REVISTA DE MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA (10). Páginas 5–18.
Diciembre de 2010. ISSN: 1886-516X. D.L: SE-2927-06.
URL: http://www.upo.es/RevMetCuant/art40.pdf

The Similarity between the Square of the Coefficient of Variation and the Gini Index of a General Random Variable

González Abril, Luis
Departamento de Economía Aplicada I
Universidad de Sevilla (España)
Correo electrónico: luisgon@us.es

Velasco Morente, Francisco
Departamento de Economía Aplicada I
Universidad de Sevilla (España)
Correo electrónico: velasco@us.es

Gavilán Ruiz, José Manuel
Departamento de Economía Aplicada I
Universidad de Sevilla (España)
Correo electrónico: gavi@us.es

Sánchez-Reyes Fernández, Luis María
Departamento de Economía Aplicada I
Universidad de Sevilla (España)
Correo electrónico: luiss-rf@us.es

ABSTRACT

In this paper, several identities concerning expectation, variance, covariance, cumulative distribution functions, the coefficient of variation, and the Lorenz curve are obtained and they are used in establishing theoretical results. Furthermore, a graphical representation of the variance is proposed which, together with the aforementioned identities, enables the square of the coefficient of variation to be considered as an equality measure in the same way as is the Gini index. A study of the similarities between the theoretical expression of the Gini index and the square of the coefficient of variation is also carried out in this paper.

Keywords: concentration measures; cumulative distribution function; Lorenz curve; mean difference.
JEL classification: C100; C190.
MSC2010: 62-09; 62P20; 91B02.

Artículo recibido el 12 de abril de 2010 y aceptado el 22 de octubre de 2010.

Similitud entre el cuadrado del coeficiente de variación y el índice de Gini en una variable aleatoria general

RESUMEN

En este trabajo se obtienen diversas identidades relativas a la espezanza, varianza, covarianza, función de distribución acumulada, coeficiente de variación y curva de Lorenz que se usarán para obtener resultados teóricos interesantes. Se construye, además, una representación gráfica de la varianza, la cual, utilizando las propiedades obtenidas, nos indica que el cuadrado del coeficiente de variación se puede considerar como una medida de igualdad, de igual forma que se considera al índice de Gini. En este artículo también se lleva a cabo un estudio de las similitudes entre la expresión teórica del índice de Gini y el cuadrado del coeficiente de variación.

Palabras clave: medidas de concentración; función de distribución; curva de Lorenz; diferencia media.
Clasificación JEL: C100; C190.
MSC2010: 62-09; 62P20; 91B02.

1 INTRODUCTION

Powerful tools, which are specifically designed for certain increasingly difficult problems, are currently under development. Nevertheless, it is not always necessary to design new tools, but to give a new interpretation to other known tools. Thus, there are easy relationships between the main characteristics of a random variable which are widely known but remained unused. In this paper, several identities are obtained from a very simple but powerful result. One particular result leads us to study the square of the coefficient of variation and the Gini index.

The Gini index or Gini coefficient (Gini 1912) is perhaps one of the main inequality measures in the discipline of Economics and it has been applied in many studies. Furthermore, this index can be used to measure the dispersion of a distribution of income, or consumption, or wealth, or a distribution of any other kind (Xu 2004) since, from the statistical point of view, it is a function of the mean difference. Its attractiveness to many economists is that it has an intuitive geometric interpretation, that is, it can be defined as twice a ratio of two regions defined by the line of perfect equality (45-degree line) and the Lorenz curve in the unit box. Furthermore, it is an important component of the Sen index of poverty intensity (Xu and Osberg 2002).

There are two main different approaches for analyzing theoretical results of the Gini index: the one is based on discrete distributions; the other on continuous distributions. Both approaches can be unified (Dorfman 1979), but for some purposes the continuous formulation is more convenient, yielding insights that are not as accessible when the random variable is discrete (Yitzhaki and Schechtman 2005). For this reason, a continuous formulation is considered in this paper.

The major drawback when the Gini index is used is that two very different distributions can have the same value of this index and, therefore, it is not possible to declare which distribution is more equitable. This problem has been faced in the literature by means of stochastic dominance (Fishburn 1980) and inverse stochastic dominance (Muliere and Scarsini 1989). It is worth noting that a more general study is carried out in (Núñez 2006), where several approaches are presented. In this paper, to avoid this situation, it is proved that the square of the coefficient of variation can be thought of as the ratio of the area that lies between the curve of equality and the Lorenz curve in the same way as can the Gini index and, therefore, it can be used as "the most natural" measure to discriminate between two distributions when their Gini indices are the same. Let us note that the square of the coefficient of variation[1] was firstly proposed as a transfer measure in (Shorrocks and Foster 1987) and later in (Davies and Hoy 1994), another possibilities were set up in (Ramos and Sordo 2003). Furthermore, it will also be shown that both coefficients have

[1]The main drawback of this coefficient is that it is very sensitive to extreme values (Bartels 1977).

a similar definition. Hence, by using the definition of the coefficient of variation, the Gini index can be defined for any random variable with a non-zero expectation and not only for non-negative expectations.

