SORT 41 (1) January-June 2017, 55-72

On a property of Lorenz curves with monotone elasticity and its application to the study of inequality by using tax data

M.A. Sordo, A. Berihuete, C.D. Ramos and H.M. Ramos*

Abstract

The Lorenz curve is the most widely used graphical tool for describing and comparing inequality of income distributions. In this paper, we show that the elasticity of this curve is an indicator of the effect, in terms of inequality, of a truncation of the income distribution. As an application, we consider tax returns as equivalent to the truncation from below of a hypothetical income distribution. Then, we replace this hypothetical distribution by the income distribution obtained from a general household survey and use the dual Lorenz curve to anticipate this effect.

MSC: 91B02.

Keywords: Lorenz curve, tax data, truncation, inequality.

1. Introduction

Tax data are commonly used sources of information in the analysis of income distributions. For example, Piketty and Saez (2003) used tax data to study the concentration of income within the top 10 percent of the distribution with higher incomes in the United States and regularly release reports with the latest available data. Atkinson, Piketty and Saez (2011) provided a comparative study of top incomes covering a wide variety of countries by using tax data. More recently, Saez and Zucman (2014) expanded these works to examine trends in wealth concentration. Tax returns, like other administrative sources, often provide more accurate and complete data for the population under study than other surveys (see Stone et al., 2015).

Research based on data from income tax returns focuses on people who file taxes. However, not everyone is required to file an income tax return every year. In general,

*Dpto. de Estadística e I.O., Universidad de Cádiz, Spain. mangel.sordo@uca.es, angel.berihuete@uca.es, carmen.ramos@uca.es, hector.ramos@uca.es
Received: September 2015
Accepted: November 2016

people who do not file are those whose income falls below certain thresholds. The exclusion of non-filers is the most significant limitation of this source if the target population is the entire population in a country. In this case, data from tax returns produce an obvious systematic error or bias in estimating characteristics related to the size of income, such as the disposable income or per capita income. The question that we investigate in this paper is whether there are a systematic error in evaluating, from tax data, characteristics related to the inequality of the income distribution for the entire population. The approach adopted here is to consider tax returns as equivalent to a truncation from below on a hypothetical distribution that would be obtained if everyone would pay taxes. Since this hypothetical distribution is unrealistic, in practice we replace it by the income distribution obtained from a general household survey. It is shown that the effect of income truncation (at any level) by itself does not necessarily introduce a bias in one direction or the other and that when it does, it depends on the shape of the Lorenz curve associated to this distribution.

Specifically, let X be the random variable that describes the true income distribution, let F be its cumulative distribution function and assume that X has a finite mean $\mu > 0$. The most widely used graphical tool for describing and comparing inequality of income distributions is the Lorenz curve[1]. For the income random variable X, the Lorenz curve is defined by

$$L(p) = \frac{1}{\mu}\int_0^p F^{-1}(t)dt, \qquad 0 \leq p \leq 1 \tag{1}$$

where we denote by F^{-1} the inverse of F defined by $F^{-1}(t) = \inf\{x : F(x) \geq t\}$, with $0 < t < 1$. For each p in $(0,1)$, the function $L(p)$ is the cumulative percetage of total income held by individuals having the $100p\%$ lowest incomes. In this paper, the main result is depicted in terms of the dual Lorenz curve $\overline{L}(p)$, which is a reverse-mirror image of the Lorenz curve reflected through the diagonal 45 degree line. It is defined by

$$\overline{L}(p) = 1 - L(1-p),\, 0 \leq p \leq 1, \tag{2}$$

and represents the proportion of total income that accrues to individuals having the $100p\%$ largest incomes. Both curves are non-decreasing and differentiable almost everywhere, with $L(0) = \overline{L}(0) = 0$ and $L(1) = \overline{L}(1) = 1$. If the distribution function F is continuous and strictly increasing, then $L(p)$ and $\overline{L}(p)$ are strictly increasing and continuously differentiable functions of p. The Lorenz curve induces a partial ordering (denoted $\leq_L$, see Arnold, 1987) on the class of income random variables, ordering them in terms of inequality. Given two income random variables X and Y with Lorenz curves $L_X(p)$ and $L_Y(p)$, respectively, X is less unequal than Y (denoted by $X \leq_L Y$) if and only

1. Lorenz curves are used in many diverse fields, other than income distributions, such as informetrics (see Sarabia, Prieto and Trueba, 2012), demography (see Ramos et al., 2013) or risk measurement (see Greselin and Zitikis, 2015) among others.

if

$$L_X(p) \geq L_Y(p) \text{ for all } 0 \leq p \leq 1 \tag{3}$$

or, equivalently, if

$$\overline{L}_X(p) \leq \overline{L}_Y(p) \text{ for all } 0 \leq p \leq 1.$$

Under a progressive tax structure there exists a tax-free threshold t below which people do not pay personal income tax. Thus, tax return distribution is obtained by a lower truncation of X at income t. For each $t > 0$, denote by $X_{(t,\infty)} = \{X \mid X \geq t\}$ the corresponding lower truncated random variable. In Section 2, we give conditions to compare, in terms of inequality, the income random variables X and $X_{(t,\infty)}$, without needing to know the distribution function of X. Thus, unlike other authors that have previously studied this topic (Bhattacharya, 1963; Moothathu, 1986; Belzunce, Candel and Ruiz, 1995, 1998), our conditions can be directly verified from the Lorenz curve of X. As we explain in Section 2, our results can also be stated in terms of the monotonicity of the function

$$e(t) = E\left[\frac{X}{t} | X > t\right]$$

which represents the expected proportional income to t, for incomes greater than t.

