MEASURING THE DEPENDENCE BETWEEN DIMENSIONS OF WELFARE: A STUDY BASED ON SPEARMAN'S FOOTRULE AND GINI'S GAMMA

ANA PÉREZ
Departamento de Economía Aplicada and IMUVA, Valladolid, Spain,
Instituto Flores de Lemus, Madrid, Spain,
Avda. Valle Esgueva 6, 47011, Valladolid, Spain
perezesp@eaee.uva.es

MERCEDES PRIETO ALAIZ
Departamento de Economía Aplicada, Valladolid, Spain,
Avda. Valle Esgueva 6, 47011, Valladolid, Spain
prietoal@eaee.uva.es

January 20, 2016

Abstract

Welfare is multidimensional as it involves not only income, but also education, health or labour. The composite indicators of welfare are usually based on aggregating somehow the information across dimensions and individuals. However, this approach ignores the relationship between the dimensions being aggregated. To face this goal, in this paper we analyse the dependence between the dimensions included in the Human Development Index (HDI), namely income, health and schooling, through three copula-based measures of multivariate association: Spearman's footrule, Gini's gamma and Spearman's rho. We discuss their properties and prove new results on Spearman's footrule. The copula approach focuses on the positions of the individuals across dimensions, rather than the values that the variables attain for such individuals. Thus, it allows for more general types of dependence than the linear correlation. We base our study on data from 1980 till 2014 for the countries included in the 2015 Human Development Report. We find out that though the overall HDI has increased over this period, the dependence between its dimensions remains high and nearly unchanged so that the richest countries tend to be also the best ranked in both health and education.

Keywords: concordance, multivariate association, human development, copula, Spearman's rho.

1 Introduction

There is a general agreement that welfare is a multidimensional concept encompassing not only income, but also non-monetary dimensions such as education, labour and health; see, for instance, Kolm [1], Atkinson and Bourbignon [2] and Sen [3]. In turn, welfare depends on the level achievements in each dimension and also on two other factors related to the spread across the individual achievements of each dimension (Kolm [1]) and the degree of association across dimensions of welfare. As Atkinson and Bourguignon [2] point out, for two joint distributions with the same marginals but different degree of association, less association between dimensions of welfare is socially preferred. For instance, a world where one country is top-ranked in all welfare dimensions, another country is second-ranked in all welfare dimensions and so forth, will enjoy lower welfare than a world where the positions of the countries across dimensions are less correlated. This criterion was formalized by Tsui [4] who called it *correlation increasing majorization* and was further discussed by Bourbignon and Chackravarty [5] who propose a stronger version called *correlation increasing switch.* The relevance of incorporating dimensional interactions when measuring welfare has also been pointed out by Duclos *et al.* [6], Decancq *et al.* [7] and Seth [8]. In this framework, we face the problem of measuring the dependence between the three dimensions included in the most commonly used index of welfare, namely the HDI. These dimensions are income, health and education.

In the bivariate case, there are several measures of dependence. The most well-known is Pearson's linear correlation coefficient, which is suitable for linear dependence and elliptical data. However, empirical research shows that, in welfare economics, the data seldom belongs to this class. In this case, copula-based measures, such as Spearman's rho, Spearman's footrule and Gini's gamma (among others), would be more appropriate. These measures focus on the positions of the individuals across dimensions, rather than on the specific values that the corresponding variables attain for such individuals. Therefore, they are invariant to increasing and continuous transformations of the underlying variables and allow us to capture more general types of dependence than the linear correlation. The copula approach also enables generalizing these measures to a multivariate framework, where more than two variables are involved. In this framework, the pairwise dependence is not sufficient, in general, to infer properties of global dependence. For instance, some examples exist of variables which are pairwise independent but are multivariate dependent; see Nelsen [9] and Nelsen and Úbeda-Flores [10]. Therefore, in a multivariate setting, using measures of multivariate (rather than only bivariate) dependence is essential.

In welfare contexts, copula-based multivariate methods have been recently employed by Decancq [11] and Pérez and Prieto [12]. The former applies both the multivariate Kendall's tau and Spearman's rho in Nelsen [13] to illustrate how the dependence between the dimensions of well-being included in the HDI has evolved in Russia between 1995 and 2015. Pérez and Prieto [12] consider other multidimensional versions of Spearman's rho proposed by Nelsen [9], Nelsen and Úbeda-Flores [10] and García *et al.* [14], that are capable of revealing some other forms of dependence. These measures are then used to study how the dependence between the dimensions included in the AROPE (At Risk Of Poverty and Exclusion) rate, i.e. income, material needs and work intensity, has evolved in Spain over the period 2009-2013.

The contribution of this paper is twofold. First, we focus on two copula-based measures of multivariate association that have been scarcely used in practice, namely Gini's gamma (also known as Gini's rank association coefficient) and Spearman's footrule. We prove some new results concerning the latter and we fit them into the unifying framework of the Average Orthant Dependence measures, so that comparisons are easier. The measures considered can be regarded as L_1-distance alternatives to the well-known Spearman's rho coefficient. Hence, they have some appealing properties although they suffer from some minor disadvantages; see Genest *et al.* [15] and the references therein. Second, in our empirical application, we employ these two coefficients, as well as a multivariate Spearman's rho, to analyse how the association among the dimensions of the HDI has evolved in the world from 1980 till 2014. We use data from the countries included in the 2015 Human Development Report (see UNDP [16]). As far as we know, this is the first time that these measures are applied in a welfare context. Hence, some new insights on the topic are provided.

The paper is organized as follows. Section 2 introduces the copula and its main properties and recall its relationship with the concepts of concordance and orthant dependence. In Section 3 we describe Spearman's footrule, both in their population and empirical versions, and we prove some new results. Section 4 is devoted to Gini's gamma coefficient. It describes its main properties and compares it with both Spearman's footrule and Spearman's rho. Section 5 illustrates the use of these three coefficients to measure the evolution of the dependence among the dimensions of the HDI over the period 1980-2014. Section 6 closes the paper with a summary of our main conclusions.

