the evolution of the mean equivalised disposable income and the Gini index from 2008 to 2020 and clearly reflects this abrupt crisis. Fig. 7 also shows the hard climb towards a recovery of the pre-crisis level, which took more than 9 years (IMF [18]). Although in 2017 Spanish GDP went beyond its pre-crisis peak of 2007 and many indicators reflected the recovery, it was generally agreed that the country was more unequal than in 2008 (The Economist [11]). Thus, it is interesting to analyze the evolution of inequality in Spanish society from 2008 onwards (see Blavier [5]), in particular to compare the distribution of income between 2008 (held fixed) and the following years. Microdata from INE cover up to year 2020 (included).

We have computed the empirical inequality index $\hat{\mathcal{I}}$ for the equivalised disposable income in Spain in 2008 (X_1) and in any of the years in the span 2009–2019 (X_2). The resulting indices (see Fig. 8) show the devastating effects of the crisis on the distribution of income. From 2011 to 2017 the indices of the corresponding years were either on the frontier L_1 (years 2011 and 2016) or very near it, meaning the curves were strictly ordered $\ell_1 \geq \ell_2$ (or almost so). Indeed, the curve ℓ_2 for the rest of the years from 2011 to 2017 is below ℓ_1 except for a rightmost interval contained in [0.8,1] where $\ell_1(t) < \ell_2(t)$. Income distribution in 2008 is therefore *almost* more equitable than that of the years 2012, 2013... (in the sense defined by Zheng [36]). This would support the generalized social perception that the 2008 crisis in Spain struck not only the lowest income class but also the middle class (see Alonso et al. [2]), broadening the gap between both groups and the richest (last income decile). The indices $\hat{\mathcal{I}}$ corresponding to years 2018 and 2019 are remarkably near the index of 2009, that is, income distribution slowly approached the pre-crisis level. Note that, although the Gini indices are nearly the same in 2008 and 2019, the Lorenz curves are still not coincident, as $\hat{\mathcal{I}}$ (for 2008 and 2019) is not $(0, 0)$. Indeed, in 2019 the poorest half of society had a lower proportion of the total income than in 2008.

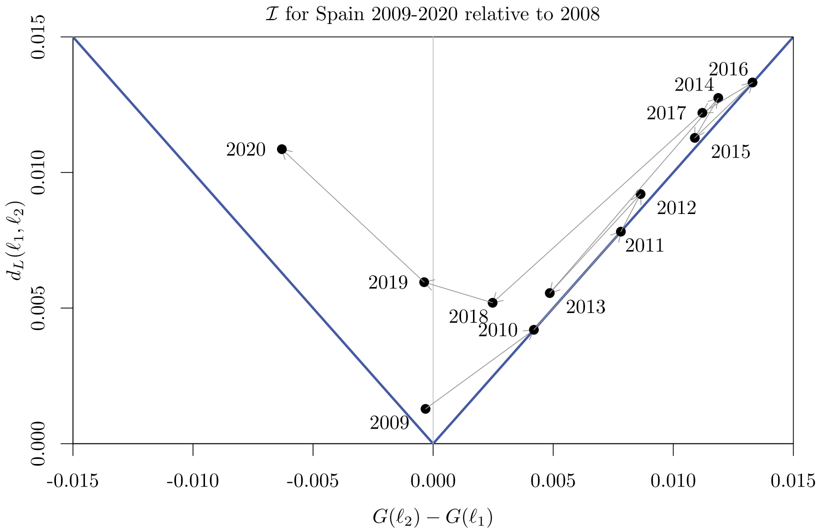

Fig. 8. The inequality index $\hat{\mathcal{I}}$ for Spanish equivalised disposable income corresponding to year 2008 (X_1) and each of the years in the span 2009–2020 (X_2).

We finally give some insight on the surprising value of $\hat{\mathcal{I}}$ for 2020, that would indicate a reduction in Spanish income inequality with respect to 2008, in spite of the economic effects of the Covid-19 pandemic on the world economy and, particularly, on the Spanish one (see, e.g., IMF [19]). As explained by the INE in a press release, the LCS collects information from a sample of households regarding their living conditions at the time of the interview (fourth trimester of the year) as well as their income in the previous year. Thus, the impact of the pandemic on the LCS-2020 is only partially reflected.

5.4. A confidence region for the inequality index $\mathcal{I}(\ell_1, \ell_2)$

In this section, we employ a standard bootstrap scheme (see Efron [12]), a widely used statistical technique based on plug-in estimation and resampling, combined with a non-parametric set estimation technique to obtain a confidence region for the bidimensional relative inequality index $\mathcal{I}$ in (23).

Let $x_{j1}, \ldots, x_{jn_j}$ denote the observations from population X_j for $j = 1, 2$. Based on the samples we construct a confidence region for the population index $\mathcal{I}(\ell_1, \ell_2)$ at the confidence level $1 - \alpha$ in the following way. First, we extract B bootstrap samples from each of the original samples $\{x_{ji}\}_{i=1}^{n_j}$, $j = 1, 2$, and we compute the corresponding bootstrap version of the empirical inequality index:

$$\begin{array}{ccccc} \text{Original samples} & & \text{Bootstrap samples} & & \text{Bootstrapped indices} \\ x_{11}, \ldots, x_{1n_1} & \longrightarrow & \left.\begin{array}{c} x_{11}^{*b}, \ldots, x_{1n_1}^{*b} \\ x_{21}^{*b}, \ldots, x_{2n_2}^{*b} \end{array}\right\} & \longrightarrow & \hat{\mathcal{I}}^{*b},\ b = 1, \ldots, B \\ x_{21}, \ldots, x_{2n_2} & \longrightarrow & & & \end{array}$$

Bootstrap usually gives good results with commonly large sample sizes available in inequality analysis and machine learning settings. After obtaining the bootstrap sample $\hat{\mathcal{I}}^{*1}, \ldots, \hat{\mathcal{I}}^{*B}$ of empirical inequality indices, we use the local convex hull (LoCoH) (also called k-nearest neighbour convex hull in the literature of set estimation, see Getz and Wilmers [16]) to construct a confidence region for $\mathcal{I}(\ell_1, \ell_2)$. The LoCoH provides results that adapt well to the bootstrap sample and to the shape of the region Δ where $\mathcal{I}$ takes values. The construction of the LoCoH is as follows. For a fixed integer $k > 0$, we construct the convex hull of each $\hat{\mathcal{I}}^{*b}$ and its $k - 1$ nearest neighbours. Then these hulls are ordered according to their area, from smallest to largest. The LoCoH is the polygonal region that results of progressively taking the union of the hulls

Table 1
Values of k and $p(k)$ for Spanish income data in 2008 and 2019 (in bold the values of k and $p(k)$ such that this proportion is closest to 95%).

k	100	200	300	400	**500**	600	700	800	900
$p(k_{0.05})$	0.923	0.932	0.935	0.936	**0.942**	0.933	0.938	0.934	0.938

from the smallest upwards, until a proportion $1-\alpha$ of $\hat{\mathcal{I}}^{*b}$'s is included in the region. We use the notation $\hat{S}(k) = \text{LoCoH}_k(\hat{\mathcal{I}}^{*1}, \ldots, \hat{\mathcal{I}}^{*B})$ to denote the resulting confidence region.

Selecting an adequate value for the number of neighbours k is a relevant matter, as it critically affects the shape of the LoCoH, but no automatized selection procedures have yet been considered in the literature. We propose to use a leave-one-out scheme to select the "optimal" k among the values in a grid $k_1, \ldots, k_M$. The idea is to determine, for each $b = 1, \ldots, B$, the LoCoH $\hat{S}^{(b)}(k) = \text{LoCoH}_k(\hat{\mathcal{I}}^{*\beta}, \beta = 1, \ldots, B, \beta \neq b)$ based on the sample of bootstrapped inequality indices from which the b-th one has been removed. Then, we compute the proportion of times that $\hat{S}^{(b)}(k)$ contains the left-out $\hat{\mathcal{I}}^{*b}$

$$p(k) = \sum_{b=1}^{B} 1_{\hat{S}^{(b)}(k)}(\hat{\mathcal{I}}^{*b})/B.$$

Our proposal for choosing the number of neighbours in the LoCoH is

$$k_\alpha = \underset{k \in \{k_1, \ldots, k_M\}}{\arg\min} |p(k) - (1-\alpha)|. \tag{26}$$

We use the procedure described above to compute a bootstrap confidence region for the inequality index $\mathcal{I}$ corresponding to two of the samples considered in Section 5.3. We have chosen the equivalised disposable income of Spain in 2008 ($n_1 = 12987$) and in 2019 ($n_2 = 15861$). The Gini index in this case takes the value 32.9% for both years, and the empirical bidimensional inequality index is $\hat{\mathcal{I}} = (-10^{-4}, 0.006)$. As the components of the index are not both close to the origin (0,0), we can conclude that the distribution of income was not exactly the same in 2008 and 2019. Indeed, this is confirmed by the two empirical Lorenz curves and their difference (re-scaled by the maximum absolute difference to improve the visualization) plotted in Fig. 9: in 2019 the poorest half of the Spanish population had a smaller cumulative proportion of income than in 2008. Regarding the proximity of the two Lorenz curves in Fig. 9(a) and the low values of the two components of $\hat{\mathcal{I}}$, let us note that Lorenz curves for the income of the same country in two different years are never radically different.

