

Bivariate Lorenz curves: a review of recent proposals

Sarabia, José María (sarabiaj@unican.es)

*Departamento de Economía, Universidad de Cantabria
Avda. de los Castros s/n, 39005 Santander, Spain*

Jordá, Vanesa (jordav@unican.es)

*Departamento de Economía, Universidad de Cantabria
Avda. de los Castros s/n, 39005 Santander, Spain*

RESUMEN

La extensión de la curva de Lorenz al caso bidimensional y a dimensiones superiores a dos no es un problema trivial. Existen en la literatura tres propuestas de curvas de Lorenz bidimensionales. La primera de estas definiciones fue propuesta por Taguchi (1972a,b). A continuación Arnold (1983) estableció una segunda definición, que es una extensión natural de la curva de concentración. Esta definición no ha recibido mucha atención en la literatura económica. Finalmente, Koshevoy y Mosler (1996) introdujeron la tercera de las definiciones, haciendo uso del concepto de zonoide. Recientemente, Sarabia y Jordá (2013, 2014) han propuesto varias clases paramétricas de curvas de Lorenz bivariadas haciendo uso de la definición de Arnold.

En el presente trabajo se revisan estas tres definiciones. Se estudia el origen de cada una de ellas, así como sus principales propiedades. Para el caso de la curva de Arnold, se presentan algunas formas paramétricas propuestas para el caso bidimensional, con diferentes estructuras de dependencia y diferentes tipos de marginales. El trabajo termina comentando las extensiones de estas familias a dimensiones superiores a dos, y sus aplicaciones al estudio del bienestar considerando de forma conjunta varios atributos.

Palabras clave: curvas de Lorenz bivariadas, distribución de Sarmanov-Lee, índice de Gini bivariado, dominación estocástica, indicadores de bienestar.

Área temática: Aspectos cuantitativos de problemas económicos y empresariales.

ABSTRACT

The extension of the Lorenz curve to the bidimensional case and dimensions higher than two is not trivial. Three different proposals can be found in the literature. The first definition was proposed by Taguchi (1972a,b). Thereafter, Arnold (1983) developed a second definition which was a natural extension of the concentration curve. This proposal has not received much attention in the economic literature. Finally, Koshevoy and Mosler (1996) introduced the third definition using the concept of Lorenz Zonoid. Recently, Sarabia and Jordá (2013, 2014) proposed a number of parametric classes of bivariate Lorenz curves based on the Arnold's definition.

In this paper, the three existing definitions are revisited. We investigate their origin as well as the main properties of each of them. In the case of the Arnold's definition, some parametric forms are presented for the bivariate case with different dependence structures and different marginals. We finish this work by examining briefly the extension of these families to dimensions higher than two and describing their applications to well-being data when several attributes are jointly considered.

1 INTRODUCTION

In the last years, there is an increasing concern about the poor performance of GDP per capita as an indicator of well-being. In this sense, quality of life is seen as a multidimensional process which involves more than just economic aspects. The discontent with the hegemony of income for measuring human well-being has led to several attempts to develop a more comprehensive indicator using a multiattribute framework (see Gadrey and Jany-Catrice (2006) for a review).

It is worth nothing that we should account for two types of inequality in a multidimensional context. On the one hand, disparities within each dimension have to be assessed as in the unidimensional case. The second type of inequality is related

to the degree of correlation among variables, which would reflect the fact that a country where the same citizen is the first ranked individual in all of attributes is more unequal than a country where different individuals occupy the first position in different variables. An increasing number of multidimensional inequality measures which considers disparities within and across variables have been developed (Atkinson, 2003; Atkinson and Bourguignon, 1982; Kolm, 1977; Maasoumi, 1986; Tsui, 1995; 1999; Seth, 2011).

However, as in the unidimensional case, multidimensional inequality measures only provide summarized information about the evolution of disparities in well-being. If no dominance relationships can be achieved, some parts of the distribution may present different trends than those obtained using inequality measures. In unidimensional environments, the Lorenz curve provides relevant insights about the evolution of different parts of the distribution and it has been widely used for studying economic inequality as well as the distribution of non-income variables. In this context, the extension of the univariate Lorenz curve to higher dimensions is not an obvious task. The three existing definitions were proposed by Taguchi (1972a,b), Arnold (1983) and Koshevoy and Mosler (1996), who introduced the concepts of Lorenz zonoid and Gini zonoid index.

In this paper, the three existing definitions are revisited. We investigate their origin as well as the main properties of each of them. In the case of the Arnold's definition, some parametric forms are presented for the bivariate case with different dependence structures and different marginals. We finish this work by examining briefly the extension of these families to dimensions higher than two and describing their applications to well-being data when several attributes are jointly considered.

The contents of the paper are the following. In Section 2 we present preliminary results including the definition of univariate Lorenz and concentration curves. The three different definitions of bivariate Lorenz curves are studied and compared in

Section 3. In Section 4 we review the bivariate Sarmanov-Lee Lorez curve, where we review some closed expression for the bivariate curve and the corresponding bivariate Gini index, according to the Arnold's (1987) definition. The bivariate Gini index can be decomposed in two factors is given, which correspond to the equality within and between variables. Different models of bivariate Lorenz curves are reviewed in Section 5. Stochastic ordering are discussed in Section 6. Extension to higher dimensions are discussed in Section 7. An application to measurement multi-dimensional inequality in well-being is given in Section 8. Finally, some conclusions are given in Section 9.