The rest of the paper is organized as follows: Section 2 presents a result which forms the basis of later developments since it provides identities on probability theory. Notes on mean difference, independence, covariance, and variance are given in Section 3. In Section 4, two equality measures of a non-negative random variable, the Gini index, and the square of the covariation coefficient, are obtained from the previous identities and a relationship between variance, expectation, the cumulative distribution function and the Lorenz curve is given, which provides us with a graphical interpretation of the variance. The identities are generalized and the Gini index is considered for any random variable. Finally, conclusions are drawn.

2 MAIN RESULT

Let us see a simple but important result:

Theorem 1 *Let $g(x)$ be a function such that $\int_{-\infty}^{\infty} |x|^r \ |g(x)| \ dx < \infty$ for $r = 0, 1$. Hence*

$$\int_{-\infty}^{\infty} x\, g(x) dx = \int_{0}^{\infty} \left(G^*(x) + G^*(-x)\right) dx \tag{1}$$

where

$$G^*(x) = G^*_{g(\cdot)}(x) = I(x) \int_{x}^{\infty} g(u)\, du - I(-x) \int_{-\infty}^{x} g(u)\, du, \tag{2}$$

and $I(x) = I_{(0,+\infty)}(x)$ is the indicator function of the interval $(0, +\infty)$.

Proof. It is straightforward by integration by parts. ■

Its generalization to two variables is an immediate consequence of this result.

Corollary 2 *Let $g(x, y)$ be a function such that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x|^r |y|^s \, |g(x, y)| \, dxdy < \infty$, for $r, s = 0, 1$. Hence*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy g(x, y) dxdy = \int_{0}^{\infty}\int_{0}^{\infty} \left(G^*(x, y) + G^*(x, -y) + G^*(-x, y) + G^*(-x, -y)\right) dxdy \tag{3}$$

where $G^*(x, y) = G^*_{G^*_{g(x,\cdot)}(y)}(x)$. ■

The expression of $G^*(x)$ is useful to simplify the thesis of Corollary 2; nevertheless an even simpler expression can be used. If $G = \int_{-\infty}^{\infty} g(u)du$ and $G(x) = \int_{-\infty}^{x} g(u)du$ are defined, then (1) can be written as:

$$\int_{-\infty}^{\infty} x\, g(x) dx = \int_{0}^{\infty} \left(G - G(x) - G(-x)\right) dx. \tag{4}$$

Let $g(x)$ and $g(x,y)$ be the marginal probability density function (pdf) of a random variable X and the joint pdf of a continuous random vector (X,Y), respectively. Hence, from (2) and (3):

$$G^*(x) = \begin{cases} -F_X(x) & \text{if } x < 0 \\ 1 - F_X(x) & \text{if } x > 0 \end{cases} \tag{5}$$

and

$$G^*(x,y) = F(x,y) - I(x)F_Y(y) - I(y)F_X(x) + I(x)I(y), \tag{6}$$

where $F(x,y)$ is the a joint cumulative distribution function (cdf) of (X,Y), and $F_X(x)$ and $F_Y(y)$ are marginal cdfs of X and Y, respectively. Therefore, since E(X) is the expectation of X and σ_{XY} is the covariance of (X,Y), the following result can be stated:

Lemma 3 *Let (X,Y) be a continuous random vector with $\sigma_{XY} < \infty$. Hence*

$$E(X) = \int_0^\infty (1 - F_X(x) - F_X(-x))\, dx, \tag{7}$$

$$E(XY) = \int_0^\infty \int_0^\infty (1 - F_X(x) - F_Y(y) - F_X(-x) - F_Y(-y) + \cdots \\ \cdots + F(x,y) + F(-x,y) + F(x,-y) + F(-x,-y))\, dx\, dy, \tag{8}$$

$$\sigma_{XY} = \int_{-\infty}^\infty \int_{-\infty}^\infty (F(x,y) - F_X(x)F_Y(y))\, dx\, dy. \tag{9}$$