Inequality is not the only characteristic of interest of income distributions. Another important aspect of the concentration of incomes is related to the notion of relative deprivation, which is based on the perception that an individual makes about his social status in a population. In order to compare distributions in terms of deprivation, the starshaped order (see Shaked and Shantikumar, 2007) and the expected proportional shortfall order (Belzunce et al., 2012, 2013) can be considered. In Section 3 we study the effect of truncations on these orderings. It will be shown that the effect of truncations on the expected proportional shortfall order depends, like in the case of the Lorenz order, on the elasticity of the Lorenz curve. In Section 4 we review some parametric models for the Lorenz curve that satisfy the conditions stated in previous sections. Finally, in Section 5, we illustrate the usefulness of our results by a descriptive study based on real data drawn from the survey EU-SILC 2010.

2. Lorenz ordering of truncated random variables

We show in this section that the effect of truncation on the inequality depends on the sensitivity of the dual Lorenz $\overline{L}(p)$ with respect to a change in p, that is, on its elasticity $\varepsilon_{\overline{L}}(p)$, defined by

$$\varepsilon_{\overline{L}}(p) = \frac{d\log\overline{L}(p)}{d\log p} = \frac{pL'(1-p)}{1-L(1-p)},\ 0 < p < 1. \tag{4}$$

We have the following result.

Theorem 2.1. *Let X be an income random variable with Lorenz curve L and let $t_0 \in [0,1]$. Then*

$$X_{(t,\infty)} \leq_L X_{(t',\infty)} \; (\geq_L) \text{ for all } t_0 < t < t' \tag{5}$$

if and only if $\varepsilon_{\overline{L}}(p)$ is increasing (decreasing) in the interval $(0, 1 - F(t_0))$.

Proof. We give the proof for the case $\leq_L$ (the proof for the $\geq_L$ case is analogous). For each $t > 0$, denote by $F_{(t,\infty)}(x)$ the distribution function of $X_{(t,\infty)}$ given by

$$F_{(t,\infty)}(x) = \begin{cases} 0 & x < t \\ \frac{F(x)-F(t)}{1-F(t)} & x \geq t \end{cases}. \tag{6}$$

Let $L_{(t,\infty)}(p)$ be the Lorenz curve of $X_{(t,\infty)}$. By using (1) and (6) it is easy to see that

$$L_{(t,\infty)}(p) = \frac{L[(1-F(t))\,p + F(t)] - L(F(t))}{1 - L(F(t))},\; 0 \leq p \leq 1. \tag{7}$$

Condition (5) holds if and only if $L_{(t,\infty)}(p)$ is decreasing in $t > t_0$ or, equivalently, if

$$1 - L_{(t,\infty)}(p) \text{ is increasing in } t > t_0.$$

By making $F(t) = a$ and using (7), this is the same as

$$\frac{1 - L[(1-a)\,p + a]}{1 - L(a)} \text{ is increasing in } a > F(t_0)$$

which, by differentiation, is satisfied if and only if

$$\frac{(1-p)\,L'[(1-a)p + a]}{1 - L[(1-a)p + a]} \leq \frac{L'(a)}{1 - L(a)},\; p \in [0,1],\; a > F(t_0).$$

The above inequality can be rewritten as

$$\frac{(1-a)(1-p)\,L'[(1-a)p + a]}{1 - L[(1-a)p + a]} \leq \frac{(1-a)\,L'(a)}{1 - L(a)},\;\; p \in [0,1],\; a > F(t_0),$$

which, by making $p_1 = (1-a)(1-p)$ and $p_2 = 1 - a$, is the same as

$$\frac{p_1 L'(1-p_1)}{1 - L(1-p_1)} \leq \frac{p_2 L'(1-p_2)}{1 - L(1-p_2)} \text{ whenever } 0 < p_1 \leq p_2 < 1 - F(t_0).$$

Using (4), this means that $\varepsilon_{\overline{L}}(p)$ is increasing in the interval $(0, 1 - F(t_0))$. ∎

By taking $t_0 = 0$ we obtain the following corollary.

Corollary 2.2. *Let X be an income random variable with Lorenz curve L. If* $\varepsilon_{\overline{L}}(p)$ *is increasing (decreasing), then*

$$X \leq_L X_{(t,\infty)} \ (\geq_L) \ \textit{for all } t.$$

Using (1) and (4), we see that the elasticity of the dual Lorenz curve can be written as

$$\varepsilon_{\overline{L}}(p) = \frac{F^{-1}(1-p)}{E\left[X\,|X > F^{-1}(1-p)\right]},\ 0 < p < 1. \tag{8}$$

Thus, the increasing (or decreasing) monotonicity of $\varepsilon_{\overline{L}}(p)$ in the interval $(0, 1-F(t_0))$ is equivalent to the increasing (respectively, decreasing) monotonicity of the function

$$e(t) = E\left[\frac{X}{t}\,|X > t\right]$$

in the interval (t_0, ∞). For an income t, the function $e(t)$ represents the expected proportional income to t, for incomes greater than t. This function was used by Belzunce, Candel and Ruiz (1998) to characterize the effect of truncation of a random variable X on the Lorenz curve. They say that X is DMLPRI (decreasing mean proportional residual income) if $e(t)$ is decreasing in t. From the above observation, we can equivalently say that X is DMLPRI if $\varepsilon_{\overline{L}}(p)$ is decreasing. It is worth noting that Theorem 2.1 and the rest of results in this paper involving $\varepsilon_{\overline{L}}(p)$ can be easily reformulated in terms of the curve $e(t)$.

Bhattacharya (1963) showed that the Lorenz curve of a lower truncated income distribution is independent of the point of truncation if, and only if, the incomes follow the Pareto law, with distribution function

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^a,\ \theta > 0, a > 0, x > \theta. \tag{9}$$

Now, combining the result of Bhattacharya with Theorem 2.1, we can characterize the Pareto distribution in terms of the elasticity of the dual Lorenz curve.