We have extracted $B = 1000$ bootstrap samples from each of the data sets corresponding to 2008 and 2019 and computed the resulting empirical indices $\hat{\mathcal{I}}^{*b}$, $b = 1, \ldots, 1000$. We have determined the bootstrap LoCoH confidence region $\hat{S}(k_{0.95})$, at the confidence level of 95%, for the relative inequality index $\mathcal{I}$ comparing 2008 and 2019. The value of $k_{0.95}$ was selected from the grid (100, 200, ..., 900) via the leave-one-out procedure described before and according to (26) with $1-\alpha = 0.95$. Table 1 displays the proportion $p(k)$ of times that the left-out index $\hat{\mathcal{I}}^{*b}$ is contained in the LoCoH $\hat{S}^{(b)}(k)$ constructed with the remaining 999 indices. The number $k_{0.95} = 500$ of nearest neighbours yielding the proportion $p(k)$ nearest to 0.95 results in the LoCoH $\hat{S}(500)$ of Fig. 10. Remark the contrast with Fig. 11, where we have plotted the bootstrap inequality indices corresponding to years 2008 and 2016 in Spain and the confidence region constructed using the LoCoH with $k_{0.95} = 600$.

5.5. Hypothesis tests for the inequality index $\mathcal{I}(\ell_1, \ell_2)$

Once determined a confidence region for $\mathcal{I}(\ell_1, \ell_2)$ at a level $(1-\alpha)$ (as explained in the previous section), we can use it to carry out the statistical test (with significance level α)

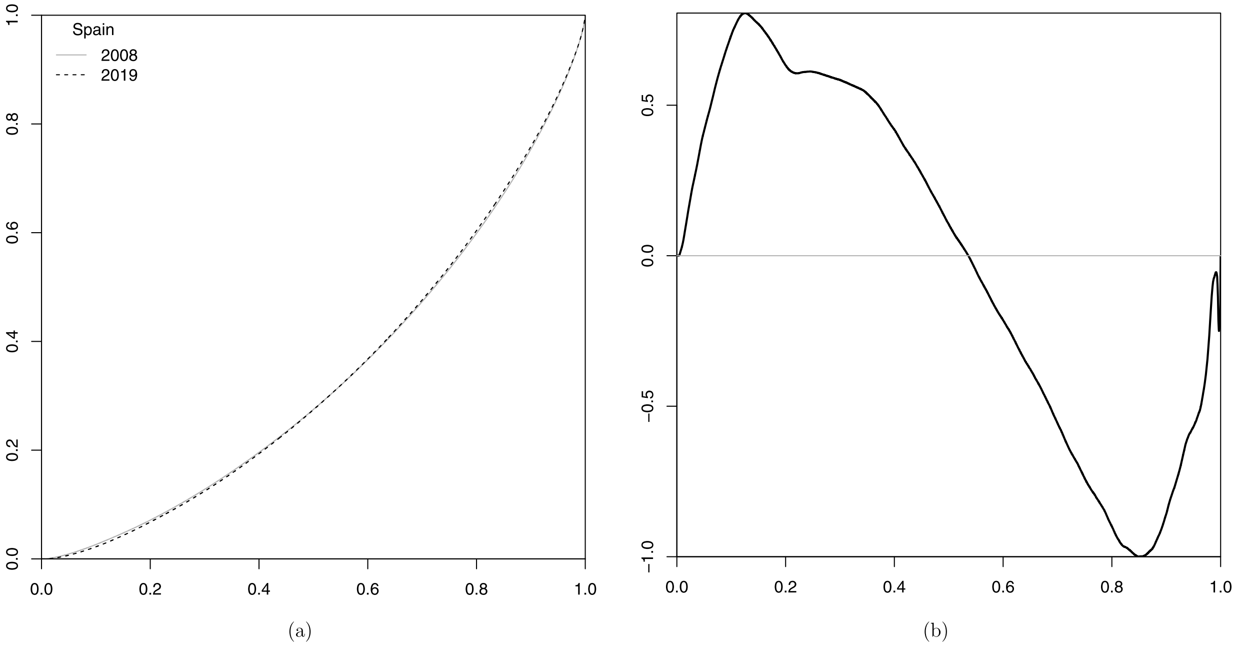

Fig. 9. (a) Empirical Lorenz curves, $\hat{\ell}_1$ and $\hat{\ell}_2$, for the equivalised household income in Spain in 2008 and 2019 respectively; (b) Difference $\hat{\ell}_1 - \hat{\ell}_2$ re-scaled by $\|\hat{\ell}_1 - \hat{\ell}_2\|_\infty$.

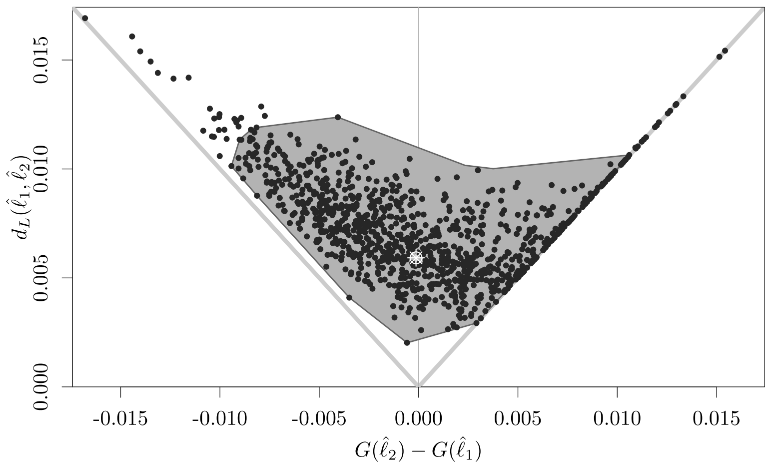

Fig. 10. Confidence region for the bidimensional relative inequality index $\mathcal{I}$ of Spain in 2008 and 2019. The crossed point in white is the empirical index $\hat{\mathcal{I}}$.

$$H_0 : \mathcal{I}(\ell_1, \ell_2) = \mathcal{I}_0 \quad \text{versus} \quad H_1 : \mathcal{I}(\ell_1, \ell_2) \neq \mathcal{I}_0,$$

where $\mathcal{I}_0$ is a known fixed value in the region Δ given in (24). The procedure to test the simple hypothesis $H_0 : \mathcal{I} = \mathcal{I}_0$ follows by the duality between confidence regions and hypothesis tests: we reject H_0 at level α whenever I_0 does not belong to the confidence region for $\mathcal{I}(\ell_1, \ell_2)$ at a level $(1 - \alpha)$.

For instance, the confidence region at level 95% in Fig. 10 does not intersect the left diagonal L_2, which means that there is evidence to reject that the relative inequality index $\mathcal{I}$ comparing 2008 and 2019 lies on any point of the left diagonal. In other words, we can affirm (with significance level 5%) that 2019 did not distribute income more fairly than 2008. However, for the same years, the confidence region $\hat{S}(500)$ intersects the right diagonal L_1, so we cannot reject that the 2019 Lorenz curve is below that of 2008. The situation is even more clear for the years 2008 and 2016 (see Fig. 11).

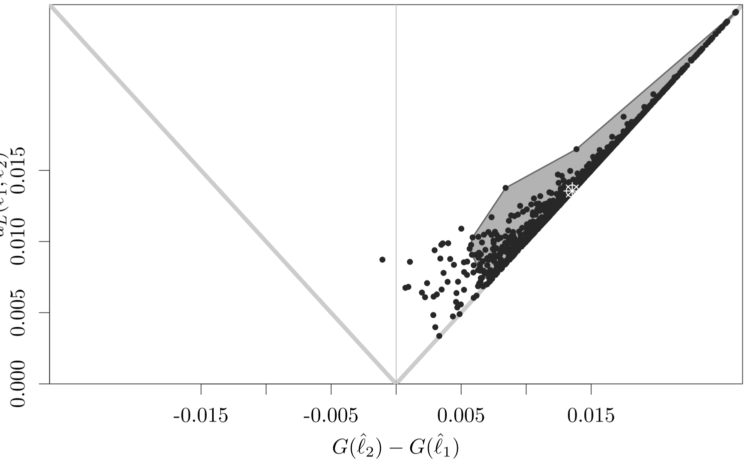

Fig. 11. Confidence region for the bidimensional relative inequality index $\mathcal{I}$ of Spain in 2008 and 2016. The crossed point in white is the empirical index $\hat{\mathcal{I}}$.

6. Proofs: extremal points and maximum distance

Here we collect the proofs of Proposition 1 and Theorems 1 and 2. First, we enumerate some regularity properties of the functions in the set $\mathcal{L}$ defined in (1) that will be useful throughout this section.

6.1. Regularity properties and compactness of $\mathcal{L}_a$

We start with a slight change in the definition of the functions in the set $\mathcal{L}$ of (1). Given $\ell \in \mathcal{L}$ we redefine the value of ℓ at 1 as $\ell(1) = \sup_{[0,1)} \ell$. This redefinition is motivated by the fact that, as shown in the following proposition, functions in $\mathcal{L}$ become continuous in $[0,1]$. In addition, the convexity of $\mathcal{L}$ and $\mathcal{L}_a$ remains true. With this definition, $\mathcal{L}$ becomes the set of convex $\ell : [0,1] \to [0,1]$ such that $\ell(0) = 0$ and $\ell(1) = \sup_{[0,1)} \ell$.

In the following proposition we denote by $W^{1,1}(0,1)$ the Sobolev space $W^{1,1}$ in the interval $(0,1)$, which is equivalent to the set of absolutely continuous functions in $[0,1]$; it is endowed with the norm

$$\|\ell\|_{W^{1,1}(0,1)} = \|\ell\| + \|\ell'\| ,$$

where ℓ' is the distributional derivative of ℓ, which coincides a.e. with the derivative of ℓ. For $\alpha \in (0,1)$ we denote by $W^{1,\infty}(0,\alpha)$ the Sobolev space $W^{1,\infty}$ in the interval $(0,\alpha)$, which is equivalent to the set of Lipschitz continuous functions in $[0,\alpha]$; it is endowed with the norm

$$\|\ell\|_{W^{1,\infty}(0,\alpha)} = \sup_{(0,\alpha)} |\ell| + \operatorname*{ess\,sup}_{(0,\alpha)} |\ell'| .$$

See, e.g., Brezis [6, Chapter 8] or Evans and Gariepy [14, Chapter 4] for the definition and properties of these spaces.