Proof. Identity (7) is obtained from identities (1) and (5). Identity (8) is given by identities (3) and (6). Identity (7) implies

$$E(X)\cdot E(Y) = \int_0^\infty \int_0^\infty (1 - F_X(x) - F_Y(y) - F_X(-x) - F_Y(-y) + F_X(x)F_Y(y) + \cdots \\ \cdots + F_X(-x)F_Y(y) + F_X(x)F_Y(-y) + F_X(-x)F_Y(-y))\, dx\, dy$$

and, therefore

$$E(XY) - E(X)\cdot E(Y) = \int_0^\infty \int_0^\infty ((F(x,y) - F_X(x)F_Y(y)) + (F(-x,y) - F_X(-x)F_Y(y)) + \cdots \\ \cdots + (F(x,-y) - F_X(x)F_Y(-y)) + (F(-x,-y) - F_X(-x)F_Y(-y)))\, dx\, dy$$

and taking into account that:

$$\int_0^\infty \int_0^\infty (F(x,-y) - F_X(x)F_Y(-y))dydx = \int_0^\infty \int_{-\infty}^0 (F(x,y) - F_X(x)F_Y(y))dydx,$$
$$\int_0^\infty \int_0^\infty (F(-x,y) - F_X(-x)F_Y(y))dxdy = \int_{-\infty}^0 \int_0^\infty (F(x,y) - F_X(x)F_Y(y))dxdy,$$
$$\int_0^\infty \int_0^\infty (F(-x,-y) - F_X(-x)F_Y(-y))dxdy = \int_{-\infty}^0 \int_{-\infty}^0 (F(x,y) - F_X(x)F_Y(y))dxdy,$$

then (9) is obtained. ■

Let us see, in the next section, how Lemma 3 is useful in establishing theoretical results.

3 NOTES ON RANGE, MEAN DIFFERENCE, INDEPENDENCE, AND COVARIANCE OF RANDOM VARIABLES

Note 4 *In fact, result (7) can easily be generalized as follows:*

$$E(X^{2r+1}) = (2r+1)\int_0^\infty x^{2r}\cdot(1-F_X(x)-F_X(-x))\,dx, \qquad \forall r=0,1,2,...$$

and, if X is non-negative, then

$$E(X^{r+1}) = (r+1)\int_0^\infty x^{r}\cdot(1-F_X(x))\,dx, \qquad \forall r=0,1,2,...$$

That is, the r-th moment about the origin of a non-negative random variable can be obtained from the cdf $F(x)$ directly instead of from the pdf $f(x)$. ▲

Note 5 *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (iid) random variables with the same distribution as X. If the transformations given by $U_n = max\{X_1, X_2, \ldots, X_n\}$ and $V_n = min\{X_1, X_2, \ldots, X_n\}$ are considered, then their cdfs are: $F_{U_n}(u) = F^n(u)$ and $F_{V_n}(v) = 1-(1-F(v))^n$.*

By using (7), $E(V_n) = \int_0^\infty (-1+(1-F(x))^n + (1-F(-x))^n)\,dx$, and $E(U_n) = \int_0^\infty (1-F^n(x)-F^n(-x))dx$. Hence,

$$E(U_n - V_n) = \int_{-\infty}^{\infty}(1-F^n(x)-(1-F(x))^n)dx.$$

Furthermore, as a particular case, the mean difference of two iid random variables, $\Delta = E(|X_1 - X_2|)$, can be written as:

$$\Delta = E(U_2 - V_2) = \int_{-\infty}^{\infty}(1-F^2(x)-(1-F(x))^2)dx = 2\int_{-\infty}^{\infty}F(x)(1-F(x))dx.$$

▲

Note 6 *Usually, the covariance is defined as $Cov(X,Y) = \mathrm{E}\,[(X-E(X))\cdot(Y-E(Y))]$ and an interpretation of its meaning with respect to the independence or dependence between X and Y is given a posteriori. From (9), it is possible to give a new introduction to covariance as follows: Given a random vector (X,Y), the variables X and Y are said to be independent if $F(x,y) = F_X(x)\,F_Y(y)$, for every $x, y \in \mathbb{R}$. Hence, there is dependence between X and Y if any $x, y \in \mathbb{R}$ exist such that $F(x,y) - F_X(x)\,F_Y(y) \neq 0$. Therefore, a first measure of dependence or covariation between two random variables can be considered as:*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(F(x,y)-F_X(x)\,F_Y(y))\,dx\,dy,$$

which is named "covariance" between X and Y, and denoted by $Cov(X,Y)$. Once the moments of a random vector are defined, then it can be proved that $Cov(X,Y) = \mathrm{E}\,[(X-E(X))\cdot(Y-E(Y))]$. Thus, covariance is introduced from the concept of independence. ▲