2 PRELIMINARY RESULTS

2.1 Univariate Lorenz and Concentration Curves

We denote the class of univariate distributions functions with positive finite expectations by \mathcal{L} and denote by \mathcal{L}_+ the class of all distributions in \mathcal{L} with $F(0) = 0$ corresponding to non-negative random variables. We use the following definition by Gastwirth (1971).

Definition 1 *The Lorenz curve L of a random variable X with cumulative distribution function $F \in \mathcal{L}$ is,*

$$L(u; F) = \frac{\int_0^u F^{-1}(y) dy}{\int_0^1 F^{-1}(y) dy} = \frac{\int_0^u F^{-1}(y) dy}{E[X]}, \quad 0 \leq u \leq 1, \quad (1)$$

where

$$\begin{aligned} F^{-1}(y) &= \sup\{x : F(x) \leq y\}, & 0 \leq y < 1, \\ &= \sup\{x : F(x) < 1\}, & y = 1, \end{aligned}$$

is the right continuous inverse distribution function or quantile function corresponding to F .

An alternative parametric definition of the Lorenz curves is possible. For a random variable X in \mathcal{L}_+ with CDF F , we define its first moment distribution function $F_{(1)}$ as,

$$F_{(1)}(x) = \frac{\int_0^x t dF(t)}{\int_0^\infty t dF(t)}, \quad 0 \leq x \leq \infty.$$

Then, using previous definition, the set of points comprising the Lorenz curve is,

$$\{(F(x), F_{(1)}(x)), \quad x \in (0, \infty]\}. \tag{2}$$

Now, if we set a simple change of variable, we can obtain another alternative expression of the Lorenz curve given by,

$$L(u; F) = F_{(1)}(F^{-1}(u)), \quad 0 \leq u \leq 1. \tag{3}$$

An important advantage of definition (2) is that the quantile function is not involved.

Now, we introduce the concept of concentration curve introduced by Kakwani (1977). Let $g(x)$ be a continuous function of x such that its first derivative exists and $g(x) \geq 0$. If the mean $E_F[g(X)]$ exists, then one can define

$$L_g(y; F) = \frac{\int_0^y g(x) dF(x)}{E_F[g(X)]}$$

where $y = g(x)$ and $f(x)$ and $F(x)$ are respectively the probability density function (PDF) and the cumulative distribution function (CDF) of the random variable X . The implicit relation between $L_g(g(x); F)$ and $F(x)$ will be called the concentration curve of the function $g(X)$. The concentration curve admits the simple implicit representation,

$$L_g(u; F) = \frac{1}{E_F[g(X)]} \int_0^u g[F^{-1}(t)] dt,$$

which will be used in the following results.

3 THE THREE DIFFERENT DEFINITIONS OF BIVARIATE LORENZ CURVES

As was mentioned in Marshall, Olkin and Arnold (2011), the extension of the LC to higher dimensions was complicated for several reasons. The usual definition of the Lorenz curve involved either order statistics or the quantile function of the corresponding CDF and either of which concepts has a simple multivariate analog.

In this section we discuss the three definitions proposed in the literature for constructing a bivariate Lorenz curve. These definitions were suggested by Taguchi (1972a,b), Arnold (1983, 1987) and Koshevoy and Mosler (1996). We consider the bivariate definition, which can be extend to higher dimensions.

3.1 The Taguchi's (1972) definition of bivariate Lorenz curve

This definition was the first proposal in the literature of bivariate Lorenz curve.

Definition 2 *The Lorenz surface of F_{12} is the set of points $(s, t, L(s, t))$ in \mathbb{R}_+^3 defined by,*

$$s = \int_0^u \int_0^v f_{12}(x_1, x_2) dx_1 dx_2, \quad t = \frac{\int_0^u \int_0^v x_1 f_{12}(x_1, x_2) dx_1 dx_2}{E[X_1]},$$

$$L(s, t; F_{12}) = \frac{\int_0^u \int_0^v x_2 f_{12}(x_1, x_2) dx_1 dx_2}{E[X_2]}.$$

This definition was investigated by Taguchi (1972a,b, 1988).

3.2 Arnold's (1983, 1987) definition of bivariate Lorenz curve

The following definition was proposed by Arnold (1983, 1987) and it is an extension of (1) quite natural to higher dimensions. Let $\mathbf{X} = (X_1, X_2)^\top$ be a bivariate

random variable with bivariate probability distribution function F_{12} on \mathbb{R}_+^2 having finite second and positive first moments. We denote by F_i , $i = 1, 2$ the marginal CDF corresponding to X_i , $i = 1, 2$ respectively.

