UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

DEPARTMENT OF ECONOMICS

PHD PROGRAM IN

METHODS AND MODELS FOR ECONOMIC DECISIONS

# ESSAYS ON MULTIDIMENSIONAL POVERTY MEASUREMENT AND THE DEPENDENCE AMONG WELL-BEING DIMENSIONS

Kateryna Tkach

A dissertation

Submitted to Department of Economics

For the Degree of Doctor of Philosophy at the

University of Insubria

Supervisor: Prof. Chiara Gigliarano

ACADEMIC YEAR 2015/2016

# Abstract

Evaluating the welfare of nations is high on the research agenda of the economists, practitioners and policy-makers. The literature contributions of the last decades triggered a multivariate perception of the well-being, which is suggested to go beyond the GDP, and created a need for more complex approaches to evaluate the welfare as well as poverty.

The first essay investigates the approaches to multivariate poverty measurement and focuses on the composite index approach and the steps involved in it. An important aspect of the multivariate perspective in well-being is the dependence among the underlying indicators. There is a growing evidence in the literature that well-being dimensions are interrelated. This dependence among attributes matters for multidimensional poverty measurement, since income is no longer the only indicator to be considered. However, the reviewed approaches to multivariate poverty measurement do not commonly capture this interdependence. The second essay suggests a copula function as a flexible tool to estimate the dependence among welfare variables. Moreover, it proposes to incorporate the evaluated dependence in the composite indicator. The trade-off among attributes, which is established via the weighting of dimensions, is identified as a possible channel to include the interdependence in the composite indicator.

The third essay of this dissertation defines bivariate and multivariate copula-based measures of dependence and applies them using the recent data from the EU-SILC. The results suggest that key dimensions of well-being, i.e. income, education and health, are positively interdependent. In addition, the strength of pairwise and multivariate dependence reinforced in the post-crises period in some European countries. Finally, the last essay proposes a new class of the copula-based multidimensional poverty indices by innovating over the weighting approach. The weighting scheme proposed in this dissertation incorporates the estimated copula-based dependence and contains necessary normative controls to be chosen by the practitioner. The findings of the last essay suggest that the overall poverty is driven not only by the individual shortfalls, but also by the degree of interdependence among well-being indicators.

Considering the proposed copula-based weighting scheme and the proposal of the new class of copula-based poverty indices, this dissertation contributes to the multivariate poverty measurement by suggesting the channel to enclose the dependence structure in the composite indicators. The proposed copula-based methodology will advance the multidimensional poverty analysis and the poverty-reducing policy, which can be designed to address the problem of interdependence of individual achievements.

*Keywords*: multidimensional poverty measurement, composite indicators, weighting, copula function, copula-based dependence

# Acknowledgements

First of all, I would like to thank Prof. Chiara Gigliarano for her excellent supervision and the motivating research environment that I had during the PhD program. I am deeply grateful for her guidance and valuable advices that I received during the entire period of my doctoral research.

My special gratitude belongs to Prof. Lidia Ceriani and Prof. Mariateresa Ciommi for their willingness to act as referees of this doctoral dissertation. Their helpful comments, in-depth questions and suggestions helped to develop this dissertation. Moreover, I would like to thank the PhD Board and PhD students of the University of Insubria for their comments and helpful discussions during the period of the program.

Last, but not least, I want to express my deepest gratitude to my parents and grandparents, who supported me throughout this challenging doctoral research project abroad.

# Contents

# List of Tables

# List of Figures

# General Introduction

Income or consumption are commonly applied as the proxies of welfare in poverty measurement. However, in the last decades the research paradigm on well-being and poverty had been shifted from the univariate context to the multivariate one. The economists and practitioners commonly suggest the multidimensional character of these notions, which are not comprehensively reflected by pecuniary indicators as earnings or consumption. A wide stream of literature emphasizes the multidimensionality of welfare and poverty (Atkinson and Bourguignon, 1982; Sen, 1999; Atkinson, 2003; Bourguignon and Chakravarty, 2003; Chakravarty, 2009; Stiglitz et al., 2009).

Accepting the advocated trajectory regarding the multidimensional nature of poverty raises a reasonable question on how this phenomenon should be measured and monitored over time. Going beyond income in measuring poverty implies that several well-being dimensions are considered by the practitioner to assess the overall level of poverty. As a result, the information from several indicators should be combined to evaluate the complex phenomenon. The literature on poverty measurement distinguishes two opposite approaches, namely the dashboard of indicators and the composite index approach. While the dashboard does not include the aggregation of information from several well-being dimensions, the composite index does the opposite. Therefore, any multidimensional poverty index aggregates the information from several well-being dimensions to form an overall level of poverty. While the debate of aggregation versus non-aggregation is not expected to be resolved easily, in this dissertation we focus on the composite index approach for measuring multidimensional poverty.

The construction of the multivariate poverty index includes several key steps, namely the identification, the weighting and the aggregation. Since the underling pillars are aggregated to form the composite indicator of poverty, a relative trade-off among the considered indicators should be established. From a technical viewpoint, the trade-off is defined by assigning a positive weight to dimensions and multiplying each indicator by it. Therefore, the contribution of the underlying indicators is governed by weights. There is a variety of methods to construct dimensional weights. For instance, the weighting scheme can be objective, if it is based on statistical techniques, or subjective, if the opinion of stakeholders defines the trade-off among dimensions (Maggino, 2017). Decancq and Lugo (2013) provides a slightly different classification of weights, which are grouped into data-driven, normative and hybrid classes. Although the selection of specific weighting approach influences the outcome of the composite index and the ranking of countries (Decancq et al., 2013b), there is no widely accepted method to select the optimal weighting.

Naturally, when constructing multidimensional poverty index the practitioner chooses multiple dimensions to represent the complex phenomenon. In line with common acceptance of the multivariate nature of welfare, the majority of empirical applications assume that poverty is represented by the shortfalls in three "core" dimensions, namely income, education and health. From the other side, the literature suggests that the achievements in key dimensions are often interrelated. In particular, the shortfalls in different attributes are frequently experienced by the same individuals. We believe that this interdependence among dimensions should be considered, while developing a weighting scheme. There is a current lack of research on how the interdimensional dependence should be handled in the context of the composite indicator approach. Some scholar may argue that the least dependent dimensions should be assigned the highest weights due to additional information they provide, while the others may support an opposite view. Therefore, the focal point of this dissertation is the selection of weighting scheme in the context of the interdependent well-being dimensions.

The aim of this dissertation is to measure the dependence among key well-being dimensions using copula function and to propose a new multidimensional poverty index that incorporates the estimated dependence among the pillars. Given the specified aim the objectives of the dissertations are the following. Firstly, our purpose is to study the dependence among the major well-being dimensions, i.e. income, education and health, in the European countries using parametric and nonparametric copula families. Our second objective is to monitor the evolution of the interrelation of individual performances in the selected dimensions in the pre-crises and post-crises periods. The last purpose is to propose a new multidimensional poverty index with the copula-based weights.

The dissertation consists of three methodologically linked Chapters. The first one does a systematic review of literature on the approaches to poverty measurement and the concepts of dependence, introducing copula-based dependence measures as well. The second Chapter studies the interdependence among key well-being dimensions and its evolution in the European countries by applying copula-based measures of dependence. The third Chapter proposes a new class of copula-based multidimensional poverty indices and applies the proposed index to the selected European countries. The dissertation is completed by general conclusion that summarizes main findings of the thesis.

# Chapter 1

# Literature review on multidimensional poverty measurement

## 1.1 Introduction

A large stream of literature has suggested a multidimensional nature of well-being and a complexity of the phenomenon. Researchers commonly advocate that the welfare is mirrored in multiple attributes, which besides income are related to human capital, longevity, political power, safety and environment etc. A paradigm shift in the well-being literature has raised a relevant question of how welfare and its insufficient level, namely poverty, should be measured. Works by Sen (1976); Atkinson and Bourguignon (1982); Atkinson (2003); Bourguignon and Chakravarty (2003); Stiglitz et al. (2009); Alkire and Foster (2011) are key landmark studies in the field that have reinforced a multidimensional perception of well-being and poverty.

The first question that appears in poverty analysis is "Poverty of what?" or the choice of relevant dimension(s) of well-being to represent the phenomenon (Decancq et al., 2013a). It is the initial step in poverty research, regardless of chosen concept, i.e. univariate or multivariate. In the former framework income and consumptions are the most commonly used proxies of welfare (Duclos et al., 2006). By contrast, in multidimensional analysis monetary proxies are complemented by non-monetary ones. Therefore, the choice of attributes becomes a crucial task due to a variety of possible indicators.

The capability approach by Sen (1999) is an attempt to go beyond income in assessing well-being from theoretical viewpoint. This approach is based on the notions of *functioning*, which represents an achievement of a person, and *capabilities* - a set of possible functionings an individual has the freedom to choose from. In turn, Nussbaum (2011) presents the list of ten central capabilities to be secured by the government to its citizens. However, the capability approach does not provide a "universally-relevant" set of dimensions that should be considered in empirical studies (Alkire, 2007). Therefore, the choice of relevant attributes can be governed by the experts' opinion, the data availability and the statistical techniques (Aaberge and Brandolini, 2015). For instance, the Multidimensional Poverty Index (MPI)[1], which was designed through a cooperation between the United Nations Development Program and the Oxford Poverty and Human Development Initiative, summarizes deprivations in the following well-being dimensions: living standard, education and health (UNDP, 2016). Similarly, the Human Development Index (HDI)[2], which aims at synthesizing information on the achievements of nation in three dimensions, focuses on the living standard, knowledge and health. A broader set of dimensions is suggested in the report of Stiglitz et al. (2009), who propose to summarize well-being in the following domains: material living standards, health, education, personal activities including work, political voice and governance, social connections and relationships, environment and insecurity.

The selection of pillars and the corresponding indicators to represent the underlying phenomenon is one of several crucial tasks in multidimensional poverty measurement. Other key steps in poverty analysis include the identification and the aggregation stages, which were originally suggested by Sen (1976). Both steps are applicable for univariate and multivariate poverty measures. The idea of identification is to separate poor individuals from non-poor ones by establishing a certain threshold. In case of univariate poverty the identification is done by drawing a poverty line and verifying if individual achievement lies below or above it (Sen, 1976; Bourguignon and Chakravarty, 2003). The multidimensional poverty framework requires more advanced strategy for identifying poor individuals. In particular, a dimension-specific cut-off is specified to identify a deprivation in each attribute (Chakravarty, 2009). Additionally, the multidimensional poverty criterion is mandatory for detecting multidimensionally poor citizens (Alkire and Foster, 2011). The mentioned criterion establishes minimum

[1]For the details on the MPI methodology, including indicators that represent each dimension, corresponding weights and the aggregation procedure see Alkire and Jahan (2018).

[2]See UNDP (2016) for more details on the steps associated with the index construction.

number of deprivations, an individual should experience to be identified as multidimensionally poor.

The identification phase is typically followed by weighting and aggregation. Regarding the latter, there are two approaches in literature: a dashboard of indicators and a composite index methods. The dashboard approach does not involve an aggregation of information from several well-being dimensions. By contrast, this method proposes to monitor a list of indicators over time keeping dimensions separately (Ravallion, 2011). For instance, the National Statistics Office of the UK monitors nation's well-being through a dashboard of indicators, which includes measures for living standard, education, health, job satisfaction and environment among others (Randall et al., 2019). In Italy the National Institute of Statistics measures equitable and sustainable well-being (BES) using 12 dimensions, which comprise, inter alia, health, eduction and training, economic and subjective well-being (ISTAT, 2018).

As opposite to the dashboard approach, the composite indicator synthesises information from several dimensions into a single number by following certain normalisation, weighting and aggregation rules (Nardo et al., 2008). Composite indicators aim at illustrating a complex phenomenon by incorporating different aspects of it represented by the underlying dimensions. The HDI and the MPI mentioned earlier are widespread examples of the synthetic indicator method. While each aggregation approach has its pros and cons, there is still a debate whether the composite indicators are the appropriate statistics to monitor welfare and poverty or the dashboard of indicators, considering the best available data, should be preferred (Ravallion, 2011). An important limiting factor of the dashboard method is that it omits information on the joint distribution of well-being indicators (Ferreira, 2011; Decancq et al., 2013a). In other words, if the same individuals have insufficient achievements in multiple dimensions cannot be ascertained using the dashboard of indices. In general, the composite index approach summarizes the information on achievements in several well-being attributes without incorporating the dependence structure that exists among dimensions.

Considering several approaches for measuring multivariate poverty proposed in the literature, the aim of this Chapter is to provide a systematic literature review on multidimensional poverty measurement with a focus on the composite indicator approach. This Chapter implies the following objectives. The first purpose is to discuss the identification and the aggregation steps that are involved in the construction of multidimensional poverty indices. The second objective is to review the weighting schemes that are applied in the multidimensional poverty indices. The last purpose of this Chapter is to discuss the properties relevant for multidimensional poverty measures.

The rest of the Chapter is structured as follows. In Section 2 we review the fundamental steps involved in the construction of multidimensional poverty index. In addition, the properties applicable to multidimensional poverty measures are discussed. In turn, Section 4 contains concluding remarks.

## 1.2 Approaches to multidimensional poverty measurement

The dashboard of indicators and the composite indicator are two approaches to multidimensional poverty measurement (Figure 1.1), each of them representing an alternative view on the aggregation procedure. In particular, the dashboard approach collects and monitors socio-economic indicators without aggregating the information from several well-being dimensions. This approach has the following advantages (Ravallion, 2011; Stiglitz et al., 2009):

- extreme variety of indicators that can be covered: among others, the living standard, life satisfaction, employment and unemployment rates, life expectancy, real GDP per capita are the examples of indicators that can be considered to represent multidimensional well-being and poverty (Kurkowiak et al., 2015);
- transparency for policy-makers and other stakeholders: since the dashboard approach provides non-aggregated indicators, all dimensions can be analysed separately by the practitioners and policy-makers; the latter can establish policy priorities and assess the results of poverty-reducing measures dimension-by-dimension (Ravallion, 2011; Decancq et al., 2013a).

However, the absence of an insight into the dependence among dimensions (Stiglitz et al., 2009) and an omission of information regarding the joint distribution of well-being attributes are the limitations of the approach. Moreover, the dashboard method does not allow a straightforward comparison across countries in terms of well-being and poverty due to a variety of considered indicators. This complex snapshot of welfare complicates the interpretation and the ranking of countries.

The composite index approach makes synthesis of well-being information into a single value. Hence, it allows unambiguous comparisons across space and time. The composite indicator approach is associated with some uncertainties related to its construction. In particular, the normalization of original variables, the choice of weighting scheme and the choice of aggregation procedure (Saisana et al., 2005; Nardo et al., 2008; Decancq et al., 2013a). The uncertainty related to dimensional weights means that different sets of weights influence the outcome of the composite index, however, there is no unanimity among researchers about the optimal weighting approach. In general, an equal weighting scheme is a common choice due to the interpretation simplicity.