Note 7 *From (9),* $Var(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F(x,y) - F_X(x)F_X(y))\, dxdy$*, where* $Var(X)$ *denotes the variance of* X*, and as* $F(x,y) = P[X \leq x, X \leq y] = P[X \leq min(x,y)]$*, then the variance can be rewritten as:*

$$Var(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F_X(min(x,y))\,(1 - F_X(max(x,y)))\, dx\, dy,$$

and, therefore, it is straightforward to prove, by taking the properties of the cdf into account, that:

$$\frac{1}{2}\Delta^2 = 2\left(\int_{-\infty}^{\infty} F(x)(1-F(x))\, dx\right)^2 \leq Var(X) \leq \left(\int_{-\infty}^{\infty} \sqrt{F(x)}\sqrt{1-F(x)}\, dx\right)^2$$

which provides us with a lower and an upper bound of the variance. ▲

4 GINI INDEX, COEFFICIENT OF VARIATION, AND A GRAPHICAL REPRESENTATION OF THE VARIANCE

Let X be a non-negative continuous random variable with cdf $F(x)$, pdf $f(x)$ and finite variance. From Note 4, the expectation of X can be written as $E(X) = \mu = \int_0^{\infty}(1 - F(x))dx$. Furthermore, the Lorenz function $L(x) = \frac{1}{\mu}\int_0^x t\, f(t)\, dt$ can be considered analogous to a cdf of a non-negative random variable U_X, and by considering $g(x) = \frac{1}{\mu}x\, f(x)$ in (1), then:

$$\int_0^{\infty} x\, g(x)\, dx = \int_0^{\infty} (1 - L(x))\, dx \Rightarrow E(X^2) = \mu E(U_X) \Rightarrow Var(X) = \mu\,(E(U_X) - \mu)\,.$$

However, $E(U_X) - \mu = \int_0^{\infty} (F(x) - L(x))\, dx$. Therefore,

$$\int_0^{\infty} (F(x) - L(x))\, dx = \frac{Var(X)}{E(X)}. \tag{10}$$

It should be pointed out that result (10) provides us with a relationship between some of the most important characteristics of a non-negative random variable: the expectation, the variance, the cumulative distribution function and the Lorenz curve. Moreover, result (10) gives a new interpretation of the variance of a non-negative random variable as the product of μ and the area enclosed by the cdf $F(x)$ and the Lorenz curve $L(x)$, that is, the variance is the product of A ($= E(X)$) and B in Figure 1.

Let us now introduce an equality measure from the area enclosed between the curve given by $y = F(x)$ and $y = L(x)$, that is, area B in Figure 1. From the previous result, $E(U_X) = \mu + \frac{Var(X)}{\mu}$; it follows that area B is equal to $E[U_X] - \mu$. In order to eliminate the units of the variable and to achieve a relative measure this value is divided by μ, thereby obtaining $\frac{B}{\mu} = E(\frac{1}{\mu}U_X - 1)$. From (10) (let us denote μ by μ_X),

$$CV^2(X) = E\left(\frac{1}{\mu_X}U_X - 1\right), \tag{11}$$

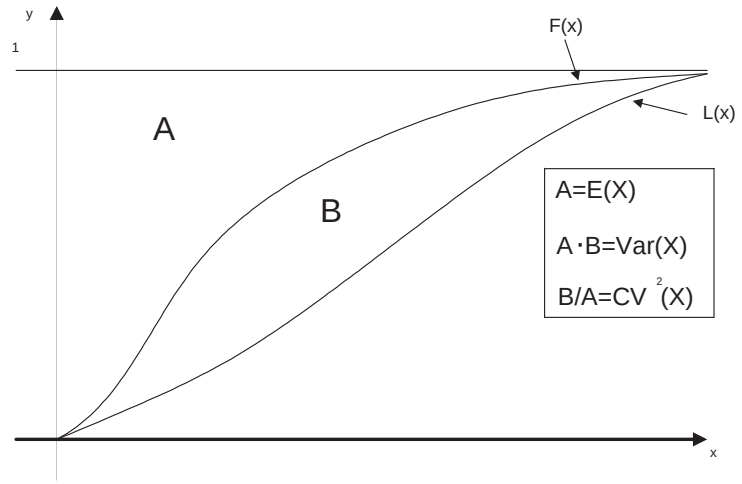

Figure 1: Graphical representation of the mean, the variance and the square of the coefficient of variation of a non-negative random variable.

where $CV(X)$ is the coefficient of the variation of X. Hence, the square of the coefficient of variation has an intuitive geometric interpretation as the ratio of two regions.