Definition 3 *The Lorenz surface of F_{12} is the graph of the function,*

$$L(u_1, u_2; F_{12}) = \frac{\int_0^{s_1} \int_0^{s_2} x_1 x_2 dF_{12}(x_1, x_2)}{\int_0^\infty \int_0^\infty x_1 x_2 dF_{12}(x_1, x_2)}, \quad (4)$$

where

$$u_1 = \int_0^{s_1} dF_1(x_1), \quad u_2 = \int_0^{s_2} dF_2(x_2), \quad 0 \leq u, v \leq 1.$$

The two-attribute Gini-Arnold index $GA(F_{12})$ is defined as,

$$GA(F_{12}) = 4 \int_0^1 \int_0^1 [u_1 u_2 - L(u_1, u_2; F_{12})] du_1 du_2, \quad (5)$$

where the egalitarian surface is given by $L_0(u_1, u_2; F_0) = u_1 u_2$. Previous definition has not been explored in detail in the literature. We highlight some of its properties:

1. The marginal Lorenz curves can be obtained as $L(u_1; F_1) = L(u_1, \infty; F_{12})$ and $L(u_2; F_2) = L(\infty, u_2; F_{12})$.
2. The bivariate Lorenz curve does not depend on changes of scale in the marginals.
3. If F_{12} is a product distribution function, then

$$L(u_1, u_2; F_{12}) = L(u_1; F_1)L(u_2; F_2),$$

which is just the product of the marginal Lorenz curves.

4. We denote by F_a the one-point distribution at $a \in \mathbb{R}_+^2$, that is, the egalitarian distribution at a . Then, the egalitarian distribution has bivariate Lorenz curve $L(u_1, u_2; F_a) = u_1 u_2$.

5. In the case of a product distribution, the two-attribute Gini-Arnold defined in (5) can be written as,

$$1 - GA(F_{12}) = [1 - G(F_1)][1 - G(F_2)].$$

The Arnold's Lorenz curve can be evaluated implicitly in some relevant bivariate families of income distributions.

3.3 The Koshevoy and Mosler (1996) Lorenz Curve

The following definition of Lorenz curve was initially proposed by Koshevoy (1995), where this author identifies the suitable definition. Then, the rest of results were obtained by Koshevoy and Mosler (1996, 1997) and Mosler (2002). Denote by \mathcal{L}_+^2 de set of all 2-dimensional non-negative random vectors X with finite positive marginal expectations. Let $\Psi^{(2)}$ denote the class of all measurable functions from \mathbb{R}_+^2 to $[0, 1]$.

Definition 4 Let $\mathbf{X} \in \mathcal{L}_+^2$. The Lorenz zonoid $L(\mathbf{X})$ of the random vector $\mathbf{X} = (X_1, X_2)^\top$ with distribution F_{12} is defined as,

$$\begin{aligned} L(\mathbf{X}) &= \left\{ \left(\int \psi(x) dF_{12}(x), \int \frac{x_1 \psi(x)}{E[X_1]} dF_{12}(x), \int \frac{x_2 \psi(x)}{E[X_2]} dF_{12}(x) \right) : \psi \in \Psi^{(2)} \right\}, \\ &= \left\{ \left(E[\psi(X)], \frac{E[X_1 \psi(X)]}{E[X_1]}, \frac{E[X_2 \psi(X)]}{E[X_2]} \right) : \psi \in \Psi^{(2)} \right\}. \end{aligned}$$

The Lorenz zonoid is a convex American football-subset of the 3-dimensional unit cube that includes the points $(0, 0, 0)$ and $(1, 1, 1)$. Extension to higher dimensions is quite direct. However, the computation of these formulas in parametric income distributions is not easy.

3.4 Comparison of the three definitions

In this section highlight the strengths and weaknesses of the three definitions of Lorenz curves.

Taguchi's definition

The strengths of this definition are the following

- It is easy to use in parametric families.
- It is related to some previous attempts by Lunetta (1972a,b).

On the other hand, the weakness of this definition are:

- The definition is not symmetric.
- Extension to higher dimensions does not look simple.

The Arnold's definition

The strengths of this definition are the following

- It is easy to use in broad classes of parametric families in dimension two (Sarabia and Jordá, 2014).
- It is related to some economic concepts in multidimensional utility theory (Muller and Trannoy, 2011).
- Extensions to higher dimensions (greater than two) are direct.
- Inequality indices can be obtained in a simple way in dimension two.

On the other hand, the weakness are:

- The computation of inequality measures for dimensions higher than two are difficult.
- The computation in dimensions higher than two for parametric families is complicated.

The Koshevoy and Mosler's definition

The strengths of this definition are:

- The extension to higher dimensions is direct.
- The economic interpretation is simple,

and the weakness are:

- The computation of the curve for dimensions two and higher than two are complicated.
- The computation in parametric families is not simple.

4 THE BIVARIATE SARMANOV-LEE CURVE

In this section we review the so-called bivariate Sarmanov-Lee Lorenz curve. Many of our results are based on a explicit expression of the bivariate Lorenz curve defined in (4). The bivariate Lorenz curve (4) admits the following simple explicit representation (see Sarabia and Jordá, 2014)

Lemma 1 *The bivariate Lorenz curve can be written in the explicit form,*

$$L(u_1, u_2; F_{12}) = \frac{1}{E[X_1 X_2]} \int_0^{u_1} \int_0^{u_2} A(x_1, x_2) dx_1 dx_2, \quad 0 \leq u_1, u_2 \leq 1, \quad (6)$$

where

$$A(x_1, x_2) = \frac{F_1^{-1}(x_1) F_2^{-1}(x_2) f_{12}(F_1^{-1}(x_1), F_2^{-1}(x_2))}{f_1(F_1^{-1}(x_1)) f_2(F_2^{-1}(x_2))}. \quad (7)$$