2 Preliminary concepts and notation

2.1 Copulas

A $d-$dimensional copula C is a multivariate distribution function on $\mathbf{I}^d$, with $\mathbf{I}=[0,1]$, whose univariate marginal distributions are uniform on $\mathbf{I}$. Thus, each $d-$copula C can be associated with a $d-$dimensional random variable $\mathbf{U}=(U_1,...,U_d)$ such that U_i is $U(0,1)$ for $i=1,2,...,d$ and C is the joint distribution function of $\mathbf{U}$. Moreover, C is also an aggregation function since it is nondecreasing in each variable and fulfills the boundary conditions

$$C(0,...,0)=0 \text{ and } C(1,...,1)=1.$$

From the statistical point of view, the importance of copulas comes up in the Sklar's theorem (Sklar [17]). This theorem establishes that, if $\mathbf{X}=(X_1,...,X_d)$ is a $d-$dimensional random variable with joint distribution function F and univariate marginal distribution functions $F_1,\ldots,F_d$, then there exists a copula

$C : \mathbf{I}^d \to \mathbf{I}$ such that, for all $(x_1, \ldots, x_d) \in R^d$, F can be represented as

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{1}$$

If the margins $F_1, ..., F_d$ are all continuous, the copula C in (1) is unique; otherwise C is uniquely determined on $RanF_1 \times ... \times RanF_d$. Conversely, if C is a $d-$copula and $F_1, ..., F_d$ are univariate distribution functions, then the function F defined in (1) is a joint distribution function with margins $F_1, ..., F_d$. Therefore, the copula C allows to link any multivariate distribution function and its margins. Throughout this paper, we assume that $F_1, ..., F_d$ are all continuous.

Any copula C satisfies the Fréchet-Hoeffding bounds inequality

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u}),$$

for every $\mathbf{u} = (u_1, ..., u_d) \in \mathbf{I}^d$, where $W(\mathbf{u}) = max(u_1 + \cdots + u_d - d + 1, 0)$ and $M(\mathbf{u}) = \min(u_1, ..., u_d)$. M is always a copula and represents maximal dependence, i.e. the case where each component X_i of $\mathbf{X}$ is almost surely a strictly increasing function of every other component X_j. However, W is only a copula if $d = 2$, in which case it represents perfect negative dependence. If the variables $X_1, ..., X_d$ are independent, the copula of $\mathbf{X}$ will be the independent copula defined as $\Pi(\mathbf{u}) = u_1 \cdots u_d$. Another important function associated with a copula C is the survival function $\overline{C} : \mathbf{I}^d \to \mathbf{I}$ defined as

$$\overline{C}(\mathbf{u}) = p(\mathbf{U} > \mathbf{u}) = p(U_1 > u_1, \ldots, U_d > u_d).$$

In general, $\overline{C}$ is not a copula; see Nelsen [18] for a survey on the copula theory.

2.2 Concordance

Roughly speaking, concordance of random variables refers to their tendency to take all large values together or all small values together. More precisely, two observations (x_1, x_2) and (x_1', x_2') from a bidimensional random variable are concordant if $\{x_1 < x_1',\ x_2 < x_2'\}$ or $\{x_1 > x_1',\ x_2 > x_2'\}$ and they are discordant if $\{x_1 < x_1',\ x_2 > x_2'\}$ or $\{x_1 > x_1',\ x_2 < x_2'\}$. In higher dimensions, two observations $\mathbf{x} = (x_1, ..., x_d)$ and $\mathbf{x}' = (x_1', ..., x_d')$ from a $d-$dimensional continuous random vector are concordant if $\mathbf{x} < \mathbf{x}'$ or $\mathbf{x} > \mathbf{x}'$, where $\mathbf{x} < \mathbf{x}'$ ($\mathbf{x} > \mathbf{x}'$) denote the component-wise inequality $x_i < x_i'$ ($x_i > x_i'$), for all $i = 1, ..., d$. However, discordance does not generalize to dimensions $d \geq 3$. Thus, in the multivariate case, some dependence properties are based on the probability of concordance alone which can be written in terms of copulas as follows.

Let $\mathbf{X}$ and $\mathbf{X}'$ be two independent $d-$dimensional continuous random vectors with common univariate margins and $d-$copulas C and C', respectively, and let Q_d' be the probability of concordance between $\mathbf{X}$ and $\mathbf{X}'$. Nelsen [13] shows that

$$Q_d'(C, C') = p(\mathbf{X} > \mathbf{X}') + p(\mathbf{X} < \mathbf{X}') = \int_{\mathbf{I}^d} [C(\mathbf{u}) + \overline{C}(\mathbf{u})] dC'\mathbf{u}).$$

The function Q_d' is symmetric in its arguments, i.e. $Q_d'(C, C') = Q_d'(C', C)$, and is easily evaluated for pairs of the copulas M and Π. In particular, it can be shown that $Q'_d(M,M) = 1$, $Q'_d(M,\Pi) = 2/(d+1)$ and $Q'_d(\Pi,\Pi) = 1/2^{d-1}$. Moreover, as Behboodian *et al.* [19] point out, the function Q'_d can be extended to the case in which either C or C' is a measurable function from $\mathbf{I}^d$ to $\mathbf{I}$, as it will happen when $C = W$, for instance.

Some well-known measures of association can be written in terms of Q'_d. For instance, the multivariate population version of Spearman's rho for a random vector $\mathbf{X} = (X_1, ..., X_d)$ with $d-$copula C proposed by Nelsen [13] is defined as

$$\rho_d(C) = \frac{(d+1)}{2^d - (d+1)}[2^{d-1}Q'_d(C,\Pi) - 1]. \tag{2}$$

Thus, in a sense, Spearman's rho measures the concordance relationship between the distribution of $\mathbf{X}$ as represented by their copula C, and independence as represented by the copula Π. Noticeably, ρ_d takes the value 1 when $C = M$ and the value 0 when $C = \Pi$, but its lower bound is $[2^d - (d+1)!]/[d!(2^d - d - 1)]$.