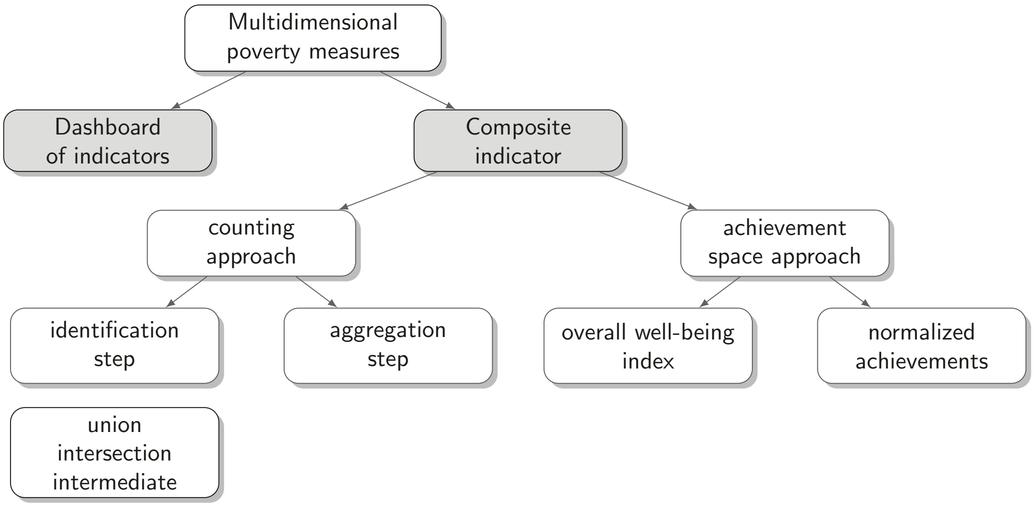



Figure 1.1: A classification of approaches to multidimensional poverty measurement

In the composite index approach there exist two options: either focusing on individual achievements or on shortfalls. Within the achievement space method one methodology is related to aggregation of achievements of each individual to form an overall cardinal index of well-being (Bourguignon and Chakravarty, 2003). When the overall well-being is obtained, the general poverty threshold is established, so that an individual is identified as poor if his overall well-being falls below the specified cut-off (Duclos et al., 2006; Alkire and Foster, 2011). Maasoumi and Lugo (2008) proposed the multidimensional poverty index that is based on the overall achievements and the aggregated poverty line. In the overall well-being index the aggregation is first made dimension-wise for each individual, which is followed by the identification of multidimensionally poor (Duclos et al., 2006; Maasoumi and Lugo, 2008; Alkire and Foster, 2011). Therefore, in this approach the aggregation is done before the identification of poor; moreover, the deprivations in each dimensions are not detected. In other words, the two steps of poverty measurement are reversed compared to the classical proposal of Sen (1976), who suggested the identification to be done prior to the aggregation.

The second method in the achievement space approach is referred to as normalized achievements. In this method each individual achievement is compared with the dimension-specific cut-off and only the deprived achievements are aggregated. The identification rules relevant for this approach are different from those applicable counting methodology, which is discussed in details in the next subsection.

### 1.2.1 Identification step and counting approach

Before proceeding with the discussion we introduce some necessary notations. The size of the population is represented by $n$ and the number of well-being attributes is represented by $d$. The achievements matrix $X$ with dimensions $n \times d$ summarizes the realization of achievements in the society. The typical element of $X$, $x_{ij} \in R$, describes the performance of individual $i$ in well-being dimension $j$. Every row of matrix $X$ shows the achievements of individual $i$ in $d$ dimensions of well-being, while every column corresponds to the distribution of achievements in dimension $j$ by all representatives of the considered society. A multidimensional case reduces to unidimensional one when the number of well-being attributes is equal to one:

$$x_{ij} = x_{i1} = x_i \quad wlog \tag{1.1}$$

The vector of dimension-specific thresholds is contained in $z = (z_1, z_2, z_3, \cdots, z_d) \in Z$, where $Z \in R^d$ 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 with $\sum_{j=1}^{d} w_j = 1$, with $w_j > 0$ being the 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 rich or non-deprived in dimension $j$ (if $x_{ij} \geq z_j$). Finally, $P(X; z)$ is a multidimensional poverty index.

Having introduced the basic notations it is necessary to clarify *poverty* and *deprivation* terms. In unidimensional case these terms coincide: if individual income $x_i$ falls below the poverty line $z$, then he is deprived with respect to income and is identified as poor. Therefore, deprivation and poverty are synonymous in the unidimensional case. In multidimensional framework it is no longer true. Individual $i$ can be deprived with respect to one or several dimensions, but not identified as multidimensionally poor. The identification step in multidimensional case is based on several criteria that are explained later in this section.

**Definition 1.2.1.** Let $\varphi_{ij}$ be a deprivation identification function based on dimension-specific cut-off such that

$$\varphi_{ij} = \mathbb{1}_{\{x_{ij} < z_j\}} \tag{1.2}$$

which identifies individual $i$ as deprived in dimension $j$ if his achievements in that dimension falls below the corresponding threshold.

Applying the identification function to the original data matrix returns a $0-1$ deprivations matrix with dimensions $n \times d$. If all well-being dimensions are represented by continuous variables, then a relative shortfall from dimension-specific cut-off $z_j$ can be computed as well.

**Definition 1.2.2.** A normalized poverty gap of individual $i$ in dimension $j$ is given by (Alkire and Foster, 2011):

$$\begin{aligned} \phi_{ij} &= \varphi_{ij} \cdot \frac{z_j - x_{ij}}{z_j} \\ &= \mathbb{1}_{\{x_{ij} < z_j\}} \cdot \frac{z_j - x_{ij}}{z_j} \end{aligned} \tag{1.3}$$

Alkire and Foster (2011) and de la Vega (2010) suggest that dimension-specific cut-offs contained in vector $z$ are necessary but not sufficient for identifying multidimensionally poor individuals. Additional criterion should be introduced, namely the number of dimensions in which person is deprived. As a consequence, this identification method is also referred to as counting approach.

**Definition 1.2.3.** Let $c_i$ be a deprivation-counting function given by:

$$c_i = \sum_{j=1}^{d} \varphi_{ij} \tag{1.4}$$

Therefore, $c_i$ gives the number of dimensions, in which an individual $i$ experiences a deprivation.

If a person is not deprived in any of the dimensions, then the function is equal to zero. The maximum value of $c_i$ corresponds to the number of dimensions, meaning that an individual is deprived in all of them. In the counting approach deprivation-counting function together with identification criteria are necessary for making distinction between multidimensionally poor and non-poor[3].

[3]For the robustness on the choice of identification criterion see de la Vega (2010), who derives dominance conditions for a set of identification cut-offs.

In a benchmark case of the counting approach, all weights attached to the *deprivations* across dimensions are assumed to be equal: $w_j = 1$ for $j = 1, \dots, d$ (Alkire et al., 2015). However, in order to model different importance of deprivations across $d$ number of dimensions, various weights can be applied. In this case, in the equation (1.4) each deprivation will be multiplied by its relative weight. We need to proceed with caution since weights applied in the identification step have to be distinguished from ones applied in the aggregation step. If the former affect the deprivation-counting function and determine the relative importance of *deprivation* in a certain dimension, the latter, instead, are a part of aggregation procedure and determine the relative trade-off between *dimensions* of well-being.

In counting approach there are three criteria to identify multidimensionally poor: *union*, *intersection* and *intermediate* (Atkinson, 2003; Duclos et al., 2006; Chakravarty, 2009; Alkire and Foster, 2011; Aaberge and Brandolini, 2015). These notions were first introduced in the literature on multidimensional poverty by Atkinson (2003). Union criterion identifies an individual as multidimensionally poor if his achievement falls below an established threshold in *at least* one well-being dimension (i.e. $c_i \geq 1$). According to union approach the absence of deprivation in every dimension is essential for being non-poor in the multivariate context. The choice of union identification criterion is driven by the purposes of an empirical application, for instance when all the considered dimensions should be emphasized as equally desirable for the society, this method is an appropriate choice. In general, this criterion can identify the major part of population as multidimensionally poor, especially when the number of attributes is high (Alkire and Foster, 2011).

At the other extreme, intersection approach identifies an individual as multidimensionally poor if he is deprived in *all* dimensions (i.e. $c_i = d$). This method is helpful to identify the most deprived part of the population, but it overlooks the information on deprived in several (but not all) dimensions. This identification approach narrows down the number of multidimensionally poor when the list of considered attributes is long. A schematic example of two identification approaches is illustrated in Figure 1.3 in Appendix.

An alternative method is intermediate approach to identification, which identifies an individual as multidimensionally poor if the number of his deprivations lies in between its minimum value (i.e. equal to 1) and maximum value (i.e. equal to the number of attributes) (Permanyer, 2014). Intermediate criterion is specified as follows (Alkire and Foster, 2011): the researcher chooses the across-dimension cut-off $k$, with $1 \leq k \leq$

$d$. Individual $i$ is identified as multidimensionally poor if he is deprived in at least $k$ dimensions (i.e. $c_i \geq k$). Intermediate approach includes both *union* (i.e. when $k = 1$) and *intersection* (i.e. when $k = d$) criteria as special cases.

**Definition 1.2.4.** Let $\zeta_i$ be the function that identifies multidimensionally poor using the intermediate criterion:

$$\zeta_i = \mathbb{1}_{\{c_i \geq k\}} \tag{1.5}$$

If the value of deprivation-counting function $c_i$ is higher or equal to the established intermediate criterion $k$, then individual $i$ is said to be multidimensionally poor; otherwise he is non-poor.

The choice of identification criterion is a normative one, which depends on the specific context of multidimensional poverty measurement. Among other things, the selection of identification approach depends on the shortlisted well-being attributes, the number of indicators, the weights attached to the deprivations as well as on the purposes of an empirical application (Tsui, 2002; Alkire and Foster, 2011). Moreover, Alkire and Foster (2011) suggest that the across-dimension cut-off $k$ is related to the policy targets and its context. For instance, if the policy goal is to address the most deprived part of the population and lift them out from poverty, then the identification should be done to focus on this group. On the contrary, if the target group of poverty-reducing policy is wider, then the identification criterion should be specified at several attributes rather then all of them.

As it was illustrated previously in the classification of the approaches to multidimensional poverty measurement (Figure 1.1), counting method is not the only available methodology. Another possibility is to focus on achievements rather than shortfalls (see the work by de la Vega and Aristondo (2012) for the discussion the focus on achievements and shortfalls in the context of inequality measurement). Some scholars (Tsui, 2002; Chakravarty, 2009) transform the entries of the original matrix $X$ into the relevant achievements, which are then employed in the aggregation step. The identification function applied within the normalized achievements approach differs from one relevant for counting approach.

**Definition 1.2.5.** Let $\tilde{x}_{ij}$ be a normalized achievement of individual $i$ in dimension $j$ such that

$$\tilde{x}_{ij} = \begin{cases} x_{ij} & \text{if } x_{ij} < z_j \\ z_j & \text{if } x_{ij} \geq z_j \end{cases} \tag{1.6}$$

where an individual achievement remains unchanged for all $x_{ij}$ that are below the corresponding cut-off; otherwise $x_{ij}$ is substituted by the value of the poverty line $z_j$.

Since counting and normalized achievements approaches employ different deprivation identification functions, the aggregation methods in two approaches differ as well. Section on the aggregation step contains a non-exhaustive list of some commonly applied multidimensional poverty indices.

### 1.2.2 Weighting

The next step in the multidimensional poverty measurement, which follows the identification of poor, is commonly related to choosing the dimensional weights. In particular, each indicator and subindicator are assigned a positive number and are multiplied by it. A sum of weighted individual deprivations in well-being indicators is known as a deprivation score.

In the composite indicators weights are used to govern the input of each well-being indicator into the overall value of the index. We highlight that dimensional weights do not measure an importance of the underlying variables in the sense of contributing to the overall index (see Paruolo et al. (2013) and Schlossarek et al. (2019) for the discussion on nominal weights in the composite indicators and the importance of underling variables). There are several approaches to weighting of dimensions in the composite indicators. Decancq and Lugo (2013) provide the following classification of dimensional weights: data-driven, normative and hybrid. Data-driven weights are commonly based on the statistical tools, such as Principal Component Analysis (PCA) or Data Envelopment Analysis (DEA); these weighting schemes are empirically defined by the distribution of achievements in the analysed society. As a consequence, these weighting approaches are suggested to be objective (Maggino, 2017). In brief, PCA reduces the dimensionality of the original data by computing a linear combinations (principal components) of variables, which are able to explain the most of the observed variance (Greco et al., 2019). In turn, DEA is a nonparametric measure of efficiency. This approach establishes dimensional weights that maximize country's performance considering the selected benchmark (Nardo et al., 2008). To summarize, the choice of optimal weights using purely statistical tools should be done with care. Without normative controls the weighting procedure can be potentially misleading and the established weights can diverge from the public opinion on the optimal trade-off among dimensions (Maggino, 2017).

The other group of weights contain normative weighting schemes, which are based on the explicit value judgement regarding the trade-off among dimensions. As a special case, this group includes equal weighting scheme. Another example of normative weighting is budget allocation process. The central idea of this approach is the following: the experts allocate a certain number of points among dimensions, while the weights are computed as the average of the experts' opinion (Greco et al., 2019; Maricic et al., 2019). This method requires that the selected stakeholders have relevant expertise and possess diverse backgrounds. Moreover, the number of dimensions should not exceed ten for obtaining optimal weighting (Saisana and Tarantola, 2002). Another normative weighting method is the Analytical Hierarchy Process, which is applied for multi-attribute decision-making and was originally proposed by Saaty (1987). The trade-off between dimensions is established by the experts, who do a pairwise comparison of indicators. In particular, the experts assess the strength of importance in each bivariate comparison using a semantic scale from 1 to 9, where 1 means that two indicators are equally important, while 9 indicates that on e of them is extremely more important than the other one (Nardo et al., 2008). This method is a suitable choice when the nu,ber of dimensions does not exceed ten. Since the assessment of the trade-off is subjective by nature, an estimation of the consistency of experts' judgement is required. It can be computed using the consistency ratio, which should be below 0.1 (Nardo et al., 2008).

The last group, namely hybrid weights, combines the features of previous two classes. In other words, hybrid weighting scheme combines an empirical evaluation of data with the normative controls over the weights (Maricic et al., 2019). For instance, stated preferences weights belong to this group, since the opinion about the trade-off between dimensions is expressed by the respondents from the analysed society (Decancq and Lugo, 2013).