Now, let $\mathbf{X} = (X_1, X_2)^\top$ be a bivariate Sarmanov-Lee (SL) distribution with joint PDF (Lee, 1996, Sarmanov, 1966),

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) \{1 + w \phi_1(x_1) \phi_2(x_2)\}, \quad (8)$$

where $f_1(x)$ and $f_2(y)$ are the univariate PDF marginals, $\phi_i(t)$, $i = 1, 2$ are bounded nonconstant functions such that,

$$\int_{-\infty}^{\infty} \phi_i(t) f_i(t) dt = 0, \quad i = 1, 2,$$

and w is a real number which satisfies the condition $1 + w\phi_1(x_1)\phi_2(x_2) \geq 0$ for all x_1 and x_2 . We denote $\mu_i = E[X_i] = \int_{-\infty}^{\infty} tf_i(t)dt$, $i = 1, 2$, $\sigma_i^2 = var[X_i] = \int_{-\infty}^{\infty} (t - \mu_i)^2 f_i(t)dt$, $i = 1, 2$ and $\nu_i = E[X_i\phi_i(X_i)] = \int_{-\infty}^{\infty} t\phi_i(t)f_i(t)dt$, $i = 1, 2$. Properties of this family has been explored by Lee (1996). Moments and regressions of this family can be easily obtained. The product moment is $E[X_1X_2] = \mu_1\mu_2 + w\nu_1\nu_2$, and the regression of X_2 on X_1 is given by

$$E[X_2|X_1 = x_1] = \mu_2 + w\nu_2\phi_1(x_1).$$

The copula associated to (8) is,

$$C(u_1, u_2; w, \phi) = u_1u_2 + \int_0^{u_1} \int_0^{u_2} \tilde{\phi}_1(s_1)\tilde{\phi}_2(s_2)ds_1ds_2,$$

where $\tilde{\phi}_i(s_i) = \phi_i(F_i^{-1}(s_i))$, $i = 1, 2$, and $F_i(x_i)$ are the CDF of \mathbf{X} . The PDF of the copula associated to (8) is,

$$c(u_1, u_2; w, \phi) = \frac{\partial C(u_1, u_2; w, \phi)}{\partial u_1 \partial u_2} = 1 + w\phi_1(F_1^{-1}(u_1))\phi_2(F_2^{-1}(u_2)).$$

Note that (8) and its associated copula has two components: a first component corresponding to the marginal distributions and the second component which defines the structure of dependence, given by the parameter w and the functions $\phi_i(u)$, $i = 1, 2$. These two components will be translated to the structure of the associated bivariate Lorenz curve, and the corresponding bivariate Gini index.

The Sarmanov-Lee copula has several advantages: its joint PDF and CDF are quite simple; the covariance structure in general is not limited and its different probabilistic features (moments, conditional distributions...) can be obtained in a explicit form. On the other hand, the SL distribution includes several relevant special cases including the classical Farlie-Gumbel-Morgenstern distribution, and the variations proposed by Huang and Kotz (1999) and Bairamov and Kotz (2003).

4.1 The bivariate SL Lorenz Curve

The bivariate SL Lorenz curve was defined by Sarabia and Jordá (2013) and (2014) and was obtained using (8) in definition (4)

Theorem 1 *Let $\mathbf{X} = (X_1, X_2)^\top$ a bivariate Sarmanov-Lee distribution with joint PDF (8) with non-negative marginals satisfying $E[X_1] < \infty$, $E[X_2] < \infty$ and $E[X_1X_2] < \infty$. Then, the bivariate Lorenz curve is given by,*

$$L_{SL}(u_1, u_2; F_{12}) = \pi L(u_1; F_1)L(u_2; F_2) + (1 - \pi)L_{g_1}(u_1; F_1)L_{g_2}(u_2; F_2), \quad (9)$$

where

$$\pi = \frac{\mu_1\mu_2}{E[X_1X_2]} = \frac{\mu_1\mu_2}{\mu_1\mu_2 + w\nu_1\nu_2},$$

and $L(u_i; F_i)$, $i = 1, 2$ are the Lorenz curves of the marginal distribution X_i , $i = 1, 2$ respectively, and $L_{g_i}(u_i; F_i)$, $i = 1, 2$ represent the concentration curves of the random variables $g_i(X_i) = X_i\phi_i(X_i)$, $i = 1, 2$, respectively.

The interpretation of (9) is the following: the bivariate Lorenz curve can be written as a convex linear combination of two components: (a) a first component corresponding to the product of the marginal Lorenz curves (marginal component) and a second component corresponding to the product of the concentration Lorenz curves (structure dependence component).

4.2 Bivariate Gini index

The following result provides a convenient write of the two-attribute bivariate Gini defined in (5). This expression permits a simple decomposition of the equality in two factors (see Sarabia and Jordá, 2014).