2.3 Orthant dependence

Let $\mathbf{X} = (X_1, ..., X_d)$ be a d-dimensional continuous random variable with marginals $F_1, .., F_d$ and copula $C : \mathbf{I}^d \to \mathbf{I}$ such that C is the joint distribution function of $\mathbf{U} = (U_1, ..., U_d)$, where $U_i = F_i(X_i)$ for $i = 1, 2, ..., d$. Then:

- $\mathbf{X}$ is positively *lower orthant dependent* (PLOD) if $C(\mathbf{u}) \geq \Pi(\mathbf{u})$, for each $\mathbf{u} \in \mathbf{I}^d$, that is, if the probability that the variables $X_1, ..., X_d$ take simultaneously small values is at least as great as it would be were they independent.
- $\mathbf{X}$ is positively *upper orthant dependent* (PUOD) if $\overline{C}(\mathbf{u}) \geq \overline{\Pi}(\mathbf{u})$, for each $\mathbf{u} \in \mathbf{I}^d$, that is, if the probability that the variables $X_1, ..., X_d$ take simultaneously large values is at least as great as it would be were they independent.
- $\mathbf{X}$ is positively *orthant dependent* (POD) if both inequalities hold.

The corresponding negative concepts are defined by reversing the sense of the inequalities. For $d = 2$, PLOD and PUOD are the same and reduce to POD. Obviously, the same reduction occurs with the analogous negative concepts.

In this framework, the differences $[C(\mathbf{u}) - \Pi(\mathbf{u})]$ and $[\overline{C}(\mathbf{u}) - \overline{\Pi}(\mathbf{u})]$ can be regarded as measures of "local" lower and upper orthant dependence, respectively. Accordingly, Dolati and Úbeda-Flores [20] introduce the following function:

$$D(C, C') = \int_{\mathbf{I}^d} [C(\mathbf{u}) - \Pi(\mathbf{u}) + \overline{C}(\mathbf{u}) - \overline{\Pi}(\mathbf{u})] dC'(\mathbf{u}),$$

that represents the $C'-$average of orthant dependence (AOD), i.e., the expectation of $[C(\mathbf{U}) - \Pi(\mathbf{U}) + \overline{C}(\mathbf{U}) - \overline{\Pi}(\mathbf{U})]$ taken with respect to copula C'. Moreover, they also define a general class of AOD measures of multivariate association as

$$\omega_d(C) = \alpha_d D(C, C'),$$

where $\alpha_d = 1/D(M, C')$, and establish conditions on C' for ω_d to be a measure of multivariate concordance. Noticeably, the function $D(C, C')$ can be easily re-written in terms of the probability of concordance as

$$D(C, C') = Q'_d(C, C') - Q'_d(\Pi, C'). \tag{3}$$

Therefore, the AOD measure ω_d can also be regarded as a normalized probability of concordance between C and C' given by

$$\omega_d(C) = \frac{Q'_d(C, C') - Q'_d(\Pi, C')}{Q'_d(M, C') - Q'_d(\Pi, C')}.$$

The coefficient ρ_d in (2) is an example of an AOD measure of multivariate association since it can be written as

$$\rho_d(C) = \frac{Q'_d(C, \Pi) - Q'_d(\Pi, \Pi)}{Q'_d(M, \Pi) - Q'_d(\Pi, \Pi)} = \alpha_d \ D(C, \Pi), \tag{4}$$

with $\alpha_d = 1/D(M, \Pi) = [2^{d-1}(d+1)]/[2^d - (d+1)]$. Moreover, ρ_d is also a measure of multivariate concordance according to the axiomatic definitions in Dolati and Úbeda-Flores [20] and Taylor [21].

In a welfare context, were the vector $\mathbf{X}$ represents the dimensions of welfare, the coefficient ρ_d has a very interesting interpretation (see Decancq [11]) as the normalized probability that a randomly selected individual from the distribution of $\mathbf{X}$ outranks or is outranked by a randomly selected individual from a reference distribution with independent dimensions. Therefore, the more dependence there is between the dimensions of welfare, the higher this normalized probability and so the higher the value of ρ_d.

In next sections we introduce multivariate versions of other not so well-known measures of association, like Spearman's footrule and Gini's gamma, and we describe their relationship with the concepts introduced in this section.

3 A multivariate version of Spearman's footrule

The population version of Spearman's footrule for a random vector (X_1, X_2) with copula C, that will be denoted by φ_2, is given by (see Nelsen [22])

$$\varphi_2 = 1 - 3 \int\int_{\mathbf{I}^2} |u_1 - u_2| dC(u_1, u_2).$$

This coefficient is similar to Spearman's ρ_2 in (2) which can be re-written as

$$\rho_2 = 1 - 6 \int\int_{\mathbf{I}^2} (u_1 - u_2)^2 dC(u_1, u_2).$$

Whereas the former is based on absolute differences, the latter employs squared differences. Because of that, Spearman's footrule has been viewed as an L_1

alternative to Spearman's rho. Moreover, the Spearman's footrule φ_2 can also be written in terms of the probability of concordance as follows

$$\varphi_2 = 6 \int\int_{\mathbf{I}^2} M(u_1, u_2) dC(u_1, u_2) - 2 = 3Q_2'(C, M) - 2.$$

This coefficient takes the values: 1 when $C = M$; 0 when $C = \Pi$; and $-1/2$ (rather than -1) when $C = W$. Hence, unlike ρ_2, it does not behave in a symmetric fashion with respect to the upper and lower Fréchet-Hoeffding bounds.

The multivariate version of the Spearman's footrule for a random vector $\mathbf{X} = (X_1, ..., X_d)$ with $d-$copula C, defined by Úbeda-Flores [23], is given by

$$\varphi_d(C) = \frac{(d+1)}{(d-1)} Q_d'(C, M) - \frac{2}{(d-1)}. \tag{5}$$

Thus, in a sense, φ_d can be regarded as a normalized concordance "distance" between the distribution of $\mathbf{X}$ as represented by their copula C, and maximal dependence as represented by the copula M.

