International Journal of Business and Management Study – IJBMS
Volume 1 : Issue 3 ***[ISSN 2372 –3955]***
Publication Date : 30 September, 2014

Modeling multidimensional inequality in well-being: the case of Spain

José María Sarabia, Vanesa Jordá

Abstract—**The aim of this paper is to assess the evolution of multidimensional inequality in well-being using Lorenz curves. Closed expressions for the bivariate Lorenz curve defined by Arnold (1983) are given. We assume a relevant type of models based on the class of distributions with given marginals described by Sarmanov and Lee (Lee, 1996; Sarmanov, 1966). This specification of the bivariate Lorenz curve can be easily interpreted as a convex linear combination of products of classical and concentrated Lorenz curves (Sarabia and Jordá, 2014). Using this methodology, we present a closed expression for the bivariate Gini index (Arnold, 1987) in terms of the classical and concentrated Gini indices of the marginal distributions, which are modeled using a convenient model. This index is especially useful and can be decomposed in two factors, corresponding to inequality within variables and the degree of correlation between dimensions (Sarabia and Jordá, 2014). Finally, we illustrate all the previous methodology by analyzing multidimensional inequality in well-being in Spain during the period 2004-2012. We focus on three dimensions, namely income, health and education. Our results point out that inequality levels decreased over the study period, especially for non-income components.**

Keywords— ***bi*variate Lorenz curve; Sarmanov-Lee distribution; bivariate Gini index; well-being.**

I. Introduction

It has been repeatedly argued that the GDP per capita is a poor indicator to evaluate the levels of well-being. Quality of live involves other aspects which are completely ignored by economic variables. Along this line, several composite indicators have been proposed to provide a more comprehensive picture of human well-being (see Gadrey and Jany-Catrice (2006) for a review). This conception of quality of life requires the development of new tools to measure inequality in a multidimensional framework. In this sense, an increasing number of inequality indices can be found in the literature. However, these measures inform about the evolution of well-being distribution in aggregate terms. If dominance relationships cannot be achieved, some parts of the distribution may exhibit an opposite trends to those pointed out by inequality measures. In this context, the Lorenz curve provides valuable insights about the evolution of distributional patterns.

José María Sarabia
Department of Economics, University of Cantabria
Santander (Spain)

Vanesa Jordá
Department of Economics, University of Cantabria
Santander (Spain)

Using the definition proposed by Arnold (1983), we obtain closed expressions for the bivariate Lorenz curve. We study a relevant type of models based on the class of bivariate distributions with given marginals described by Sarmanov and Lee (Lee, 1996; Sarmanov, 1966). This model yields a convenient expression of the bivariate Lorenz curve, which can be easily interpreted as a convex linear combination of products of classical Lorenz curves and concentration curves. A closed expression for the bivariate Gini index (Arnold, 1987) is also given. This index is especially useful, and can be decomposed in two factors, corresponding to the equality within and between variables. We present the moments estimates of this curve when the beta distribution is assumed for the marginals. To illustrate all the previous methodology, we analyze multidimensional inequality in well-being in Spain during the period 2004-2012. Using data from the European Union Statistics on Income and Living Conditions (EU-SILC) (Eurostat, 2013), we investigate the evolution of disparities in three dimensions of quality of life, namely income health and education.

II. Methodology: The bivariate Sarmanov-Lee Lorenz curve

The methodology of this work is based on Sarabia and Jordá (2014). We start with the concepts of Lorenz curve and concentration curve (Kakwani, 1977) for the univariate case. Denote the class of univariate distribution functions with positive finite expectations by $\mathscr{L}$ and denote by $\mathscr{L}+$ the class of all distributions in $\mathscr{L}$ with $F(0) = 0$, corresponding to non-negative random variables. We use the following definition by Gastwirth (1971).

Definition 3.1. *The Lorenz curve L of a random variable X with cumulative distribution function $F \in \mathscr{L}$ is,*

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

where

$$F^{-1}(y)=\sup\{x:F(x)\le y\},\qquad 0\le y\le 1,$$

and $F^{-1}(y)=\sup\{x:F(x)<1\}$ if y=1 *is the right continuous inverse distribution function or quantile function corresponding to F.*

Definition 3.2. *Let g(x) be a continuous function of x such that its first derivative exists and g(x) ≥ 0. If the mean* $E_F\left[g(X)\right]$ *exits, then the concentration curve is defined by*

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

Analogously to the Lorenz curve, we can use an index from the concentration curve which, in contrast to the Gini index, can be negative if the area above the egalitarian line is greater than the area below the egalitarian line. According to Kakwani (1977), if $g(x) \geq 0$ for all x, then the concentration index is positive and equal to the Gini index of $g(x)$. If $g(x) < 0$ for all x, the concentration curve lies completely above the egalitarian line and, consequently, its associated concentration index is negative and equal to minus the Gini index of $g(x)$. Finally, if $g(x)$ is not monotonic, the concentration index is ranged from minus the Gini index of $g(x)$ and the Gini index of $g(x)$.