Proposition 5. *Let $a \in [0,1]$ and $\ell \in \mathcal{L}_a$. Then*

(a) The function ℓ is non-decreasing, absolutely continuous in $[0,1]$ and Lipschitz in $[0,\alpha]$ for each $\alpha \in (0,1)$. Moreover,

$$\|\ell\|_{W^{1,1}(0,1)} \le 1 + \frac{1-a}{2},$$

and for each $\alpha \in (0,1)$,

$$\|\ell\|_{W^{1,\infty}(0,\alpha)} \leq 1 + \frac{1-a}{(1-\alpha)^2}.$$

(b) The function ℓ' *is locally of bounded variation, the right derivative* $\ell'(x^+)$ *exists for all* $x \in [0,1)$ *and is non-decreasing. Moreover,* $\ell'(0^+) \geq 0$.

(c) ℓ'' *is a non-negative Radon measure.*

Proof. Convex functions are locally Lipschitz (see Evans and Gariepy [14, Theorem 6.3.1] or Simon [30, Theorem 1.19]), have a first derivative locally of bounded variation (see Evans and Gariepy [14, Theorem 6.3.3]) and have a second derivative in the sense of distributions (see Evans and Gariepy [14, Theorem 6.3.2] or Simon [30, Theorem 1.29]), which in fact is a non-negative Radon measure. Further, the right derivative $\ell'(x^+)$, exists for all $x \in [0,1)$ and is non-decreasing (see Simon [30, Theorem 1.26]). As $\ell(0) = 0$ and $\ell \geq 0$, we necessarily have that $\ell'(0^+) \geq 0$, so $\ell'(x^+) \geq 0$ for all $x \in [0,1)$. By the version of the fundamental theorem of calculus for convex functions (see Simon [30, Theorem 1.28]), ℓ is non-decreasing. In addition, the derivative of ℓ exists a.e. and coincides a.e. with the right derivative, so $\ell' \geq 0$ a.e. In particular,

$$\|\ell'\| = \int_0^1 \ell'(t)\,\mathrm{d}t = \ell(1) - \ell(0) \leq 1.$$

We conclude that $\|\ell\|_{W^{1,1}(0,1)} \leq 1 + \frac{1-a}{2}$.

On the other hand, we observe that the affine function $s : [0,1] \to \mathbb{R}$ given by $s(x) = \ell'(\alpha^+)(x-\alpha) + \ell(\alpha)$ is a supporting line of ℓ at the point $(\alpha, \ell(\alpha))$. By convexity, we hence have that $s \leq \ell$ and then,

$$\frac{1-a}{2} = \|\ell\| \geq \int_\alpha^1 \ell(t)\,\mathrm{d}t \geq \int_\alpha^1 s(t)\,\mathrm{d}t = \ell'(\alpha^+)\frac{(1-\alpha)^2}{2} + \ell(\alpha)(1-\alpha) \geq \ell'(\alpha^+)\frac{(1-\alpha)^2}{2}.$$

We conclude that

$$\operatorname*{ess\,sup}_{(0,\alpha)} \ell' \leq \ell'(\alpha^+) \leq \frac{1-a}{(1-\alpha)^2}.$$

Consequently, $\|\ell\|_{W^{1,\infty}(0,\alpha)} \leq 1 + \frac{1-a}{(1-\alpha)^2}$. □

The set $\mathcal{L}_a$ is clearly convex. We are now ready to prove Proposition 1.

Proof of Proposition 1 in Section 3 (Compactness of $\mathcal{L}_a$ and $\mathcal{R}_b$). Let $\{\ell_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}_a$. By Proposition 5, $\{\ell_n\}_{n\in\mathbb{N}}$ is bounded in $W^{1,1}(0,1)$, so by the Rellich–Kondrachov theorem (see Brezis [6, Theorem 8.8]), there exists a subsequence (not relabelled) and an $\ell \in L^1$ such that $\ell_n \to \ell$ in L^1 as $n \to \infty$. This also implies that $G(\ell) = a$.

On the other hand, for each $\alpha \in (0,1)$ we have by Proposition 5 (a) that $\{\ell_n\}_{n\in\mathbb{N}}$ is bounded in $W^{1,\infty}(0,\alpha)$, so by the Ascoli–Arzelà theorem (see Brezis [6, Theorems 4.25 and 8.8]), for a further subsequence, $\ell_n \to \ell$ uniformly in $[0,\alpha]$ as $n \to \infty$. In particular, $\ell(0) = 0$ and $0 \leq \ell \leq 1$ in $[0,\alpha]$. As the pointwise limit of convex function is a convex function, we obtain that ℓ is convex in $[0,\alpha]$. Therefore, $0 \leq \ell \leq 1$ in $[0,1)$ and ℓ is convex in $[0,1)$. We redefine $\ell(1)$ as $\ell(1) = \sup_{[0,1)}$, so that ℓ becomes continuous in $[0,1]$. We also obtain that $0 \leq \ell \leq 1$ in $[0,1]$ and ℓ is convex in $[0,1]$. Therefore, $\ell \in \mathcal{L}_a$ and the proof is finished. □

6.2. Proof of Theorem 1 in Section 3 (extreme points of $\mathcal{L}_a$)

We will use an alternative description of the elements in $\mathcal{L}_a$ in terms of positive measures concentrated on the interval $(0,1)$. The main idea is based on the following fact: any curve $\ell \in \mathcal{L}_a$ is univocally determined by its second derivative, ℓ'', together with the conditions $\ell(0) = 0$ and $G(\ell) = a$ (or, equivalently, $\|\ell\| = \frac{1-a}{2}$).

Given $a \in [0,1]$, we denote by $\mathcal{M}_a$ the set of non-negative Radon measures μ concentrated on the interval $(0,1)$ and such that

$$\int_0^1 (1-s)^2 \,\mathrm{d}\mu(s) \leq 1-a \quad \text{and} \quad \int_0^1 s(1-s)\,\mathrm{d}\mu(s) \leq a. \tag{27}$$

Proposition 6. *For $a \in [0,1]$, the map $T_a : \mathcal{L}_a \to \mathcal{M}_a$ defined by $T_a(\ell) = \ell''$ is an affine isomorphism with inverse $T_a^{-1} : \mathcal{M}_a \to \mathcal{L}_a$ given by*

$$T_a^{-1}\mu(x) = \left[1 - a - \int_0^1 (1-s)^2 \,\mathrm{d}\mu(s)\right] x + \int_0^x (x-s)\,\mathrm{d}\mu(s), \qquad x \in [0,1]. \tag{28}$$

Proof. First we see that the map T_a is well defined. Given $\ell \in \mathcal{L}_a$, we have from Proposition 5 that ℓ'' is a non-negative Radon measure. As ℓ' is locally of bounded variation,

$$\ell'(t) = \ell'(0^+) + \int_0^t \mathrm{d}\ell''(s), \qquad \text{a.e. } t \in (0,1).$$

As ℓ is locally Lipschitz and $\ell(0) = 0$, for $x \in [0,1]$, we have that

$$\begin{aligned} \ell(x) &= \int_0^x \ell'(t)\,\mathrm{d}t = \int_0^x \left[\ell'(0^+) + \int_0^t \mathrm{d}\ell''(s)\right] \mathrm{d}t \\ &= \ell'(0^+)x + \int_0^x (x-s)\,\mathrm{d}\ell''(s), \end{aligned} \tag{29}$$

where for the last equality we have used Fubini's theorem. Integrating in $x \in (0,1)$ equality (29) (and by Fubini's theorem again) we obtain the restriction

$$\frac{1-a}{2} = \|\ell\| = \frac{\ell'(0^+)}{2} + \frac{1}{2}\int_0^1 (1-s)^2 \,\mathrm{d}\ell''(s). \tag{30}$$

As $\ell'(0^+) \geq 0$, from (30) we directly obtain the first inequality of (27). On the other hand, (29) and (30) show that

$$\ell(x) = \left[1 - a - \int_0^1 (1-s)^2 \,\mathrm{d}\ell''(s)\right] x + \int_0^x (x-s)\,\mathrm{d}\ell''(s), \quad x \in [0,1]. \tag{31}$$

Hence, imposing $\ell(1) \leq 1$ we have the second inequality of (27).

Now, for $\mu \in \mathcal{M}_a$, we define ℓ as in the right-hand side of (28) and we will check that $\ell \in \mathcal{L}_a$. First, $\ell(0) = 0$ and

$$\int_0^1 \ell(x)\,\mathrm{d}x = \frac{1}{2}\left[1-a-\int_0^1 (1-s)^2\,\mathrm{d}\mu(s)\right] + \int_0^1\int_s^1 (x-s)\,\mathrm{d}x\,\mathrm{d}\mu(s) = \frac{1-a}{2}.$$

Further, thanks to the second inequality of (27), we obtain that

$$\ell(1) = 1-a+\int_0^1 s(1-s)\,\mathrm{d}\mu(s) \le 1.$$

By Leibniz integral rule (differentiation under the integral sign), it can also be checked that

$$\ell'(x) = 1-a-\int_0^1 (1-s)^2\,\mathrm{d}\mu(s) + \int_0^x \mathrm{d}\mu(s), \quad \text{a.e. } x\in(0,1) \tag{32}$$

and hence

$$\ell'' = \mu \quad \text{as measures.} \tag{33}$$

As μ is positive, from (32) we have that ℓ' is essentially non-decreasing, ℓ is convex and

$$\ell'(x) \ge 1-a-\int_0^1 (1-s)^2\,\mathrm{d}\mu(s) \ge 0, \quad \text{a.e. } x\in(0,1),$$

by the first inequality of (27), so ℓ is non-decreasing. In particular, $\ell \ge 0$. This shows that $\ell \in \mathcal{L}_a$.