Úbeda-Flores [23] shows that φ_d is 0 when $C = \Pi$ and 1 when $C = M$ but its lower bound is $-1/d$. Moreover, he shows that in the trivariate case ($d = 3$), where $\mathbf{X} = (X_1, X_2, X_3)$ with copula C, the coefficient $\varphi_3(C)$ can be written as

$$\varphi_3(C) = \frac{\varphi_{12} + \varphi_{13} + \varphi_{23}}{3}, \tag{6}$$

where φ_{ij} denotes the corresponding Spearman's footrule for the bivariate random variable (X_i, X_j), with $1 \le i < j \le 3$.

We next show that φ_d can be regarded as an AOD multivariate measure of association. However, as Dolati and Úbeda-Flores [20] point out, φ_d is not a measure of multivariate concordance according to their axiomatic definition.

Proposition 1 *Let C be a d-copula. Then $\varphi_d(C)$ is an AOD measure of association given by*

$$\varphi_d(C) = \alpha_d \ D(C, M), \tag{7}$$

with $\alpha_d = 1/D(M, M)$.

Proof. Recall that $Q_d'(\Pi, M) = 2/(d+1)$. Thus, using (3) and (5) we have

$$\varphi_d(C) = \frac{(d+1)}{(d-1)} \left[Q_d'(C, M) - Q_d'(\Pi, M)\right] = \frac{(d+1)}{(d-1)} D(C, M).$$

But $D(M, M) = Q_d'(M, M) - Q_d'(\Pi, M) = 1 - 2/(d+1) = (d-1)/(d+1)$, thus

$$\varphi_d(C) = \frac{Q_d'(C, M) - Q_d'(\Pi, M)}{Q_d'(M, M) - Q_d'(\Pi, M)} = \frac{D(C, M)}{D(M, M)},$$

and the result follows. ■

In practice, the copula C is unknown and $\varphi_d(C)$ must be estimated from the data. Therefore, a sample version of $\varphi_d(C)$ is required. Let $\{(X_{1j}, ..., X_{dj})\}_{j=1,...,n}$

be a sample of n serially independent random vectors from the d-dimensional vector $\mathbf{X} = (X_1, ..., X_d)$ and let R_{ij} be the *rank* of X_{ij} among $\{X_{i1}, ..., X_{in}\}$, with $i = 1, ..., d$ and $j = 1, ..., n$. Úbeda-Flores [23] proposes the following estimator of φ_d:

$$\widehat{\varphi}_d = 1 - \frac{d+1}{d-1}\frac{1}{n^2-1}\sum_{j=1}^{n}\left[\max_{1\le i\le d}(R_{ij}) - \min_{1\le i\le d}(R_{ij})\right]. \tag{8}$$

Genest *et al.* [15] show that, under fairly general conditions, $\widehat{\varphi}_d$ is an asymptotically unbiased estimator of φ_d and it is asymptotically normally distributed.

Note that, in the case of perfect dependence, i.e., when the ranks in each dimension coincide, $\widehat{\varphi}_d = 1$. Also note that, when $d = 2$, the coefficient defined in (8) reduces to the sample bivariate Spearman's footrule $-$ usually denoted as f_S $-$ given by

$$\widehat{\varphi}_2 = f_S = 1 - \frac{3}{n^2-1}\sum_{j=1}^{n}|R_{1j} - R_{2j}|.$$

Moreover, we next show that, in the trivariate case ($d = 3$), property (6) continues to hold for the corresponding empirical coefficients.

Proposition 2 *The sample Spearman's footrule $\widehat{\varphi}_3$ is equal to the average of the three pairwise sample Spearman's footrule coefficients, that is*

$$\widehat{\varphi}_3 = \frac{\widehat{\varphi}_{12} + \widehat{\varphi}_{13} + \widehat{\varphi}_{23}}{3}. \tag{9}$$

Proof. The trivariate sample Spearman's footrule coefficient for a tridimensional random vector (X_1, X_2, X_3) is

$$\widehat{\varphi}_3 = 1 - \frac{2}{n^2-1}\sum_{j=1}^{n}[\max(R_{1j}, R_{2j}, R_{3j}) - \min(R_{1j}, R_{2j}, R_{3j})],$$

and the pairwise sample Spearman's footrule coefficient for a pair (X_i, X_k) is

$$\widehat{\varphi}_{ik} = 1 - \frac{3}{n^2-1}\sum_{j=1}^{n}|R_{ij} - R_{kj}|. \tag{10}$$

But $|u - v| = 2\max(u, v) - (u + v)$, so that

$$\frac{\widehat{\varphi}_{12} + \widehat{\varphi}_{13} + \widehat{\varphi}_{23}}{3} = 1 - \frac{1}{n^2-1}\sum_{j=1}^{n}(|R_{1j} - R_{2j}| + |R_{1j} - R_{3j}| + |R_{2j} - R_{3j}|) =$$

$$= 1 - \frac{2}{n^2-1}\sum_{j=1}^{n}[\max(R_{1j}, R_{2j}) + \max(R_{1j}, R_{3j}) + \max(R_{2j}, R_{3j}) - R_{1j} - R_{2j} - R_{3j}].$$

Now, it is checked readily that the term in brackets in the equation above is equal to $\{\max(R_{1j}, R_{2j}, R_{3j}) - \min(R_{1j}, R_{2j}, R_{3j})\}$. For instance, if $R_{1j} > R_{2j} > R_{3j}$,

$$\max(R_{1j}, R_{2j}, R_{3j}) - \min(R_{1j}, R_{2j}, R_{3j}) = R_{1j} - R_{3j},$$

and

$$\max(R_{1j}, R_{2j}) + \max(R_{1j}, R_{3j}) + \max(R_{2j}, R_{3j}) - R_{1j} - R_{2j} - R_{3j} =$$
$$= R_{1j} + R_{1j} + R_{2j} - R_{1j} - R_{2j} - R_{3j} = R_{1j} - R_{3j}.$$