There have been three different attempts to extend the concept of Lorenz curve to higher dimensions. The three existing definitions were proposed by Taguchi (1972a,b), Arnold (1983) and Koshevoy and Mosler (1996). Due to its suitable structure to handle with parametric models, we use the definition proposed by Arnold (1983, 1987).

Let $\mathbf{X} = (X_1, X_2)^T$ be a bivariate random variable with bivariate probability distribution function F_{12} on $\Re_+^2$ 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 3.3. *The Lorenz surface of* F_{12} *is the graph of the function*,

$$L(u_1,u_2;F_{12})=\frac{\int_0^{s_1}\int_0^{s_2} x_1x_2\,dF_{12}(x_1,x_2)}{\int_0^{\infty}\int_0^{\infty} x_1x_2\,dF_{12}(x_1,x_2)}, \qquad (1)$$

where

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

The two-attribute Gini-Arnold index is defined as,

$$GA(F_{12})=4\int_0^1\int_0^1\left[u_1u_2-L(u_1,u_2;F_{12})\right]du_1du_2, \qquad (2)$$

where the egalitarian surface is given by $L_0(u_1,u_2;F_0)=u_1u_2.$

The Arnold´s Lorenz curve (1) can be evaluated implicitly in some relevant bivariate families of distributions.

In order to obtain bivariate Lorenz curves using (1), we use flexible bivariate joint cumulative distribution functions given by F_{12}. In this work, we propose to use the distribution derived from the Sarmanov-Lee copula, which presents several advantages in relation with other models. Its joint PDF and CDF are quite simple and its different probabilistic features (moments, conditional distributions) can be obtained in an explicit form. On the other hand, the covariance structure is not limited, including correlations ranged from -1 to 1. This model considers the case of independence. Additionally, the Sarmanov-Lee distribution includes several relevant special cases including the classical Farlie-Gumbel-Morgenstern distribution.

Let $\mathbf{X} = (X_1, X_2)^T$ be a random variable that follows a Sarmanov-Lee distribution with joint PDF,

$$f(x_1,x_2)=f_1(x_1)f_2(x_2)\left\{1+\omega\,\phi_1(x_1)\phi_2(x_2)\right\}, \qquad (3)$$

where $f_1(x_1)$ and $f_2(x_2)$ are the univariate PDF marginals, $\phi_i(t),\ i=1,2$ are bounded non-constant functions such that,

$$E_{f_i}[\phi_i(X)]=0, \quad i=1,2,$$

and ω is a real number which satisfies the condition $1+\omega\phi_1(x_1)\phi_2(x_2)\geq 0$ for all x_1, and x_2.

The bivariate Sarmanov-Lee Lorenz curve is obtained using the distribution in (3) in the explicit version of the Arnold Lorenz curve (see Sarabia and Jordá, 2014).

Theorem 1 *Let* $\mathbf{X}$= $(X_1, X_2)^T$ *a bivariate Sarmanov-Lee distribution with joint PDF (3), 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,*

$$L_{SL}(u_1,u_2)=\pi L_{F_1}(u_1)L_{F_2}(u_2)+(1-\pi)L_{g_1}(u_1)L_{g_2}(u_2), \qquad (4)$$

where $\pi=\mu_1\mu_2/(\mu_1\mu_2+w\nu_1\nu_2)$, $L_{F_i}(u_i),\ i=1,2$ *are the Lorenz curves of the marginal distributions* X_i, $i=1,2$ *respectively, and* $L_{g_i}(u_i)$, $i=1,2$ *represent the concentration curves of the random variables,* $g_i(X_i)=X_i\phi_i(X_i),\ i=1,2$, *respectively.*

The meaning of Equation (4) is the following (Sarabia and Jordá, 2014): the bivariate Lorenz curve can be written as a convex linear combination of two components: 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 (dependence component).

The next result (Sarabia and Jordá, 2014) provides a convenient write of the two-attribute bivariate Gini defined in (2). This expression permits a simple decomposition of the overall equality (1 - $G(F_{12})$) in two factors: a first factor which represents the equality within variables (associated with the concept of *distribution sensitive inequality* (Kolm, 1977)) and a second factor which represents the equality between variables (related to the so-called *association sensitive inequality* (Atkinson and Bourgignon, 1982)).