In the empirical applications of the composite indicators, including the multidimensional poverty measures, the choice of dimensional weights is a necessary step. The researchers suggest that the outcome of the multidimensional poverty index depends on the normative choices including weights. Since there is no uniformly correct weighting scheme, Decancq and Lugo (2013) suggest that a set of weights should be applied and the robustness checks are required. Among others Saisana et al. (2005); Cherchye et al. (2008); Permanyer (2011, 2012); Foster et al. (2013); Athanassoglou (2015) contributed to the literature on the composite indicators' weights and the corresponding robustness analysis. For instance, applying the sensitivity analysis and Monte Carlo simulation, Saisana et al. (2005) concludes that the weighting scheme affect the eventual countries' ranking. Furthermore, Permanyer (2012) suggests that the larger is a set of applied weights, the greater is the difference in the produced ranking. Therefore, the selection of weights is associated with uncertainty to some degree.

The other aspect of weighting is the issue of dependence between the dimensions and the composite indicator, in other word, how each underlying indicators contribute to the overall index. McGillivray (1991) addresses the issue of dependence and weighting from the following perspective: he investigates the level of correlation between the overall value of the HDI and the underlying components. The results suggest high and statistically significant correlation between the indicators and outcome of the composite index. Therefore, high correlation in this cases may translate into redundancy problem: the composite indicator may not reflect properly a multivariate phenomenon it aims at describing, since it is mostly driven by one (or several) underlying indicator.

### 1.2.3 Properties for multidimensional poverty indices

Before discussing the aggregation step and some widespread multidimensional poverty indices, we review some fundamental properties relevant for poverty measurement. Following Foster (2006, p. 44), "...a key step towards justifying a particular measure of poverty is identifying the properties it satisfies". Most multidimensional properties overlap with uni-dimensional axioms. However, there are axioms that are applicable to multivariate context only. Table 1 summarizes the most important properties[4] of univariate and multidimensional poverty measures.

The first group of axioms, namely the invariance one, ensures that a poverty measure considers only relevant aspects of achievements' distribution, while it is insensitive to other ones (Foster, 2006; Chakravarty, 2009).

**Definition 1.2.6.** (Symmetry axiom) If $X = AY$, where $X$ and $Y$ are two matrices of individual achievements and $A$ is permutation matrix[5], then $P(X; z) = P(Y; z)$.

According to symmetry axiom, any switch of rows in the original matrix of achievements does not affect the poverty measure, if the vector of dimension-specific cut-offs is fixed. In other words, poverty measure depends on the individual achievements, while other characteristics of individuals as gender, age, race etc. does not influence the overall level of poverty. This property is also called anonymity axiom (Chakravarty, 2009; Alkire et al., 2015).

[4]Literature contribution to axiomatic characterization of poverty measures was made, among others, by Bourguignon and Chakravarty (1999), Chakravarty and Silber (2008), Chakravarty (2009), de la Vega (2010), Alkire and Foster (2011), Bossert et al. (2013)

[5]Permutation matrix is a square matrix that contains single "1" in every row and column and zeros as the rest of entries. Source: Bronshtein et al. (2007)

Table 1.1: Groups of axioms relevant for unidimensional and multidimensional poverty measures

| Group of properties | Univariate measures | Multivariate measures |
|---|---|---|
| Invariance | Symmetry | Symmetry |
| | Replication invariance | Replication invariance |
| | Scale invariance | Scale invariance |
| | Focus | Poverty focus |
| | | Deprivation focus |
| Dominance | Monotonicity | Monotonicity |
| | | Dimensional Monotonicity |
| | Transfer | Multidimensional transfer |
| Subgroup Axioms | Subgroup decomposability | Subgroup decomposability |
| | Subgroup consistency | Subgroup consistency |
| Technical properties | Continuity | Continuity |
| | Normalization | Normalization |
| | Non-triviality | Non-triviality |

Sources: Tsui (2002); Bourguignon and Chakravarty (2003); Foster (2006); Chakravarty (2009); Alkire et al. (2015).

**Definition 1.2.7.** (Replication invariance) If $Y$ is obtained from $X$ by replication of its rows, then $P(Y; z) = P(X; z)$.

According to replication invariance, the magnitude of poverty does not change if rows of the original data matrix are replicated a fixed number of times, while deprivation cut-offs remain the same. The replication of population is useful for comparing multidimensional poverty over time given some fluctuation in the population size and across countries with different number of citizens (Bourguignon and Chakravarty, 1999).

**Definition 1.2.8.** (Scale invariance) If both $X$ and $z$ are multiplied by a diagonal matrix $\Delta$, where $\Delta = diag(\delta_1, \delta_2, \delta_3, ...\delta_n)$ and $\delta_i > 0$, then $P(X\Delta; z\Delta) = P(X; z)$.

According to scale invariance, scale transformation of the achievements matrix and the vector of cut-offs does not affect multidimensional poverty index (Chakravarty and Silber, 2008). In other words, a change of unit measurement of well-being indicator does not affect the poverty index if the corresponding dimensional cut-off is adjusted (Alkire et al., 2015).

Unlike inequality measurement which considers the whole population, poverty measurement is focused on the bottom of the achievements' distribution. Therefore, focus property is of great significance for poverty indices. Since in the multidimensional context poverty and deprivation have different meanings, Alkire et al. (2011) distinguish two types of focus property, namely *poverty focus* and *deprivation focus.*

**Definition 1.2.9.** (Poverty focus) If $Y$ is obtained from $X$ by an improvement of any achievement of non-poor individual, then $P(Y;z) = P(X;z)$.

The poverty focus axiom (also referred to as weak focus axiom) requires a poverty indicator to be independent from any improvements of non-poor citizens.

**Example 1.2.1.** Let $X = \begin{bmatrix} 3 & 5 & 7 \\ 2 & 5 & 8 \\ 6 & 8 & 9 \end{bmatrix}$ and $z = \begin{pmatrix} 4 & 6 & 9 \end{pmatrix}$. Here individuals 1 and 2 are multidimensionally poor and an individual 3 is non-poor using any of the proposed identification criteria. If $Y = \begin{bmatrix} 3 & 5 & 7 \\ 2 & 5 & 8 \\ 6 & 8 & \mathbf{10} \end{bmatrix}$, then the overall poverty remains unchanged.

□

**Definition 1.2.10.** (Deprivation focus) If $Y$ is obtained from $X$ by an improvement in a non-deprived dimension[6], then $P(Y;z) = P(X;z)$.

Under deprivation focus a poverty index does not change, if an increase in the non-deprived dimensions occurs. Deprivation focus property is also known as strong focus axiom (Bourguignon and Chakravarty, 2003).

**Example 1.2.2.** Let $X = \begin{bmatrix} 3 & 5 & 7 \\ 2 & 7 & 8 \\ 6 & 8 & 9 \end{bmatrix}$ and $z = \begin{pmatrix} 4 & 6 & 9 \end{pmatrix}$. If $Y = \begin{bmatrix} 3 & 5 & 7 \\ 2 & \mathbf{8} & 8 \\ 6 & 8 & 10 \end{bmatrix}$, then poverty does not change.

□

The second group of axioms are related to the dominance properties, which include monotonicity and transfer axioms. Monotonicity was originally introduced to unidimensional poverty measures, which requires that an increase in any achievement of a poor individual should be reflected at the aggregated level (Foster, 2006). Formulating it differently, poverty should not increase, if a poor individual experiences an improvement in the considered well-being attribute (Bourguignon and Chakravarty, 1999). In the multidimensional context, Alkire et al. (2011) distinguish between monotonicity and dimensional monotonicity properties, according to each type of improvement experienced by the poor:

[6]A dimension is non-deprived when the individual achievement is above the dimension-specific cut-off

- an improvement in attribute $j$ such that $x'_{ij} < z_j$, which does not eliminate this deprivation
- an improvement of attribute $j$ such that $x'_{ij} < z_j$, which eliminates the deprivation

**Definition 1.2.11.** (Monotonicity) If a multidimensionally poor individual experiences an improvement in his deprived dimension such that $x_{ij} < x'_{ij} < z_j$, then the overall poverty should decrease.

A multidimensional poverty measure satisfies monotonicity if an improvement in a deprived dimension $x_{ij}$ of the multidimensionally poor is reflected in corresponding decrease of poverty index.

**Example 1.2.3.** Let $X = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 3 \\ 1 & 2 & 2 \end{bmatrix}$ and $z = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix}$. If a multidimensionally poor individual $x_3$, who is identified as poor according to any identification criterion, experiences an improvement in the third dimension such that the new achievements matrix is the following $Y = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 3 \\ 1 & 2 & \mathbf{3} \end{bmatrix}$, then $P(Y;z) < P(X;z)$.

□

**Definition 1.2.12.** (Dimensional monotonicity) If $Y$ is obtained from $X$ by the second-type improvement, which removes the deprivation of a poor individual, then $P(Y;z) < P(X;z)$.

Dimensional monotonicity is satisfied, if an improvement of a poor individual removes the considered deprivation implies a corresponding decrease of multidimensional poverty measure.

**Example 1.2.4.** Let the original achievements matrix be defined as $X = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 3 \\ 1 & 2 & 2 \end{bmatrix}$ and the vector of dimension-specific cut-offs be given by $z = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix}$. If multidimensionally poor individual $x_3$, who is identified poor according to any of the mentioned criteria, experiences an improvement in the third well-being dimension such that the new achievements matrix is $Y = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 3 \\ 1 & 2 & \mathbf{5} \end{bmatrix}$, then $P(Y;z) < P(X;z)$. □

The second subgroup of properties included in the dominance class of axioms is transfer property, which is associated with inequality-sensitive poverty measurement. Let us first provide the definition of transfer between two individuals.

**Definition 1.2.13.** (Progressive or Pigou–Dalton transfer) Let individual individual achievements be given by $x_1$ and $x_2$ such that $x_1 < x_2 \leq z$; $t$ is a progressive transfer if $x_1 < x_1 + t < x_2 - t < x_2 \leq z$ for all $t \in [0, (x_2 - x_1)/2]$ (Foster, 2006; Castagnoli and Muliere, 1989).

Initially, transfer axiom was formulated by Sen (1976) for univariate poverty: if there is a progressive transfer from someone who is better-off to someone worse-off, then poverty should correspondingly decrease. In the unidimensional context this property requires to put more weight on someone who is poorer given certain poverty line (Foster, 2006). In other words, according to transfer axiom poverty measure should be sensitive to inequality among the poor. In multidimensional context the situation with transfers becomes more complicated, since several dimensions are now considered. Before turning to transfer principle, we need to clarify when the multidimensional distribution of achievements is considered more equal.

**Definition 1.2.14.** (Uniform majorization principle) Let X be a matrix of achievements $X = [\ldots]$ and $A$ be a bistochastic matrix and not a permutation matrix such that $a_{ij} \geq 0$ and $\sum_i a_{ij} = \sum_j a_{ij} = 1$ for all $i, j$. If $X^{'} = AX$, then $X^{'}$ is more equal than $X$ (Kolm, 1977).

A multidimensional distribution of well-being is said to be more equal (or less concentrated) if *averaging* procedure was applied to the original data matrix (Kolm, 1977; Chakravarty, 2009). Averaging is a transformation of the original achievements matrix, when its rows are replaced by their convex combinations (Bourguignon and Chakravarty, 2003).

**Example 1.2.5.** Let $X$ be the matrix of achievements $X = [\ldots]$ and the vector of cut-offs $z = \begin{pmatrix} 3 & 6 & 9 \end{pmatrix}$. Let also $A$ be a bistochastic matrix. If $Y = AX$, then $Y$ is obtained from $X$ by averaging of achievements among the poor, while the achievements of non-poor remain unchanged:

$$Y = AX = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.75 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 5 & 7 \\ 2 & 3 & 4 \\ 4 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1.25 & 4.5 & 6.25 \\ 1.75 & 3.5 & 4.75 \\ 4 & 8 & 10 \end{bmatrix}$$

□

**Definition 1.2.15.** (Multidimensional transfer) If $Y$ is obtained from $X$ by the uniform majorization among the poor such that $Y = AX$, where $Y = [\ldots]$ and $X = [\ldots]$ are matrices of achievements and $A$ is a bistochastic matrix, then $P(Y; z) \leq P(X; z)$.

Therefore, multidimensional transfer principle can formulated as follows: if the distribution of achievements among the poor becomes less unequal, then multidimensional poverty should decrease or at least stay at the same level (Bourguignon and Chakravarty, 1999; Alkire et al., 2015).

The third group of axioms, which is reviewed here, concerns subgroup analysis. In particular, this group includes subgroup decomposability and subgroup consistency as its axioms.

**Definition 1.2.16.** (Subgroup decomposability) Let $X^1, X^2, \ldots, X^m$ be matrices of achievements that correspond to population subgroups $n_1, n_2, \ldots, n_m$ such that $\sum_{i=1}^{m} n_i = n$. Then $P(X; z) = \sum_{i=1}^{m} \frac{n_i}{n} P(X^i; z)$.

The groups in the population can be formed according to the socio-economic or geographic characteristics. Subgroup decomposability means that the overall poverty level is a weighted sum of poverty in several homogeneous groups of citizens, while weights are computed as the share of each group in the total population (Bourguignon and Chakravarty, 1999; Foster, 2006; Chakravarty, 2009). This property is identical in univariate and multivariate frameworks.

**Definition 1.2.17.** (Subgroup consistency) Let $n_1$ and $n_2$ be population subgroups such that $\sum(n_1 + n_2) = n$ and let $X$ and $Y$ be corresponding matrices of achievements at time $t$. Let matrices of achievements at $t + 1$ be $X^1$ and $Y^1$, while the population size in each group and the vector of cut-offs $z$ keep being fixed. If $P(X^1; z) > P(X; z)$ and $P(Y^1; z) = P(Y; z)$, then $P(X^1; Y^1; z) > P(X; Y; z)$.

According to subgroup consistency property, any change of the poverty magnitude in a population subgroup should be reflected on the aggregated level (Foster, 2006; Alkire et al., 2015).

The last group of properties considers some technical requirements to poverty indices. In particular, it comprises continuity, normalization and non-triviality.

**Definition 1.2.18.** (Continuity) $P(X; z)$ is continuous in $(X; z)$ (Chakravarty, 2009).

Continuity axiom requires poverty measure to be continuous over all incomes or achievements. Continuity rules out over-sensitivity of poverty measure towards minor observational errors (Bourguignon and Chakravarty, 2003).

**Definition 1.2.19.** (Normalization) If $x_{ij} \geq z_j$ for all $i$ and $j$, then $P(X;z) = 0$ for any $(X;z)$ and if $x_{ij} = 0$ for all $i$ and $j$, then $P(X;z) = 1$ (Chakravarty and Silber, 2008; Rippin, 2010).

According to normalization, poverty index has to be bounded within the interval $[0,1]$ (Alkire and Foster, 2011). Finally, non-triviality axiom requires poverty measure to take at least two different values (Alkire et al., 2015).