Corollary 2.3. *Let X be an income random variable with Lorenz curve L. Then, X follows the Pareto distribution if and only if* $\overline{L}(p)$ *has a constant elasticity.*

Similar results can be stated for upper truncations. If we denote by $X_{(0,s)} = \{X \mid X \leq s\}$ the upper truncated random variable at income s, it can be shown that the corresponding Lorenz curve $L_{(0,s)}(p)$ satisfies

$$L_{(0,s)}(p) = \frac{L(F(s)p)}{L(F(s))},\ 0 \leq p \leq 1. \tag{10}$$

The comparison of upper truncations of a random variable X is characterized in terms of the elasticity of the Lorenz curve

$$\varepsilon_L(p) = \frac{d\log L(p)}{d\log p} = \frac{pL'(p)}{L(p)},\ 0 < p < 1.$$

The proof of the following result follows the same lines as the proof of Theorem 2.1 and therefore it is omitted.

Theorem 2.4. *Let X be an income random variable with Lorenz curve L and let $s_0 \in [0,1]$. Then*

$$X_{(0,s)} \leq_L X_{(0,s')}\ (\geq_L)\ \textit{for all}\ s < s' < s_0 \tag{11}$$

if and only if $\varepsilon_L(p)$ is increasing (decreasing) in $[F(s_0), 1]$.

Moothathu (1986) showed that the Lorenz curve is unchanged by upper truncation if, and only if, incomes follow a power law, with distribution function

$$F(x) = \left(\frac{x}{\lambda}\right)^a,\ \lambda > 0, a > 0, 0 < x < \lambda. \tag{12}$$

The combination of this result with Theorem 2.4 let us characterize the power distributions in terms of the elasticity of the Lorenz curve.

Corollary 2.5. *Let X be an income random variable with Lorenz curve L. Then, X follows the power distribution if and only if $L(p)$ has a constant elasticity.*

3. The effect of truncations on the starshaped order and the expected proportional shortfall order

In Section 2 we have shown that the effect of truncation on the inequality depends on the elasticities of the Lorenz curve $L(p)$ and its dual $\overline{L}(p) = 1 - L(p)$. However, the Lorenz curve is not the only tool for comparing income distributions in terms of concentration. The Lorenz order is a pure inequality order, in the sense that it is consistent with the well-known Pigou–Dalton Transfer Principle, which demands that a transfer from a richer person to a poorer person of less than the difference in their income unambiguously reduces inequality. When we compare income distributions in terms of relative status of people or relative deprivation (rather than in terms of inequality), some other orderings, such as the starshaped order and the expected proportional shortfall order, can also be considered (see Shaked and Shantikumar, 2007, and Belzunce et al., 2012, 2013, 2016, for properties and applications of these orders) and it is of interest to investigate whether similar results for truncated distributions can be given taking into account the elasticity of some related functions like, for example, the quantile function.

First, we define these orders.

Definition 3.1. *Given two income random variables X and Y, with distribution functions F and G, respectively, then:*

(i) We say that X is smaller than Y in the starshaped order (denoted by $X \leq_* Y$) if

$$\frac{G^{-1}(p)}{F^{-1}(p)} \text{ is increasing in } p \in (0,1).$$

(ii) We say that X is smaller than Y in the expected proportional shortfall order (denoted by $X \leq_{ps} Y$) if

$$E\left[\left(\frac{X-F^{-1}(p)}{F^{-1}(p)}\right)^+\right] \leq E\left[\left(\frac{Y-G^{-1}(p)}{G^{-1}(p)}\right)^+\right] \text{ for } p \in (0,1),$$

where $(x)^+ = x$ if $x \geq 0$ and $(x)^+ = 0$ if $x < 0$.

It can be shown (see Theorem 2.11 in Belzunce et al., 2012) that $X \leq_{ps} Y$ if and only if

$$\int_p^1 \left[\frac{F^{-1}(t)}{F^{-1}(p)}\right] dt \leq \int_p^1 \left[\frac{G^{-1}(t)}{G^{-1}(p)}\right] dt, \text{ for all } p \in (0,1). \tag{13}$$

On the other hand, it is well-known that

$$X \leq_* Y \Longrightarrow X \leq_{ps} Y \Longrightarrow X \leq_L Y.$$

The next result shows that the effect of truncations on the starshaped order depends on the the elasticities of the quantile function and the inverse of the survival function.

Theorem 3.2. *Let X be an absolutely continuous income random variable with distribution function F and survival function $\overline{F} = 1 - F$. Denote by $\varepsilon_{F^{-1}}(p)$ the elasticity of the quantile function $F^{-1}(p)$ and by $\varepsilon_{\overline{F}^{-1}}(p)$ the elasticity of the inverse survival function $\overline{F}^{-1}(p)$. Then*

(i) $X_{(t,\infty)} \leq_* X_{(t',\infty)}$ $(\geq_*)$ for all $t < t'$ if and only if $\varepsilon_{\overline{F}^{-1}}(p)$ is increasing (decreasing) in $p \in (0,1)$.

(ii) $X_{(0,s)} \leq_* X_{(0,s')}$ $(\geq_*)$ for all $s < s'$ if and only if $\varepsilon_{F^{-1}}(p)$ is increasing (decreasing) in $p \in (0,1)$.