Theorem 2 *Let $\mathbf{X} = (X_1, X_2)^\top$ be a bivariate Sarmanov-Lee distribution with bivariate Lorenz curve $L(u, v; F_{12})$. The two-attribute bivariate Gini index defined in*

(5) is given by,

$$1 - G(F_{12}) = \pi[1 - G(F_1)] \cdot [1 - G(F_2)] + (1 - \pi)[1 - G_{g_1}(F_1)] \cdot [1 - G_{g_2}(F_2)], \quad (10)$$

where $G(F_i)$, $i = 1, 2$ are the Gini indices of the marginal Lorenz curves, and $G_{g_i}(F_i)$, $i = 1, 2$ represent the concentration indices of the concentration Lorenz curves $L_{g_i}(u_i, F_i)$, $i = 1, 2$.

Then the overall equality (OE) given by $1 - G(F_{12})$ can be decomposed in two factors,

$$OE = EW + EB, \quad (11)$$

where

$$\begin{aligned} OE &= 1 - G(F_{12}), \\ EW &= \pi[1 - G(F_1)][1 - G(F_2)], \\ EB &= (1 - \pi)[1 - G_{g_1}(F_1)][1 - G_{g_2}(F_2)], \end{aligned}$$

and EW represents the equality within variables and the second factor EB represent the equality between variables. Note that the decomposition is well defined (since $0 \leq OE \leq 1$ and $0 \leq EW \leq 1$ and in consequence $0 \leq EB \leq 1$) and EB includes the structure of dependence of the underlying bivariate income distribution through the functions g_i , $i = 1, 2$.

5 Bivariate Lorenz curve Models

In this section we reviewed some models based on the previous methodology.

5.1 Bivariate Pareto Lorenz curve based on the FGM family

The following model was considered by Sarabia and Jordá (2014). Let $\mathbf{X} = (X_1, X_2)^\top$ be a bivariate FGM with classical Pareto marginals (Arnold, 1983). The

bivariate Lorenz curve associated to this family is,

$$L_{FGM}(u_1, u_2; F_{12}) = \pi_w L(u_1; \alpha_1) L(u_2; \alpha_2) + (1 - \pi_w) L_{g_1}(u_1; \alpha_1) L_{g_2}(u_2; \alpha_2), \quad (12)$$

where $L(u_i; \alpha_i) = 1 - (1 - u_i)^{1-1/\alpha_i}$, $0 \leq u \leq 1$, $i = 1, 2$ and $L_{g_i}(u_i; \alpha_i) = 1 - (1 - u_i)^{1-1/\alpha_i} [1 + 2(\alpha_i - 1)u_i]$, $0 \leq u \leq 1$, $i = 1, 2$ and $\pi_w = \frac{(2\alpha_1 - 1)(2\alpha_2 - 1)}{(2\alpha_1 - 1)(2\alpha_2 - 1) + w}$.

The bivariate Gini index can be written as,

$$G(\alpha_1, \alpha_2) = \frac{(3\alpha_1 - 1)(3\alpha_2 - 1)(2\alpha_1 + 2\alpha_2 - 3) + [h(\alpha_1, \alpha_2)]w}{(3\alpha_1 - 1)(3\alpha_2 - 1)[(1 - 2\alpha_1)(1 - 2\alpha_2) + w]},$$

where $h(\alpha_1, \alpha_2) = -3 - 4\alpha_1^2(\alpha_2 - 1)^2 + (5 - 4\alpha_2)\alpha_2 + \alpha_1(5 + \alpha_2(8\alpha_2 - 7))$.

5.2 Bivariate Sarmanov-Lee Lorenz curves with Beta and GB1 Marginals

This model was considered by Sarabia and Jordá (2013). Let $X_i \sim \mathcal{B}e(a_i, b_i)$, $i = 1, 2$ be two classical beta distributions with PDF,

$$f_i(x_i; a_i, b_i) = \frac{x_i^{a_i-1}(1-x_i)^{b_i-1}}{B(a_i, b_i)}, \quad 0 \leq x_i \leq 1, \quad i = 1, 2$$

where $B(a_i, b_i) = \Gamma(a_i)\Gamma(b_i)/\Gamma(a_i + b_i)$, for $i = 1, 2$. This distribution has been proposed as a model of income distribution by McDonald (1984) and more authors. If we consider the mixing functions $\phi_i(x_i) = x_i - \mu_i$, where $\mu_i = E[X_i] = a_i/(a_i + b_i)$, $i = 1, 2$, the bivariate SL distribution is,

$$f_{12}(x_1, x_2) = f_1(x_1; a_1, b_1) f_2(x_2; a_2, b_2) \left\{ 1 + w \left(x_1 - \frac{a_1}{a_1 + b_1} \right) \left(x_2 - \frac{a_2}{a_2 + b_2} \right) \right\}, \quad (13)$$

where w satisfies $\frac{-(a_1+b_1)(a_2+b_2)}{\max\{a_1a_2, b_1b_2\}} \leq w \leq \frac{(a_1+b_1)(a_2+b_2)}{\max\{a_1b_2, a_2b_1\}}$. An interesting property of this family is that it can be expressed as a linear combination of products of univariate beta densities.