It is worth noting that the construction in Figure 1 is similar to that of the Gini index. In order to study this similarity, the transformation $U = F(X)$ is considered, and the Lorenz curve can be written as $L(u) = \frac{1}{\mu}\int_0^u F^{-1}(t)dt$, $0 \leq u \leq 1$, where F^{-1} is the left inverse of F. Hence, the area enclosed between the curve given by $y = u$ and $y = L(u)$, that is, area B in Figure 2, is an equality measure. In the same way as for $L(x)$, the $L(u)$ function can be considered analogous to a cdf of a non-negative random variable $U_{F(X)}$, and from (7), $E(U_{F(X)}) = \int_0^1 (1 - L(u))\, du = \int_0^1 (u - L(u))\, du + \frac{1}{2} \leq 1$ (note that $U = F(X)$ is a uniform distribution and $F_U(u) = u$, $0 < u < 1$). Hence, $0 \leq E(U_{F(X)}) - \frac{1}{2} = \int_0^1 (u - L(u))\, du \leq \frac{1}{2}$, and multiplying by 2 in order to normalize this expression, results in $0 \leq E(2U_{F(X)} - 1) = 2\int_0^1 (u - L(u))\, du \leq 1$. Furthermore, it is well-known that the Gini index is $IG(X) = 2\int_0^1 (u - L(u))\, du$ and that $E(U) = E(F(X)) = \mu_{F(X)} = \frac{1}{2}$, and hence a similar expression of the square of the coefficient of variation (11) is given by the Gini index:

$$IG(X) = E\left(\frac{1}{\mu_{F(X)}} U_{F(X)} - 1\right). \tag{12}$$

Hence, the Gini index can be seen as a "normalization" of the square of the coefficient of variation, by using the transformation $U = F(X)$, from (11) and (12). Therefore, the square of the coefficient of variation of X is an equality measure in the same as is the Gini index.

Another two similar expressions, which are straightforward to obtain, for $IG(X)$ and $CV^2(X)$, are given in the following:

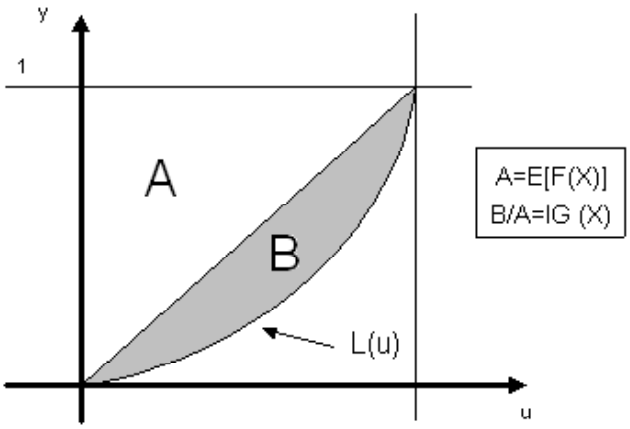

Figure 2: Graphical representation of the Gini index of a non-negative random variable.

In terms of integrals:

$$\begin{aligned} IG(X) &= \frac{1}{E(U)} \int_0^1 (u - L(u))\, du, \\ CV^2(X) &= \frac{1}{E(F^{-1}(U))} \int_0^1 (u - L(u))\, dF^{-1}(u), \qquad \text{(from (10))}. \end{aligned}$$

In terms of covariances:

$$\begin{aligned} IG(X) &= Cov\left(\frac{X}{\mu_X}, \frac{F(X)}{\mu_{F(X)}}\right), \text{ (given in (Lerman and Yitzhaki 1984))}. \\ CV^2(X) &= Cov\left(\frac{X}{\mu_X}, \frac{X}{\mu_X}\right). \end{aligned}$$

Note 8 *It is worth bearing in mind that the square of the coefficient of variation, as an inequality measure of a distribution of income (or consumption, or wealth, or a distribution of any other kind), verifies the four properties which are generally postulated in the economic literature on inequality (for the sake of simplicity let us interpret this coefficient on countries): Anonymity (it does not matter who the high and low earners are); Scale Independence (it does not consider the size of the economy, the way it is measured, or whether it is a rich or poor country on average); Population Independence (it does not matter how large the population of the country is); and Transfer Principle (if an income less than the difference is transferred from a rich person to a poor person, then the resulting distribution is more equal) (Dalton 1920).* ▲

Example 4.1 *If $X \in U(a,b)$ (Uniform distribution), then $F(x) = u = \frac{x-a}{b-a}$, with $a \leq x \leq b$ and $dF^{-1}(u) = (b-a)du$. Hence,*

$$\begin{aligned} IG(X) &= 2\int_0^1 (u - L(u))\, du = \frac{2}{b-a} \int_0^1 (u - L(u))\, dF^{-1}(u) = \frac{2}{b-a}\mu CV^2(X) \\ &= \frac{b+a}{b-a} CV^2(X) = \frac{1}{3} \cdot \frac{b-a}{b+a}. \end{aligned}$$