Proof. Let f be density function of X. In order to prove (i), observe that

$$\varepsilon_{\overline{F}^{-1}}(p) = \frac{-p}{f\left(\overline{F}^{-1}(p)\right)\overline{F}^{-1}(p)}, \text{ for } p \in (0,1).$$

Belzunce, Candel and Ruiz (1995, Theorem 3) showed that $X_{(t,\infty)} \leq_* X_{(t',\infty)}$ $(\geq_*)$ for all $t < t'$ if and only if the function

$$\frac{xf(x)}{1-F(x)} \text{ is decreasing (increasing).} \tag{14}$$

By making the change $x = \overline{F}^{-1}(p) = F^{-1}(1-p)$, we see that (14) is equivalent to say that

$$\frac{\overline{F}^{-1}(p)\, f\left(\overline{F}^{-1}(p)\right)}{p} \text{ is increasing (decreasing),}$$

which holds if, and only if, $\varepsilon_{\overline{F}^{-1}}(p)$ is increasing (decreasing). Part (ii) is proven similarly by using Theorem 4 of Belzunce, Candel and Ruiz (1995). ■

Next we show that the effect of trunctions on the expected proportional shortfall order depends, like in the case of the Lorenz order, on the elasticities of the Lorenz curve an its dual.

Theorem 3.3. *Let X be an absolutely continuous income random variable with distribution function F and survival function $\overline{F} = 1 - F$. Then*

(i) $X_{(t,\infty)} \leq_{ps} X_{(t',\infty)}$ $(\geq_{ps})$ for all $t < t'$ if and only if $\varepsilon_{\overline{L}}(p)$ is increasing (decreasing) in $p \in (0,1)$.

(ii) $X_{(0,s)} \leq_{ps} X_{(0,s')}$ $(\geq_{ps})$ for all $s < s'$ if and only if $\varepsilon_L(p)$ is increasing (decreasing) in $p \in (0,1)$.

Proof. We only prove the case $\leq_{ps}$ of part (i), the case $\geq_{ps}$ and part (ii) are proven similarly. First observe from (8) that $\varepsilon_{\overline{L}}(p)$ can be written as

$$\varepsilon_{\overline{L}}(p) = \frac{pF^{-1}(1-p)}{\int_{1-p}^{1} F^{-1}(t)dt}, \quad 0 < p < 1. \tag{15}$$

Let $F_{(t,\infty)}(x)$ be the distribution function of $X_{(t,\infty)}$ given by (6) and let

$$F_{(t,\infty)}^{-1}(u) = F^{-1}\left[(1-F(t))\,u + F(t)\right],\ u \in (0,1), \tag{16}$$

be the corresponding quantile function. Suppose that

$$X_{(t,\infty)} \leq_{ps} X_{(t',\infty)} \text{ for all } t < t'$$

or equivalently, using (13), that

$$\int_p^1 \left[\frac{F^{-1}_{(t,\infty)}(u)}{F^{-1}_{(t,\infty)}(p)}\right] du \leq \int_p^1 \left[\frac{F^{-1}_{(t',\infty)}(u)}{F^{-1}_{(t',\infty)}(p)}\right] du, \text{ for all } p \in (0,1). \tag{17}$$

From (16) we see that (17) is equivalent to

$$\frac{\int_p^1 F^{-1}\left[(1-F(t))\,u+F(t)\right] du}{F^{-1}\left[(1-F(t))\,p+F(t)\right]} \leq$$

$$\frac{\int_p^1 F^{-1}\left[(1-F(t'))\,u+F(t')\right] du}{F^{-1}\left[(1-F(t'))\,p+F(t')\right]},\ 0<t<t'<1,\ p \in (0,1). \tag{18}$$

A change of variable shows that (18) holds if and only if

$$\frac{\int_{(1-F(t))p+F(t)}^1 F^{-1}(x)\,dx}{(1-F(t))\,F^{-1}\left[(1-F(t))\,p+F(t)\right]} \leq \tag{19}$$

$$\frac{\int_{(1-F(t'))p+F(t')}^1 F^{-1}(x)\,dx}{(1-F(t'))\,F^{-1}\left[(1-F(t'))\,p+F(t')\right]},\ 0<t<t'<1,\ p \in (0,1).$$

Substituting $v=(1-F(t))\,p+F(t)$ and $u=(1-F(t'))\,p+F(t')$ we see that (19) is satisfied if and only if

$$\frac{\int_v^1 F^{-1}(x)\,dx}{(1-v)\,F^{-1}(v)} \leq \frac{\int_u^1 F^{-1}(x)\,dx}{(1-u)\,F^{-1}(u)} \text{ for all } 0<v<u<1 \tag{20}$$

or, equivalently, if (15) is decreasing in p. ■

4. Some models with $\varepsilon_{\overline{L}}(p)$ monotone

From the results in previous sections, the monotonicity of the elasticity of the dual Lorenz curve of a population may indicate a possible underestimation of the inequality (as measured by the Lorenz curve) and the feeling of relative deprivation (as measured by the expected proportional shortfall function) as reported by tax returns. The economic literature contains many parametric models for the Lorenz curve (see, for example, the papers by Kakwani and Podder, 1973; Rasche et al., 1980; Gupta, 1984; Aggarwal, 1984; Arnold, 1986; Arnold et al., 1987; Villaseñor and Arnold, 1989; Basmann et al., 1990; Ortega et al., 1991; Ryu and Slottje, 1996; Sarabia, 1997; Sarabia, Castillo and Slottje, 1999, 2001; Sarabia and Pascual, 2002; Rohde, 2009; Wang, Smyth and Ng, 2009; Sarabia et al., 2010 and Sordo, Navarro and Sarabia, 2014). In this section we collect some models such that the dual Lorenz curve has monotone elasticity.

