MULTIDIMENSIONAL POVERTY MEASUREMENT: DEPENDENCE BETWEEN WELL-BEING DIMENSIONS USING COPULA FUNCTION

Kateryna Tkach, Chiara Gigliarano

1. Introduction

Well-being is suggested to be multidimensional and to consist of both monetary and non-monetary attributes. Recent studies have investigated the aspect of this multidimensionality from various perspectives, i.e. from the discussion of key dimensions to be considered to the extension of univariate poverty measures into multidimensional setting (Bourguignon and Chakravarty, 2003; Alkire and Foster, 2011). Although current approaches to well-being indices have mostly focused on disjoint performance of individuals in several dimensions (e.g. the Human Development Index, OECD Better Life Index), recent literature has pointed out the relevance of the dependence between dimensions (Decancq and Lugo, 2012; Decancq, 2014).

This paper focuses on the dependence between three dimensions of well-being, namely income, education and health, using cross-sectional data from the EU-SILC referred to the year 2015. We employ copula function (Nelsen, 2006) and copula-based measures of concordance, namely Kendall's and Spearman's rank correlation coefficients, to measure the dependence between specified dimensions.

The rest of the paper is structured as follows: Section 2 provides a brief overview about copula and copula-based dependence measures, Section 3 shows data and the results of parametric and nonparametric estimation of Spearman's rho and Kendall's tau coefficients. Section 4 describes copula-based weighting scheme in multidimensional poverty measure and, finally, Section 5 concludes.

2. Methodology

Let us introduce necessary notations here. Let $F(x)$ and $G(y)$ denote marginal distributions of random variables X and Y and $H(x,y) = P[X \leq x, Y \leq y]$ their joint distribution function. Finally, let $C(u,v)$ with $(u,v) \in [0,1]^2$ denote bivariate

copula function and $C(\boldsymbol{u})$ with $\boldsymbol{u} = (u_1, \dots, u_d) \in [0,1]^d$ be d-dimensional copula. Copula is a function that separates the dependence behavior from the marginal distributions. The main definition about the copula function is the following.

If $\mathrm{H}(x, y)$ is a joint distribution function with uniform margins F and G, then there exists a 2-dimensional copula $\mathrm{C}: [0,1]^2 \longrightarrow [0,1]$ such that (Nelsen, 2006)

$$\mathrm{H}(x, y) = \mathrm{C}\big(F(x), G(y)\big) \tag{1}$$

Parametric copula families can be classified into elliptical and Archimedean ones[1]. Gaussian and t-copula belong to the elliptical group, while Frank and Gumbel copulas come from the Archimedean family. Each copula allows modelling various dependence structures (e.g. symmetric dependence in tails, strong dependence in upper tail etc.).

There exist several approaches to copula inference that can be grouped into parametric and semi-parametric ones. Classical fully-parametric method is the maximum likelihood estimator (MLE). The MLE is the preferred first option due to its optimality properties (Kojadinovic and Yan, 2010). However, previous statement is true if the marginal distributions are specified correctly. As argued by Kim et al. (2007), fully parametric estimators (including the MLE) of copula parameter might be biased due to the misspecification of margins.

An alternative approach belongs to semi-parametric group and is known as pseudo-maximum-likelihood (PML) estimator discussed by Genest et al. (1995). Following the PML approach, marginal distributions are estimated non-parametrically by their empirical cumulative distribution functions. On the second step, the copula dependence parameter is estimated by maximizing the pseudo-loglikelihood function

$$logL(\theta) = \sum log\left(c_\theta\big(\hat{U}, \hat{V} \big| \theta\big)\right) \tag{2}$$

where c_θ is a copula density function, θ is a copula parameter to be estimated, $\hat{U}$ and $\hat{V}$ are rank-transformed pseudo-observations on the unit interval $[0,1]$. In case ties occur, the average rank is assigned to each element.