Proceeding in a similar way for the other possible cases, the result follows. ■

4 A multivariate version of Gini's gamma

Another measure of association closely related to Spearman's rho and Spearman's footrule is the Gini's gamma. The bivariate population version of Gini's gamma is based on the L_1-distance from any point (u_1, u_2) in the square $\mathbf{I}^2$ to the principal and secondary diagonals of $\mathbf{I}^2$. In particular, given a random vector (X_1, X_2) with copula C, its Gini's gamma coefficient is defined as

$$\gamma_2 = 2 \int\int_{\mathbf{I}^2} (|1 - u_1 - u_2| \; - |u_1 - u_2|) dC(u_1, u_2).$$

Its relationship to Spearman's ρ_2 shows up when this is re-written as

$$\rho_2 = 3 \int\int_{\mathbf{I}^2} [(1 - u_1 - u_2)^2 - (u_1 - u_2)^2] dC(u_1, u_2);$$

see Nelsen [13]. Thus, Gini's gamma (like Spearman's footrule) employs absolute differences rather than squared differences and so it can also be regarded as an L_1 competitor to Spearman's rho.

To further enhance the differences and similarities between these three coefficients, let us note that bivariate Gini's gamma can also be written in terms of the probability of concordance as

$$\gamma_2 = 4 \int\int_{\mathbf{I}^2} [(M(u_1, u_2) + W(u_1, u_2)] dC(u_1, u_2) - 2 = 2[Q_2'(C, M) + Q_2'(C, W)] - 2;$$

see Nelsen [22]. Like Spearman's ρ_2, the coefficient γ_2 takes the value 1 when $C = M$, the value 0 when $C = \Pi$ and the value -1 when $C = W$. Nevertheless, whereas bivariate Spearman's ρ_2 and Spearman's footrule φ_2 measure concordance relationships between copula C and Π (independence) and copula C and M (perfect positive dependence), respectively, bivariate Gini's gamma measures a concordance relationship between copula C and monotone dependence, as represented by the copulas M (perfect positive dependence) and W (perfect negative dependence). In higher dimensions ($d \geq 3$), this interpretation does not hold since W is no longer a copula and the concept of perfect negative dependence is not that clear.

Behboodian *et al.* [19] works out an alternative expression of γ_2 given by

$$\gamma_2 = 4 \; [Q_2'(C, A) - Q_2'(\Pi, A)],$$

where A is the average of the upper and lower Fréchet-Hoeffding bounds, i.e. $A = (M + W)/2$. Based on this expression, they define the multivariate population version of Gini's gamma associated with a $d-$copula C as

$$\gamma_d(C) = \alpha_d\,[Q'_d(C, A) - Q'_d(\Pi, A)], \tag{11}$$

where $\alpha_d = 1/[Q'_d(M, A) - Q'_d(\Pi, A)]$ with:

$$Q'_d(M, A) = 1 - \sum_{i=1}^{d-1} \frac{1}{4i},$$

$$Q'_d(\Pi, A) = \frac{1}{d+1} + \frac{1}{2(d+1)!} + \sum_{i=0}^{d} (-1)^i \binom{d}{i} \frac{1}{2(i+1)!}.$$

The coefficient γ_d in (11), like ρ_d and φ_d, takes the value 1 when $C = M$ (perfect dependence) and the value 0 when $C = \Pi$ (independence). Moreover, taking into account equations (3) and (11) it turns out that the multivariate Gini's gamma coefficient γ_d can be written as

$$\gamma_d(C) = \frac{Q'_d(C, A) - Q'_d(\Pi, A)}{Q'_d(M, A) - Q'_d(\Pi, A)} = \alpha_d\, D(C, A), \tag{12}$$

where $\alpha_d = 1/D(M, A)$. Therefore, γ_d is also an AOD measure of association; compare expression (12) with the AOD representations of the coefficients ρ_d and φ_d in (4) and (7), respectively. Whereas ρ_d and φ_d are AOD measures with respect to copulas Π and M, respectively, the multivariate Gini's γ_d is an AOD measure of association with respect to A. Thus, the Gini's coefficient γ_d can be interpreted as a normalized average "distance" between the copula C and the average of the upper and lower Fréchet-Hoeffding bounds.

As to our knowledge, it is not determined whether γ_d satisfies the axioms established in Dolati and Úbeda-Flores [20] and Taylor [21] for measures of multivariate concordance, since A is not even a copula. This topic is left for further research.

In the trivariate case, Behboodian *et al.* [19] show that the trivariate Gini's gamma for (X_1, X_2, X_3) is just the average of the three corresponding pairwise Gini's coefficients, that is

$$\gamma_3(C) = \frac{\gamma_{12} + \gamma_{13} + \gamma_{23}}{3}, \tag{13}$$

where γ_{ik} denotes the pairwise Gini's gamma for the bivariate random variable (X_i, X_k), with $1 \le i < k \le 3$. Furthermore, Behboodian *et al.* [19] provides the sample version of the coefficient γ_d and show that, in the trivariate case, the relationship (13) also holds for the corresponding sample versions, that is

$$\widehat{\gamma}_3 = g_3 = \frac{\widehat{\gamma}_{12} + \widehat{\gamma}_{13} + \widehat{\gamma}_{23}}{3}, \tag{14}$$

where $\widehat{\gamma}_{ik}$ is the bivariate sample Gini's gamma for the pair (X_i, X_k), namely

$$\widehat{\gamma}_{ik} = \frac{1}{\lfloor n^2/2 \rfloor} \sum_{j=1}^{n} (|n + 1 - R_{ij} - R_{kj}| - |R_{ij} - R_{kj}|), \tag{15}$$

where $\lfloor x \rfloor$ denotes de integer part of x.

Noticeably, properties (13) and (14) are shared by both multivariate Spearman's footrule (see Section 3) and multivariate Spearman's rho (see Dolati and Úbeda-Flores [20]). Hence, in the trivariate case, the sample version of the multivariate Spearman's rho in (2) can also be computed as the average of the well-known pairwise sample Spearman's rho coefficients.

5 Empirical application

In this section, we apply the copula-based coefficients of multivariate association described in previous sections to measure how the dependence among dimensions of welfare has evolved in the world over the period 1980-2014.