4.1. Power Lorenz curve

The power Lorenz curve is given by $L(p) = p^k$, with $k \geq 1$ and its dual is given by

$$\overline{L}(p) = 1 - (1-p)^k \, , \; k \geq 1. \tag{21}$$

For $k = 1$, we have $\varepsilon_{\overline{L}}(p) = \varepsilon_L(p) = k$. In order to show that (21) is decreasing for $k > 1$, note that

$$\varepsilon_{\overline{L}}(p) = \frac{kp\,(1-p)^{k-1}}{1-(1-p)^k}.$$

Differentiating with respect to p, it is not hard to see that

$$\varepsilon'_{\overline{L}}(p) \leq 0 \text{ if and only } 1 - pk \leq (1-p)^k \, . \tag{22}$$

Now, define the auxiliary function

$$h(p) = (1-p)^k - (1-pk) \, , \;\; p \in [0,1] \, , \; k > 1.$$

It is easy to see that h is increasing on $[0,1]$. Since $h(0) = 0$, it follows that $h(p) \geq 0$ for every $p \in [0,1]$. This implies that $(1-p)^k \geq 1 - pk$ for every p in $[0,1]$ and from (22) it follows that $\varepsilon_{\overline{L}}(p)$ is decreasing.

4.2. Distorted Lorenz curves

Sordo, Navarro and Sarabia (2014) considered a general method of modeling a family of Lorenz curves by distorting a baseline Lorenz curve, L, as follows

$$L_h(p) = h(L(p)), 0 \leq p \leq 1, \tag{23}$$

where h is a convex distortion function (that is, an increasing function from $[0,1]$ to $[0,1]$ such that $h(0) = 0$ and $h(1) = 1$) and showed that a large number of parametric models for the Lorenz curve adopt the form (23). In this section we provide conditions on the distortion h under which the elasticity of $\overline{L}_h$ (the dual of the distorted curve L_h) inherits the monotonicity of the elasticity of $\overline{L}$ (the dual of the initial curve L). In that follows, denote by $\overline{h}(p) = 1 - h(1-p)$, for $0 \leq p \leq 1$ (observe that $h(p)$ is a convex distortion function if and only if $\overline{h}$ is a concave distortion function).

Theorem 4.1. *Let $L(p)$ be a Lorenz curve and let h be a convex distortion function. Let $L_h(p)$ be a distorted Lorenz curve (DLC) of the form (23) and let $\overline{L}_h(p)$ be its dual. Then*

$$\varepsilon_{\overline{L}_h}(p) = \varepsilon_{\overline{L}}(p) \cdot \varepsilon_{\overline{h}}(\overline{L}(p)) \tag{24}$$

Proof. From (4) and (23) we obtain

$$\varepsilon_{\overline{L}_h}(p) = \frac{pL_h'(1-p)}{1-L_h(1-p)} = \frac{p\,h'(L(1-p))\,L'(1-p)}{1-h(L(1-p))} \text{ for } 0<p<1.$$

Using that $L(1-p) = 1-\overline{L}(p)$ and rearranging the expression above we obtain

$$\varepsilon_{\overline{L}_h}(p) = \frac{p\,L'(1-p)}{\overline{L}(p)} \cdot \frac{\overline{L}(p)\,h'(1-\overline{L}(p))}{1-h(1-\overline{L}(p))}$$

which is (24). ■

Corollary 4.2. *If $\overline{L}$ and $\overline{h}$ have increasing (respectively, decreasing) elasticities, then $\overline{L}_h$ has increasing (respectively, decreasing) elasticity.*

Next, we give some examples of families of DLC of the form (23) such that $\varepsilon_{\overline{L}_h}(p)$ is monotone.

4.2.1. The class $L_\delta(p) = 1-[1-L(p)]^\delta$

The dual of the convex distortion function $h(t) = 1-(1-t)^\delta$, $0<\delta\leq 1$ has constant elasticity. Therefore, if $L(p)$ is a baseline Lorenz curve such that $\varepsilon_{\overline{L}}(p)$ is increasing (respectively, decreasing) it follows from Corollary 4.2. that the dual of a DLC of the form

$$L_\delta(p) = 1-[1-L(p)]^\delta, 0<\delta\leq 1,$$

has increasing (respectively decreasing) elasticity.

4.2.2. The hierachical class of Sarabia et al. (1999)

Let h be the convex distortion function defined by $h(t) = t^k, k\geq 1$. We know, from Section 4.1, that $\varepsilon_{\overline{h}}(p)$ is decreasing. Given a baseline Lorenz curve $L(p)$, Sarabia et al. (1999) considered a hierachical class of Lorenz curves of the form

$$L_k(p) = [L(p)]^k, k\geq 1$$

If $\varepsilon_{\overline{L}}(p)$ is decreasing, it follows from Corollary 4.2 that the elasticity of $\overline{L}_k(p)$ is decreasing. As a consequence, the elasticity of the curve $\overline{L}_{k,\delta}(p)$, where

$$L_{k,\delta}(p) = \left[1-(1-p)^{\delta}\right]^{k}, k \geq 1, 0 < \delta \leq 1, \tag{25}$$

is also decreasing (the curve (25) is one of the Lorenz curves in the Pareto hierarchy considered by Sarabia et al., 1999).

4.2.3. The class $L_\theta(p) = \dfrac{\theta L(p)}{1-(1-\theta)L(p)}$

Let h be the convex distortion function given by $h(t) = \dfrac{\theta t}{1-(1-\theta)t}$, with $0 < \theta \leq 1$. It is easy to prove that the elasticity of $\overline{h}$ is the function

$$\varepsilon_{\overline{h}}(t) = \frac{\theta}{\theta + (1-\theta)t},$$

which is decreasing in the interval $(0,1)$. Therefore, it follows from Corollary 4.2 that if $\overline{L}$ has decreasing elasticity, the family of DLC of the form

$$L_\theta(p) = \frac{\theta L(p)}{1-(1-\theta)L(p)}, 0 \leq p \leq 1, 0 < \theta \leq 1,$$

considered by Sordo, Navarro and Sarabia (2014) has decreasing elasticity.