Let us now turn to copula-based dependence measures. Bivariate Kendall's τ and Spearman's ρ coefficients written in terms of copula are given by

[1] Elliptical copulas are derived from elliptical distributions characterized by radial symmetry. Gaussian copula and t-copula are typical examples of this group. Archimedean copulas form another parametric family of copulas that are based on the generator function (Nelsen, 2006). While elliptical copula functions model symmetric behavior in tails of distribution, Archimedean copulas, instead, allow a wider range of dependence structures.

$$\tau_K = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 \tag{3}$$

$$\rho_S = 12\int_0^1\int_0^1 uv\, dC(u,v) - 3 \tag{4}$$

Six copula families together with the corresponding methods to get Kendall's τ and Spearman's ρ coefficients are defined in Table 1. Both Kendall's and Spearman's coefficients are normalized on the interval $[-1;1]$, where these extremes correspond to countermonotonic and comonotonic random variables respectively, while zero stays for independent ones.

Table 1–*Kendall's τ, Spearman's ρ coefficients and their definitions in terms of copula*

Copula	$C(u,v,\theta)$	τ_K	ρ_S
Gaussian	$\Phi\rho\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)$	$\frac{2}{\pi}\arcsin\theta$	$\frac{6}{\pi}\arcsin\left(\frac{\theta}{2}\right)$
t-copula	$T_{\rho z}\left(T_z^{-1}(u),T_z^{-1}(v)\right)$	$\frac{2}{\pi}\arcsin\theta$	$\frac{6}{\pi}\arcsin\left(\frac{\theta}{2}\right)$
Frank[1]	$\frac{1}{\theta}ln\left(1+\frac{\left(e^{\theta u}-1\right)\left(e^{\theta v}-1\right)}{e^{\theta}-1}\right)$	$1-\frac{4}{\theta}\left(1-D_1(\theta)\right)$	$1-\frac{12}{\theta}\left(D_2(-\theta)-D_1(-\theta)\right)$

[1]*Both correlation coefficients based on Frank copula rely on Debye function given by* $\mathrm{D_k}(\theta)=\frac{\mathrm{k}}{\theta^{\mathrm{t}}}\int_0^{\theta}\frac{\mathrm{t}^{\mathrm{k}}}{\mathrm{e}^{\mathrm{t}}-1}\mathrm{dt}$, *where* $\mathrm{k}=1{,}2$ *(see Genest (1987) and Nelsen (2006) for details)*
Sources: Frees and Valdez (1998); Huard et al. (2006); Nikoloulopoulos and Karlis (2008); Joe (2015)

Possible d-dimensional generalizations of both rank correlation measures were discussed by Schmid and Schmidt (2007) and Blumentritt and Schmid (2014) among others. Multidimensional extensions of Spearman's ρ_S and Kendall's τ_K coefficients are defined as follows

$$\rho_S = \frac{d+1}{2^d-(d+1)}\cdot\left\{2^d\int_{[0,1]^d} C(\boldsymbol{u})d\boldsymbol{u} - 1\right\} \tag{5}$$

$$\tau_K = \frac{1}{2^{d-1}-1}\cdot\left\{2^d\int_{[0,1]} C(\boldsymbol{u})dC(\boldsymbol{u}) - 1\right\} \tag{6}$$

with $\boldsymbol{u}=(u_1,\dots,u_d)\in[0,1]^d$. When $d = 2$, equations (5) and (6) coincide with bivariate versions of coefficients given in (3) and (4). For further details and other multivariate extensions see Schmid and Schmidt (2007) and Genest et al. (2011). In the d-dimensional case upper bound and independence benchmark are maintained. However, both coefficients do not approach -1 as lower bound. This feature is explained by the fact that perfect negative dependence is not possible if the number

of dimensions $d > 2$. Therefore, both multivariate coefficients have the following bounds (Nelsen, 1996): *(i)* $\frac{2^d-(d+1)!}{d!\{2^d-(d+1)\}} \leq \rho_S \leq 1$; and *(ii)* $\frac{-1}{2^{d-1}-1} \leq \tau_K \leq 1$.