The LC of the univariate classical beta distribution is given by (Sarabia, 2008),

$$L(u_i; F_i) = G_{\mathcal{B}e(a_i+1, b_i)} [G_{\mathcal{B}e(a_i, b_i)}^{-1}(u_i)], \quad i = 1, 2, \quad (14)$$

where $G_{\mathcal{B}e(a,b)}(z)$ represents the CDF of a classical Beta distribution. On the other hand, concentration curve can be written as,

$$L_{g_i}(u_i; F_i) = \frac{E[X_i^2]G_{\mathcal{B}e(a_i+2,b_i)}(G_{\mathcal{B}e(a_i,b_i)}^{-1}(u_i)) - E[X_i]^2L(u_i; F_i)}{\text{var}[X_i]}, \quad i = 1, 2. \quad (15)$$

Note that $\nu_i = E_{F_i}[X_i\phi_i(X_i)] = \text{var}[X_i]$, $i = 1, 2$. If we combine (14) with (15) in (9), we obtain the bivariate beta Lorenz curve. This model can be extended easily to the SL distribution with generalized beta of the first type (GB1) marginals (McDonald, 1984).

5.3 Distributions with lognormal marginals

The lognormal distribution plays an important role in the analysis of income and wealth data. There are several classes of bivariate distributions with lognormal distributions. Perhaps, the most natural definition of bivariate lognormal distribution is given in terms of a monotone marginal transformation of the classical bivariate normal distribution. Other alternatives have been described by Sarabia et al (2007).

5.4 Other classes of bivariate Lorenz curves

Other alternative families of bivariate Lorenz curves can also be considered, including models with marginals specified in terms of univariate Lorenz curves (see Sarabia et al., 1999) and models based on mixture of distributions (see Sarabia et al, 2005), which permits to incorporate heterogeneity factors in the inequality analysis. Other classes of bivariate Lorenz curves can be defined in conditional terms.

6 STOCHASTIC ORDERINGS

In this section we review some stochastic orderings related with the Lorenz curves previously defined. The usual Lorenz ordering is given in the following definition.

Definition 5 Let $X, Y \in \mathcal{L}$ with CDFs F_X and F_Y and Lorenz curves $L(u; F_X)$ and $L(u; F_Y)$, respectively. Then, X is less than Y in the Lorenz order denoted by $X \preceq_L Y$ if $L(u; F_X) \geq L(u; F_Y)$, for all $u \in [0, 1]$.

A similar definition for concentration curves can also be considered.

In the multivariate case, we denote by \mathcal{L}_+^k the set of all k -dimensional nonnegative random vectors \mathbf{X} and \mathbf{Y} with finite marginal expectations, that is $E[X_i] \in \mathbb{R}_{++}$. We define the following orderings (Marshall, Olkin and Arnold, 2011; Sarabia and Jordá, 2014)

Definition 6 Let $\mathbf{X}, \mathbf{Y} \in \mathcal{L}_+^k$, and we define the orders

- (i) $\mathbf{X} \preceq_L \mathbf{Y}$ if $L(u; F_{\mathbf{X}}) \geq L(u; F_{\mathbf{Y}})$,
- (ii) $\mathbf{X} \preceq_{L1} \mathbf{Y}$ if $E \left[g \left(\frac{X_1}{E[X_1]}, \dots, \frac{X_k}{E[X_k]} \right) \right] \leq E \left[g \left(\frac{Y_1}{E[Y_1]}, \dots, \frac{Y_k}{E[Y_k]} \right) \right]$, for every continuous convex function $g : \mathbb{R}^k \rightarrow \mathbb{R}$ for which expectations exists,
- (iii) $\mathbf{X} \preceq_{L2} \mathbf{Y}$ if $\sum_{i=1}^k a_i X_i \preceq_L \sum_{i=1}^k a_i Y_i$ for every $a \in \mathbb{R}^k$
- (iv) $\mathbf{X} \preceq_{L3} \mathbf{Y}$ if $\sum_{i=1}^k b_i X_i \preceq_L \sum_{i=1}^k b_i Y_i$ for every $b \in \mathbb{R}_+^k$
- (v) $\mathbf{X} \preceq_{L4} \mathbf{Y}$ if $X_i \preceq_L Y_i$, $i = 1, 2, \dots, k$

It can be verified (Marshall, Olkin and Arnold, 2011) that $\preceq_{L1} \Rightarrow \preceq_{L2}$ and $\preceq_{L2} \Leftrightarrow \preceq_L \Rightarrow \preceq_{L3} \Rightarrow \preceq_{L4}$. We have the next theorem.

Theorem 3 Let $\mathbf{X}, \mathbf{Y} \in \mathcal{L}_+^2$ with the same Sarmanov-Lee copula. Then, if $X_i \preceq_L Y_i$, and $X_i \preceq_{Lg_i} Y_i$ $i = 1, 2$, then $\mathbf{X} \preceq_L \mathbf{Y}$.

7 EXTENSIONS TO HIGHER DIMENSIONS

In this section we review extensions of the definitions to dimensions higher than two. First, we consider the general definition of Lorenz curve (4).

7.1 The Arnold's Lorenz Curve

Let $\mathbf{X} = (X_1, \dots, X_m)^\top$ be a random vector in \mathcal{L}_+^m with joint CDF $F_{12\dots m}(x_1, \dots, x_m)$. The multivariate Arnold's Lorenz curve can be defined as,

$$L(\mathbf{u}; F_{12\dots m}) = \frac{\int_0^{s_1} \dots \int_0^{s_m} \prod_{i=1}^m x_i dF_{12\dots m}(x_1, \dots, x_m)}{E \left[\prod_{i=1}^m X_i \right]},$$

where $u_i = \int_0^{s_i} dF_i(x_i)$, $i = 1, 2, \dots, m$ and $0 \leq u_1, \dots, u_m \leq 1$.