5.1 Data and variables

The dimensions of welfare we have selected are those included in the most commonly used index of welfare, namely the HDI. These dimensions are health, education and income. Health is approximated by life expectancy at birth, indicating number of years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout the infant's life. Education is characterized by two indicators: mean years of schooling for adult (average number of years of education received by people ages 25 and older) and expected years of schooling for children (number of years of schooling that a child of school entrance age can expect to receive if prevailing patterns of age-specific enrolment rates persist throughout the child's life). Finally, income is approximated by Gross National Income per capita in 2011 Purchasing Power Parity terms. The data of the four indicators are available at the web page http://hdr.undp.org and cover yearly data over the period 1980–2014 with discontinuous intervals up to 2005.

Following the methodology of UNDP [16], we normalize the four indicators to put on a common [0,1] scale, in the following way:

$$Normalized\ Indicator = \frac{(actual\ value -\ minimum\ value)}{(maximum\ value - minimum\ value)}.$$

The minimum and maximum values, fixed goalposts, can be interpreted for each indicator as "natural zero" and "aspirational goals", respectively (for further details see UNDP [16] and Anand and Sen [24]). To have only one index for the education dimension, the arithmetic mean of the two normalized indicators of education described above is taken. Finally, to get a consistent time series over the full period of analysis, we have computed all the indicators using the methodology of UNDP [16]. Thus, the HDI is calculated through the geometric mean of the three dimensional indices.

The unit of our analysis is the country and the countries considered are those included in the UNDP [16]. From 2010 till 2015 we have full data for 187 countries. Nevertheless, there are missing data for some countries in some years before 2010. For instance, in 1980 and 1985 we lack data from 63 countries, but over the period 2005- 2009 the missing data reduce to 13 countries.

5.2 Results

We first analyse the evolution of the distribution of the HDI. In order to do that, we plot in Figure 1 the corresponding boxplots for the years considered. The box-and-whisker plot (called simply a boxplot) is a method of displaying data that consists of a box extending from the first quartile ($Q1$) to the third quartile ($Q3$), with an added line at the median and a dot at the mean and two whiskers extending from $Q1$ to $Q1 - 1.5(Q3 - Q1)$ and from $Q3$ to $Q3 + 1.5(Q3 - Q1)$; see Tukey [25]. Figure 1 shows that, in terms of the HDI, there has been a steady increase of welfare around the world over the last thirty five years though this increase has not been homogeneous over the full period. In particular, we observe a remarkable increase of the HDI over the period 1980-2005, where the world average HDI rose from 0.543 in 1980 to 0.650 in 2005. This upward trend continued mildly till 2014, when the average stood at 0.694. This overall increase reflects aggregate improvements in the three dimensions that make up the HDI (health, education and income).

INSERT FIGURE 1 AROUND HERE

The question arises whether this increase in welfare is accompanied by an increase or a decrease of the interdependence among the three dimensions of the HDI. As we point out in the introduction, incorporating dimensional interactions becomes relevant in measuring welfare. Furthermore, since we face a multidimensional problem, measures of multivariate dependence should be used to grasp the wide spectrum of possible bivariate and multivariate relationships. To face this goal we apply the measures of multivariate association described in sections 3 and 4. To compute the multivariate sample versions of Spearman's footrule and Gini's gamma we apply formulae (9)-(10) and (14)-(15), respectively. As a benchmark, we also compute the sample version of multivariate Spearman's rho as the average of the three pairwise sample Spearman's rho coefficients. The results obtained are displayed in Figure 2.

INSERT FIGURE 2 AROUND HERE

Several conclusions emerge from this figure. First, the three coefficients are all positive and moderately high revealing an important degree of association between the three dimensions of welfare. This means that the richest countries tend to be also the healthiest and best educated ones, and the poorest countries tend to be also those with the lowest health and the least educated ones. Noticeably, this conclusion is robust to the measure used. Another remarkable feature from Figure 2 is that the values of the three measures have barely changed since 1980, indicating that no improvements in lessening the multivariate dependence between income, health and education have been achieved over the last years. It is also worth noting that Spearman's rho is always higher than the two other measures, regardless of the year considered. Whereas Spearman's footrule has kept around 0.59 and Gini's gamma around 0.67, Spearman's rho has been about 0.81. Recall that these measures can take different values as they capture distinct aspects of the multivariate dependence between the three dimensions. In fact, it could be concluded that the joint distribution of the three development dimensions is further from independence (as the Spearman's rho reveals) than from the maximal dependence (as captured by the Spearman's footrule).

To complement these results, Table 1 displays, for each of the three measures considered, the corresponding possible pairwise coefficients to measure bivariate association between: health and education (indexed as 12), health and income (indexed as 13) and education and income (indexed as 23).

[INSERT TABLE 1 AROUND HERE]

In this table, we first observe that all the pairwise coefficients are positive, regardless of the measure used and the pair of variables being associated. This confirms that, given a pair of dimensions, countries that are top-ranked in one dimensions are more likely to be also top-ranked in the other. The positive association between health and income is consistent with other studies, like Preston [26], Sen [27] and Deaton [28]. The positive association between health and education is also reported by Deaton [28], Grossman [29] and Cutler and Lleras-Muney [30]. Evidence on the positive relationship between education and income dates back to the seminal works of Becker ([31], [32]), Schultz [33] and Denison [34]. Recent works on the topic include Barro [35] and Francis and Iyare [36], among others. Table 1 also reveals that the temporal pattern of the bidimensional coefficients depends on the dimensions in hand and the period considered. From 2005 onwards, all measures keep nearly constant but the degree of association between health and education is always lower than the degree of association between income and education or income and health, which are both nearly the same over this period, regardless of the measure used. By contrast, the patters between 1980 and 2005 are quite different. Whereas association between health and education decreased sharply in this period, the association between income and health and income and education both increase though at different rates. Again, this conclusion is robust to the measure used.