### 1.2.4 Aggregation step

The aggregation methods applied in multidimensional poverty indices are commonly based on additive, multiplicative or mixed rules approaches. In this subsection we review a non-exhaustive list of poverty indices existing in univariate and multivariate contexts. One of the well-known proposals to poverty measurement was done by Foster–Greer–Thorbecke (1984). This class of indices was originally introduced to measure univariate poverty. Since an extension of the FGT class of indices was proposed in literature, we provide an original formulation of this index. The generalized FGT class of poverty measures in univariate context is defined as follows

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z - x_i}{z}\right)^\alpha \tag{1.7}$$

where $n$ is the size of the population, $q$ stays for the number of poor individuals, $z$ is a poverty line ($z > 0$), $x_i$ is an achievement of poor individual in the considered well-being dimension and $\alpha$ is a poverty aversion parameter. For different values of $\alpha$ different versions of index can be obtained (Foster et al., 1984):

- when $\alpha = 0$, $P_\alpha$ coincides with the headcount ratio,
- when $\alpha = 1$, $P_\alpha$ coincides with the income-gap index,
- when $\alpha = 2$, $P_\alpha$ is an index sensitive to inequality among the poor.

The generalization of the FGT family to the multidimensional framework was proposed by Bourguignon and Chakravarty (2003) and is defined as follows:

$$P_\alpha^\theta(X;z) = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{d} w_j\left[\text{ Max }\left(\frac{z_j - x_{ij}}{z_j}, 0\right)\right]^\theta\right]^{\alpha/\theta} \tag{1.8}$$

where $w_j > 0$, $\sum_{j=1}^{d} w_j = 1$, represents a weight attached to the $j$-th well-being dimension, $\theta > 1$ measures the proximity between dimensions or, in other words, the elasticity of substitution, while $\alpha$ is a positive parameter behaving in the same way as in the FGT class of measures and is again interpreted as an inequality aversion parameter.

Another class of multidimensional poverty indices was proposed by Tsui (2002) and is formulated as follows

$$P(X;z) = \frac{1}{n}\sum_{i=1}^{n}\left[\prod_{j=1}^{d}\left(\frac{z_j}{\tilde{x}_{ij}}\right)^{a_j} - 1\right] \tag{1.9}$$

where $\tilde{x}_{ij} = \text{Min}\{x_{ij}, z_j\}$ is a normalized achievement of individual $i$ in well-being dimension $j$, $a_j \geq 0 \,\forall\, j$ is a parameter, whose value should guarantee that $\prod_{j=1}^{d}\left(\frac{z_j}{\tilde{x}_{ij}}\right)^{a_j}$ is convex with respect to its argument[7].

The multidimensional generalization of Watts index[8] should be mentioned as well. Watts multidimensional poverty index (Chakravarty, 2009) is given by

$$P(X;z) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} w_j \log\left(\frac{z_j}{\tilde{x}_{ij}}\right) \tag{1.10}$$

where $\tilde{x}_{ij} = \text{Min}\{x_{ij}, z_j\}$ is a normalized achievement.

The multidimensional poverty indices discussed so far imply the union criterion to multidimensional poverty. The last family of poverty indices reviewed in this section was proposed by Alkire and Foster (2011). These poverty measures are formulated with the intermediate identification criterion:

$$P(X;z) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} w_j g_{ij}^{\alpha}(k) \tag{1.11}$$

where $g_{ij}^{\alpha}$ is a deprivation of individual $i$ in well-being dimension $j$, $k$ is the intermediate criterion to the identification of poor, while $\alpha$ is a non-negative parameter can take the following values:

- for $\alpha = 0$, the measure is a weighted adjusted headcount ratio, which can be applied for both ordinal and cardinal variables,
- for $\alpha = 1$, the measure is adjusted poverty gap,
- for $\alpha = 2$, the measure is adjusted squared poverty gap.

The properties that characterize multidimensional poverty measures reviewed in this section are summered in Table 1.2.

[7] The requirement of convexity is necessary for this multidimensional poverty index to satisfy the multidimensional transfer principle. For details and proof of the proposition see Tsui (2002).

[8] For uni-dimensional version of index and its axiomatic characterization see Zheng (1993)

Table 1.2: Multidimensional poverty measures and axioms they satisfy

| Axiom | Bourguignon and Chakravarty (2003) | Tsui 2002 | Multidimensional Watts index | Alkire and Foster (2011) |
|---|---|---|---|---|
| Symmetry | S | S | S | S |
| Replication Invariance | S | S | S | S |
| Poverty Focus | S | S | S | S |
| Deprivation Focus | S | S | S | S |
| Scale invariance | S | S | S | S |
| Monotonicity | S for $\alpha > 0$ | S for $\alpha \geq 0$ | S | S for $\alpha > 0$ |
| Dimensional Monotonicity | S for $\alpha > 0$ | S for $\alpha \geq 0$ | S | S |
| Multidimensional Transfer | S for $\alpha \geq 0$ | S | S | NS for $\alpha \geq 0$ |
| Subgroup decomposability | S | S | S | S |
| Subgroup consistency | S | S | S | S |
| Continuity | S | S | S | S[1] |
| Normalization | S | NS[2] | NS[2] | S |
| Non-triviality | S | S | S | S |

*Note.* S = satisfied axiom, NS = not satisfied axiom.
[1] Satisfied when $\alpha > 0$ and union criterion is applied.
[2] A Lower bound zero is satisfied, but an upper bound is not fixed at 1.
Sources: Tsui (2002); Bourguignon and Chakravarty (2003); Rippin (2010); Alkire and Foster (2011)

## 1.3 Concluding remarks

In this Chapter we perceive poverty as a multidimensional phenomenon and contribute to the literature by reviewing the approaches to multidimensional poverty measurement. The discussion starts with defining two opposite methods to multivariate poverty measurement, namely the dashboard of indicators and the composite index approach, together with their pros and cons giving a comprehensive picture of methods already existing in literature. Having described the features of each method we focus on the composite index methodology.

We discuss three essential steps involved in the construction of multidimensional poverty indices, namely the identification, the weighting and the aggregation. In particular, we address three approaches to the identification of multidimensionally poor, which are relevant for counting approach, and identify their advantages and drawbacks. We proceed with the weighting step and review the approaches to establishing weights of well-being dimensions, namely data-driven, normative and hybrid. Finally, in the aggregation step we review several multidimensional poverty measures providing the list of axioms, satisfied by each class of indices.

## 1.4 Appendix

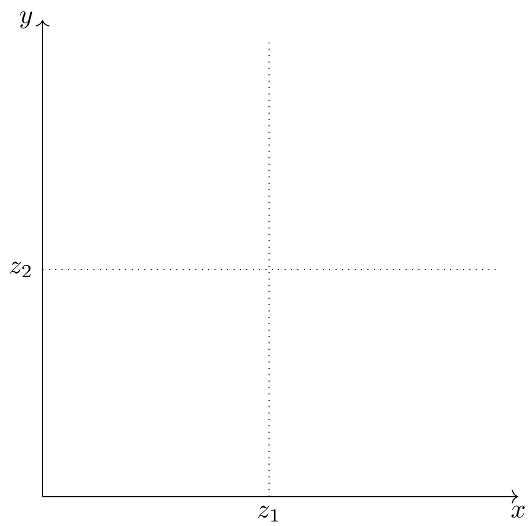



Figure 1.2: Example: a simple model of well-being represented by two attributes ($x$ and $y$) and the respective cut-offs ( $z_1$ and $z_2$)

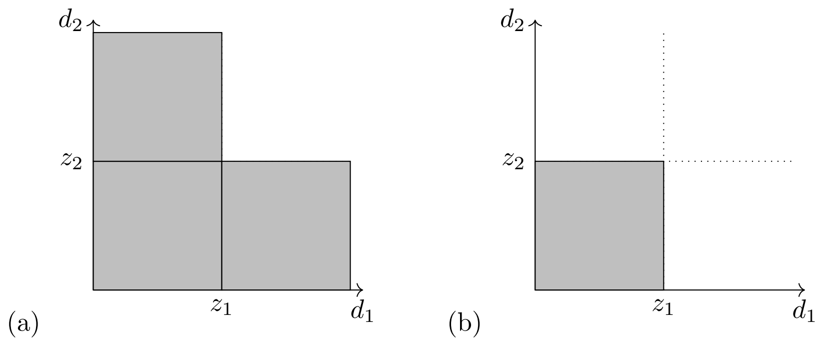



Figure 1.3: Union (a) and intersection (b) approaches to the identification of multidimensionally poor in the case of 2 well-being attributes. Grey areas on Figure illustrate multidimensionally poor according to each criterion.

# Chapter 2

# Literature review on copula functions

## 2.1 Introduction

It is widely acknowledged that well-being of a nation goes beyond the GDP. The economists commonly agree on the multidimensional nature of well-being and propose several approaches to assess the progress of the nation in terms of both objective and subjective metrics of welfare. In particular, there are two opposite view on how the information from several welfare dimensions should be summarized, namely the dashboard of indicators and the composite index methods. The dashboard is a non-aggregating approach, which monitors multidimensional well-being through a variety of indicators, while the composite index methodology measures multivariate phenomena by aggregating the progress of the society in multiple dimensions.

Although in recent years the composite indicators have become common measures of multivariate phenomena, this approach has both supporters and the sceptical audience. The opponents criticize composite indices due to lack of transparency and the subjective decisions involved in their methodology. Another shortcoming associated with the approach is related to the joint distribution of welfare indicators. The well-being dimensions commonly demonstrate a certain degree of dependence, that is the achievements of individuals tend to be associated. Therefore, the wealthier citizens are more inclined towards higher education and better health status compared to their somewhat deprived peers. In general, composite indices do not capture the dependence structure, which is present among well-being dimensions. By construction, the dashboard approach does not shed light on the interdependence among indicators as well since it does not involve the aggregation step. Therefore, a middle ground between the two extremes, namely the dashboard and the composite index approaches, should be found.

The research question we are dealing with is related to establishing the trade-off among well-being dimensions, which are interrelated with each other. There is a growing evidence in literature that the individual achievements in key dimensions are interconnected. By core well-being dimensions we understand, besides income, also education and health. For instance, better educated citizens commonly report to have better self-perceived health, while the earnings and the educational attainment that an individual possess are interdependent upon each other (Oreopoulos and Salvanes, 2011; Montenegro and Patrinos, 2014). Other empirical works have investigated the dependence among dimensions that form a composite indicator, in particular the dependence among the ingredients of the HDI. In particular, Pèrez and Prieto-Alaiz (2016) have shown that the multivariate dependence among the pillars of this considered composite indicator remain highly dependent, despite an improvement of the overall index during the last decades. Similarly, Decancq (2014) shows an analogous result regarding the relation among income, education and health, which became more dependent in Russia during its transition from a planned economy to a market economy.

We believe that the interdependence among dimensions is an essential factor, which should be considered while developing a composite indicator and defining a trade-off among the underlying dimensions. At the first step in resolving the problem of dependence, the degree of interrelation among attributes needs to be estimated. In the welfare context measuring the dependence is not a straightforward task since most well-being indicators are described by ordinal variables. The widespread linear correlation coefficient captures only linear dependence among variables and can produce misleading results in the context of welfare data (Pèrez and Prieto-Alaiz, 2016). Therefore, a tool, which can capture different dependence structures, is required for the specified objective. A flexible statistical tool that is able to capture the dependence among well-being dimensions is a *copula* function (Atkinson, 2011). In brief, a copula function together with the marginal distributions fully characterizes the joint distribution of two random variables (Nelsen, 2006). The applications of copula function into well-being framework are still rare (see the works of Quinn (2007); Bonhomm and Robin (2009); Decancq (2014); Pèrez and Prieto-Alaiz (2016) for the applications of copula function with welfare variables).

At the second step the estimated dependence should be incorporated in the composite indicator. A possible channel to include the dependence structure in the composite indicator is by defining a proper weighting scheme based on copula. As already defined in Chapter 1, the trade-off among well-being indicators is modelled by the researcher, who chooses the weighting approach. Therefore, a copula function plays an important role in the context of estimating the dependence among welfare variables and the proposal of a new weighting approach for the composite indicators.

The rest of the Chapter is structured as follows. In Section 2 we discuss the fundamental theorems in the copula function theory as well as copula-based dependence measures. In addition, some parametric families of copula function are provided. In Section 3 we summarize the recent application of copula with the welfare data. Finally, Section 4 contains concluding remarks.

## 2.2 Copula function

In this Section we summarize the theorems and definitions related to the copula function concept. Along with copula theory we classify the dependence concepts and define copula-based measures of interdependence, which are applicable to welfare indicators. Finally, an overview of some bivariate parametric copula functions is given.

### 2.2.1 Theorems and properties

We begin with introducing necessary notations and basic definitions. Let $x$ and $y$ be realized values of two random variables $X$ and $Y$, while $F(x) = P[X \leq x]$ and $G(y) = P[Y \leq y]$ denote their marginal distribution functions. Let also $H(x, y) = P[X \leq x, Y \leq y]$ define the joint distribution function. These preliminary notions are necessary for introducing copula function. In probability theory copula is used to describe the dependence structure present between random variables. Therefore, copula is an essential instrument considering the purpose of this dissertation, namely to estimate the dependence among welfare indicators. Let us now provide a definition of copula function.

**Definition 2.2.1.** (Mari and Kotz, 2001; Nelsen, 2006) A $n$-dimensional copula is a function from unit $n$-cube $[0, 1]^n$ to unit interval $[0, 1]$ that separates the dependence behaviour from the marginal distributions.

It is clear from the definition 2.2.1 that copula is a joint cumulative distribution function, whose marginal distributions are uniform on $[0, 1]$ (Schweizer, 1991; Cherubini et al., 2004). Many empirical applications of a copula function are based on the Sklar's theorem, which is the central one in the theory of copula.

**Theorem 2.2.1.** Sklar's theorem (Nelsen, 2006) If $H$ is a joint distribution function with uniform marginal distributions $F$ and $G$, then there exists a 2-dimensional copula $C_{XY}$ from unit square $[0, 1]^2$ to unit interval $[0, 1]$ such that for all $x$ and $y$

$$H(x, y) = C_{XY}(F(x), G(y)) \tag{2.1}$$

If $F(x)$ and $G(y)$ are continuous, then $C_{XY}$ is a unique copula. If marginal distributions are not continuous, then it is uniquely determined on $RanF \times RanG$, where $RanF$ is the range of the marginal distribution function $F$. This theorem can be extended to $n$ dimensions. Let $C(u, v)$ denote a bivariate copula function with uniform margins $u$ and $v$.

Some properties of a joint distribution are employed in the context of copula and, therefore, are reviewed in this subsection. A 2-dimensional copula function $C(u, v)$ satisfies the following properties for every $u$ and $v$ in $[0, 1]$ (Mari and Kotz, 2001; Nelsen, 2006):

1. $C(u, 0) = P[U \leq u, V \leq 0] = 0$. Similarly for $C(0, v) = 0$. This property is known as grounded property of copula function. According to it, if any of two marginal probabilities is equal to zero, then the joint probability takes the value of zero as well.