Theorem 2 *Let* $\boldsymbol{X}=(X_1, X_2)^T$ *be a bivariate random variable that follows a Sarmanov-Lee distribution with bivariate Lorenz curve* $L(u_1,u_2;F_{12})$. *The two-attribute bivariate Gini index defined in (2) is given by,*

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

where G(Fi) i=1,2 are the Gini indices of the marginal Lorenz curves, and $G_{gi}(F_i)$, *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 into two factors (see Sarabia and Jordá, 2014),

$$OE=EW+EB, \tag{5}$$

Where $OE=1-G(F_{12})$, and,

$$EW=\pi[1-G(F_1)][1-G(F_2)],$$

$$EB=(1-\pi)[1-G_{g_1}(F_1)][1-G_{g_2}(F_2)].$$

The factor *EW* represents the equality within variables and the second factor *EB* represents the equality between variables which includes the structure of dependence of the underlying bivariate distribution through the functions g_i, i =1, 2. Therefore, *EB* informs about the degree of association between dimensions, which plays a main role in the assessment of multidimensional disparities (Duclos et al., 2011; Kovacevic, 2010; Seth, 2013).

III. The bivariate Lorenz curve for a class of well-being indices

In this section we consider a relevant model to study the distribution of well-being as a multidimensional process, taking the Human Development Index (HDI) as a theoretical benchmark. Therefore, the evaluation of well-being is based on the three dimensions considered in the HDI: income, health and educational attainment. These components, placed on a scale 0 to 1, are transformed indicators of the original variables. The composite index is constructed using a geometric mean of the three sub-indices.

Before going any further, it should be emphasized that the bivariate Lorenz curve defined in (3) is especially suitable for modeling inequality in the HDI. The construction formula of this indicator is characterized by a multiplicative scheme which is also adopted in the specification of the bivariate Lorenz curve (1). Notwithstanding the especial case of the HDI, the bivariate Lorenz curve can be used to measure inequality in other kinds of variables if the marginal distributions are satisfactorily modeled. In this case, given that the indicators considered are ranged from 0 to 1, the beta distribution seems to be an adequate model in this case. Then, we define the bivariate Lorenz curve based on the Sarmanov-Lee distribution considering the beta distribution for the marginals.

Let $X_i \sim \mathcal{B}e(a_i, b_i)$, i =1,2 be two classical beta distributions with PDF,

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

where $B(a_i,b_i)=\Gamma(a_i)\Gamma(b_i)/\Gamma(a_i+b_i)$ for i=1,2 denotes the beta function and $\Gamma(x)$ the gamma function . This distribution was proposed as a model of income distribution by McDonald (1984). 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 Sarmanov-Lee distribution is,

$$f_{12}(x_1,x_2)=f_1(x_1;a_1,b_1)f_2(x_2;a_2,b_2) \\ \left\{1+\omega\left(x_1-\frac{a_1}{a_1+b_1}\right)\left(x_2-\frac{a_2}{a_2+b_2}\right)\right\}, \tag{6}$$

where ω satisfies,

$$\frac{-(a_1+b_1)(a_2+b_2)}{\max\{a_1a_2,b_1b_2\}}\le\omega\le\frac{(a_1+b_1)(a_2+b_2)}{\max\{a_1a_2,b_1b_2\}}.$$

An interesting property of this family is that it can be expressed as a linear combination of products of univariate beta densities.

The univariate Lorenz curve of the classical beta distribution if given by (Sarabia, 2008),

$$L_{F_i}(u_i)=H_{Be(a_i+1,b_i)}[H^{-1}_{Be(a_i,b_i)}(u_i)],\quad i=1,2,$$ where

$H_{Be(a,b)}=(1/B(a,b))\int_0^z t^{a-1}(1-t)^{b-1}dt$ represents the CDF of the classical beta distribution. The concentration curve can be written as,

$$L_{g_i}(u_i,F_i)=\frac{E[X_i^2]G_{Be(a_i+2,b_i)}[G^{-1}_{Be(a_i,b_i)}(u_i)]-E[X_i]^2L(u_i,F_i)}{var[X_i]},\ i=1,2$$

A. Estimation methods

Let $\mathbf{X}=(X_1, X_2)^{\mathrm{T}}$ be a bivariate distribution with joint PDF given by Equation (6). Let $(x_{11}, x_{21}),\ldots,(x_{1n}; x_{2n})$ a sample of size n from (6). For the estimation of the parameters (a_1; b_1; a_2; b_2; ω), we proceed in two steps:

1. Estimation of the marginal distributions. We define,

$$m_i=\frac{1}{n}\sum_{j=1}^{n}x_{ij},\ s_i^2=\frac{1}{n}\sum_{j=1}^{n}(x_{ij}-m_i)^2,\ i=1,2.$$

TABLE I. DECOMPOSITION OF EQUALITY USING THE SARMANOV-LEE DISTRIBUTION WITH BETA MARGINALS

	Education/Health			Education/Income			Health/Income		
	Overall	*Within*	*Between*	*Overall*	*Within*	*Between*	*Overall*	*Within*	*Between*
2004	0.7807	0.7072	0.0736	0.6475	0.4855	0.1620	0.5983	0.5855	0.0128
2005	0.7848	0.7304	0.0545	0.6757	0.5105	0.1652	0.6121	0.5984	0.0137
2006	0.7908	0.7312	0.0597	0.6825	0.5140	0.1685	0.6165	0.6026	0.0138
2007	0.7958	0.7376	0.0582	0.6816	0.5155	0.1660	0.6110	0.6053	0.0057
2008	0.8136	0.7492	0.0645	0.7057	0.5269	0.1787	0.6482	0.6240	0.0242
2009	0.8192	0.7465	0.0727	0.7132	0.5219	0.1913	0.6482	0.6153	0.0329
2010	0.8198	0.7494	0.0704	0.7102	0.5257	0.1846	0.6442	0.6190	0.0252
2011	0.8113	0.7541	0.0572	0.7216	0.5398	0.1817	0.6499	0.6332	0.0167
2012	0.8233	0.7527	0.0706	0.7314	0.5478	0.1836	0.6553	0.6427	0.0126
2004-2012	0.0545	0.0644	-0.0400	0.1295	0.1284	0.1329	0.0951	0.0976	-0.0188

Then, the point estimates of the couples (a_i, b_i), i = 1, 2 are,

$$\hat{a}_i = \frac{m_i(m_i - m_i^2 - s_i^2)}{s_i^2}, \quad i = 1,2,$$

$$\hat{b}_i = \frac{(1-m_i)(m_i - m_i^2 - s_i^2)}{s_i^2}, \quad i = 1,2.$$

2. Estimation of the structure of dependence. The estimate of ω is based on the sample relation $\rho = \omega\sigma_1\sigma_2$. Then, if r denotes the sample linear correlation coefficient, and s_i, i = 1, 2, the sample standard deviation of the marginal distributions X_i, i = 1, 2, the point estimate of ω is,

$$\hat{\omega} = \frac{r}{s_1 \cdot s_2}.$$

IV. Results

In this section we illustrate all the previous methodology by analyzing the evolution of multidimensional inequality in well-being in Spain for the period 2004-2012. The sample is drawn from the EU-SILC database (Eurostat, 2013), which includes data on gross individual income, years of schooling and the self-perception of health status.

Table 1 presents the evolution of overall equality (Eq. (5)) and its decomposition in the equality within variables and the degree of correlation between them. Our results point out that the distribution of the two non-income variables is the most equal over the whole period. The joint distribution of income and each of the other two variables presents higher levels of inequality, especially in the case of health.

The evolution of bidimensional inequality shows that the distribution of education and income has seen the highest increase of equality, due to the improvement of equality within-variables and the decrease of the correlation between these dimensions. The join distribution of income and health also increased substantially its levels of equality. In this case, the improvement was mainly driven by the increase of the equality within variables, while correlation between dimensions seems to play little role. Finally, we observe that equality increased by 5 percent in the case of the distribution of non-income variables. The raise of equality was prompted by the improvement of the equality within-variables, but the increase of the correlation between dimensions slowed down the equalization of this distribution.

The previous trends describe the evolution of bivariate inequality in aggregate terms. However, it is possible that some parts of the distribution exhibit the opposed patterns to those concluded by using inequality measures. Fig. 1 shows the bivariate Lorenz curves of these distributions in the first and the last year of the study period. These estimates suggest that almost the whole distribution became less unequal over the period 2004-2012 given that the Lorenz curve in 2012 lies above the Lorenz curve in 2004 for all pairs of variables. However, these curves seem to cross at the lower end of the distribution, thus pointing out that the poorest, least educated and least healthy individuals had a more equal situation in 2004. Therefore, the multidimensional conception of well-being makes the extension of the Lorenz curve to higher dimensions essential to analyze the internal dynamics of well-being distribution and to offer a complete perspective of the evolution of disparities in quality of life.