The m -dimensional version of the Sarmanov-Lee distribution is defined as,

$$f(x_1, \dots, x_m) = \left\{ \prod_{i=1}^m f_i(x_i) \right\} \{1 + R_{\phi_1 \dots \phi_m \Omega_m}(x_1, \dots, x_m)\}, \quad (16)$$

where

$$\begin{aligned} & R_{\phi_1 \dots \phi_m \Omega_m}(x_1, \dots, x_m) \\ &= \sum_{1 \leq i_1 \leq i_2 \leq m} w_{i_1 i_2} \phi_{i_1}(x_{i_1}) \phi_{i_2}(x_{i_2}) \\ &= \sum_{1 \leq i_1 \leq i_2 \leq i_3 \leq m} w_{i_1 i_2 i_3} \phi_{i_1}(x_{i_1}) \phi_{i_2}(x_{i_2}) \phi_{i_3}(x_{i_3}) \\ &+ \dots \\ &+ w_{12\dots m} \prod_{i=1}^m \phi_i(x_i), \end{aligned}$$

and $\Omega_m = \{w_{i_1 i_2}, w_{i_1 i_2 i_3}, \dots, w_{12\dots m}\}$, $k \geq 2$ and $m \geq 3$. The set of real numbers Ω_m is such that $1 + R_{\phi_1 \dots \phi_m \Omega_m}(x_1, \dots, x_m) \geq 0$, $\forall (x_1, \dots, x_m) \in \mathbb{R}^m$.

Using expression (16), the m -variate Lorenz curve takes de form,

$$L(\mathbf{u}; F_{12\dots m}) = w_0 \prod_{i=1}^m L(u_i; F_i) + \sum_{i_1, \dots, i_k} w_{i_1 \dots i_k} \prod_{j=i_1}^{i_k} L_{g_j}(u_j; F_j) \prod_{j=i_k+1}^{i_m} L(u_j; F_j) \\ + w_{12\dots m} \prod_{i=1}^m L_{g_i}(u_i; F_i),$$

where $g_i(x_i) = x_i \phi_i(x_i)$, $i = 1, 2, \dots, m$. Similar expression for the multivariate Gini is also possible.

7.2 The Koshevoy and Mosler LC

One of the main advantages of the Koshevoy and Mosler definition is the easiness to extend to higher dimensions. Let \mathcal{L}_+^m the set of all m -dimensional nonnegative random vectors $\mathbf{X} = (X_1, \dots, X_m)^\top$ with finite and positive marginal expectations, and let $\Psi^{(m)}$ the class of all measurable functions from \mathbb{R}_+^m to $[0, 1]$. The Lorenz zonoid $L(\mathbf{X})$ of the random vector \mathbf{X} with joint CDF F is,

$$L(\mathbf{X}) = \left\{ \left(\int \psi(\mathbf{x}) dF(\mathbf{x}), \int \frac{x_1 \psi(\mathbf{x})}{E(X_1)} dF(\mathbf{x}), \dots, \int \frac{x_m \psi(\mathbf{x})}{E(X_m)} dF(\mathbf{x}) : \psi \in \Psi^{(m)} \right) \right\}, \\ = \left\{ \left(E\psi(\mathbf{X}), \frac{E(X_1 \psi(\mathbf{X}))}{E(X_1)}, \dots, \frac{E(X_m \psi(\mathbf{X}))}{E(X_m)} : \psi \in \Psi^{(m)} \right) \right\}. \quad (17)$$

In consequence, the Lorenz zonoid is thus a convex football shaped subset of the $(m + 1)$ -dimensional unit cube that includes the points $(0, \dots, 0)$ and $(1, \dots, 1)$ (see Marshall, Olkin and Arnold, 2011)

8 APPLICATIONS: MULTIDIMENSIONAL INEQUALITY IN WELL-BEING

The study of multidimensional inequality in well-being has received an increasing amount of attention in recent years. The discontent with the GDP per capita as

an indicator of well-being has led to the consideration of other dimensions such as health or education, which are considered as important as income in the assessment of quality of life. The new conception of well-being requires the development of multivariate tools to measure inequality within the multi-attribute framework. In this context, the multidimensional Lorenz curves are necessary to study distributional dynamics of quality of life in different parts of the distribution.

In Sarabia and Jordá (2013), the Sarmanov-Lee Lorenz curve was applied to study inequality in well-being, taking the Human Development Index as a theoretical benchmark. Then, the beta distribution was considered a convenient model for the marginals in this case given that it is ranged from 0 to 1. The results pointed out that bidimensional inequality decreased during the last three decades. However, inequality measures only offer summarized information of the evolution well-being distribution and hence some internal dynamics can be masked. The estimation of bivariate Lorenz curves showed a decrease in disparities in well-being for the whole distribution except at the lower tail. The poorest, least educated and least healthy countries presented a more unequal situation at the end of the study period. It should be noted that the contrasting patterns at different parts of the distribution cannot be concluded using inequality indices. As a result, the extension of the Lorenz curves to higher dimensions stands as essential to assess the evolution of disparities in well-being.