Multivariate extensions from equations (5) and (6) are estimated nonparametrically by the empirical copula for $d \geq 2$ (Schmid and Schmidt, 2007). The interpretation of multivariate versions of Spearman's ρ and Kendall's τ coefficients in the well-being context is the following. Both measures compare the specified society with a reference society. For Spearman's coefficient a society with independent dimensions serves as the reference point, while Kendall's measure considers a society with the same level of dependence as a reference.

3. Results of copula estimation

3.1. Data description

This analysis aims to understand the degree of dependence between the dimensions of well-being and reflect it into multidimensional poverty measure by using an appropriate weighting scheme.

In this study, we focus on the following dimensions of well-being: income, education and health. Dimensions chosen in this study are the ones considered in the Human Development Index and in the Multidimensional Poverty Index. We will investigate the correlation between each pair of dimensions.

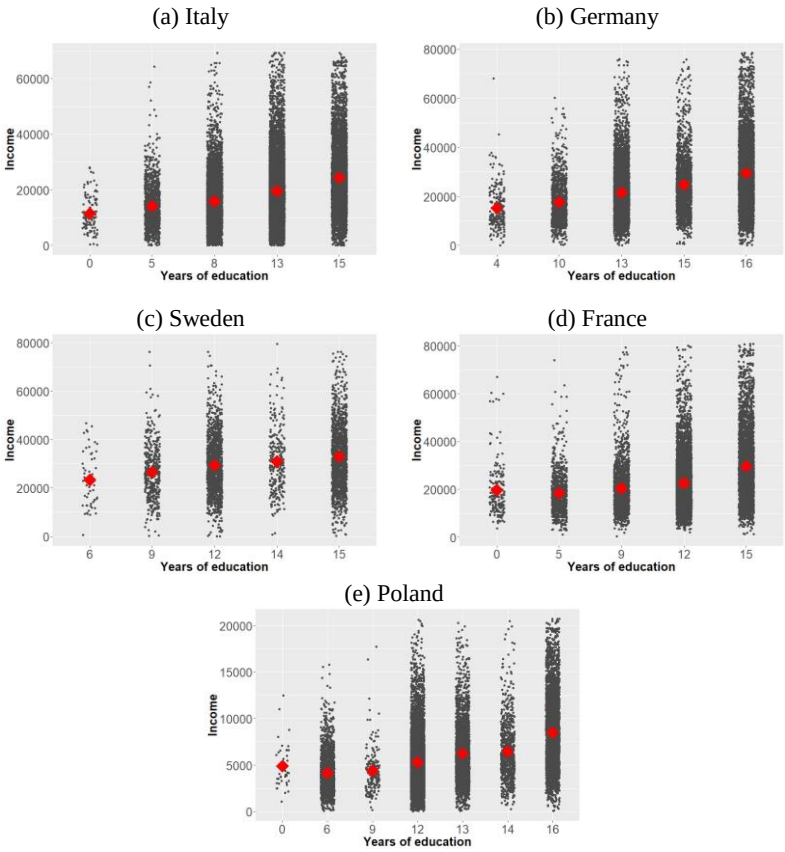
To measure the dependence between dimensions, we use the data of the EU-SILC referred to the year 2015. In our study we consider the following countries of the EU: Italy, Germany, Sweden, France and Poland abbreviated as IT, DE, SE, FR and PL respectively. Our sample includes individuals aged from 16 to 65. We consider every individual aged 16 years or more as an adult, following the approach of the EU-SILC. We exclude young adults, who are currently involved into educational programs since they have not achieved their highest educational level yet.