Finally, we prove that the maps T_a and (28) are mutually inverse. Given $\ell \in \mathcal{L}_a$, if we apply first T_a and then (28) we get ℓ back thanks to (31). Conversely, given $\mu \in \mathcal{M}_a$, if we apply first (28) and then T_a we recover μ by (33). Since T_a is affine, the proof is concluded. □

Next we calculate $\mathrm{Ext}(\mathcal{M}_a)$. We denote by δ_x the Dirac measure at $x\in[0,1]$.

Proposition 7. *For $a\in[0,1]$, we have that*

$$\begin{aligned}\mathrm{Ext}(\mathcal{M}_a) = &\{0\} \cup \left\{\frac{1-a}{(1-x_1)^2}\delta_{x_1} : x_1\in(0,a]\right\} \cup \left\{\frac{a}{x_2(1-x_2)}\delta_{x_2} : x_2\in(a,1)\right\}\\ &\cup \left\{\frac{x_2-a}{(1-x_1)(x_2-x_1)}\delta_{x_1} + \frac{a-x_1}{(x_2-x_1)(1-x_2)}\delta_{x_2} : x_1\in(0,a),\ x_2\in(a,1)\right\}.\end{aligned}$$

Proof. The proof is divided into several smaller results.

Step 1: The null measure $\mu\equiv 0 \in \mathrm{Ext}(\mathcal{M}_a)$.

This is direct as all the measures in $\mathcal{M}_a$ are non-negative.

Step 2: For all $x_1\in(0,a]$, the measure $\mu = \frac{1-a}{(1-x_1)^2}\delta_{x_1} \in \mathrm{Ext}(\mathcal{M}_a)$.

Clearly, $\mu\in\mathcal{M}_a$. Assume that $\mu = t_1\mu_1 + t_2\mu_2$, for some $t_1,t_2>0$ with $t_1+t_2=1$ and $\mu_1,\mu_2\in\mathcal{M}_a$. Then $\mu_i = \beta_i\delta_{x_1}$ and, due to (27),

$$\beta_i \ge 0, \quad \beta_i(1-x_1)^2 \le 1-a, \quad \beta_i x_1(1-x_1) \le a, \qquad i=1,2.$$

Thus,

$$\frac{1-a}{(1-x_1)^2} = t_1\beta_1 + t_2\beta_2 \le t_1\frac{1-a}{(1-x_1)^2} + t_2\frac{1-a}{(1-x_1)^2} = \frac{1-a}{(1-x_1)^2},$$

so, necessarily, $\beta_1 = \beta_2 = \frac{1-a}{(1-x_1)^2}$, and, hence, $\mu_1 = \mu_2$. Therefore, $\mu \in \mathrm{Ext}(\mathcal{M}_a)$.

Step 3: For all $x_2 \in (a, 1)$, *the measure* $\frac{a}{x_2(1-x_2)}\delta_{x_2} \in \mathrm{Ext}(\mathcal{M}_a)$.

The proof is similar to the one of the previous step and it is therefore omitted.

Step 4: If for some $x_1 \in (0, a]$ *and* $\alpha \in \mathbb{R} \setminus \{0, \frac{1-a}{(1-x_1)^2}\}$, $\mu = \alpha\delta_{x_1} \in \mathcal{M}_a$, *then* $\mu \notin \mathrm{Ext}(\mathcal{M}_a)$.

The fact $\mu \in \mathcal{M}_a$ implies that $0 < \alpha < \frac{1-a}{(1-x_1)^2}$. Therefore, for $\varepsilon > 0$ small enough, we have that $\mu \pm \varepsilon\delta_{x_1} \in \mathcal{M}_a$ since both are positive measures and the restrictions (27) are satisfied; indeed,

$$\int_0^1 (1-s)^2 \,\mathrm{d}(\mu \pm \varepsilon\delta_{x_1})(s) = (\alpha \pm \varepsilon)(1-x_1)^2 < 1-a$$

and

$$\int_0^1 s(1-s)\,\mathrm{d}(\mu \pm \varepsilon\delta_{x_1}) = (\alpha \pm \varepsilon)x_1(1-x_1) < \frac{1-a}{1-x_1}x_1 \le a.$$

Finally, we can write $\mu = \frac{1}{2}(\mu + \varepsilon\delta_{x_1}) + \frac{1}{2}(\mu - \varepsilon\delta_{x_1})$, and, hence, $\mu \notin \mathrm{Ext}(\mathcal{M}_a)$.

Step 5: If for some $x_2 \in (a, 1)$ *and* $\alpha \in \mathbb{R} \setminus \{0, \frac{a}{x_2(1-x_2)}\}$, $\mu = \alpha\delta_{x_2} \in \mathcal{M}_a$, *then* $\mu \notin \mathrm{Ext}(\mathcal{M}_a)$.

The proof is similar to that of Step 4 and it is left to the reader.

Step 6: For all $x_1 \in (0, a)$ *and* $x_2 \in (a, 1)$, $\mu = \frac{x_2-a}{(1-x_1)(x_2-x_1)}\delta_{x_1} + \frac{a-x_1}{(x_2-x_1)(1-x_2)}\delta_{x_2} \in \mathrm{Ext}(\mathcal{M}_a)$.

It is immediate to check that

$$\int_0^1 (1-s)^2 \,\mathrm{d}\mu(s) = 1-a \quad \text{and} \quad \int_0^1 s(1-s)\,\mathrm{d}\mu(s) = a.$$

Therefore, $\mu \in \mathcal{M}_a$. Moreover, if $\mu = t_1\mu_1 + t_2\mu_2$ for some $t_1, t_2 > 0$ with $t_1 + t_2 = 1$ and $\mu_1, \mu_2 \in \mathcal{M}_a$, then

$$\int_0^1 (1-s)^2 \,\mathrm{d}\mu_i(s) = 1-a \quad \text{and} \quad \int_0^1 s(1-s)\,\mathrm{d}\mu_i(s) = a, \qquad i = 1, 2.$$

Furthermore, as $\mu_i = \sum_{j=1}^2 \beta_{ij}\delta_{x_j}$ for some $\beta_{ij} \ge 0$ (for $i, j = 1, 2$), we have that

$$\sum_{j=1}^2 \beta_{ij}(1-x_j)^2 = 1-a \quad \text{and} \quad \sum_{j=1}^2 \beta_{ij}x_j(1-x_j) = a, \qquad i = 1, 2,$$

and, hence,

$$\beta_{i1} = \frac{x_2 - a}{(1 - x_1)(x_2 - x_1)}, \quad \beta_{i2} = \frac{a - x_1}{(x_2 - x_1)(1 - x_2)}, \qquad i = 1, 2.$$

Therefore, $\mu_1 = \mu_2$. Consequently, $\mu \in \text{Ext}(\mathcal{M}_a)$.

Step 7: If for some $\alpha_1, \alpha_2 > 0$ *and* $0 < x_1 < x_2 < 1$ *with*

$$\alpha_1 \neq \frac{x_2 - a}{(1 - x_1)(x_2 - x_1)} \quad \textit{or} \quad \alpha_2 \neq \frac{a - x_1}{(x_2 - x_1)(1 - x_2)}$$

the measure $\mu = \sum_{i=1}^{2} \alpha_i \delta_{x_i} \in \mathcal{M}_a$*, then* $\mu \notin \text{Ext}(\mathcal{M}_a)$.

By (27), we have that

$$\sum_{i=1}^{2} \alpha_i (1 - x_i)^2 \leq 1 - a \quad \text{and} \quad \sum_{i=1}^{2} \alpha_i x_i (1 - x_i) \leq a. \tag{34}$$

If both inequalities in (34) were equalities, we necessarily have that

$$\alpha_1 = \frac{x_2 - a}{(1 - x_1)(x_2 - x_1)} \quad \text{and} \quad \alpha_2 = \frac{a - x_1}{(x_2 - x_1)(1 - x_2)},$$

against our assumption. Therefore, at least one of the two inequalities of (34) is strict. If $\sum_{i=1}^{2} \alpha_i (1 - x_i)^2 < 1 - a$, then we consider the signed measure defined by

$$\mu_0 = x_2(1 - x_2)\delta_{x_1} - x_1(1 - x_1)\delta_{x_2}.$$

Then, it is straightforward to check that, for small enough $\varepsilon > 0$, $\mu \pm \varepsilon\mu_0 \in \mathcal{M}_a$ and

$$\mu = \frac{1}{2}(\mu + \varepsilon\mu_0) + \frac{1}{2}(\mu - \varepsilon\mu_0). \tag{35}$$

Therefore, $\mu \notin \text{Ext}(\mathcal{M}_a)$.

If, instead, $\sum_{i=1}^{2} \alpha_i x_i (1 - x_i)^2 < a$, we then consider the signed measure

$$\mu_0 = (1 - x_2)^2 \delta_{x_1} - (1 - x_1)^2 \delta_{x_2}.$$

Again, we have that, for small enough $\varepsilon > 0$, $\mu \pm \varepsilon\mu_0 \in \mathcal{M}_a$ and equality (35) holds. We conclude that $\mu \notin \text{Ext}(\mathcal{M}_a)$.