As a final comment, let us recall that Spearman's rho coefficients are based on squared differences of ranks, while Gini's gamma and Spearman's footrule coefficients are based on absolute differences. Hence, Gini's gamma and Spearman's footrule are more robust than Spearman's rho to higher discrepancies between the ranks in different dimensions for a given country. For instance, Equatorial Guinea belongs to the 1st decile in life expectancy (position 15 in this dimension) but it is within the 30% of countries with the highest income (position 134 in this dimension). Therefore, the influence of a country like this one is more harmful to $\widehat{\rho}_{12}$ (and consequently to $\widehat{\rho}_3$) than to the corresponding gamma and footrule coefficients, that keep more robust to such a big difference between the ranks.

6 Conclusions

In this paper we have focused on two copula-based multivariate measures of association that have been scarcely used in practice, namely Gini's gamma and Spearman's footrule. As a benchmark, we have also considered a multivariate generalization of the well-known Spearman's rho rank association coefficient. We have revised their main properties and discussed their similarities and differences and we have also provided new results for the multivariate Spearman's footrule. In our empirical application, we have applied these three measures to analyse how the dependence among dimensions of welfare has evolved in the world over the period 1980-2014. The dimensions considered are those included in the HDI, that is, income, health and education, and the countries considered are those included in the 2015 Human Development Report.

Our main conclusion is that, although the welfare, as measured by the HDI, has increased over the period considered, the degree of multivariate dependence among income, health and schooling has kept positive and high and nearly unchanged. Moreover, this result is robust to the measure of multivariate dependence used. This means that small (high) values of the three welfare dimensions tend to occur together, so that the richest countries tend to be also the healthiest and the most educated ones, while the poorest countries tend to have simultaneously the lowest level of both education and health. Therefore, even though the overall welfare is steadily improving, it is very unlikely for a country that is low-ranked in one dimension of welfare to get a higher position in the other two dimensions and vice-versa, specially if the country is low ranked in income. Moreover, when looking at the bidimensional relationships, our results reveal that, over the last decade, the degree of association between health and education has become lower than the degree of association between income and education or income and health.

Acknowledgments

Financial support from the Comunidad de Castilla y León under project VA066U13 is gratefully acknowledged by the first author. The second author acknowledges financial support from the Spanish Government under project ECO2012-32178. We also thank P.C. Alvarez for helping us with R graphics. The usual disclaimers apply.

References

[1] S. Kolm, “Multidimensional equalitarianisms”, *Quaterly Journal Economic* **91** (1977) 1-13.

[2] A.B. Atkinson and F.Bourguignon, "The comparison of multidimensioned distributions of economic status", *Review of Economic Studies* **12** (1982) 183-201.

[3] A.K. Sen, *Commodities and capabilities* (North Holland, Amsterdam, 1985).

[4] K.Y. Tsui, "Multidimensional inequality and multidimensional generalized entropy measures: an axiomatic derivation" *Soc. Choice Welf.* **16** (1999) 145–157.

[5] F. Bourguignon and S.R. Chakravarty, "The measurement of multidimensional poverty", *Journal of Economic Inequality* **1** (2003) 25–49.

[6] J.Y. Duclos, D. Sahn and S. Younger, "Robust multidimensional poverty comparisons", *Economic Journal* **116** (2006) 943-968.

[7] K. Decancq, A. Decoster and E. Schokkaert, "The evolution of world inequality in well-being", *World Dev.* **37** (2009) 11–25.

[8] S. Seth, "Inequality, interactions, and human development", *J. Hum. Dev.* **10** (2009) 375–396.

[9] R.B. Nelsen, "Nonparametric measures of multivariate association", in *Distributions with given marginals and related topics*, eds. L. Rüschendorf, B. Schweizer and M. D. Taylor (Hayward, CA: Institute of Mathematical Statistics, 1996) pp. 223–232.

[10] R.B. Nelsen and M. Úbeda-Flores, "Directional dependence in multivariate distributions", *Ann Inst Stat Math* **64** (2012) 677-685.

[11] K. Decancq, "Copula-based measurement of dependence between dimensions of well-being", *Oxford Economic Papers* **66** (2013) 681-701.

[12] A. Pérez and M. Prieto, "Measuring dependence between dimensions of poverty in Spain: An approach based on copulas", *Proc. 2015 Conference of the International Fuzzy System Association and the European Society for Fuzzy Logic and Technology*, Gijón, Spain, July 2015, pp. 734–740.

[13] R.B. Nelsen, "Concordance and copulas: A survey", in *Distributions with Given Marginals and Statistical Modelling*, eds. C.M. Cuadras, J. Fortiana and J.A. Rodriguez-Lallena (Kluwer, Dordrecht, 2002) pp. 169-178.

[14] J.E. García, V.A. González-López and R.B. Nelsen, "A new index to measure positive dependence in trivariate distributions", *Journal of Multivariate Analysis* **115** (2013) 481–495.

[15] C. Genest, J. Neslehová and N.B. Ghorbal, "Spearman's footrule and Gini's gamma: a review with complements", *Journal of Nonparametric Statistics* **22** (2010) 937-954.

[16] UNDP, *Human Development Report 2015* (Oxford University Press, New York, 2015)

[17] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges", *Pub. Inst. Statist. Univ. Paris* **8** (1959) 229-231.

[18] R.B. Nelsen, *An Introduction to Copulas*, 2nd ed. (Springer, New York, 2006).

[19] J. Behboodian, A. Dolati and M. Úbeda-Flores, "A multivariate version of Gini's rank association coeficient", *Statistical Papers* **48** (2007) 295-304.

[20] A. Dolati and M. Úbeda-Flores, "On measures of multivariate concordance", *Journal of Probability and Statistical Science* **4** (2006) 147-163.

[21] M.D. Taylor, "Multivariate measures of concordance", *Ann. Inst. Statist. Math.* **59** (2007) 789–806.

[22] R.B. Nelsen, "Concordance and Gini's measure of association", *Nonparametric Statistics* **9 (**1998) 227-238.

[23] M. Úbeda-Flores, "Multivariate versions of Bloomqvist's beta and Spearman's footrule", *Ann. Inst. Statist. Math.* **57** (2005) 781-788.

[24] S. Anand and A. Sen, "The income component of the Human Development Index", *Journal of Human Development and Capabilities* **1** (2000) 83–106.