As indicator for income dimension we apply equivalised disposable income constructed by OECD-modified scale. Individuals with the top 1% of income are excluded from the sample. As indicator for education we choose the number of years of education. However, since the latter variable is not present in the survey, we construct it in the following way: we assign to each individual the number of years required in the specific country to complete the highest educational level attained (Meschi and Scervini, 2014). The health indicator is represented by the self-assessed general health status.

3.2. Pairwise dependence

The first pair of well-being dimensions we focus on is income-education. The highest education attainment considered in our sample is the tertiary one corresponding to 15-16 years of total duration of schooling. The concordance between income and years of education across countries is shown in Figure 1. For each country its population is divided into groups according to the total years of schooling and the equivalised disposable income is plotted for each group. As Figure 1 highlights, the mean of income increases when years of education increase. The steepest increase of income is observed for the post-secondary tertiary and non-tertiary education in all countries, except Sweden, where the association between the variables is the lowest.

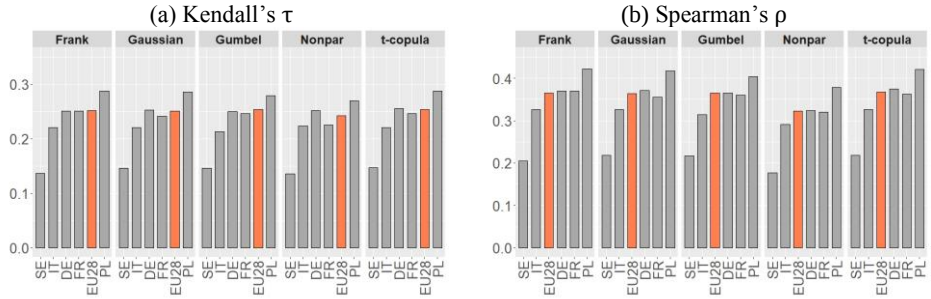
Figure 1 – *Scatterplot of income versus years of education*



Note: red diamonds represent mean of income in each educational group

In the next step of the analysis, we consider bivariate versions of Kendall's and Spearman's correlation coefficients discussed earlier. Figure 2 reports estimates of both measures grouped according to the underlying copula family. The average dependence across 28 countries of the European Union (red bar in the figure) serves as a benchmark.

Figure 2 – *Copula-based correlation coefficients for income and education dimensions*