Step 8: If $\mu \in \mathcal{M}_a$ *is supported in more than two points, then* $\mu \notin \text{Ext}(\mathcal{M}_a)$.

In this case, there exist Borel disjoint sets $A_i \subset (0, 1)$ such that $\mu|_{A_i} \neq 0$, for $i = 1, 2, 3$. Let $(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3 \setminus \{(0, 0, 0)\}$ be such that

$$\sum_{i=1}^{3} \alpha_i \int_{A_i} (1 - s)^2 \, \mathrm{d}\mu(s) = 0 \quad \text{and} \quad \sum_{i=1}^{3} \alpha_i \int_{A_i} s(1 - s)^2 \, \mathrm{d}\mu(s) = 0.$$

We consider the signed measure $\mu_0 = \sum_{i=1}^{3} \alpha_i \mu|_{A_i}$. For $\varepsilon > 0$ small enough, define μ^+ and μ^- as $\mu^\pm = \mu \pm \varepsilon\mu_0$. Then $\mu = \frac{1}{2}\mu^+ + \frac{1}{2}\mu^-$ with $\mu^\pm \neq \mu$. Moreover, $\mu^\pm$ are positive measures since

$$\mu^{\pm} = \mu|_{A_1^c \cap A_2^c \cap A_3^c} + \sum_{i=1}^{3} (1 \pm \varepsilon\alpha_i)\mu|_{A_i},$$

where A^c stands for the complement of the set A in $(0, 1)$. In fact, $\mu^{\pm} \in \mathcal{M}_a$ since

$$\int_0^1 (1-s)^2 \, \mathrm{d}\mu^{\pm}(s) = \int_0^1 (1-s)^2 \, \mathrm{d}\mu(s) \leq 1-a \text{ and } \int_0^1 s(1-s) \, \mathrm{d}\mu^{\pm}(s) = \int_0^1 s(1-s) \, \mathrm{d}\mu(s) \leq a.$$

Therefore, we conclude that $\mu \notin \mathrm{Ext}(\mathcal{M}_a)$.

The eight steps above complete the proof. $\square$

Step 8 of the previous proof is related to the works by Winkler [33] and Pinelis [26], where they analyze the set of extreme points of subset of measures defined through some inequalities.

Proof of Theorem 1 in Section 3 (Extreme points of $\mathcal{L}_a$). By Proposition 6, we have the equality

$$\mathrm{Ext}(\mathcal{L}_a) = T_a^{-1}(\mathrm{Ext}(\mathcal{M}_a)).$$

Now, we can use Proposition 7 to determine the set $\mathrm{Ext}(\mathcal{L}_a)$. By (28) and Proposition 7, we obtain three families of extreme curves in $\mathcal{L}_a$. First, for $x_1 \in (0, a]$, let $\ell^a_{x_1} = T_a^{-1}\left(\frac{1-a}{(1-x_1)^2}\delta_{x_1}\right)$ and $\ell^a_0 = T_a^{-1}(0)$. More explicitly, we have that, for $x_1 \in [0, a]$

$$\ell^a_{x_1}(x) = \frac{1-a}{(1-x_1)^2} \max\{0, x - x_1\}, \quad x \in [0, 1].$$

Second, for $x_2 \in (a, 1)$, we set $m^a_{x_2} = T_a^{-1}\left(\frac{a}{x_2(1-x_2)}\delta_{x_2}\right)$. We obtain that

$$m^a_{x_2}(x) = \frac{1}{x_2}\left[(x_2 - a)x + \frac{a}{1-x_2}\max\{0, x - x_2\}\right], \quad x \in [0, 1].$$

Finally, for $x_1 \in (0, a)$ and $x_2 \in (a, 1)$, let $n^a_{x_1,x_2} = T_a^{-1}\left(\frac{x_2-a}{(1-x_1)(x_2-x_1)}\delta_{x_1} + \frac{a-x_1}{(x_2-x_1)(1-x_2)}\delta_{x_2}\right)$. In this case we have that

$$n^a_{x_1,x_2}(x) = \frac{1}{x_2 - x_1}\left[\frac{x_2 - a}{1 - x_1}\max\{0, x - x_1\} + \frac{a - x_1}{1 - x_2}\max\{0, x - x_2\}\right], \quad x \in [0, 1].$$

These curves admit the characterization as piecewise affine functions given in (12).

Therefore, the proof of Theorem 1 is complete. $\square$

6.3. *Proof of Theorem 2 in Section 4 (maximal distance)*

Here we compute the exact value of $M(a, b)$ in (15). The proof of this theorem is long and we have divided it into several results. It is based on following proposition.

Proposition 8. *For $a, b \in [0, 1]$, let $M(a, b)$ be defined in (15). We have that*

$$M(a, b) = \max\{d_{\mathrm{L}}(\ell_1, \ell_2) : \ell_1 \in \mathrm{Ext}(\mathcal{L}_a) \text{ and } \ell_2 \in \mathrm{Ext}(\mathcal{L}_b)\}. \tag{36}$$

Proof. The distance $d_{\mathrm{L}} : \mathcal{L}_a \times \mathcal{L}_b \to \mathbb{R}$ is a convex and continuous functional in L^1. Further, by Proposition 1, the convex sets $\mathcal{L}_a$ and $\mathcal{L}_b$ are compact in L^1. Therefore, by Bauer's maximum principle (see, e.g., Phelps [25, Proposition 16.6]), the maximum of d_{L} is attained at the set $\mathrm{Ext}(\mathcal{L}_a \times \mathcal{L}_b) = \mathrm{Ext}(\mathcal{L}_a) \times \mathrm{Ext}(\mathcal{L}_b)$. □

Proposition 8 together with Theorem 1 reduce the calculation of $M(a, b)$ to a finite-dimensional problem, in fact, to several problems of dimension at most 4. Although in principle these problems can be solved using elementary analytic techniques, the computations are extremely cumbersome. For this reason, in the following we present several auxiliary results to simplify the calculations.

6.3.1. Proof of Proposition 2 in Section 4 (maximality property)

Let $a \in [0, 1]$ and $\ell \in \mathcal{L}_a$. To prove the inequalities in (21), we will first show that there exist $c^+, c^- \in [0, 1]$ (depending on ℓ) such that

$$\ell(x) \le \ell_a^+(x) \text{ if } 0 \le x \le c^+ \quad \text{and} \quad \ell(x) \ge \ell_a^+(x) \text{ if } c^+ \le x \le 1, \tag{37}$$

$$\ell_a^-(x) \le \ell(x) \text{ if } 0 \le x \le c^- \quad \text{and} \quad \ell_a^-(x) \ge \ell(x) \text{ if } c^- \le x \le 1. \tag{38}$$

To check (37), we note that the right derivative of ℓ at 0, $\ell'(0^+)$, is necessarily less than or equal to $1 - a$. Further, $\ell'(0^+) = 1 - a$ if and only if $\ell = \ell_a^+$, as in this case ℓ_a^+ is a supporting line of ℓ at 0. If $\ell'(0^+) < 1 - a$, then $\ell - \ell_a^+$ is a continuous and convex function in $[0, 1)$, starting at 0, with negative derivative at 0, and zero integral. Hence, there exists c^+ satisfying (37). To prove (38), observe that $\ell_a^- - \ell \le 0$ in $[0, a)$. Also, $\ell(a) = 0$ if and only if $\ell = \ell_a^-$. If $\ell(a) > 0$, then $\ell_a^- - \ell$ is a continuous and concave function in $[a, 1)$ such that $\ell_a^-(1) - \ell(1^-) \ge 0$. As $\ell_a^- - \ell$ has zero integral, there exists $c^- \in (a, 1)$ satisfying (38).

Now, for $0 \le t \le c^+$, by (37), we directly have that $\int_0^t \ell_a^+(x)\,\mathrm{d}x \ge \int_0^t \ell(x)\,\mathrm{d}x$. For $c^+ \le t \le 1$, as ℓ and ℓ_a^+ have the same integral in $(0, 1)$, we have that

$$\begin{aligned}
\int_0^t \ell_a^+(x)\,\mathrm{d}x &= \int_0^1 \ell_a^+(x)\,\mathrm{d}x - \int_t^1 \ell_a^+(x)\,\mathrm{d}x \\
&\ge \int_0^1 \ell(x)\,\mathrm{d}x - \int_t^1 \ell(x)\,\mathrm{d}x \\
&= \int_0^t \ell(x)\,\mathrm{d}x.
\end{aligned}$$

An analogous reasoning shows the second inequality in (21) and the proof is complete. □

6.3.2. The case in which $\ell - m$ has at most one sign change

Firstly, we carry out an analysis of the maximum value of the distance $d_{\mathrm{L}}(\ell, m)$ (for $\ell \in \mathcal{L}_a$ and $m \in \mathcal{L}_b$) when the number of crossing points of the functions ℓ and m is less than one. In the following lemma we consider the simpler case in which the Lorenz curves are ordered.

Lemma 1. *For $a, b \in [0, 1]$, let $\ell \in \mathcal{L}_a$ and $m \in \mathcal{L}_b$. We have that $d_{\mathrm{L}}(\ell, m) = |a - b|$ if and only if $\ell \le m$ or $m \le \ell$. In addition,*

$$\min\{d_{\mathrm{L}}(\ell, m) : \ell \in \mathcal{L}_a \text{ and } m \in \mathcal{L}_b\} = |a - b|.$$

Proof. By (9) and the triangular inequality, we always have that

$$|a-b| = 2|\|\ell_1\| - \|\ell_2\|| \leq 2\|\ell_1 - \ell_2\| = d_{\mathrm{L}}(\ell_1, \ell_2).$$

Moreover, if the inequality above is an equality, we conclude that ℓ and m are ordered since $|\ell - m| - ||\ell| - |m||$ is a continuous and non-negative function on $(0,1)$. The last part follows from the fact that $d_{\mathrm{L}}(\ell_a^+, \ell_b^+) = |a-b|$, where ℓ_a^+ and ℓ_b^+ are defined as in (16). □

Observe that the minimal distance is attained when the curves $\ell \in \mathcal{L}_a$ and $m \in \mathcal{L}_b$ are (pointwise) ordered, i.e., when the underlying variables are ordered in the Lorenz sense.