4.2.4. Wang-Smyth-Ng model

Let h be the convex distortion function defined by

$$h_{\beta,\gamma}(t) = 1-(1-t)\exp[-\gamma[1-(1-p)^{1/\beta}]], 0 < \beta \leq 1, \gamma > 0.$$

The elasticity of $\overline{h}_{\beta,\gamma}$, given by

$$\varepsilon_{\overline{h}_{\beta,\gamma}}(t) = 1 + \frac{\gamma}{\beta} t^{1/\beta},$$

is increasing in $t \in (0,1)$. From Corollary 4.2, if $\overline{L}$ has an increasing elasticity then

$$h_{\beta,\gamma}(L(p)), 0 < \beta \leq 1, \gamma > 0, \tag{26}$$

has also increasing dual elasticity. The family (26) was considered by Sordo, Navarro and Sarabia (2014). In particular, by taking $L(p) = 1-(1-p)^{\beta}$, we obtain the class of Lorenz curves suggested by Wang, Smyth and Ng (2009). All the curves in this class have increasing elasticity.

5. An illustration using real data

Personal income can be measured using different sources of information. In Europe, the main source is the European Union Survey of Income and Living Conditions (EU-SILC) conducted by the Central Statistics Office. Alternative sources include, among other surveys and administrative data (such as those from Social Security records), data from tax income returns. In Spain, for example, taxation microdata are available under request from the Institute of Fiscal Studies (IFE), an institution attached to the Ministry of the Economy through the State Secretariat for Taxation and Budgets. Although, undoubtedly, EU-SILC and tax income returns taken together complement each other, any analysis of inequality based on the separate interpretation of data from tax returns requires caution because these data exclude people with very low taxable income. If we ignore, for the sake of argument, some issues related to the nature of data[2], the study population (or tax filers) becomes a subset, obtained via lower truncation, of a hypothetical reference population which is the same as the reference population of EU-SILC. A possible underestimation of inequality as reported by tax returns may be anticipated using a simple visual of the elasticity plot of the adjusted dual Lorenz curve for this reference population.

In order to illustrate this issue, we have carried out a descriptive study of the function $\varepsilon_{\overline{L}}(p)$ using data from the EU-SILC 2010 survey, which provides income data of 225,987 households and covers 29 European countries. The variable under study is the "total disposable income of the household", adjusted to take into account that we are dealing with individuals who are members of households of different size and composition (we make this adjustment employing the modified OECD equivalence scale). The unit of analysis chosen is the individual; the income assigned to each individual is the total income of the household to which they belong, adjusted according to the equivalence scale to ensure comparability (see Eurostat, 2010).

Taking in mind the expression (8), we have computed the function $\varepsilon_{\overline{L}}(p)$ from data in the following way. If $x_{(i)}$ denotes the i-th ordered income in the sample of size n, and ω_i denotes its corresponding sample weight[3], for $i = 1,2,\ldots,n-1$, we calculate the points

2. For example, EU-SILC refers to individuals living in households and tax income returns refer to taxpayers. We are deliberately ignoring that members of the same family or household may file separate tax returns.

3. Due to the use of sophisticated sampling techniques of stratification, rotation and non-response adjustment, microdata provided by the EU-SILC survey are weighted according to specific sample designs.

$$\left(\frac{\sum_{k=1}^{i} \omega_k}{\sum_{k=1}^{n} \omega_k}, \frac{x_{(j)} \sum_{k=1}^{i} \omega_k}{\left(\sum_{k=1}^{j} \omega_k - \sum_{k=i+1}^{n} \omega_k \right) x_{(j)} + \sum_{k=j+1}^{n} \omega_k x_{(k)}} \right) \quad (27)$$

where j is the index such that $\sum_{k=1}^{j-1} \omega_k < \sum_{k=i+1}^{n} \omega_k \leq \sum_{k=1}^{j} \omega_k$ when $j \geq 2$ and $j = 1$ in case $0 < \sum_{k=i+1}^{n} \omega_k \leq \omega_1$. Observe that this set of points can be considered as an analog estimation of the graph of the elasticity of the dual Lorenz curve associated to the income