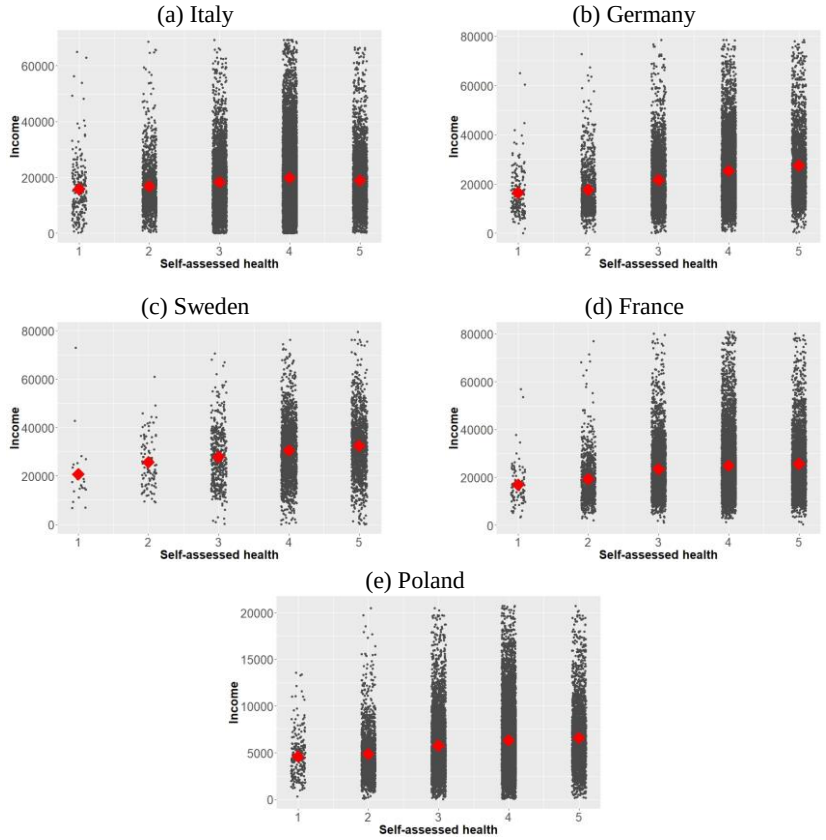


The highest level of dependence between income and education is observed in Poland, which also outranks the EU average dependence for both parametric and nonparametric estimation, while Sweden has the bottom position in the ranking. Both results are robust against the choice of the copula, suggesting that selection of the copula family does not affect the top and bottom positions in the ranking, while for intermediate positions fluctuations might occur. In particular, Germany and France exchange their positions in ranking due to the change of underlying copula.

We now consider income and health as the next pair of dimensions. The scatterplot displayed in Figure 3 illustrates the concordance between equivalised disposable income and self-assessed health. In Italy the association is minimal, while in Germany and Sweden it is more pronounced. Overall, the income mean increases less dramatically with the improvement of health status than with the increase of years of education.

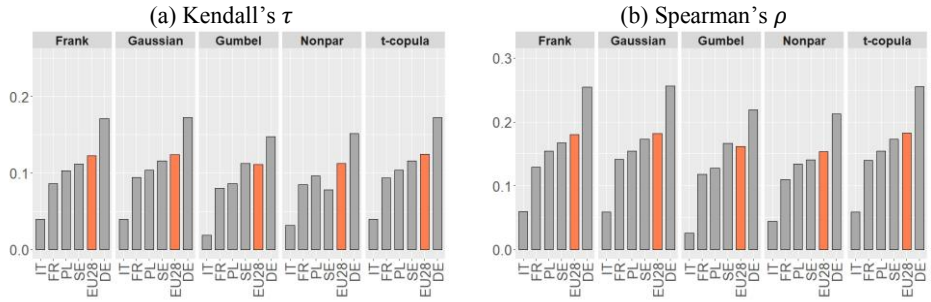
The ranking of countries according to the level of dependence between income and health status is shown in Figure 4 This pairwise dependence is generally lower than the one between income and education for all countries. In Germany, individuals who are better-off in terms of income also report better health status. In Italy, instead, the correlation between ranks in income and health dimensions is the lowest among the countries considered, indicating the weakest dependence between these two well-being domains.

Figure 3 - *Scatterplot of income versus health status*



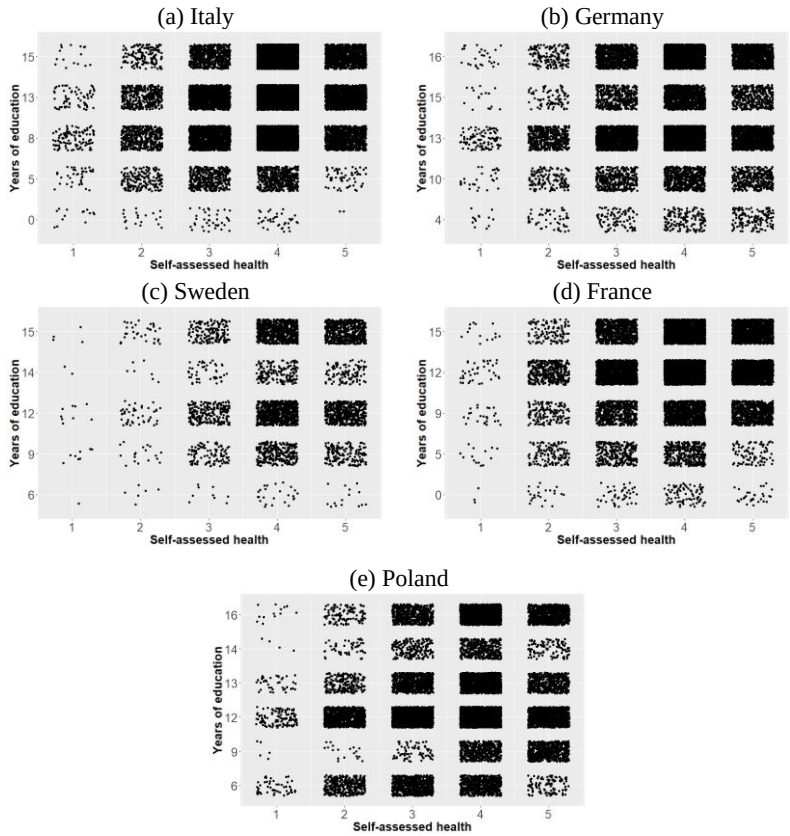
Note: red diamonds represent mean of income for each health status