■

The major drawback (when the Gini index is used) is that there are non-negative random variables X and Y such that $IG(X) = IG(Y)$ and, therefore, it is impossible to quantify which distribution is more equitable. To avoid this situation, and by following the above results, the most natural solution is obtained by calculating the square of the coefficient of variation. Let us see an example:

Example 4.2 *Let $X \in U(\frac{1}{49}, 1)$. The square of the coefficient of variation is straightforward to calculate: $CV^2(X) = \frac{1}{3}\frac{(b-a)^2}{(b+a)^2} = 0.3072$, and, from Example 4.1, the Gini index is $IG(X) = \frac{1}{3}\frac{b-a}{b+a} = \frac{8}{25}$.*
Let us consider the random variable Y with values and probabilities given by $\{0, 0.5, 1\}$ and $\{0.2, 0.6, 0.2\}$, respectively. In this case, the Gini index is $IG(Y) = \frac{8}{25} = IG(X)$, nevertheless, $CV^2(Y) = 0.4000$ is greater than $CV^2(X)$. Thus, it can be concluded that the distribution of X is more equitable than the distribution of Y. ■

Another expression with regard to the integrals can be given: Let X_1 and X_2 be independent and identically distributed (iid) random variables with the same distribution as X, then:

$$\begin{aligned} \int_0^1 (u - L(u))\, du &= \frac{E\,|X_1 - X_2|}{2\mu}; \\ \int_0^1 (u - L(u))\, dF^{-1}(u) &= \frac{E(X_1 - X_2)^2}{2\mu}. \end{aligned}$$

The main advantage of the Gini index over the square of the coefficient of variation is that the Gini index is bounded, that is, $0 \leq IG(X) \leq 1$ while the square of coefficient of variation has no upper bound, that is, $0 \leq CV^2(X)$. Nevertheless, the Gini index is only defined for non-negative random variables and this condition is not required by the coefficient of variation. In both cases, by the definition of the $L(\cdot)$ function, it is necessary that $\mu \neq 0$.

The condition $X \geq 0$ leads to a bounded Gini index, but it is also possible to define the Gini index for any X random variable. This is studied in the following section.

5 THE GINI INDEX OF ANY RANDOM VARIABLE

Let X be a continuous random variable with cdf $F(x)$, pdf $f(x)$, $\mu \neq 0$ and finite variance. Clearly, the Lorenz function, $L(x) = \frac{1}{\mu}\int_{-\infty}^{x} t\, f(t)\, dt$, cannot be considered as analogous to a cdf of a random variable since $L(x)$ can take negative values. Nevertheless, it is possible to consider $g(x) = \frac{1}{\mu} x\, f(x)$ in (1) and hence, by using (4):

$$E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x) dx = \mu \int_0^{\infty} (1 - L(x) - L(-x))\, dx.$$

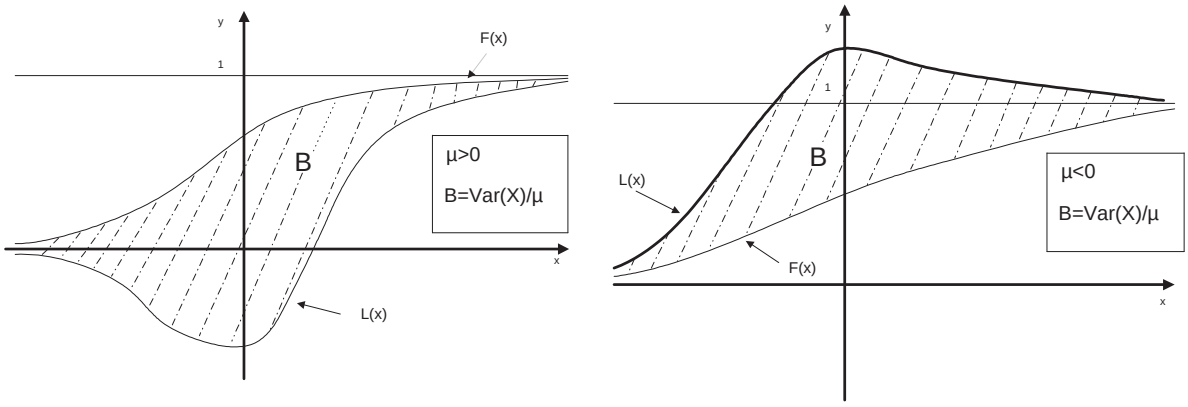

Figure 3: Graphical representation of $B = Var(X)/E(X)$ from the curves $y = F(x)$ and $y = L(x)$.