2. $C(u, 1) = P[U \leq u, V \leq 1] = P[U \leq u] = u$. Analogously with $C(1, v) = v$. According to this property, if the probability of any marginal outcome is equal to one, then the joint probability equals to the probability of the remaining uncertain outcome.

3. For every $u_1, u_2, v_1, v_2$ in $[0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

   $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$. This property is called rectangular inequality, which means that copula $C$ is 2-increasing.

Properties 1 and 2 hold also for n-dimensional copulas, while the third property in multivariate case claims that copula $C$ is $n$-increasing.

Since marginal distributions take values in the interval $[0, 1]$, copula as a joint cumulative distribution function has upper and lower bounds called Fréchet-Hoeffding bounds (Nelsen, 2006). We now provide the theorem related to these bounds.

**Theorem 2.2.2.** (Schweizer, 1991) Let $C$ be a 2-dimensional copula. Then for all $(u, v)$ in $[0, 1]^2$ the following inequality holds

$$max(u + v - 1, 0) \leq C(u, v) \leq min(u, v) \tag{2.2}$$

where left-hand side of the inequality is the lower bound $C_L$ of a 2-dimensional copula (the lower bound $C_L$ is not a copula for $n \geq 3$), while right-hand side of the inequality is the upper bound $C_U$ of a 2-dimensional copula ($C_U$ is always a copula).

In the empirical applications copula function can be parametrized (Trivedi and Zimmer, 2007):

$$H(x, y) = C_{XY}(F(x), G(y); \rho) \tag{2.3}$$

where $\rho$ is a dependence parameter between marginal distributions $F(x)$ and $G(y)$. Thus, $\rho$ is a scalar for a bivariate copula and a vector of parameters for the multivariate case. Copula is also called dependence function because it separates marginal distributions from the notion of dependence (Cherubini et al., 2004; Trivedi and Zimmer, 2007).

Let us now present theorem on the Lipschitz condition of uniform continuity for 2-dimensional copula.

**Theorem 2.2.3.** (Schweizer, 1991) If $C$ is a 2-dimensional copula, then for all $(u_1, u_2), (v_1, v_2)$ it satisfies the Lipschitz condition

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1| \tag{2.4}$$

which means that $C$ is continuous in both $u$ and $v$.

This theorem also holds when $n \geq 3$, so that $C$ is continuous in all its arguments. Properties 1-3 together with continuity imply that the graph of 2-dimensional copula is a "continuous surface over the unit square that contains the skew quadrilateral whose, vertices are (0,0,0), (1,0,0), (1,1,1) and (0,1,0). This surface is bounded below by the two triangles that together make up the surface of $C_L$ and above by the two triangles that make up the surface of $C_U$" (Schweizer, 1991). Figure 2.1 illustrates a graph of 2-dimensional Gaussian copula in the unit square.

Another property of copula function, which is relevant for the empirical applications is the invariance property discussed below.

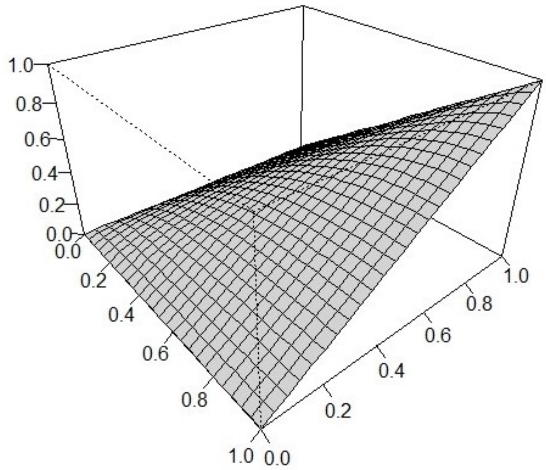



Figure 2.1: A cumulative distribution function of Gaussian copula with the dependence parameter $\rho = 0.8$

**Theorem 2.2.4.** (Schweizer and Wolff, 1981; Nelsen, 2006; Trivedi and Zimmer, 2007) Let $X$ and $Y$ be random variables with the continuous marginal distributions $F(x)$ and $G(y)$ respectively. Let also $\gamma$ and $\phi$ be strictly increasing on $RanX$ and $RanY$, then

$$C_{\gamma(X)\phi(Y)} = C_{XY} \tag{2.5}$$

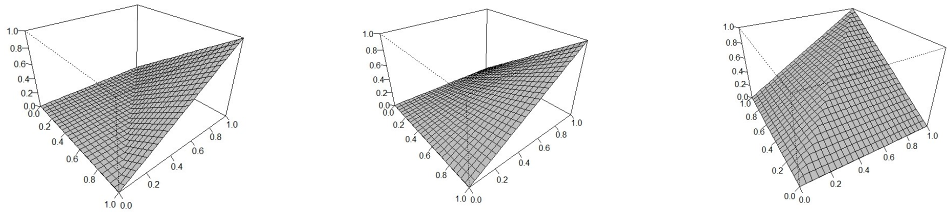
According to Theorem 2.2.4, copula remains invariant under strictly increasing transformations of the marginal distributions, even if $\gamma$ and $\phi$ ($\gamma \neq \phi$) affect differently $X$ and $Y$ (Cherubini et al., 2004). Trivedi and Zimmer (2007) suggest that the same copula function can be applied with the joint distributions of $(X, Y)$ and $(lnX, lnY)$.

In the context of dependence between the two continuous random variables $X$ and $Y$ copula $C$ is associated with the following fundamental properties (Schweizer and Wolff, 1981; Schweizer, 1991; Trivedi and Zimmer, 2007; Cherubini et al., 2004):

1. $X$ and $Y$ are *independent* if and only if $C(F(x), G(y)) = F(x)G(y)$. In other words, $X$ and $Y$ are independent if their dependence structure is modelled by a product copula $C_P$ (see Figure 2.2 for the product copula and the Fréchet-Hoeffding upper and lower bounds). Therefore, the product copula is an important benchmark of independence between two random variables.

2. $X$ and $Y$ are *comonotonic* or perfectly positively dependent if $C(F(x), G(y)) = C_U$. In other words, $X$ is said to be almost surely an increasing function of $Y$ if and only if their copula is equal to the Fréchet-Hoeffding upper bound.

3. $X$ and $Y$ are *countermonotonic* or perfectly negatively dependent if $C(F(x), G(y)) = C_L$. Formulating it differently, $X$ is said to be almost surely a decreasing function of $Y$ if and only if their copula is equal to the Fréchet-Hoeffding lower bound.



(a) $C_L$ (b) $C_P$ (c) $C_U$

Figure 2.2: Lower Fréchet-Hoeffding bound (a), product copula (b) and upper Fréchet-Hoeffding bound (c)

In some empirical applications the researcher can be interested in the survival time of individuals or firms. For this reason, we should introduce an univariate and a joint survival functions. Let $\overline{F}(x)$ be the univariate survival function such that $\overline{F}(x) = P[X > x]$ and $\overline{H}(x, y) = P[X > x, Y > y]$ be the joint survival function. The typical focus of survival analysis is time to the certain event, e.g. the duration of unemployment or time needed to find the job after graduating the university. In the multivariate survival models the practitioner can be interested in assessing the dependence among times of several events. Let us now define a survival copula that is the approach to analyse the joint survival times.

**Theorem 2.2.5.** (Nelsen, 2006) Let $(X, Y)$ be a pair of random variables with the marginal distributions $F(x)$ and $G(y)$ and the joint distribution function $H(x, y)$. Let $\overline{H}(x, y)$ be the joint survival function with the univariate survival margins $\overline{F}(x)$ and $\overline{G}(y)$ of $X$ and $Y$. Then the survival copula $\overline{C}$ is given by

$$\overline{H}(x, y) = \overline{C}(\overline{F}(x), \overline{G}(y)) \tag{2.6}$$

From Theorem 2.2.5 it follows that the joint survival function can be represented in terms of its marginal survival functions and the corresponding survival copula $\overline{C}$, which is analogous to the relationship between the marginal distributions and the joint distribution in the Sklar's theorem. Similarly, the survival copula $\overline{C}$ satisfies the upper and lower Fréchet-Hoeffding bounds $C_L < \overline{C} < C_U$ (Cherubini et al., 2004).

### 2.2.2 Concepts and measures of dependence

Since the empirical applications of copulas aim at capturing the dependence structure among variables, some concepts and measures of dependence including those based on copula should be reviewed here. Let $\delta$ be a measure of the dependence between pair of continuous random variables $(X, Y)$. The properties of the dependence measure were formulated by Embrechts et al. (2002); Balakrishnan and Lai (2009); Schweizer and Wolff (1981) among others. The axioms provided here were proposed by Schweizer and Wolff (1981):

1. Completeness: $\delta(X, Y)$ is defined for any $X$ and $Y$.

2. Symmetry: $\delta(X, Y) = \delta(Y, X)$.

3. Normalization: $0 \leq \delta(X, Y) \leq 1$.

4. $\delta(X, Y) = 0$ if and only if $X$ and $Y$ are independent.

5. $\delta(X, Y) = 1$ if $(X, Y)$ are comonotonic. According to Schweizer and Wolff (1981), a dependence measure is normalized on the interval $[0, 1]$, but this property can be extended to include the negative dependence. For instance, the dependence measure can be normalized on the interval $[-1, 1]$ and $\delta(X, Y) = -1$ for countermonotonic $(X, Y)$.

6. Let $T_1$ and $T_2$ be strictly monotonic transformation on $RanX$ and $RanX$, then $\delta(T_1(X), T_2(Y)) = \delta(X, Y)$.

7. If the joint distribution of $X$ and $Y$ is a bivariate normal with the correlation coefficient $\rho$, then $\delta(X, Y)$ is a strictly increasing function of $|\rho|$.

The provided list of properties can be too strong for some dependence measures (Schweizer and Wolff, 1981), while adding other properties can contradict the requirement of the existing ones (Embrechts et al., 2002).

Random variables $X$ and $Y$ are not independent if $C(F(x), G(y)) \neq F(x)G(y)$. Different concepts of the dependence are defined in the literature, e.g. the linear dependence, the concordance, the tail dependence etc. Two random variables $X$ and $Y$ are said to be associated if $cov(X, Y) \geq 0$ (Esary et al., 1967). A measure of association assigns a numeric value to the degree of dependence between random variables (Gibbons, 1993). The terms association and dependence will be used interchangeably thereafter.

One of the commonly applicable measures of linear dependence is Pearson correlation coefficient. For a pair of random variables $(X, Y)$ Pearson correlation coefficient is defined as

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.7}$$

where $cov(X,Y) = E[XY] - E[X]E[Y]$, $\sigma_X$ is the standard deviation of $X$ and $\sigma_Y$ is the standard deviation of $Y$. Correlation coefficient $\rho_{XY}$ is normalized on the interval $[-1, 1]$, symmetric and invariant under linear transformation (Trivedi and Zimmer, 2007). Lower and upper bounds of correlation coefficient correspond to maximal negative and maximal positive dependence respectively (Mari and Kotz, 2001).

However, if $\rho_{XY} = 0$, it does not imply in general that $X$ and $Y$ are independent. Zero correlation means that $cov(X,Y) = 0$ and implies independence only if $(X, Y)$ have bivariate normal distribution. For instance, if we assume that $X \sim N(0,1)$ and $Y = X^2$, then $cov(X,Y)$ is equal to zero, although $X$ and $Y$ are clearly dependent. Following Trivedi and Zimmer (2007), random variables are independent if $cov(\phi_1(X), \phi_2(Y))$ is equal to zero for $\phi_1$ and $\phi_2$ being any functions. This requirement does not hold for the linear correlation coefficient, which is its limiting factor as a measure of the dependence. In addition, the linear correlation coefficient is not invariant under non-linear strictly increasing transformations $T$: $\rho_{XY}(T(X), T(Y)) \neq \rho_{XY}(X,Y)$ (Embrechts et al., 2002).

Having discussed the linbear correlation coefficient and its drawbacks, we now define a concept of concordance and discuss some measures of dependence based on it.

**Definition 2.2.2.** (Nelsen, 2006) Let $(x_1, y_1)$ and $(x_2, y_2)$ be two observations of a pair of continuous random variables $(X, Y)$. $X$ and $Y$ are said to be concordant if $x_1 < x_2$ and $y_1 < y_2$ or $(x_1 - x_2)(y_1 - y_2) > 0$. Alternatively, they are said to be discordant, if $x_1 < x_2$ and $y_1 > y_2$ or $(x_1 - x_2)(y_1 - y_2) < 0$.

Concordance is a form of dependence, according to which large values of random variable $X$ are associated with large values of random variable $Y$. When this definition is reversed - large values of random variable $X$ are associated with small values of random variable $Y$ - we receive a definition of discordance (Trivedi and Zimmer, 2007). Both Spearman's $\rho$ and Kendall's $\tau$ measures are based on the concept of concordance and assess the monotonic relationships between $(X, Y)$ (Nelsen, 2006; Trivedi and Zimmer, 2007; Cherubini et al., 2004). Let us give a definition of population version of Kendall's $\tau_K$ measure.

**Definition 2.2.3.** Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be two independent and identically distributed random vectors with the same joint distribution $H$ and copula $C$. Then the population version of Kendall's $\tau_K$ measure is given by

$$\tau_K = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \tag{2.8}$$

In other words, Kendall's $\tau_K$ coefficient is obtained as the difference between the probability of concordance and the probability of discordance between $(X_1, Y_1)$ and $(X_2, Y_2)$. For a sample of $n$ pairs of observations from the vector $(X, Y)$ the sample version of Kendall's $\tau_K$ coefficient is formulated as follows

$$\tau_K = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \tag{2.9}$$

where $n_c$ gives the number of concordant pairs and $n_d$ stays for the number of discordant pairs. We now provide an illustrative example to show the performance of the sample version of Kendall's $\tau_K$ coefficient.

**Example 2.2.1.** Let the pairs of observations from two random variables $(X, Y)$ be given by

$$(1, 3), (2, 1), (5, 6), (15, 16), (9, 10)$$

To estimate the dependence with the sample version of Kendall's $\tau_K$ measure the original observations of both variables are transformed into ranks from 1 to $m$, where 1 corresponds to the highest value in the set of observations, while $m$ stays for the lowest value ($m$ coincides with the number of observations of each random variable). The obtained pairs of ranks $(\text{rank } X_i, \text{rank } Y_i)$ are the following: $(5, 4), (4, 5), (3, 3), (1, 1), (2, 2)$. In the next step the ranks of either random variable are sorted in the increasing order keeping the original pairs connected. As a result, we obtain the number of concordant and discordant pairs in the considered sample:

| | rank $X$ | rank $Y$ | $C$ | $D$ |
|---|---|---|---|---|
| | 1 | 1 | 4 | 0 |
| | 2 | 2 | 3 | 0 |
| | 3 | 3 | 2 | 0 |
| | 4 | 5 | 0 | 1 |
| | 5 | 4 | - | - |
| Total | - | - | 9 | 1 |

Finally, the sample version of Kendall's $\tau_K$ coefficient is estimated using the formula from (2.9):

$$\tau_K = \frac{9-1}{\frac{1}{2}5(5-1)} = 0.8$$

□

We now provide a definition of Kendall's $\tau_K$ coefficient based on copula function. In the bivariate case the definition of Kendall's $\tau_K$ is formulated as follows.