Figure 4 – *Copula-based correlation coefficients for income and health dimensions*



The last bivariate analysis that we consider here is between education and health. The relationship between the two dimensions is shown in Figure 5. The subjective health status is plotted over years of schooling confirming that the variables are concordant. The majority of observations are concentrated in the right upper corner of the plot, demonstrating a positive association between the dimensions in Poland, Italy, France and Germany.

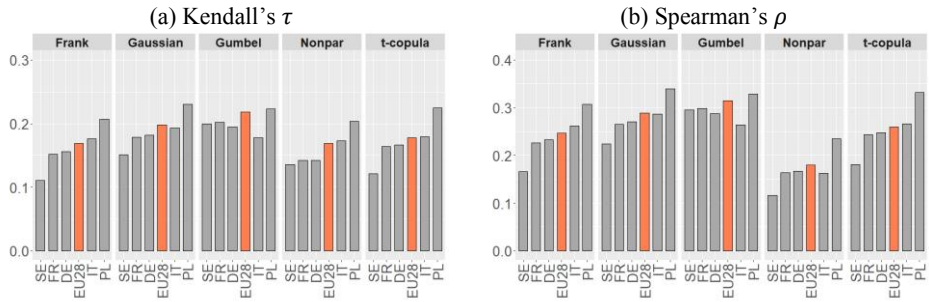
Figure 5 – *Scatterplot of education versus health status*



Estimates of copula-based dependence measures, reported in Figure 6, confirm the results discussed previously. Poland clearly outperforms rest of the countries in terms of dependence between education and health. The underlying copula does not cause a downward shift of Poland from its top position. Italy has the second position in ranking, although the latter is sensitive to the change of copula function. Similarly, France and Germany may exchange their positions in the ranking due to

the change in estimation procedure. Finally, for Sweden the estimates of correlation coefficients are the lowest in 4 out of 5 cases.

Figure 6 – *Copula-based correlation coefficients for education and health dimensions*



4. Copula-based weights in multidimensional poverty measurement

We have just shown that the performance of individuals in well-being dimensions correlates. We hence propose here to include in the multidimensional poverty measurement a weighting scheme that captures this correlation by using copula-based dependence measure. Before proceeding with it, we are introducing some notations. The size of the population is represented by n and the number of well-being dimensions is represented by d. The achievements matrix X with dimensions $n \times d$ summarizes distribution of attributes within the population. The typical element of X, x_{ij}, represents an achievement of individual i in well-being dimension j. Every row of matrix X shows achievements of individual i across all dimensions, while every column corresponds to distribution of dimension j across total population. The vector of dimension-specific thresholds is $z = (z_1, z_2, z_3, \cdots, z_d) \in Z$, where Z is a set of all possible real valued d-dimensional vectors z. Let $w = (w_1, w_2, \cdots, w_d)$ be a vector of weights and $\sum_{j=1}^{d} w_j = 1$, where $w_j > 0$ is a weight assigned to dimension j. An individual i is said to be deprived in dimension j if $x_{ij} < z_j$. Otherwise, individual is referred to as non-deprived (if $x_{ij} \geq z_j$). Finally, $P(X; z)$ is a multidimensional poverty measure. Let g_{ij}^{α} be a matrix of deprivations defined as follows:

$$g_{ij}^{\alpha} = \left[Max\left(\frac{z_j - x_{ij}}{z_j}, 0 \right) \right]^{\alpha} \tag{7}$$

with $\alpha \geq 0$ being a parameter of poverty aversion. If $\alpha = 0$, g_{ij}^{0} becomes a $0-1$ matrix of deprivations, for $\alpha = 1$ it is a matrix of poverty gaps and, finally, for $\alpha = 2$ it becomes a matrix of squared gaps. Following the counting approach introduced by Alkire and Foster (2011) with $\alpha = 0$, an individual is identified as multidimensionally poor if the total number of deprivations experienced by this individual is higher than the inter-dimensional cut-off k, with $1 \leq k \leq d$. Finally, the poverty measure with copula-based weights is computed as follows:

$$P(X,z) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} w_j g_{ij}^{\alpha}(k) \tag{8}$$

where matrix $g_{ij}^{\alpha}(k)$ summarizes deprivations of individuals using cut-offs for each dimension and identifies multidimensionally poor with inter-dimensional cut-off k. Here we propose copula-based weights for $\mathrm{d} = 3$ be computed as follows:

$$w_1 = \frac{\rho_{12}^{\theta_{12}} + \rho_{13}^{\theta_{13}}}{2\left(\rho_{12}^{\theta_{12}} + \rho_{13}^{\theta_{13}} + \rho_{23}^{\theta_{23}}\right)} \tag{9}$$

where ρ_{12}, ρ_{13} and ρ_{23} are copula-based correlation coefficients between two well-being dimensions and $\theta_{12}, \theta_{13}, \theta_{23} \geq 1$ are positive parameters that model the elasticity of substitution between respective dimensions of well-being. The higher the value of the elasticity parameter, the lower the level of substitution. When it reaches its minimum value (e.g. $\theta_{12} = 1$), the two dimensions are considered as complementary. Weights for dimensions 2 and 3 can be computed analogously. Dimensional weights constructed in this way belong to a data-driven group (Decancq and Lugo, 2013). Data-driven approach to weights does not imply a normative judgment about a trade-off between dimensions. Instead, the dimensions are weighted according to the distribution of individual performances in a society.

5. Concluding remarks

In this paper we have measured the dependence between well-being dimensions using copula-based dependence measures. This methodology has been applied to a selection of European countries using EU-SILC data for the year 2015. Results suggest that estimated correlation coefficients differ from the independence benchmark for the most pairwise comparisons and copula families employed. The highest dependence is observed between educational and income dimensions. We have then proposed to use copula-based weights in multidimensional poverty measures to make them sensitive to the distribution of individual ranks across

dimensions. This approach allows discriminating societies according to the level of dimensional correlation. Information on the correlation of deprivations enables more efficient allocation of resources by poverty-reducing policy.

References

ALKIRE S., FOSTER J. 2011. Counting and Multidimensional Poverty Measurement, *Journal of Public Economics*, Vol. 95, No. 7-8, pp. 476-487.

BLUMENTRITT T., SCHMID F. 2014. Nonparametric Estimation of Copula-Based Measures of Multivariate Association from Contingency Tables, *Journal of Statistical Computation and Simulation*, Vol. 84, No. 4, pp. 781-797.

BOURGUIGNON F., CHAKRAVARTY S. R. 2003. The Measurement of Multidimensional Poverty, *The Journal of Economic Inequality*, Vol. 1, No. 1, pp. 25-49.