The second result (Lemma 3 below) analyzes the maximum distance between pairs of curves with only one sign switch. We need the following auxiliary lemma.

Lemma 2. *For any convex function $\varphi : [0,1] \to \mathbb{R}$, we have that*

$$\int_0^1 \varphi \,\mathrm{d}\ell_a^- \leq \int_0^1 \varphi \,\mathrm{d}\ell \leq \int_0^1 \varphi \,\mathrm{d}\ell_a^+. \tag{39}$$

Proof. For this proof we consider the original definition of $\mathcal{L}$ given by (1), i.e., any $\ell \in \mathcal{L}$ satisfies $\ell(1) = 1$ instead of $\ell(1) = \sup_{[0,1)} \ell$. In this way, each $\ell \in \mathcal{L}_a$ is itself a distribution function of a random variable, say X_ℓ, concentrated on the interval $[0,1]$. The inequalities in (21) together with the fact that all variables X_ℓ ($\ell \in \mathcal{L}_a$) have the same expectation imply (39); see [28, Theorem 3.A.1]. □

We now establish the maximum value of the Lorenz distance for curves with one crossing point.

Lemma 3. *Let $a, b \in [0,1]$ and consider $\ell \in \mathcal{L}_a$, $m \in \mathcal{L}_b$ and $\ell_a^\pm$, $\ell_b^\pm$ the curves defined in (16). We have that*

(a) If for some $t_0 \in (0,1)$ we have $\ell \leq m$ in $[0,t_0]$ and $\ell \geq m$ in $[t_0,1]$, then $d_{\mathrm{L}}(\ell, m) \leq d_{\mathrm{L}}(\ell_a^-, \ell_b^+)$.
(b) If for some $t_0 \in (0,1)$ we have $\ell \geq m$ in $[0,t_0]$ and $\ell \leq m$ in $[t_0,1]$, then $d_{\mathrm{L}}(\ell, m) \leq d_{\mathrm{L}}(\ell_a^+, \ell_b^-)$.

Moreover, it holds that

$$d_{\mathrm{L}}(\ell_a^-, \ell_b^+) = d_{\mathrm{L}}(\ell_a^+, \ell_b^-) = \frac{(1-a)b^2 + (1-b)a^2}{a+b-ab}. \tag{40}$$

Proof. As in Lemma 2, in this proof we consider the definition of $\mathcal{L}$ given by (1), i.e., any $\ell \in \mathcal{L}$ satisfies $\ell(1) = 1$. Thus, each function in $\mathcal{L}$ is a distribution function of a random variable concentrated on $[0,1]$.

We only show *(a)*, as the proof of *(b)* is analogous. Let us consider the function $s_{t_0} : [0,1] \to \mathbb{R}$ defined as

$$s_{t_0}(x) = t_0 - x + 2\max\{0, x - t_0\}.$$

Clearly, the function s_{t_0} is convex, Lipschitz and its distributional derivative is given by $s'_{t_0} = -1_{[0,t_0]} + 1_{(t_0,1]}$, where 1_A stands for the indicator function of the set A. By integration by parts in the Lebesgue–Stieltjes integral, we obtain that

$$\int_0^1 s_{t_0} \,\mathrm{d}(m - \ell) = \int_0^1 (\ell - m) \,\mathrm{d}s_{t_0} = \int_0^{t_0} (m - \ell) + \int_{t_0}^1 (\ell - m) = \|\ell - m\|.$$

As s_{t_0} is convex, by Lemma 2, we conclude that

$$\int_0^1 s_{t_0}\,\mathrm{d}(m-\ell) \le \int_0^1 s_{t_0}\,\mathrm{d}\ell_b^+ - \int_0^1 s_{t_0}\,\mathrm{d}\ell_a^- .$$

Now, from the expression of the Wasserstein distance between probability distributions on the line (see Vallender [31]) and by virtue of the Kantorovich–Rubinstein duality (see Villani [32, eq. (6.3)]), we have that

$$\left\|\ell_b^+ - \ell_a^-\right\| = \sup\left\{\int_0^1 f\,\mathrm{d}(\ell_b^+ - \ell_a^-) : f \in W^{1,\infty}(0,1),\ \|f'\|_\infty \le 1\right\}.$$

Putting together the relations above we obtain that

$$\|\ell - m\| = \int_0^1 s_{t_0}\,\mathrm{d}(m-\ell) \le \int_0^1 s_{t_0}\,\mathrm{d}(\ell_b^+ - \ell_a^-) \le \left\|\ell_b^+ - \ell_a^-\right\|.$$

Therefore, part *(a)* of this lemma is proved.

To finish, we note that the curves ℓ_a^+ and ℓ_b^- have one sign switch at the point $t_0 = a/(a+b-ab)$. The Lorenz distance in (40) can be directly computed by elementary geometry (as twice the sum of the areas of two triangles; see Fig. 5),

$$d_{\mathrm{L}}(\ell_a^-, \ell_b^+) = a(1-b)t_0 + b(1-t_0) = \frac{(1-a)b^2 + (1-b)a^2}{a+b-ab},$$

which completes the proof of the lemma. □

6.3.3. Symmetry properties of the set $\mathcal{L}_a$

Another observation that simplifies to a great extend the calculations is a symmetry reasoning. Given the graph of a function in $\mathcal{L}_a$, its symmetry along the line $y = 1 - x$ corresponds to the graph of another curve in $\mathcal{L}_a$. The precise definition, statement and proof are as follows.

Lemma 4. *For $\ell \in \mathcal{L}$, define $\ell^{-1} : [0,1] \to [0,1]$ as in (6). Let us consider the function $\tilde{\ell} : [0,1] \to [0,1]$ given by*

$$\tilde{\ell}(x) = 1 - \ell^{-1}(1-x), \quad x \in [0,1].$$

We have that the map $\ell \mapsto \tilde{\ell}$ is a bijective isometry from $(\mathcal{L}, d_{\mathrm{L}})$ to $(\mathcal{L}, d_{\mathrm{L}})$ whose inverse is itself. That is, for any ℓ, $m \in \mathcal{L}$, we have that $d_{\mathrm{L}}(\ell, m) = d_{\mathrm{L}}(\tilde{\ell}, \tilde{m})$.

Moreover, for $a \in [0,1]$, the map $\ell \mapsto \tilde{\ell}$ is also a bijective isometry from $(\mathcal{L}_a, d_{\mathrm{L}})$ to $(\mathcal{L}_a, d_{\mathrm{L}})$.

Proof. As in the proof of Lemma 3, we consider the original definition of $\mathcal{L}$ given by (1), i.e., any $\ell \in \mathcal{L}$ satisfies $\ell(1) = 1$.

Let $\ell \in \mathcal{L}$. To check the first assertion we will first show that $\tilde{\ell} \in \mathcal{L}$. Let us consider $x_\ell = \sup\{x \in [0,1] : \ell(x) = 0\}$. We observe that $\ell : (x_\ell, 1) \to (0, \ell(1^-))$ is a bijection and

$$\ell^{-1}(y) = \begin{cases} 0 & \text{if } y = 0, \\ \left(\ell|_{(x_\ell,1)}\right)^{-1}(y) & \text{if } y \in (0, \ell(1^-)), \\ 1 & \text{if } y \in [\ell(1^-), 1]. \end{cases} \tag{41}$$

Indeed, by Proposition 5, ℓ is non-decreasing. We shall see that $\ell : [x_\ell, 1] \to \mathbb{R}$ is strictly increasing. Assume, by contradiction, that there exist x_1, x_2 with $x_\ell \leq x_1 < x_2 \leq 1$ and $\ell(x_1) = \ell(x_2)$. Then, the constant function $\ell(x_2)$ is a supporting line of ℓ at $(x_2, \ell(x_2))$, so, by convexity, $\ell \geq \ell(x_2)$. Since $x_2 > x_\ell$, by definition of x_ℓ we have that $\ell(x_2) > 0$. In particular $\ell(0) > 0$, which is a contradiction. We conclude that $\ell : [x_\ell, 1] \to \mathbb{R}$ is strictly increasing. By Proposition 5, $\ell|_{[0,1)}$ is continuous and, in particular, $\ell : (x_\ell, 1) \to (0, \ell(1^-))$ is a bijection. Moreover, $\ell|_{[0,x_\ell]} = 0$ and $\ell(1) = 1$. With these properties of ℓ, it is immediate to check that ℓ^{-1} is given as in (41). Now, from (41), we obtain that

$$\tilde{\ell}(x) = \begin{cases} 0 & \text{if } x \in [0, 1 - \ell(1^-)], \\ 1 - \left(\ell|_{(x_\ell,1)}\right)^{-1}(1-x) & \text{if } x \in (1 - \ell(1^-), 1), \\ 1 & \text{if } x = 1. \end{cases} \tag{42}$$

The inverse of an increasing convex function is concave (see Simon [30, Example 1.6]). Hence, the function ℓ^{-1} in (41) is concave, and this implies that $\tilde{\ell}$ in (42) is convex. In particular, it follows that $\tilde{\ell} \in \mathcal{L}$.