**Definition 2.2.4.** (Schweizer and Wolff, 1981) Let $(X, Y)$ be a pair of independent continuous random variables with the marginal distributions $F(x)$ and $G(y)$ and the joint distributions $H$. Then Kendall's $\tau_K$ coefficient is given by

$$\begin{aligned} \tau_K &= 4 \int_0^1 \int_0^1 H(x, y)\, dH(x, y) - 1 \\ &= 4 \int_0^1 \int_0^1 C(u, v)\, dC(u, v) - 1 \end{aligned} \tag{2.10}$$

Kendall's $\tau_K$ is normalized so that $-1 \leq \tau_K \leq 1$. For a pair of continuous random variables the lowest value of $\tau_K$ defines the countermonotonic random variables, while the highest value characterizes the comonotonic ones. In terms of copula

$$\tau_K = \begin{cases} -1 & \text{iif } C = C_L \\ 1 & \text{iif } C = C_U \end{cases} \tag{2.11}$$

Likewise Spearman's $\rho_S$ measure is based on the concept of concordance. A population version of this measure is provided below.

**Definition 2.2.5.** Let $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ be independent and identically distributed random vectors with the marginal distributions $F$ and $G$ and the joint distribution $H$. Then Spearman's $\rho_S$ coefficient is computed as follows

$$\rho_S = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]) \tag{2.12}$$

Similarly to Kendall's $\tau_K$ coefficient, Spearman's $\rho_S$ is obtained a difference between the probability of concordance and the probability of discordance between two pairs of random vectors $(X_1, Y_1)$ and $(X_2, Y_3)$, where the latter is the pair of independent random variables. It can be substituted by $(X_3, Y_2)$ without affecting the inequality above (Kruskal, 1958). If $X$ and $Y$ are independent, then Spearman's $\rho_S$

returns the value of zero. Spearman's $\rho_S$ is normalized on the same interval as Kendall's $\tau_K$ and satisfies the conditions from equation (2.11). Spearman's $\rho_S$ can be formulated using copula function as well.

**Definition 2.2.6.** (Joe, 1997) Let $(X, Y)$ be a pair of continuous random variables with the joint distribution $H$ and the univariate margins $F$ and $G$. Then Spearman's $\rho_S$ coefficient is obtained as

$$\begin{aligned}\rho_S &= 12\int_0^1\int_0^1 F(x)G(y)\,dH(x,y) - 3\\ &= 12\int_0^1\int_0^1 uv\,dC(u,v) - 3\end{aligned} \tag{2.13}$$

Since Kendall's $\tau_K$ and Spearman's $\rho_S$ are invariant under strictly increasing transformations and attain Fréchet-Hoeffding upper and lower bounds, they are suggested to be more flexible measures of dependence, which are suitable for non-normally distributed random variables (Joe, 1997). Moreover, the rank transformation of the original data by the proposed coefficients make them suitable for capturing more general types of dependence than linear correlation (Pèrez and Prieto-Alaiz, 2016).

Another concept of dependence is the positive quadrant dependence (PQD), which was originally formulated by Lehmann (1966).

**Definition 2.2.7.** (Lehmann, 1966) Let $(X, Y)$ be a pair of random variables with the joint distribution $H$. A pair $(X, Y)$ is defined as positively quadrant dependent if

$$P(X \le x, Y \le y) \ge P(X \le x)P(Y \le y) \qquad \text{for all } x, y \text{ in } R^2 \tag{2.14}$$

The condition provided in equation (2.14) is equivalent to the following definition (Joe, 1997):

$$P(X > x, Y > y) \ge P(X > x)P(Y > y) \qquad \text{for all } x, y \text{ in } R^2 \tag{2.15}$$

The condition of the PQD requires that probability of $(X, Y)$ to take large or small values simultaneously is greater or equal to the same probability in the case of independence (Nelsen, 2006). If the main inequality in equations (2.14) and (2.15) is reversed, then $(X, Y)$ are negatively quadrant dependent (NQD) (Joe, 1997). Let us now provide a definition of positive quadrant dependence in terms of copula function.

**Definition 2.2.8.** (Cherubini et al., 2004) Two continuous random variables $X$ and $Y$

are positively quadrant dependent if and only if

$$C(u,v) \geq uv \qquad \text{for all } (u,v) \text{ in } [0,1]^2 \tag{2.16}$$

In other words, a couple of random variables $X$ and $Y$ respects the requiems of the PQD if their copula is greater or equal to the product copula. There is a geometric interpretation of the PQD. If a pair of random variables $(X, Y)$ is positevely quadrant dependent, then the graph of the linking copula lies above or exactly on the graph of product copula $C_P$ (Nelsen, 2006). An important consequence of the PQD is illustrated in theorem (2.2.6).

**Theorem 2.2.6.** (Cherubini et al., 2004) If $X$ and $Y$ are continuous random variables that are positively quadrant dependent, then Kendall's $\tau_K$, Spearman's $\rho_S$ measures and the linear correlation coefficient take non-negative values:

$$\tau_{XY} \geq 0, \quad \rho_S \geq 0, \quad \rho_{XY} \geq 0 \tag{2.17}$$

The definition of the PQD between a pair of random variables can be extended to the multivariate context. The definition below contains the multidimensional extension of the PQD concept.

**Definition 2.2.9.** (Joe, 1997; Mari and Kotz, 2001) Let $X = (X_1, X_2, \ldots, X_n)$ with $(n > 2)$ be a vector of random variables. $X$ is defined as positively upper orthant dependent if the following condition holds:

$$P(X_i > x_i, i = 1, \ldots, n) \geq \prod_{i=1}^{n} P(X_i > x_i) \qquad \text{for all } x = (x_1, x_2, \ldots, x_n) \text{ in } R^n \tag{2.18}$$

while it is positively lower orthant dependent if

$$P(X_i \leq x_i, i = 1, \ldots, n) \geq \prod_{i=1}^{n} P(X_i \leq x_i) \qquad \text{for all } x = (x_1, x_2, \ldots, x_n) \text{ in } R^n \tag{2.19}$$

Conditions from equations (2.18) and (2.19) coincide in the bivariate case $(n = 2)$, while this statement does not hold in the multivariate framework (Joe, 1997).

The last type of dependence that is reviewed in this subsection is the tail dependence. The tail dependence assesses the probability that extreme values of random variables $X$ and $Y$ are interconnected. For instance, $X$ and $Y$ are asymptotically dependent if the following condition holds:

$$P(X > x|Y > y) \longrightarrow c > 0 \qquad \text{for } x, y \longrightarrow \infty$$

Otherwise, if the mentioned probability goes to zero, $X$ and $Y$ are defined as asymptotically independent (Heffernan, 2000). Therefore, the definition of the tail dependence is based on the conditional probability. Let us now recall the definition of it. The conditional probability that $X$ takes a value less or equal to $x$ given the probability of $Y$ to take a value less or equal to $y$ is defined as follows:

$$P(X \leq x|Y \leq y) = \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)} \tag{2.20}$$

Similarly the conditional probability of survival for two random variables is given by

$$P(X > x|Y > y) = \frac{P(X > x, Y > y)}{P(Y > y)}\ , \tag{2.21}$$

where the probability of survival $Pr(Y > y) = 1 - G(y)$. Both equations (2.20) and (2.21) can be reformulated using copula function:

$$\begin{aligned} P(U \leq u|V \leq v) &= \frac{C(u,v)}{v} \\ P(U > u|V > v) &= \frac{\overline{C}(u,v)}{1-v} \end{aligned} \tag{2.22}$$

Now we are ready to define measures of tail dependence.

**Definition 2.2.10.** (Joe, 1997; Cherubini et al., 2004) Let $(U, V)$ be a pair of uniformly distributed random variables on the unit square, then the coefficients of the lower tail dependence $\lambda_L$ andthe upper $\lambda_U$ tail dependence are defined as

$$\begin{aligned} \lambda_L &= \lim_{u,v \to \infty} \frac{C(u,v)}{v} \\ \lambda_U &= \lim_{u,v \to 0} \frac{\overline{C}(u,v)}{1-v} \end{aligned} \tag{2.23}$$

with $\lambda_U \in [0,1]$ and $\lambda_L \in [0,1]$. Two random variables are not upper (lower) tail dependent if $\lambda_U = 0 \quad (\lambda_L = 0)$.

After recalling the fundamental definition of the conditional probability and defining two measures of the tail dependence, we formulate two conditions related to a particular monotonicity in tails.

**Definition 2.2.11.** (Cherubini et al., 2004; Nelsen, 2006) Let $(X, Y)$ be a pair of random variables. $X$ is defined to be left-tail decreasing (LTD) in $Y$ if

$$P(X \leq x | Y \leq y) \quad \text{is a non-increasing function of } y \text{ for all } x \tag{2.24}$$

and right-tail increasing (RTI) if

$$P(X > x | Y > y) \quad \text{is a non-decreasing function of } y \text{ for all } x \tag{2.25}$$

According to the LTD condition, the probability that $X$ takes small value does not increase if $Y$ is increasing. Similarly, the RTI condition requires that the probability of $X$ to take high value does not decrease if $Y$ is increasing. Let us now give a theorem that shows the relationship between the RTI and the LTD conditions and the positive quadrant dependence concept.

**Theorem 2.2.7.** (Nelsen, 2006) Let $X$ and $Y$ be continuous random variables that satisfy the requirements of either the RTI condition or LTD condition, then $X$ and $Y$ are positively quadrant dependent.

According to Theorem 2.2.7, if a pair of random variables $(X, Y)$ satisfies the condition of the tail dependence, then the condition of PQD is also satisfied. However, the converse does not hold: the PQD does not imply the tail dependence.

### 2.2.3 Parametric families of copula function

With $C(u, v, \rho)$ we denote a bivariate copula with $\rho$ parameter, which captures the dependence existing between two uniformly distributed margins. The optimal choice of (parametric) copula family is driven by the original dataset. In particular, an appropriate copula function captures the dependence in the best possible way Trivedi and Zimmer (2007). In this subsection we discuss some bivariate copula families, which are frequently applied in the empirical applications. Let us start with defining the Normal or Gaussian copula.

**Definition 2.2.12.** (Joe, 1997; Cherubini et al., 2004; Mari and Kotz, 2001) Let $\Phi_{\rho_{XY}}$ denote the joint normal distribution with the correlation coefficient $\rho_{XY}$ and let $\Phi$ define a standard normal distribution. For $0 \leq \rho_{XY} \leq 1$ and $\Phi^{-1}$ being an inverse of $\Phi$, the *Gaussian copula* is given by

$$C^{Ga}(u, v, \rho_{XY}) = \Phi\rho_{XY}(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{2.26}$$

Let $\Phi^{-1}(u) = s_1$ and $\Phi^{-1}(v) = s_2$, then equation from (2.26) can be reformulated as follows:

$$\int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho_{XY}^2}} exp\left(\frac{2\rho s_1 s_2 - s_1^2 - s_2^2}{2(1-\rho_{XY}^2)}\right) ds_1 ds_2 \tag{2.27}$$

The Gaussian copula $C^{Ga}$ has a multivariate extension and is positively ordered with respect to the correlation coefficient $\rho_{XY}$ (Cherubini et al., 2004):

$$C^{Ga}_{\rho=-1} < C^{Ga}_{\rho=0} < C^{Ga}_{\rho=1} \tag{2.28}$$

Since the Gaussian copula is parametrized by the dependence parameter $\rho_{XY}$, which exists on the interval $[-1, 1]$, the copula approaches its lower and upper Fréchet bounds.

Another popular copula family is based on the bivariate Student's $t$-distribution and will be referred to as bivariate Student's $t$-copula.

**Definition 2.2.13.** (Embrechts et al., 2002; Fan and Patton, 2014) Let $T_z$ be a standard univariate Student's distribution with $z > 2$ degrees of freedom and let $T_{\rho z}$ be a bivariate $t$-distribution with correlation $\rho_{XY}$. Then the *Student's copula* is defined by

$$C^{St}(u, v, \rho_{XY}, z) = T_{\rho z}(T_z^{-1}(u), T_z^{-1}(v)) \tag{2.29}$$

which is equivalent to the following representation

$$C^{St} = \int_{-\infty}^{T_z^{-1}(u)} \int_{-\infty}^{T_z^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho_{XY}^2}} \left(1 + \frac{s_1^2 + s_2^2 - 2\rho_{XY} s_1 s_2}{z(1-\rho_{XY}^2)}\right)^{-\frac{z+2}{2}} ds_1 ds_2 \tag{2.30}$$

The parameter $z$ is an additional parameter, which controls the number of joint extreme events in Student's copula. When $z \to \infty$, the Student's $t$-copula converges to the Gaussian copula function (Cherubini et al., 2004). However, when $z$ is decreasing, the probability of joint extreme values is higher compared to the Gaussian copula and the joint $t$-distribution becomes fat-tailed. For all values of $\rho_{XY} < 1$ the Gaussian copula demonstrates the asymptotic independence, which means that the coefficients for tail dependence are equal to zero (Embrechts et al., 2002). On the contrary, the Student's copula incorporates the property of the asymptotic dependence, which strengthens when the degrees of freedom decrease (Trivedi and Zimmer, 2007). The limitation of both the Gaussian and the Student's $t$-copula is their assumption of symmetric distribution of upper and lower tails (Fan and Patton, 2014), which is not necessarily a feature of the considered dataset.

Similarly to the Gaussian copula, the Student's $t$-copula is positively ordered with respect to $\rho_{XY}$ and attains the upper and the lower Fréchet bounds for the finite $z$ (Cherubini et al., 2004):

$$C^{St} \begin{cases} = C_L & \text{if } \rho_{XY} = -1 \\ = C_U & \text{if } \rho_{XY} = 1 \\ \neq C_P & \text{if } \rho_{XY} = 0 \end{cases} \tag{2.31}$$

Another particular family of copulas is the Archimedean copula function, which have been studied by Genest and Rivest (1993) and Nelsen (2006) among others. These copula function are widespread in the empirical applications due to their attractive properties. Therefore, we provide a brief review of the Archimedean in this subsection.