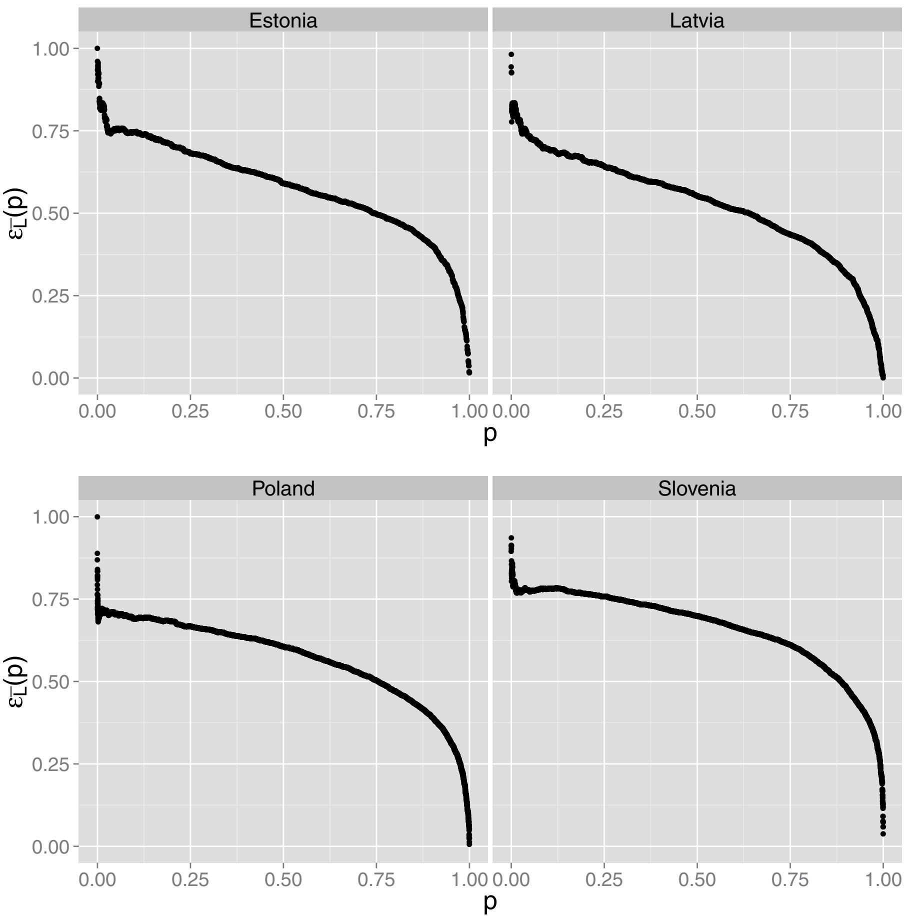

Figure 1: *$\varepsilon_{\overline{L}}(p)$ calculated for Estonia, Latvia, Poland and Slovenia. Source: Generated by authors based on data from EU-SILC 2010.*

distribution. However, since this study is purely illustrative, we have not considered the inferential properties of this estimation (and consequently, we can not discuss about the statistical significance of the results). For the sake of reproducibility, the R code used to calculate the set of points in (27) can be found at Github.[4]

From the results of this study, we conclude that the shapes of the computed elasticity curves can be grouped in basically two different types:

(a) For some countries, the dual Lorenz curve shows a decreasing elasticity. It follows from Corollary 2.2 and Theorem 3.3 that $X \geq_L X_{(t,\infty)}$ and $X \geq_{ep} X_{(t,\infty)}$ for all t, which suggests that statistics from tax returns may under-report inequality and relative deprivation (this is the case of Estonia, Latvia, Poland and Slovenia, see Figure 1).

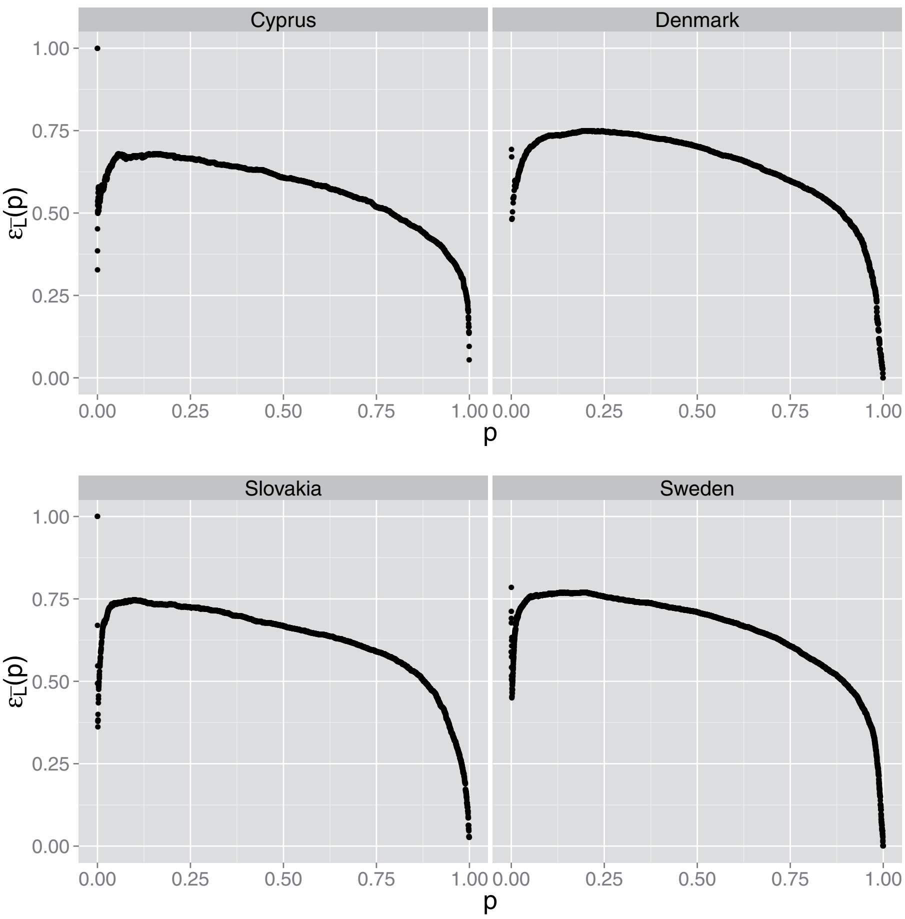

__Figure 2:__ $\varepsilon_{\overline{L}}(p)$ calculated for Cyprus, Denmark, Slovakia and Sweden. Source: Generated by authors based on data from EU-SILC 2010.

4. https://gist.github.com/AngelBerihuete/fdb11a7dc3ece81bcf5d6261a49af440

(b) For some countries the elasticity curve presents an U inverted shape (this is the case of Cyprus, Denmark, Slovakia and Sweden, see Figure 2). In Denmark, for example, $\varepsilon_{\overline{L}}(p)$ increases in $p \in (0, 0.21)$ and then decreases. From Theorem 2.1 this implies that

$$X_{(t,\infty)} \leq_L X_{(t',\infty)} \text{ for all } t, t' \text{ such that } F^{-1}(0.79) < \ t < t'.$$

Thus, for example, the inequality among the 10% richer of the population is higher than the inequality among the 20% richer. In this case, the elasticity $\varepsilon_{\overline{L}}(p)$ does not provide conclusive information on the relation, in terms of inequality, among tax filers and the entire population.

Acknowledgements

We are grateful to two anonymous referees and the associate editor for their valuable comments, which have improved the presentation and contents of this paper. Miguel A. Sordo and Carmen Ramos acknowledge the support received from Ministerio de Economía y Competitividad (Spain) under grant MTM2014-57559-P.