Next we will check that $\tilde{\tilde{\ell}} = \ell$. It can be easily seen that $\tilde{\ell}(1^-) = 1 - x_\ell$ and $x_{\tilde{\ell}} = 1 - \ell(1^-)$. Moreover, $\left(\tilde{\ell}|_{(x_{\tilde{\ell}},1)}\right)^{-1} : (0, 1 - x_\ell) \to (x_{\tilde{\ell}}, 1)$ is given by $\left(\tilde{\ell}|_{(x_{\tilde{\ell}},1)}\right)^{-1}(y) = 1 - \ell(1-y)$. Therefore, by (42), we conclude that $\tilde{\tilde{\ell}} = \ell$.

Next we need to prove that the map $\ell \mapsto \tilde{\ell}$ is an isometry. We consider $m \in \mathcal{L}$ and observe first that $\|\tilde{\ell} - \tilde{m}\| = \|\ell^{-1} - m^{-1}\|$. Further, as ℓ and m are distribution functions of random variables concentrated on the interval $[0, 1]$, by using the expression of the Wasserstein distance for real-valued random variables (see Vallender [31]), we have that $\|\ell^{-1} - m^{-1}\| = \|\ell - m\|$. Hence, we conclude that $d_{\mathrm{L}}(\ell, m) = d_{\mathrm{L}}(\tilde{\ell}, \tilde{m})$.

Finally, to check the last assertion of the lemma it is enough to verify that, for $a \in [0, 1]$ and $\ell \in \mathcal{L}_a$, one has $G(\tilde{\ell}) = a$. Let us fix $\ell \in \mathcal{L}_a$. We will equivalently show that $\|\tilde{\ell}\| = (1-a)/2$. We apply Laisant's formula (see, for instance, Parker [24]) for the integral of the inverse to obtain

$$\int_0^{\ell(1^-)} \left(\ell|_{(x_\ell,1)}\right)^{-1}(y)\,\mathrm{d}y = \ell(1^-) - \int_{x_\ell}^{1} \ell(x)\,\mathrm{d}x = \ell(1^-) - \frac{1-a}{2}.$$

Therefore,

$$\int_0^1 \ell^{-1}(y)\,\mathrm{d}y = \int_0^{\ell(1^-)} \left(\ell|_{(x_\ell,1)}\right)^{-1}(y)\,\mathrm{d}y + 1 - \ell(1^-) = 1 - \frac{1-a}{2}.$$

From (42), we finally obtain that

$$\int_0^1 \tilde{\ell}(y)\,\mathrm{d}y = 1 - \int_0^1 \ell^{-1}(y)\,\mathrm{d}y = \frac{1-a}{2},$$

which completes the proof of the lemma. □

Lemma 4 helps us to disregard some cases in the computation of the value of $M(a, b)$, as the next result shows.

Lemma 5. *For $a, b \in [0, 1]$, let $M(a, b)$ be defined in (15). We have that*

$$M(a, b) = \max\{d_1(a, b),\, d_2(a, b),\, d_3(a, b)\},$$

where

$$\begin{aligned} d_1(a,b) &= \max\{d_{\mathrm{L}}(\ell^a_{x_1}, \ell^b_{y_1}) : x_1, y_1 \in [0,a]\}, \\ d_2(a,b) &= \max\{d_{\mathrm{L}}(m^a_{x_2}, m^b_{y_2}) : x_2, y_2 \in (a,1)\}, \\ d_3(a,b) &= \max\{d_{\mathrm{L}}(\ell^a_{x_1}, m^b_{x_2}) : x_1 \in [0,a], x_2 \in (a,1)\}. \end{aligned} \tag{43}$$

Proof. We describe first how the extreme points of Theorem 1 are affected under the isometry $\ell \mapsto \tilde{\ell}$ defined in Lemma 4. From (12) and (42), we find that they are the piecewise affine functions such that

$$\tilde{\ell}^a_{x_1} : \begin{cases} 0 \mapsto 0 \\ 1 - \frac{1-a}{1-x_1} \mapsto 0 \\ 1 \mapsto 1 - x_1, \end{cases} \qquad \tilde{m}^a_{x_2} : \begin{cases} 0 \mapsto 0 \\ 1 - (x_2 - a) \mapsto 1 - x_2 \\ 1 \mapsto 1, \end{cases} \qquad \tilde{n}^a_{x_1,x_2} : \begin{cases} 0 \mapsto 0 \\ 1 - \frac{x_2-a}{1-x_1} \mapsto 1 - x_2 \\ 1 \mapsto 1 - x_1. \end{cases}$$

In other words,

$$\tilde{\ell}^a_{x_1} = \ell^a_{\frac{a-x_1}{1-x_1}}, \qquad \tilde{m}^a_{x_2} = m^a_{a+1-x_2}, \qquad \tilde{n}^a_{x_1,x_2} \notin \mathrm{Ext}(\mathcal{L}_a).$$

As $\tilde{n}^a_{x_1,x_2} \notin \mathrm{Ext}(\mathcal{L}_a)$, by Proposition 8 and since $\ell \mapsto \tilde{\ell}$ is an isometry (see Lemma 4), we can exclude the functions $n^a_{x_1,x_2}$ in the computation of the maximum given in (36). □

6.3.4. Computation of $M(a,b)$

Here, we will combine all the previous results to prove Theorem 2. We start with the following lemma.

Lemma 6. *For $a, b \in [0,1]$, let $d_1(a,b)$ and $d_2(a,b)$ be defined in (43). We have that*

$$d_1(a,b),\ d_2(a,b) \le d_{\mathrm{L}}(\ell^-_a, \ell^+_b) = d_{\mathrm{L}}(\ell^+_a, \ell^-_b),$$

where the curves $\ell^\pm_a$, $\ell^\pm_b$ are defined as in (16).

Proof. It is easy to see that the curves $\ell^a_{x_1}$ and $\ell^b_{y_1}$ cannot have two crossing points, and neither can $m^a_{x_2}$ and $m^b_{y_2}$. Therefore, the conclusion follows from Lemmas 1 and 3. □

We are then led to the computation of $d_3(a,b)$, i.e., the maximum of value of $d_{\mathrm{L}}(\ell^a_{x_1}, m^b_{x_2})$, when $x_1 \in [0,a]$ and $x_2 \in (b,1)$. This question is addressed in the following result.

Lemma 7. *For $a, b \in [0,1]$, let $d_3(a,b)$ be defined in (43). We have that*

$$d_3(a,b) \le d_{\mathrm{L}}(\ell^-_a, \ell^+_b) = d_{\mathrm{L}}(\ell^+_a, \ell^-_b).$$

Proof. Let $x_1 \in [0,a]$ and $x_2 \in (b,1)$. Thanks to Lemmas 1 and 3, we can assume that the curves $\ell^a_{x_1}$ and $m^b_{x_2}$ cross each other twice; Fig. 12 represents this situation. This happens if and only if the triangle with vertices

$$(x_1, 0), \quad (x_2, x_2 - b), \quad \left(1, \frac{1-a}{1-x_1}\right)$$

has positive orientation; equivalently, if and only if

$$(1-a)(x_2 - x_1) - (x_2 - b)(1 - x_1) + x_1(1 - x_1)(x_2 - b) > 0.$$

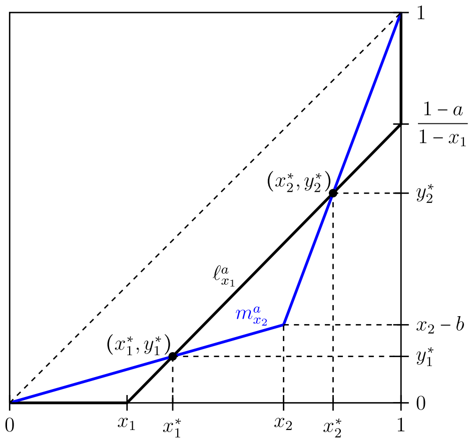

Fig. 12. A graphical representation of the functions $\ell^a_{x_1}$ (in black) and $m^a_{x_2}$ (in blue), as well as their crossing points, (x_1^*, y_1^*) and (x_2^*, y_2^*). (For the coloured version of the figure, the reader is referred to the web version of this article.)

In this case, they cross each other at the points (x_1^*, y_1^*) and (x_2^*, y_2^*), with $x_1^* \le x_2^*$ and

$$x_1^* = \frac{(1-a)x_1x_2}{(1-a)x_2 - (1-x_1)^2(x_2-b)}, \quad y_1^* = \frac{(1-a)x_1(x_2-b)}{(1-a)x_2 - (1-x_1)^2(x_2-b)},$$
$$x_2^* = \frac{(1-a)x_1x_2(1-x_2) - b(1-x_1)^2x_2}{(1-a)x_2(1-x_2) - (1-x_1)^2(x_2-b)(1-x_2) - b(1-x_1)^2},$$
$$y_2^* = \frac{(1-a)\left[x_1(x_2-b)(1-x_2) - b(x_2-x_1)\right]}{(1-a)x_2(1-x_2) - (1-x_1)^2(x_2-b)(1-x_2) - b(1-x_1)^2}.$$

For simplicity, from now on we will call $A = A(x_1, x_2) = d_{\mathrm{L}}\left(\ell^a_{x_1}, m^b_{x_2}\right)$. The value of A can be therefore calculated as twice the sum of the areas of the three triangles (see Fig. 12)

$$(0,0), (x_1, 0), (x_1^*, y_1^*), \quad (x_1^*, y_1^*), (x_2, x_2 - b), (x_2^*, y_2^*), \quad (x_2^*, y_2^*), \left(1, \frac{1-a}{1-x_1}\right), (1,1).$$

Elementary geometry shows that

$$A = x_1\, y_1^* + x_1^*(x_2 - b - y_2^*) + x_2(y_2^* - y_1^*) + x_2^*(y_1^* - (x_2 - b)) + \frac{(a-x_1)(1-x_2^*)}{1-x_1}.$$

So are led to the maximization of $A = A(x_1, x_2)$ under the constraints

$$0 \le x_1 \le a, \quad b < x_2 < 1, \quad (1-a)(x_2 - x_1) - (x_2 - b)(1 - x_1) + x_1(1-x_1)(x_2 - b) > 0. \tag{44}$$

Obviously, the supremum under (44) coincides with the maximum under the analogous inequalities of (44) but replacing the '$<$' with '$\le$', and '$>$' with '$\ge$'. Observe that A is a rational function, and, hence, the computation of its maximum in region (44) can be done by elementary techniques. Nevertheless, the computations are extremely long, so we have used the programme [34] in the rest of the proof to avoid unnecessary details.