**Theorem 2.2.8.** (Mari and Kotz, 2001; Nelsen, 2006) Let $\phi$ be a continuous decreasing convex function (called *generator*) from the unit interval [0,1] to $[0, \infty]$, such that $\phi(1) = 0$. Its pseudo-inverse is defined by

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t \leq \phi(0) \\ 0 & \phi(0) \leq t \leq \infty \end{cases} \tag{2.32}$$

Then the *Archimedean copula* is defined as follows:

$$C^A(u, v, \rho) = \phi^{[-1]}(\phi(u) + \phi(v)) \tag{2.33}$$

The pseudo-inverse of $\phi$ in the functional composition with the generator results in the identity for all $u$ and $v$ on the unit interval (Cherubini et al., 2004):

$$\phi^{[-1]}(\phi(t)) = t \tag{2.34}$$

If $\phi^{[-1]} = \phi^{-1}$, then the generator and the corresponding copula are know as strict (Nelsen, 2006).

**Theorem 2.2.9.** (Mari and Kotz, 2001; Nelsen, 2006) Let $C$ be an Archimedean copula and let $\phi$ be its generator. The generator $\phi$ is not unique, since for any positive constant $c > 0$, $c\phi$ is the generator of the same copula $C$.

One of the most famous Archimedean copulas is the Clayton copula, which is also called the Cook and Johnson copula. Let us now provide the definition of it.

**Definition 2.2.14.** (Cook and Johnson, 1981; Nelsen, 2006) The *Clayton copula* is given by

$$C^{Cl}(u, v, \rho) = (u^{-\rho} + v^{-\rho} - 1)^{-1/\rho} \tag{2.35}$$

with the generator $\phi(t) = \frac{1}{\rho}(t^{-\rho} - 1)$.

The parameter $\rho$ is the dependence parameter captured by copula. For the strict Clayton copula $\rho$ exists on the interval $(0, \infty)$ (Nelsen, 2006). Therefore, for $\rho$ close to zero, two marginal distributions are independent, whereas for $\rho$ going to infinity, the copula function approaches the upper bound $C_U$ (Trivedi and Zimmer, 2007). In the strict version the lower bound $C_L$ is not attained.

In contrast to the Gaussian and the Student's $t$-copula, the Clayton copula overcomes the limitations of symmetric distribution in tails and can capture the asymmetric dependence (Fan and Patton, 2014). The Clayton copula is useful for modelling dependence of data that demonstrates stronger left-tail dependence and weaker right-tail dependence (Nelsen, 2006; Trivedi and Zimmer, 2007).

Another family of the Archimedean copulas is the Gumbel copula, which is also referred to as the Gumbel-Hougaard copula.

**Definition 2.2.15.** (Hougaard, 1986; Mari and Kotz, 2001) The *Gumbel copula* has the following form:

$$C^{Gu}(u, v, \rho) = exp\left\{ - \left[ -(ln\, \overset{\rho}{u}) + (-ln\, \overset{\rho}{v}) \right]^{1/\rho} \right\} \tag{2.36}$$

with the generator $\phi(t) = (-lnt)^{\rho}$.

The dependence parameter in the Gumbel copula exists on the interval $[1, \infty)$, where $\rho = 1$ corresponds to the independence benchmark and $\rho = \infty$ gives the upper bound (Nelsen, 2006). The Gumbel copula has an asymmetric dependence structure as well, but unlike the Clayton copula, it shows stronger right-tail dependence and weak left-tail dependence (Marcantoni, 2014).

The last family of the Archimedian copula functions is the Frank copula, which was discussed by Genest (1987) and Nelsen (2006).

**Definition 2.2.16.** (Genest, 1987; Nelsen, 2006) The *Frank copula* takes the following form:

$$C^{Fr}(u, v, \rho) = -\frac{1}{\rho} ln\left(1 + \frac{(e^{-\rho u} - 1)(e^{-\rho v} - 1)}{e^{-\rho} - 1}\right) \tag{2.37}$$

with the generator $\phi(t) = -ln \frac{e^{-\rho t} - 1}{e^{-\rho} - 1}$.

The dependence parameter exists in the interval $(-\infty, \infty)$ including zero. This copula family is comprehensive: when $\rho$ equals $-\infty$ and $\infty$, then the copula attains the lower and the upper bounds respectively, while the case $\rho = 0$ corresponds to the independence case (Trivedi and Zimmer, 2007). Unlike previously discussed Archimedean copulas, the Frank copula represents symmetric tail dependence similarly to the Gaussian and the Students copula (Trivedi and Zimmer, 2007). However, the strongest dependence in the Frank copula is concentrated in the centre of the distribution.

## 2.3 Applications of copula in the welfare framework

The copula function is a flexible statistical tool that is commonly applied in finance and risk management for modelling the dependence (Embrechts et al., 2002; Cherubini et al., 2004; Marcantoni, 2014). However, the applications of copula with the welfare indicators are still rare. In this Section we review the most important works that have extended the application of copula function to the welfare context.

Firstly, as it was highlighted previously in this Chapter well-being dimensions are frequently interrelated. Marmot et al. (2008) suggests that health status is determined by such social determinants as education, occupation, gender and income etc. According to this report, there is a correlation between the distributions of health and wealth across countries. However, the interrelation between health and earnings is rather complex and requires further research.

A possible approach to assess the dependence among well-being dimensions is by applying copula function, the theoretical foundation of which was provided in Section 2. To the best of our knowledge, one of the first suggestions to consider the dependence between dimensions and to apply copula function for this purpose was made by Atkinson (2011). In his work he provides the example in the 2-dimensional case considering income and health status. The dependence between these well-being dimensions is referred to as the *income-health gradient*.

We recall from Chapter 1 that the aggregation in multidimensional poverty measure can be done in either sequence: first across individuals followed by the aggregation across dimensions or vice versa. According to Atkinson (2011), the choice of aggregation order is driven by the assumption regarding the impact of the dependence among dimensions on the overall value of the composite indicator. If the interrelation between income and health status should be considered in the evaluation of overall welfare, then the column-first aggregation rule should be avoided.

Further, shifting from the theoretical framework towards the empirical evaluation, the author suggests an application of copula, since it separates the marginal distributions from the dependence structure Atkinson (2011). The level of interdependence between health status and earning is expected to be country-specific. Therefore, changes in the marginal distribution of income will have diverse effects on the marginal distribution of health given a certain degree of interdependence between the two. Applying copula function in this case will shed light on the interrelation among well-being indicators. To sum up, the work by Atkinson (2011) is a useful starting point and an important conceptual argument to apply copula function with the well-being indicators.

One prominent application of copula function into well-being analysis was done by Decancq (2014). His application covers the evolution of the dependence among well-being dimensions in Russia using the data from the RLMS[1] was employed. Decancq (2014) considers the core dimensions of well-being, namely living standard, health and education; the attributes are represented by such indicators as the household disposable income, self-assessed health status and years of schooling. The well-being dimensions considered in this paper coincide with the dimensions of the HDI. The paper is based on 10 waves of the survey starting from 1995. During this period Russia experienced the transition from planned economy to market economy. The main hypothesis of the paper is that the dependence between three well-being dimensions increased. For measuring the interdependence among three dimensions of well-being author applies multivariate extensions of copula-based Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients.

To begin with, Decancq studies the time evolution of each well-being dimension separately. Additionally, he considers the development of the HDI in during the same period. The evolution of three well-being indicators demonstrate a different development patterns. However, the HDI, which aggregates the information from these dimensions, does not consider the dependence among its pillars. The estimation of multivariate versions of two copula-based dependence measures confirms the hypothesis of the increased dependence over the period under consideration. Therefore, the author concludes that the composite indicators of well-being and poverty should consider this interrelation among its pillars to capture the complex underlying phenomenon.

[1]RLMS - the Russia Longitudinal Monitoring Survey

An example of going beyond the GDP in the evaluation of nation's well-being is the paper by Kobus (2014). An important argument of the paper is that a component decomposition of the inequality index should be considered. This proposal is motivated by some policy implications. For instance, if two regions in the country have the same value of multidimensional inequality index, for the efficient government intervention it is important to know, which components contribute the most to the overall inequality. Additionally, Kobus (2014) suggests to distinguish between the contribution of dimensions from the impact of their interdependence on the overall indicator. This argument, namely the separation of the effect of within-dimension inequality from the between-dimension inequality, motivates the application of copula in the framework.

The empirical analysis was done using the US data from 1972 to 2010. The variables of interest are happiness and the health status. In order to measure the interdependence between the two indicators Kobus (2014) applies the copula-based Blomqvist's $\beta$ coefficient (for the details on this dependence measure see Schmid and Schmidt (2007b) and the references therein). The choice of this particular measure of dependence is motivated by the discrete nature of both variables, where the chance of obtaining neither concordant nor discordant pairs is high. Consequently, the requirement of normalization of the dependence measure is not fulfilled. The Blomqvist's $\beta$ coefficient does not have this limitation. According to Kobus (2014), an important advantage of using copula function in the inequality measurement is the opportunity to deal with the ordinal variables.

Another application of copula into welfare analysis was made by Bonhomm and Robin (2009), who considered the data of the Labour Force Survey in France. In their paper, the probability of transition between earnings quintiles is assumed to depend on the educational attainment. The dependence between individual earnings trajectories and the education is modelled by one-parameter Plackett's copula (this copula was originally proposed by Plackett (1965)). The intuition of copula application is the following: the higher is the dependence parameter, the lower is the mobility of earnings and vice versa.

Dearden et al. (2008) employs copula in a similar context as Bonhomm and Robin to model the lifetime earnings in the UK. They assume that th lifetime earnings follow a first-order Markov process, which means that the wage in period $t + 1$ depends on the wage from period $t$. This assumption allows authors to reconstruct the lifelong earning's trajectories from the available data. Additionally, authors model the dependence between earnings and the duration of labour market experience. For this purpose, they apply the Student's $t$-copula function.

Another application of copula was done by Quinn (2007). His contribution addresses the dependence between income and self-assessed health across countries (i.e. the income-health gradient described by Atkinson (2011)). Since copula function can join marginal distributions of different types, in this paper the dependence structure between the quantitative variable (i.e. income) and the ordinal variable (self-perceived health) is studied by applying copula function.

The population version of Kendall's $\tau$ measure can also be represented in terms of a copula function (see equation (2.10) in the section Concepts of dependence). However, it is a theoretical formulation and for empirical analysis an analytical expression is needed. Although an analytical version of Kendall's $\tau$ measure in terms of copula is not always available (dos Santos Silva and Lopes, 2008), there are some exceptions, in particular for the Archimedean class of copulas. For instance, for Clayton, Gumbel and Frank copulas the analytical form exists (see Quinn (2007) and dos Santos Silva and Lopes (2008) for details). The functional relationship between Kendall's $\tau$ measure and some Archimedean copulas is used in the paper in order to measure income-health gradient. Hence, considered countries were compared and ranked according to the level of dependence between distributions of health and income. To check the robustness of results, author compares the ranking obtained using copulas with the ranking received from the concentration index (this index is a well-developed measure of health inequality caused by the inequality of income, for the definition of the concentration index and related theorems see Kakwani (1980)).

The last application to be reviewed in this Section is the article by Kobus and Kurek (2017), who applied copula-based dependence measure to estimate the interdependence between such ordinal variables as education and mental health states. In particular, they have estimated the bounds of Kendall's $\tau_K$ measure for health categories and educational attainment. This estimation of bounds allows establishing dominance relationships between the US regions according to the education-health gradient.

## 2.4 Concluding remarks

This Chapter contributes to the literature on the dependence concepts and the copula-based measures of association. In this paper we build a methodological framework of copula function and suggest its application to the welfare data. We justify the use of copula in the well-being context by illustrating the relevance of interdependence among well-being attributes, especially in the context of the composite indicators. In particular, we suggest that the trade-off among the underlying attributes should be established considering the interrelation of individual performances across dimensions.

Since there is a growing evidence that key well-being dimensions are interconnected, the selection of indicators' weights is of particular interest in the composite index approach. Based on this fact we suggest that the trade-off among dimensions, which is defined by the researcher through assigning the weights to the attributes, should be established considering the strength of dependence. We propose copula function as a flexible statistical tool to measure the interrelation among attributes. Main theorems and properties of copula are summarized in this Chapter. Moreover, we introduce two copula-based measures of dependence, namely Spearman's $\rho_S$ and Kendall's $\tau_K$, which then can be applied to the welfare data to uncover its dependence structure. Summarizing the discussion, we propose to estimate the dependence among well-being dimensions using copula-based measures and establish the trade-off among well-being indicators of the composite indices considering this dependence.

# Chapter 3

# Measuring the dependence among well-being dimensions using copula function

## 3.1 Introduction

Well-being is suggested to be multidimensional and consist of both monetary and non-monetary attributes. While assessing their well-being respondents would characterize it as multivariate, including besides income also quality of dwelling, education, longevity etc. The level of income does not perfectly correlate with other attributes of well-being, violating the assumption of univariate approach to poverty measurement. Therefore, the level of income *per se* does not shed light on the possession of non-monetary resources by an individual. Hence, income alone reflects ones' well-being only partially, being one dimension among others representing this complex phenomenon.

Nevertheless, the empirical work related to well-being and poverty has been mostly focused on either income or consumption overlooking the multidimensionality of these notions. This choice of empirical studies is motivated by computational simplicity and an intuitive interpretation of univariate indices. Despite these obvious advantages, the analysis of individual or household well-being appears to be too restrictive if only income-related outcomes are considered.

Although a multidimensional approach has been developed and accepted by economists, there is still an open discussion on how to follow it in the empirical work.

As it was already emphasized in Chapter 1, multidimensionality of well-being and poverty can be reflected in either a dashboard of indicators or a composite index. While the former method does not imply an aggregation across indicators, the latter synthesizes information from several dimensions into a single number. Therefore, each approach provides opposite views on aggregation procedure.

In the last decades poverty measurement literature has been shifted to multidimensional framework as well. Recent studies have investigated the multidimensionality from various perspectives, i.e. from the discussion of key dimensions to be considered to the extension of univariate poverty measures into multidimensional context (Bourguignon and Chakravarty, 2003; Alkire and Foster, 2011). Besides income, the importance of such non-monetary attributes as health status, education and political power has been highlighted. Likewise, some multivariate poverty measures as extensions of univariate ones have been developed in literature (see Tsui (2002); Bourguignon and Chakravarty (2003); Alkire and Foster (2011) for the details). The axiomatic description of these poverty indices was discussed in Chapter 1.

What is typically left beyond the scope of multidimensional well-being and poverty indices is the dependence among dimensions. The following example is handy for illustrating the problem better. The Human Development Index (HDI) is a composite indicator that summarizes human capabilities on a country level. The index is composed of three dimensions: longevity, education and living standard. The HDI gives a snapshot of human development of a population and emphasizes the importance of capabilities in assessing an economic development. However, it does not give a clue about the interrelation of achievements across dimensions.