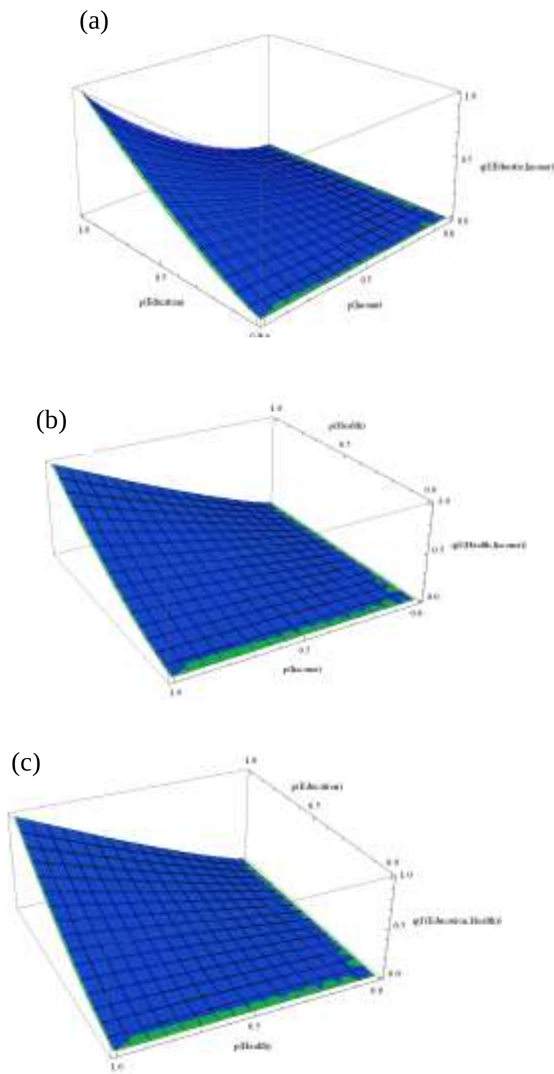

Figure 1. *Bivariate Lorenz curves of education and income (a), health and income (b) and health and education (c) in 2004 (blue) and 2012 (green).*

Acknowledgments

Authors thank the University of Cantabria (Convocatoria de Proyectos Puente) and the Ministerio de Education (FPU AP2010-4907) for partial support of this work.

References

[1] B.C. Arnold, Pareto Distributions. Maryland: International Co-operative Publishing House, 1983.

[2] B.C. Arnold, Majorization and the Lorenz Curve. Lecture Notes in Statistics 43, New York: Springer Verlag, 1987.

[3] A.B. Atkinson and F. Bourguignon, The comparison of multi-dimensioned distributions of economic status. Review of Econ. Stud., vol. 49, pp. 181–201, 1982.

[4] J.-C Duclos., D. Sahn and S. Younger, Partial multidimensional inequality orderings. J. of Public Econ., 95, pp. 225–238, 2011.

[5] Eurostat, "European Union Statistics on Income and Living Conditions". Accessed on 2nd February 2014.

[6] J. Gadrey and F. Jany-Catrice, The new Indicators of Well-being and Development. London: Macmillan, 2006.

[7] J.L. Gastwirth, "A general definition of the Lorenz curve". *Econometrica*, vol. 39, pp. 1037–1039, 1971.

[8] N.C. Kakwani, "Applications of Lorenz Curves in Economic Analysis", Econometrica, vol. 45, pp. 719–728, 1977.

[9] S.C. Kolm, "Multidimensional Equalitarianisms", Quarterly J. of Econ., vol. 91, pp. 1–13, 1977.

[10] M. Kovacevic, "Measurement of Inequality in Human Development-A Review" Human Development Research Paper No. 35. UNDP, 2010.

[11] M-L.T. Lee, "Properties of the Sarmanov Family of Bivariate Distributions", Communications in Statistics, Theory and Methods, vol. 25, pp. 1207–1222, 1966.

[12] J.M. Sarabia, "Parametric Lorenz curves: models and applications", in Modeling income distributions and Lorenz curves. Series: Economic studies in inequality social exclusion and well-being, vol. 4, D. Chotikapanich, Eds, New York: Springer-Verlag, pp. 167–190, 2008.

[13] J.M. Sarabia and V. Jordá, "Modeling Bivariate Lorenz Curves based on the Sarmanov-Lee Distribution"., IWS 2013 Book of Proceedings, 2014, in press

[14] O.V. Sarmanov, "Generalized Normal Correlation and Two-Dimensional Frechet Classes, Doklady (Soviet Mathematics), vol. 168, pp. 596–599, 1966.

[15] S. Seth, "A class of distribution and association sensitive multidimensional welfare indices". J. of Econ. Inequal., 2013.

[16] T. Taguchi, "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, vol. 24, pp. 355–382, 1972a.

[17] T. Taguchi, "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, vol. 24, pp. 599–619, 1972.