We find that the only solution of $\frac{\partial A}{\partial x_1} = \frac{\partial A}{\partial x_2} = 0$ under (44) is

$$x_1 = 1 - \sqrt{1-a}, \quad x_2 = \frac{1+b}{2},$$

which in fact requires $a + b^2 < 2b$. The corresponding value of A is

$$A\left(1-\sqrt{1-a}, \frac{1+b}{2}\right) = \frac{4\left(\sqrt{1-a}-1\right)b + a(b-2) - 4\sqrt{1-a} + b^2 + 4}{b},$$

which is seen to be less than the value of $d_{\mathrm{L}}(\ell_a^-, \ell_b^+)$ (computed in Lemma 3), thanks to the restriction $a + b^2 < 2b$.

After having checked the value of A at the critical points in the interior of region (44), we analyze the value of A on the boundary. Describing the boundary of region (44) is cumbersome since it involves several cases, according to the values of a, b. In any case, the boundary is clearly contained in the set

$$\big\{(x_1, x_2) \in [0,a] \times [b,1] : x_1 = 0 \text{ or } x_1 = a \text{ or } x_2 = b \text{ or } x_2 = 1 \\ \text{or } (1-a)(x_2 - x_1) - (x_2 - b)(1 - x_1) + x_1(1-x_1)(x_2 - b) = 0\big\}.$$

When $(1-a)(x_2 - x_1) - (x_2 - b)(1 - x_1) + x_1(1-x_1)(x_2 - b) = 0$ the curves $\ell^a_{x_1}$ and $m^b_{x_2}$ do not have a proper crossing, so, by Lemma 1, $A = |b - a|$, which does not release a maximum. Therefore, we are led to the maximization of $A(x_1, x_2)$ in the set

$$\big\{(x_1, x_2) \in [0,a] \times [b,1] : \ x_1 = 0 \text{ or } x_1 = a \text{ or } x_2 = b \text{ or } x_2 = 1\big\}$$

The value of A when $x_1 = 0$ is

$$A(0, x_2) = a + b - \frac{2ab}{a + b - ax_2},$$

which is decreasing in x_2, so the maximum is attained at $x_2 = b$ and equals

$$A(0, b) = a + b - \frac{2ab}{a + b - ab} = \frac{(1-a)b^2 + (1-b)a^2}{a + b - ab} = d_{\mathrm{L}}(\ell_a^-, \ell_b^+).$$

The value of A when $x_1 = a$ is

$$A(a, x_2) = a + b - \frac{2ab}{ax_2 + b - ab}$$

which is increasing in x_2, so the maximum is attained at $x_2 = 1$ and equals

$$A(a, 1) = \frac{(1-a)b^2 + (1-b)a^2}{a + b - ab} = d_{\mathrm{L}}(\ell_a^-, \ell_b^+).$$

The value of A when $x_2 = b$ is

$$A(x_1, b) = \frac{a^2 b - a^2 + ab^2 - b^2 + (-4ab + 2a + 2b)\, x_1 + (a + b - 2)\, x_1^2}{ab - a - b + 2x_1 - x_1^2},$$

which is decreasing in x_1, so the maximum is attained at $x_1 = 0$ and equals

$$A(0, b) = \frac{(1-a)b^2 + (1-b)a^2}{a + b - ab} = d_{\mathrm{L}}(\ell_a^-, \ell_b^+).$$

The value of A when $x_2 = 1$ is

$$A(x_1, 1) = \frac{a^2 b - a^2 + ab - b^2 + \left(a + 2b^2 - 3ab\right) x_1 + \left(a - b^2\right) x_1^2 + (b-1)\, x_1^3}{ab - a - b + 2x_1 - x_1^2},$$

which is increasing in x_1. Therefore, the maximum is attained at $x_1 = a$ and equals

$$A(a,1) = \frac{(1-a)b^2 + (1-b)a^2}{a+b-ab}.$$

This concludes the proof. $\square$

We finally observe that the proof of Theorem 2 directly follows from Lemmas 3, 5, 6, and 7.

Acknowledgments

A. Baíllo and J. Cárcamo are supported by the Spanish MCyT grant PID2019-109387GB-I00. C. Mora-Corral is supported by the Spanish MCyT grant MTM2017-85934-C3-2-P.

References

[1] C.D. Aliprantis, K.C. Border, Infinite Dimensional Analysis: A Hitchhiker's Guide, Springer, 2006.
[2] L.E. Alonso, C.J. Fernández, R. Ibáñez, "I think the middle class is disappearing": crisis perceptions and consumption patterns in Spain, Int. J. Consum. Stud. 41 (2017) 389–396.
[3] F. Andreoli, Robust inference for inverse stochastic dominance, J. Bus. Econ. Stat. 36 (1) (2018) 146–159.
[4] B.C. Arnold, J.M. Sarabia, Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics, Springer, 2018.
[5] P. Blavier, How did the great recession affect income inequality in Spain?, Econ. Sociol. Eur. Electron. Newsl. 19 (1) (2017) 7–14.
[6] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, Springer, 2011.
[7] J. de la Cal, J. Cárcamo, Inverse stochastic dominance, majorization, and mean order statistics, J. Appl. Probab. 47 (1) (2010) 277–292.
[8] F. Cowell, Measuring Inequality, Oxford University Press, 2011.
[9] D. Dentcheva, A. Ruszczyński, Inverse stochastic dominance constraints and rank dependent expected utility theory, Math. Program. 108 (2) (2006) 297–311.
[10] F.G. De Maio, Income inequality measures, J. Epidemiol. Community Health 61 (2007) 849–852.
[11] The Economist, Lessons from Spain's recovery after the euro crisis, Available: www.economist.com/leaders/2018/07/26/lessons-from-spains-recovery-after-the-euro-crisis, 26 Jul. 2018.
[12] B. Efron, Bootstrap methods: another look at the jackknife, Ann. Stat. 7 (1979) 1–26.
[13] I.I. Eliazar, I.M. Sokolov, Measuring statistical evenness: a panoramic overview, Phys. A, Stat. Mech. Appl. 391 (4) (2012) 1323–1353.
[14] L.C. Evans, R.F. Gariepy, Measure Theory and Fine Properties of Functions, CRC Press, 1992.
[15] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[16] W.M. Getz, C.C. Wilmers, A local nearest-neighbor convex-hull construction of home ranges and utilization distributions, Ecography 27 (2004) 489–505.
[17] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, second edition, Springer, New York, 2009.
[18] International Monetary Fund, Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Spain, Country Report No. 17/319, 6 Oct. 2017.
[19] International Monetary Fund, Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Spain, Country Report No. 2020/298, 13 Nov. 2020.
[20] C. Kleiber, S. Kotz, Statistical Size Distributions in Economics and Actuarial Sciences, vol. 470, John Wiley & Sons, 2003.
[21] M. Leshno, H. Levy, Preferred by "all" and preferred by "most" decision makers: almost stochastic dominance, Manag. Sci. 48 (8) (2002) 1074–1085.
[22] C.J. Lloyd, Estimation of a convex ROC curve, Stat. Probab. Lett. 59 (1) (2002) 99–111.
[23] P. Muliere, M. Scarsini, A note on stochastic dominance and inequality measures, J. Econ. Theory 49 (2) (1989) 314–323.
[24] F.D. Parker, Integrals of inverse functions, Am. Math. Mon. 62 (6) (1955) 439–440.
[25] R.R. Phelps, Lectures on Choquet's Theorem, second edition, Lecture Notes in Mathematics, vol. 1757, Springer, 2001.
[26] I. Pinelis, On the extreme points of moments sets, Math. Methods Oper. Res. 83 (3) (2016) 325–349.
[27] L.A. Prendergast, R.G. Staudte, A simple and effective inequality measure, Am. Stat. 72 (4) (2018) 328–343.
[28] M. Shaked, J.G. Shanthikumar, Stochastic Orders, Springer, 2006.
[29] A.F. Shorrocks, J.E. Foster, Transfer sensitive inequality measures, Rev. Econ. Stud. 54 (3) (1987) 485–497.
[30] B. Simon, Convexity, Cambridge University Press, 2011.
[31] S.S. Vallender, Calculation of the Wasserstein distance between probability distributions on the line, Theory Probab. Appl. 18 (4) (1974) 784–786.

[32] C. Villani, Optimal Transport: Old and New, Springer, 2009.
[33] G. Winkler, Extreme points of moment sets, Math. Oper. Res. 13 (4) (1988) 581–587.
[34] Wolfram Research, Mathematica, Inc., Version 9.0, Champaign, IL, 2012.
[35] S. Yitzhaki, E. Schechtman, The Gini Methodology: A Primer on a Statistical Methodology, Springer Science & Business Media, 2012.
[36] B. Zheng, Almost Lorenz dominance, Soc. Choice Welf. 51 (1) (2018) 51–63.
[37] C. Zoli, Inverse stochastic dominance, inequality measurement and Gini indices, J. Econ. 77 (1) (2002) 119–161.