References

Aggarwal, V. (1984). On optimum aggregation of income distribution data. *Sankhya B*, 46, 343–355.

Arnold, B.C. (1986). A class of hyperbolic Lorenz curves. *Sankhya A*, 48, 427–436.

Arnold, B.C. (1987). Majorization and the Lorenz order: A brief introduction. *Lecture Notes in Statistics,* 43, Springer-Verlag.

Arnold, B.C., Brockett, P.L., Robertson, C.A. and Shu, B.Y. (1987). Generating ordered families of Lorenz curves by strongly unimodal distributions. *Journal of Business and Economic Statistics,* 5, 305–308.

Atkinson, A.B., Piketty, T. and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature,* 49, 3–71.

Basmann, R.L., Hayes, K.J., Slottje, D.J. and Johnson,J.D. (1990). A general functional form for approximating the Lorenz curve. *Journal of Econometrics*, 43, 77–90.

Belzunce, F., Candel, J. and Ruiz, J. M. (1995). Ordering of truncated distributions through concentration curves. *Sankhya A*, 57, 375–383.

Belzunce, F., Candel, J. and Ruiz, J. M. (1998). Ordering and asymptotic properties of residuals income distributions. *Sankhya B*, 60, 331–348.

Belzunce, F., Pinar, J.F., Ruiz, J.M. and Sordo, M.A. (2012). Comparisons of risks based on the expected proportional shortfall. *Insurance: Mathematics and Economics*, 51, 292–302.

Belzunce, F., Pinar, J.F., Ruiz, J.M. and Sordo, M.A. (2013). Comparison of concentration for several families of income distributions. *Statistics and Probability Letters*, 83, 1036–1045.

Belzunce, F., Martínez-Riquelme, C., Ruiz, J.M. and Sordo, M.A. (2016). On the Comparison of Relative Spacings with Applications. *Methodology and Computing in Applied Probability*. First online. Doi:10.1007/s11009-016-9479-6

Bhattacharya, N. (1963). A property of the Pareto distribution. *Sankhya B*, 25, 195–196.

Eurostat (2010). Description of target variables: cross-sectional and longitudinal. *EU-SILC 065/2010*, Eurostat, Luxembourg.

Greselin, F. and Zitikis, R. (2015). Measuring economic inequality and risk: a unifying approach based on personal gambles, societal preferences and references. *Preprint arXiv:1508.00127.*

Gupta, M.R. (1984). Functional form for estimating the Lorenz curve. *Econometrica*, 52, 1313–1314.

Kakwani, N.C. and Podder, N. (1973). On estimation of Lorenz curves from grouped observations. *International Economic Review*, 14, 278–292.

Moothathu, T.S.K. (1986). A characterization of power distribution through a property of the Lorenz curve. *Sankhya B*, 48, 262–265.

Ortega, P., Fernández, M.A., Ladoux, M. and García, A. (1991). A new functional form for estimating Lorenz curves. *The Review of Income and Wealth*, 37, 447–452.

Piketty, T. and Saez, E. (2003). Income Inequality in the United States: 1913-1998. *Quarterly Journal of Economics*, 118, 1–39.

Ramos, H., Peinado, A. and Ollero, J. (2013). Analysis of inequality in fertility curves fitted by gamma distributions. *SORT*, 37, 233–240.

Rasche, R.H., Gaffney, J., Koo, A. and Obst, N. (1980). Functional forms for estimating the Lorenz curve. *Econometrica*, 48, 1061–1062.

Rohde, N. (2009). An alternative functional form for estimating the Lorenz curve. *Economics Letters*, 105, 61–63.

Ryu, H. and Slottje, D. (1996). Two flexible functional forms for approximating the Lorenz curve. *Journal of Econometrics*, 72, 251–274.

Saez, E. and Zucman, G. (2014). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *NBER Working Paper 20625,* http://eml.berkeley.edu/˜saez/saez-zucmanNBER14wealth.pdf.

Sarabia, J.M. (1997). A hierarchy of Lorenz curves based on the generalized Tukey's lambda distribution. *Econometric Reviews*, 16, 305–320.

Sarabia, J.M., Castillo, E., and Slottje, D.J. (1999). An ordered family of Lorenz curves. *Journal of Econometrics*, 91, 43–60.

Sarabia, J.M., Castillo, E. and Slottje, D.J. (2001). An exponential family of Lorenz curves. *Southern Economic Journal*, 67, 748–756.

Sarabia, J.M. and Pascual, M. (2002). A class of Lorenz curves based on linear exponential loss functions. *Communications in Statistics-Theory and Methods*, 31, 925–942.

Sarabia, J. M., Gómez-Déniz E., Sarabia M. and Prieto, F. (2010). Revisiting a functional form for the Lorenz curve. *Economics Letters*, 107, 249–252.

Sarabia, J. M., Prieto, F. and Trueba, C. (2012). Modeling the probabilistic distribution of the impact factor. *Journal of Infometrics*, 6, 66–79.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic orders. Series: Springer Series in Statistics*, Springer.

Sordo, M.A., Navarro, J. and Sarabia, J.M. (2014). Distorted Lorenz curves: models and comparisons. *Social Choice and Welfare*, 42, 761–780.

Stone, C., Trisi, D., Sherman, A. and DeBot, B. (2015). *A Guide to Statistics on Historical Trends in Income Inequality*, Washington, DC: Center on Budget and Policy Priorities.

Villaseñor, J.A. and Arnold, B.C. (1989). Elliptical Lorenz curves. *Journal of Econometrics*, 40, 327–338.

Wang, Z., Smyth, R. and Ng, Y.K. (2009). A new ordered family of curves with an application to measuring income inequality and poverty in rural China. *China Economic Review,* 20, 218–235.