Furthermore, by using (7):

$$E^2(X) = \mu E(X) = \mu \int_0^\infty (1 - F(x) - F(-x))\, dx,$$

and hence $Var(X) = E(X^2) - E^2(X) = \mu \int_0^\infty (F(x) - L(x) + F(-x) - L(-x))\, dx$. Therefore, as $\int_0^\infty (F(-x) - L(-x))\, dx = \int_{-\infty}^0 (F(x) - L(x))\, dx$, result (10) has been generalized. Let us write it in the form of a theorem.

Theorem 9 *Let X be a continuous random variable with cdf $F(x)$, pdf $f(x)$, $\mu \neq 0$ and finite variance. If the Lorenz function is defined as $L(x) = \frac{1}{\mu} \int_{-\infty}^x t\, f(t)\, dt$, then*

$$\int_{-\infty}^{\infty} (F(x) - L(x))dx = \frac{Var(X)}{E(X)}. \tag{13}$$

■

If $support(X) \triangleq \{x \,/\, f(x) > 0\} = (a, b)$, with $-\infty \leq a < b \leq \infty$, and $R(x) = F(x) - L(x)$ are considered for any $x \in (a, b)$ (see Figure 3), then:

1. If $\mu > 0$, then $R(x) > 0$, and the maximum is attained in $x = \mu$.
2. If $\mu < 0$, then $R(x) < 0$, and the minimum is attained in $x = \mu$.

Hence, in the same way as for the non-negative random variable X, the square of the coefficient of variation can be considered as an equality measure since:

$$0 \leq \frac{1}{\mu} \int_{-\infty}^{\infty} (F(x) - L(x))dx = \frac{1}{\mu} \int_0^1 (u - L(u))\, dF^{-1}(u) = CV^2(X).$$

The only difference between the general random variable case with regard to the non-negative random variable case is that the graphical interpretation of this coefficient as the ratio between two areas is not possible.

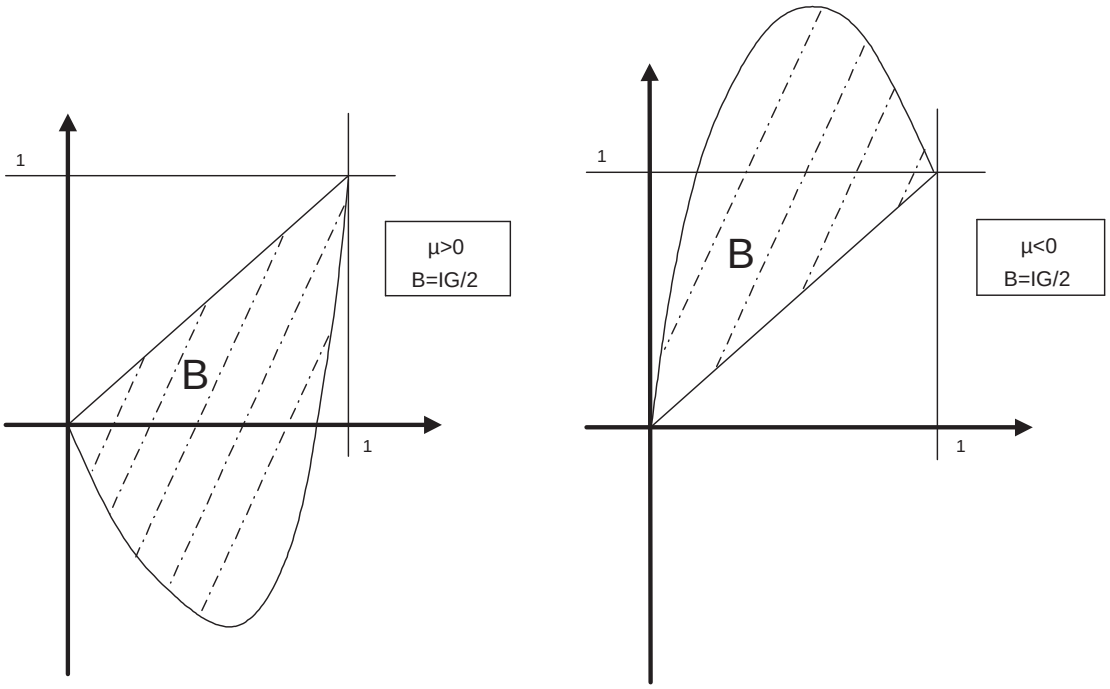

Figure 4: Graphical representation of the Gini index from the line $y = u$ and the curve $y = L(u)$.

On the other hand, the Gini index must be considered in terms of absolute values since if $\mu < 0$ then $u \leq L(u)$ (see Figure 4), that is, $IG(X) = \left|2\int_0^1 (u - L(u))\, du\right|$. The main problem is that the integral has no upper bound, although the remaining of considerations made for a non-negative random variable still remain valid for a random variable.