9 CONCLUSIONS

In this paper, the three existing definitions have been revisited. We have reviewed their origin as well as the main properties of each of them. For the Arnold's definition, some parametric forms are presented for the bivariate case with different dependence structures and different marginals based on the Sarabia and Jordá

(2014) work. We have examined briefly the extension of these families to dimensions higher than two and describing their applications to well-being data when several attributes are jointly considered.

10 REFERENCES

- Arnold, B. C. (1983). *Pareto Distributions*. International Co-operative Publishing House, Fairland, MD.
- Arnold, B.C. (1987). *Majorization and the Lorenz Curve*, Lecture Notes in Statistics 43, Springer Verlag, New York.
- Atkinson, A.B. (2003). Multidimensional deprivation: contrasting social welfare and counting approaches. *Journal of Economic Inequality*, 1, 51-65.
- Atkinson, A.B., Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status. *Review of Economics Studies*, 49, 183-201.
- Bairamov, I., Kotz, S. (2003). On a new family of positive quadrant dependent bivariate distributions. *International Mathematical Journal*, 3, 1247-1254.
- Gadrey J., Jany-Catrice, F. (2006). *The new Indicators of Well-being and Development*. Macmillan, London.
- Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, 1037-1039.
- Huang, J.S., Kotz, S. (1999). Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. *Metrika*, 49, 135-145.
- Kakwani, N.C. (1977). Applications of Lorenz Curves in Economic Analysis. *Econometrica*, 45, 719-728.

- Kolm, S.C. (1977). Multidimensional Equalitarianisms. *Quarterly Journal of Economics*, 91, 1-13.
- Koshevoy, G. (1995). Multivariate Lorenz majorization. *Social Choice and Welfare*, 12, 93-102.
- Koshevoy, G., Mosler, K. (1996). The Lorenz zonoid of a multivariate distribution. *Journal of the American Statistical Association*, 91, 873-882.
- Koshevoy, G., Mosler, K. (1997). Multivariate Gini indices. *Journal of Multivariate Analysis*, 60, 252-276.
- Lee, M-L.T. (1996). Properties of the Sarmanov Family of Bivariate Distributions, *Communications in Statistics, Theory and Methods*, 25, 1207-1222.
- G. Lunetta (1972). *Sulla concentrazione delle distribuzioni doppie*, Societa Italiana di Statistica, 27 (Riunione Scientifica, Palermo).
- G. Lunetta (1972). *Di un indice di concentrazione per variabili statistiche doppie*, Annah della Facolta di Economia e Commercio dell Universita di Catania A 18.
- Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica*, 54, 991-997.
- Marshall, A.W., Olkin, I., Arnold, B.C. (2011). *Inequalities: Theory of Majorization and Its Applications*. Second Edition. Springer, New York.
- McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52, 647-663.
- Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*. Lecture Notes in Statistics 165, Springer, Berlin.

- Muller, C., Trannoy, A. (2011). A dominance approach to the appraisal of the distribution of well-being across countries. *Journal of Public Economics*, 95, 239-246.
- Sarabia, J.M. (2008). Parametric Lorenz Curves: Models and Applications. In: *Modeling Income Distributions and Lorenz Curves*. Series: Economic Studies in Inequality, Social Exclusion and Well-Being 4, Chotikapanich, D. (Ed.), 167-190, Springer-Verlag.
- Sarabia, J.M., Castillo, E., Pascual, M., Sarabia, M. (2005). Mixture Lorenz Curves, *Economics Letters*, 89, 89-94.
- Sarabia, J.M., Castillo, E., Pascual, M., Sarabia, M. (2007). Bivariate Income Distributions with Lognormal Conditionals. *Journal of Economic Inequality*, 5, 371-383.
- Sarabia, J.M., Castillo, E., Slottje, D. (1999). An Ordered Family of Lorenz Curves. *Journal of Econometrics*, 91, 43-60.
- Sarabia, J.M., Jordá, V. (2013). Modeling Bivariate Lorenz Curves with Applications to Multidimensional Inequality in Well-Being. Fifth meeting of the Society for the Study of Economic Inequality ECINEQ Bari, Italy.
- Sarabia, J.M., Jordá, V. (2014). Bivariate Lorenz Curves based on the Sarmanov-Lee Distribution. *IWS 2013 Book Proceedings*, Springer (forthcoming).
- Sarmanov, O.V. (1966). Generalized Normal Correlation and Two-Dimensional Frechet Classes, *Doklady (Soviet Mathematics)*, 168, 596-599.
- Seth, S. (2011). A class of distribution and association sensitive multidimensional welfare indices. *Journal of Economic Inequality*, DOI10.1007/s10888-011-9210-3

- Taguchi, T. (1972a). On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two-dimensional case-I. *Annals of the Institute of Statistical Mathematics*, 24, 355-382.
- Taguchi, T. (1972b). On the two-dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to the two-dimensional case-II. *Annals of the Institute of Statistical Mathematics*, 24, 599-619.
- Taguchi, T. (1988). On the structure of multivariate concentration - some relationships among the concentration surface and two variate mean difference and regressions. *Computational Statistics and Data Analysis*, 6, 307-334.
- Tsui, K.Y. (1995). Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach. *Journal of Economic Theory*, 67, 251-265.
- Tsui, K.Y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: an axiomatic derivation. *Social Choice and Welfare*, 16, 145-157.