[25] J.W. Tukey, *Exploratory Data Analysis* (Addison-Wesley, Massachusetts, 1977).

[26] S. Preston, "The changing relation between mortality and level of economic development", *Population Studies* **29** (1975) 231–248.

[27] A. Sen, "The economics of life and death", *Scientific American* **266** (1993) 40–47.

[28] A. Deaton, "Health, Income and Inequality". *NBER Reporter*, Spring 2003.

[29] M. Grossman, "Education and non-market outcomes", in *Handbook of the Economics of Education*, eds. E. Hanushek and F. Welch (Elsevier, Amsterdam, 2006) pp. 577-633.

[30] D. Cutler, A. Lleras-Muney, "Education and health: evaluating theories and evidence", in *Making Americans Healthier: Social and Economic Policy as Health Policy*, eds. R.F. Schoeni, J.S. House, J.A. Kaplan and H. Pollack (Russell Sage Foundation, New York, 2008)

[31] G. Becker, "Investment in human capital: A theoretical analysis", *Journal of Political Economy* **70** (1962) 9 49.

[32] G. Becker, *Human Capital* (Columbia University Press, New York, 1964).

[33] T. Schultz, “Investment in Human Capital”, *American Economic Review* **51** (1961) 1–17.

[34] E. Denison, “Education, Economic Growth, and Gaps in Information”, *The Journal of Political Economy* **70** (1962) 124-128.

[35] R. Barro, “Education and Economic Growth”, in *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being. : International Symposium Report*, eds. J.F. Helliwell (OECD and Human Resources Development Canada, Quebec, 2001), pp. 13-41.

[36] B. Francis and S. Iyare, “Education and development in the caribbean: a cointegration and causality approach”, *Economics Bulletin* **15** (2006) 1-13.

Table 1. Evolution of the sample bivariate Spearman's footrule, Gini's gamma and Spearman's rho coefficients between dimensions 1 (health), 2 (education) and 3 (income) of the HDI over the period 1980-2014

Year	$\widehat{\varphi}_{12}$	$\widehat{\varphi}_{13}$	$\widehat{\varphi}_{23}$	$\widehat{\gamma}_{12}$	$\widehat{\gamma}_{13}$	$\widehat{\gamma}_{23}$	$\widehat{\rho}_{12}$	$\widehat{\rho}_{13}$	$\widehat{\rho}_{23}$
1980	0.670	0.586	0.526	0.752	0.659	0.614	0.885	0.791	0.737
1985	0.660	0.612	0.545	0.744	0.680	0.634	0.884	0.819	0.770
1990	0.561	0.661	0.541	0.657	0.728	0.623	0.813	0.856	0.769
2000	0.537	0.653	0.596	0.629	0.722	0.680	0.788	0.855	0.817
2005	0.539	0.628	0.610	0.628	0.701	0.687	0.781	0.836	0.828
2006	0.544	0.612	0.608	0.638	0.681	0.685	0.783	0.820	0.825
2007	0.542	0.611	0.614	0.633	0.677	0.692	0.783	0.818	0.829
2008	0.544	0.609	0.618	0.635	0.675	0.695	0.784	0.814	0.831
2009	0.544	0.605	0.614	0.628	0.699	0.690	0.781	0.811	0.827
2010	0.545	0.622	0.602	0.628	0.690	0.684	0.780	0.828	0.821
2011	0.544	0.624	0.614	0.626	0.691	0.692	0.778	0.831	0.827
2012	0.544	0.624	0.611	0.627	0.691	0.692	0.777	0.830	0.826
2013	0.541	0.622	0.613	0.624	0.688	0.694	0.774	0.826	0.828
2014	0.560	0.615	0.610	0.650	0.691	0.687	0.802	0.825	0.831

Figure 1: Evolution of the distribution of the HDI over the period 1980-2014

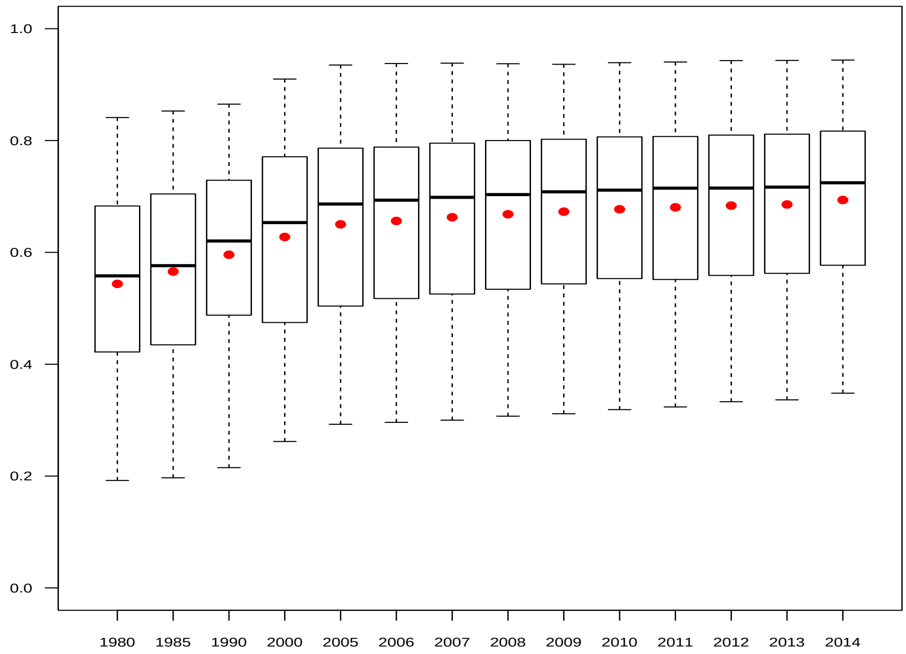

Figure 2: Evolution of the multivariate association between the dimensions of the HDI over the period 1980-2014 as measured by the coefficients Spearman's footrule (black), Gini's gamma (dark grey) and Spearman's rho (light grey) coefficients