DECANCQ K. 2014. Copula-Based Measurement of Dependence Between Dimensions of Well-Being, *Oxford Economic Papers*, Vol. 66, No. 3, pp. 681-701.

DECANCQ K., LUGO M. A. 2012. Inequality of Well-Being: A Multidimensional Approach, *Economica*, Vol. 79, No. 316, pp. 721-746.

DECANCQ K., LUGO M. A. 2013. Weights in Multidimensional Indices of Well-Being: An Overview, *Econometric Reviews*, Vol. 32, No. 1, pp. 7-34.

FREES E. W., VALDEZ E. A. 1998. Understanding Relationships Using Copulas, *North American Actuarial Journal*, Vol. 2, No. 1, pp. 1-25.

GENEST C. 1987. Frank's Family of Bivariate Distributions, *Biometrika*, Vol. 74, No. 3, pp. 549-555.

GENEST C., GHOUDI K., RIVEST L.-P. 1995. A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, *Biometrika*, Vol. 82, No. 3, pp. 543-552.

GENEST C., NEŠLEHOVÁ J., GHORBAL N. B. 2011. Estimators Based on Kendall's Tau in Multivariate Copula Models, *Australian & New Zealand Journal of Statistics*, Vol. 53, No. 2, pp. 157-177.

HUARD D., ÉVIN G., FAVRE A.-C. 2006. Bayesian Copula Selection, *Computational Statistics & Data Analysis*, Vol. 51, No. 2, pp. 809-822.

JOE H. 2015. *Dependence Modeling with Copulas*. New York: CRC Press.

KIM G., SILVAPULLE M. J., SILVAPULLE P. 2007. Comparison of Semiparametric and Parametric Methods for Estimating Copulas, *Computational Statistics & Data Analysis*, Vol. 51, No. 6, pp. 2836-2850.

KOJADINOVIC I., YAN J. 2010. Comparison of Three Semiparametric Methods for Estimating Dependence Parameters in Copula Models, *Insurance: Mathematics and Economics*, Vol. 47, No. 1, pp. 52-63.

MESCHI E., SCERVINI F. 2014. A New Dataset on Educational Inequality, *Empirical Economics*, Vol. 47, No. 2, pp. 695-716.

NELSEN R. B. 1996. Nonparametric Measures of Multivariate Association. In RÜSCHENDORF L., SCHWEIZER B., TAYLOR M. D. (Eds) *Distribution with Fixed Marginals and Related Topics*, Vol. 28, Hayward: Institute of Mathematical Statistics, pp. 223-232.

NELSEN R. B. 2006. *An Introduction to Copulas*. New York: Springer-Verlag.

NIKOLOULOPOULOS A. K., KARLIS D. 2008. Copula Model Evaluation Based on Parametric Bootstrap, *Computational Statistics and Data Analysis*, Vol. 52, pp. 3342-3353.

SCHMID F., SCHMIDT R. 2007. Multivariate Extensions of Spearman's Rho and Related Statistics, *Statistics & Probability Letters*, Vol. 77, No. 4, pp. 407-416.

SUMMARY

Multidimensional poverty measurement: dependence between well-being dimensions using copula function

In this paper one of the challenges of multidimensional poverty measurement is addressed, i.e. the dependence between well-being dimensions. To measure this dependence, copula-based correlation coefficients are employed. This methodology is applied to EU-SILC data for the year 2015. The results of estimation suggest that individual performances in different dimensions correlate, but its magnitude varies across countries. The highest dependence is observed between educational and income dimensions.

The results of the pairwise estimation of correlation coefficients are then used for developing a weighting scheme in the multidimensional poverty index. This approach allows, in particular, to incorporate this inter-dimensional dependence into a multidimensional poverty measure.

Kateryna TKACH, Department of Economics, Università degli Studi dell'Insubria, ktkach@uninsubria.it

Chiara GIGLIARANO, Department of Economics, Università degli Studi dell'Insubria, chiara.gigliarano@uninsubria.it