Sometimes the Lorenz curve is not known, and only values at certain intervals are given. In that case, the most common technique is to approximate the curve in each interval as a straight line between consecutive points, and therefore area B can be approximated with trapezoids. Thus, if $\{(F_k, L_k),\, k = 0, 1, \cdots, n\}$, where $F_0 = L_0 = 0$ and $F_n = L_n = 1$, are the known points set on the Lorenz curve, with the F_k indexed in increasing order $(F_{k-1} < F_k)$, then (Rao 1969):

$$IG = \left|\sum_{k=0}^{n-1} (F_k L_{k+1} - F_{k+1} L_k)\right|.$$

It is important to note that this expression does not depend on whether the random variable is non-negative. Let us see an example which allows us to clarify the idea of this approach.

Example 5.1 *Four players of cards start a game where each one bets* \$100 *and a debit balance is allowed. Let us consider the variable* $X_t =$ *monetary value of each player at time* t*. A possible change of the distribution of profit and loss in the game is shown in*

Table 1. In this table, it can be seen that, when $x_4 = -200$ and $t = 4$, the Gini index is greater than one, which is consistent because the situation after this sharing out is worse than the earlier case. The Gini index, in the same way as the square of the coefficient of variation, increases when the sharing out is less equitable. Therefore, there is no reason to demand non-negativity in random variables when using the Gini index in equality studies. ■

Table 1: Example of inequality in the game.

t	x_1	x_2	x_3	x_4	IG	CV^2
0	100	100	100	100	0.0000	0.0000
1	150	120	75	50	0.2188	0.1563
2	200	150	50	0	0.4375	0.6250
3	400	200	0	−200	1.2500	5.0000
4	800	100	−200	−300	2.2500	18.5000
5	1000	0	−200	−400	2.7500	29.0000

CONCLUSION

We have obtained several generalizations of some well-known results which establish relationships concerning the joint cdf, the marginal cdfs and other characteristics of a random vector (X, Y). We have proved that these relationships can be useful in theoretical considerations.

An identity which relates four of the most important characteristics of a random variable (the mean, the variance, the cumulative distribution function, and the Lorenz curve) has also been given. This result provides us with a graphical representation of the mean, the variance and the square of the coefficient of variation in the same figure. Furthermore, new expressions of the Gini index have been given, and they have their counterparts in the square of the coefficient of variation.

In this paper, by following the same interpretation as the one of the coefficient of variation, the Gini index is defined as an equality measure for random variables which do not have to be non-negative.

References

Bartels, C.P.A. (1977): Economic aspects of regional welfare. *Martinus Nijhoff Sciences Division*.

Dalton, H. (1920): Measurement of the inequality of income. *Economic Journal* 30, 348–361.

Davies, J. and Hoy, M. (1994): The normative significance of using third-degree stochastic dominance in comparing income distributions. *Journal of Economic Theory* 64, 520–530.

Dorfman, R. (1979): A formula for the Gini coefficient. *Review of Economics and Statistics* (61), 146–149.

Fishburn, P. (1980): Stochastic dominance and moments of distributions. *Mathematics of Operations Research* 5, 94–100.

Gini, C. (1912): Variabiliá e Mutabilitá. *Studi Economico-Giuridici dell'Universitá di Cagliari*. 3, 1–158.

Lerman, R. and Yitzhaki, S. (1984): A note on the calculation and interpretation of the Gini index. *Economics Letters* (15), 363–368.

Muliere, P. and Scarsini, M. (1989): A note on stochastic dominance and inequality measures. *Journal of Economic Theory* 49(2), 314–323.

Núñez, J.J. (2006): La desigualdad económica medida a través de las curvas de Lorenz. *Revista de Métodos Cuantitativos para la Economía y la Empresa* 2, 67–108.

Ramos, H. and Sordo, M.A. (2003): Dispersion measures and dispersive orderings. *Statistics and Probability Letters* 61, 123–131.

Rao, V. (1969): Two decompositions of concentration ratio. *Journal of the Royal Statistical Society, Serie A* 132, 418–425.

Shorrocks, A. and Foster, J.E. (1987): Transfer sensitive inequality measures. *Review of Economic Studies* 54(1), 485–497.

Xu, K. (2004): How has the literature on Gini's index evolved in the past 80 years? *Economics working paper*, December 2004, Dalhousie University, Canada.

Xu, K. and Osberg, L. (2002): The social welfare implications, decomposability, and geometry of the sen family of poverty indices. *Canadian Journal of Economics* 35(1), 138–152.

Yitzhaki, S. and Schechtman, E. (2005): The properties of the extended Gini measures of variability and inequality. *Technical report, Social Science Research Network*. September 2005. Available at SSRN: http://ssrn.com/abstract=815564.