Recent literature has pointed out the importance of the interdependence among dimensions in well-being, poverty and inequality measurement (Stiglitz et al., 2009; Decancq and Lugo, 2012; Decancq, 2014). The degree of association among dimensions matters in the context of multidimensional poverty due to several aspects. Firstly, a relatively low performance in income dimension may be accompanied by an insufficient educational attainment and/or poor health status. As a result, deprivations are accumulated by the same individuals, who are falling into a trap of monetary and non-monetary impoverishment. This interrelationship among well-being attributes appears in the multidimensional context and, therefore, it should not be neglected. Secondly, societies where the interrelation among dimensions is higher, are also expected to be comparatively poorer than those with lower interdependence (Duclos et al., 2006). Authors suggest that comparing societies using a dashboard of indicators may bring a researcher to the conclusion that one country dominates the other in terms of poverty, while he might conclude the opposite if applying a multivariate poverty measure. Duclos et al. (2006) motivate this contradiction by the extent of dependency among indicators and propose a dominance surface approach. The advocated method takes into account interdependence among deprivations and establishes a robust dominance relationship between countries in terms of multidimensional poverty.

Besides poverty measurement, the dependence among well-being dimensions matters for poverty-reducing policy as well. In case income and health are positively dependent, an anti-poverty government intervention may allocate resources mainly in income to improve the living standard of citizens. This intervention will normally lead to a lower income poverty within a society, but it is also expected to have a positive impact on health due to the association between the two. However, if income and health are independent, focus on improving only monetary side of welfare will not be enough for reducing multidimensional poverty. Therefore, identifying the relationship among well-being attributes is essential for promoting efficiency of poverty-reducing actions.

Previous paragraph developed an idea of data-driven aspect of dependence, namely the coherence between individual achievements in well-being attributes, and its importance for poverty measurement and poverty-reducing policy. Another aspect of interplay between dimensions is of normative origin. In particular, the elasticity of substitution between two attributes: well-being dimensions can be modelled as either (perfect) substitutes or complements. This decision defines the role of each domain for the multidimensional poverty. For instance, a deficit in the first dimension is analysed in the light of above-threshold (below-threshold) performance in the second one. Hence, we can highlight two cases: the first one, when an individual is poor in two complementary attributes, and the second one, when he underperforms in two dimensions that are substitutes. Applying the elasticity assumption gives a theoretical ground to identify multidimensional poverty in complementary dimensions as a worse scenario.

To address the issue of dependence among dimensions, we propose to investigate the association structure of key dimensions and rank societies according to the observed level of dependency. Our framework is based on the *copula* function, which allows to detach the interdependence among variables from their marginal distributions. We study the properties of copula-based dependence measures under alternative column- and row-wise distribution of achievements. This enables distinguishing the rearrangements in individual achievements according to their effect: those considered undesirable since they lead to higher dependence or, instead, beneficial due to an offsetting impact of alternative allocation of achievements on dependence level.

Despite its useful properties discussed in Chapter 2 and prevalence in finance and risk management fields, the applications of copula into well-being context are still rare. Quinn (2007) has used copula to measure the dependence between income and self-assessed health also known as income-health gradient. Dearden et al. (2008) applied the copula function to implement the simulation of lifelong earnings of individuals in the UK. A similar application was done by Bonhomm and Robin (2009), who assessed the mobility of earnings in France. Kobus (2014) suggested another application, where a decomposition of the inequality index into within-dimension inequality and across-dimension correlation components was accomplished with a copula function.

Our application of copula is inspired by papers of Decancq (2014) and Pèrez and Prieto-Alaiz (2016), who use copula-based measures of association to estimate the dependence among dimensions included in the HDI. In particular, Decancq (2014) applies multivariate copula-based correlation coefficients, i.e. Spearman's $\rho$ and Kendall's $\tau$, to study the dependence among three pillars of the HDI in the transition economy. The results suggest that during one decade Russian society became remarkably more dependent than previously. Therefore, an improvement of the HDI for Russia is accompanied by a raising association among indicator's components for the same period.

Similarly, Pèrez and Prieto-Alaiz (2016) estimate the dependence among income, education and health with three multivariate copula-based correlation measures. The authors track the dependence among selected attributes in the time interval from 1980 until 2014. The analysis includes countries covered by the Human Development Report. According to the results, world's welfare increased in the considered time period; however, the dependence among dimensions remained high. Thus, the high-income countries tend to occupy high positions in education and health and vice versa.

The aim of this paper is to measure the dependence among dimensions of well-being that are most frequently used in multidimensional well-being and poverty studies. The specified aim implies the following research objectives. Firstly, our purpose is to develop a solid theoretical foundation of bivariate and multivariate measures of dependence based on copula function. Our second objective is to estimate a pairwise and an overall dependence among the selected attributes in the European countries. The last objective is to trace the evolution of the multivariate association focusing on pre- and post-crisis time intervals.

The rest of the Chapter is structured as follows: Section 2 gives an overview of copula and copula-based dependence measures, Section 3 provides a simulation study that identifies the level of dependence under alternative distributions estimated by copula-based correlation coefficients. Section 4 describes an application to the European Union countries. Section 5 gives the concluding remarks.

## 3.2 Methodology

This section synthesizes main theoretical findings that were discussed in details in Chapter 2. The notion of copula function was introduced by Sklar (1959) in his well-known theorem that has become the central one in copula theory. Copula function "links" univariate distributions of random variables in order to obtain their joint multivariate distribution. The purpose of this section is to give an overview of theory on copula function and its estimation. After introducing the basic definitions of copula theory, bivariate copula-based measures of dependence are explored. Finally, the dependence measures based on copulas are extended to the multidimensional framework.

### 3.2.1 Copula: definition and estimation

Let us introduce necessary notations here. Let $F(x)$ and $G(y)$ denote marginal distributions of random variables $X$ and $Y$ and $H(x,y) = P\big[X \leq x, Y \leq y\big]$ their joint distribution function. Finally, let $C(u,v)$ with $(u,v) \in [0,1]^2$ denote bivariate copula function and $C(\mathbf{u})$ with $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$ be $d$-dimensional copula function. Copula is a function that separates the dependence behaviour from the marginal distributions. As already defined in Chapter 2, let us recall the definition of a bivariate copula function.

**Definition 3.2.1.** (Nelsen, 2006) If $H(x,y)$ is a joint distribution function with uniform margins $F$ and $G$, then there exists a 2-dimensional copula function $C : [0,1]^2 \longrightarrow [0,1]$ such that

$$H(x,y) = C\big(F(x), G(y)\big) \tag{3.1}$$

Intuitively, a copula function "incorporates" all the dependence existing between $x$ and $y$. In the parametric approach, this bivariate dependence is captured by one or two parameters. In case of one parameter all the existing dependence is reflected in it and the copula is defined as follows

$$H(x,y) = C\big(F(x), G(y); \rho\big) \tag{3.2}$$

where $\rho$ is the copula dependence parameter. Alternatively, the overall dependence can be split between its types, e.g. one parameter capturing the dependence in upper tail and the other one reflecting the association in lower tail. Let us now provide a definition of copula density function.

**Definition 3.2.2.** (Charpentier et al., 2007) If $C(u, v)$ is a parametric bivariate copula with uniform margins and the dependence parameter $\rho$, then the corresponding copula density is expressed by

$$c_\rho(u, v) = \frac{d^2C(u, v)}{dudv} \tag{3.3}$$

From the definition it follows that the copula density is the derivative of $C(u, v)$ with respect to its arguments. Parametric copula functions can be classified into elliptical and Archimedean groups.

Elliptical copulas are derived from the elliptical distributions characterized by radial symmetry. In this type of probability distributions the mean and the median coincide and the distribution is symmetric about this point. Archimedean copula functions form another parametric family of copulas that are based on the generator function (see Chapter 2 for details). Gaussian and t-copula are typical examples of the elliptical copulas, while Frank and Gumbel copulas come from the Archimedean family.

While elliptical copula functions model symmetric behaviour in tails of distribution, Archimedean copulas, instead, allow a wider range of dependence structures and enable the correlation in tails to differ in magnitude. Thus, Gumbel copula - a representative of Archimedean group - is useful for estimating the dependence between two variables, which are correlated in upper tail of distribution and, simultaneously, demonstrate relatively weak correlation in the lower tail (Cherubini et al., 2004). By contrast, Frank copula should be applied when correlation in both tails is relatively weak and variables tend to be associated in the middle of the distribution (Nelsen, 2006; Trivedi and Zimmer, 2007). Finally, if the dependence in tails is symmetric, either Gaussian or Student's t-copula should be applied according to the magnitude of this dependence (Trivedi and Zimmer, 2007).

There exist several approaches to copula inference that can be classified into parametric and semi-parametric ones. One of classical fully-parametric methods is the maximum likelihood estimator (MLE). The MLE is the preferred first option due to its optimality properties (Kojadinovic and Yan, 2010). However, the previous statement is true only if the marginal distributions are specified correctly. As argued by Kim et al.

(2007), fully parametric estimators (including the MLE) of copula parameters 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) discussed by Genest et al. (1995). Following the PML estimator, marginal distributions are estimated non-parametrically by their empirical cumulative distribution functions. On the second step, the copula parameters are estimated by the MLE (Kim et al., 2007). The copula dependence parameter is estimated by maximizing the pseudo-loglikelihood function

$$logL(\rho) = \sum log\big[c_\rho(\hat{U}, \hat{V}|\rho)\big], \tag{3.4}$$

where $c_\theta$ is a copula density function, $\theta$ is a copula parameter to be estimated. $\hat{U}$ and $\hat{V}$ are rank-transformed pseudo-observations on the unit interval $[0,1]$ defined as follows: $\hat{U} = \dfrac{R(x_i)}{n+1}$ with $R(x_i)$ denoting the rank of $x_i$ among $x_1, \ldots, x_n$ and similarly for $\hat{V}$. In case ties (equal ranking of elements) occur, the average rank is assigned to each element.

### 3.2.2 Bivariate copula-based measures of dependence

Let us now turn to the measures of dependence based on copula function. We begin with providing a definition of concordance.

**Definition 3.2.3.** (Nelsen, 1996, 2006) Let $(x_1, y_1)$ and $(x_2, y_2)$ be two observations of continuous random variables $\big(X, Y\big)$. These variables are concordant if they increase or decrease coherently: $x_1 < x_2$ and $y_1 < y_2$ or $x_1 > x_2$ and $y_1 > y_2$. Otherwise, if $x_1 < x_2$ and $y_1 > y_2$ or $x_1 > x_2$ and $y_1 < y_2$, $X$ and $Y$ are discordant.

Both Kendall's $\tau_K$ and Spearman's $\rho_S$ coefficients measure this type of dependence and are defined as the difference between a probability of concordance and a probability of discordance between two random variables.

**Definition 3.2.4.** Let $\big(X_1, Y_1\big)$ and $(X_2, Y_2)$ be two independent and identically distributed random vectors. Kendall's $\tau_K$ correlation measure is defined as

$$\tau_K = P\big[(X_1 - X_2)(Y_1 - Y_2) > 0\big] - P\big[(X_1 - X_2)(Y_1 - Y_2) < 0\big] \tag{3.5}$$

**Definition 3.2.5.** Let $(X_1, Y_1)$, $(X_2, Y_2)$, $(X_3, Y_3)$ be independent and identically distributed random vectors. Then Spearman's $\rho_S$ correlation coefficient is defined between two pairs of random vectors $\big(X_1, Y_1\big)$ and $\big(X_2, Y_3\big)$, where the latter pair contains independent random variables:

$$\rho_S = 3\big(P\big[(X_1 - X_2)(Y_1 - Y_3) > 0\big] - P\big[(X_1 - X_2)(Y_1 - Y_3) < 0\big]\big) \tag{3.6}$$

Equations (3.5) and (3.6) provide population versions of the considered correlation measures. Kendall's $\tau_K$ and Spearman's $\rho_S$ coefficients differ in terms of their reference point. While the former computes probabilities of concordance and discordance between pairs of random variables with common joint distribution function, in the latter one pair of random variables is independent.

We now define sample versions of both correlation measures. Let a sample contains $n$ pairs of observations from vector $\big(X, Y\big)$. Original scores are transformed into ranks $\big(\text{rank of } X_i, \text{rank of } Y_i\big)$ from 1 to $n$, where 1 is assigned to the highest value and $n$ to the lowest one in a sample. Then Kendall's $\tau_K$ and Spearman's $\rho_S$ coefficients are defined by

$$\begin{aligned} \tau_K &= \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \\ \rho_S &= 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \end{aligned} \tag{3.7}$$

where $n_c$ and $n_d$ represent the number of concordant and discordant pairs respectively, in turn $d_i^2$ is a squared difference between two ranks for each pair of observations. Since both coefficients transform original scores into ranks, they are termed rank correlation coefficients.

We emphasized earlier that Kendall's $\tau_K$ and Spearman's $\rho_S$ are defined as the difference between the probabilities of concordance and discordance between two random variables. Therefore, we reformulate the definitions of both correlation coefficients using copula function and use the terms copula-based dependence measure and rank correlation coefficient interchangeably thereafter.

**Theorem 3.2.1.** (Nelsen, 2006) Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be independent and identically distributed random vectors, whose marginal distributions are $F$ and $G$ and joint distributions $H_1$ and $H_2$ respectively. Additionally, let $C_1$ and $C_2$ denote corresponding copula functions of each vector, such that $H_1\big(x, y\big) = C_1\big(F(x), G(y)\big)$ and similarly for $C_2$. If $Q$ is the difference between probabilities of concordance and discordance as given in equation (3.5), then

$$\tau_K = Q\big(C_1, C_2\big) = 4\int_0^1\int_0^1 C_2(u,v)\, dC_1(u,v) - 1 \tag{3.8}$$

See Nelsen (2006) for a formal proof. Similarly, we can obtain a copula-based definition of Spearman's $\rho_S$.

**Theorem 3.2.2.** (Nelsen, 2006) Let $\big(X_1, Y_1\big)$ and $\big(X_2, Y_3\big)$ be independent and identically distributed random vectors. Let also $C$ be a copula function associated with the former vector and $P$ be a product copula of the latter vector that contains independent random variables. If $Q$ is the difference between probabilities of concordance and discordance from equation (3.6), then

$$\rho_S = 3Q\big(C, P\big) = 12\int_0^1\int_0^1 C(u,v)\,dudv - 3 \tag{3.9}$$

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 (definitions of comonotonic and countermonotonic variables are given in Chapter 2). Four copula families together with the corresponding methods to get Kendall's $\tau_K$ and Spearman's $\rho_S$ coefficients are defined in Table 3.1.

Table 3.1: Kendall's $\tau_K$ and Spearman's $\rho_S$ coefficients and their definitions in terms of copula

| Copula family | $C(u,v,\rho)$ | $\tau_K$ | $\rho_S$ |
|---|---|---|---|
| Gaussian | $\Phi\rho(\Phi^{-1}(u), \Phi^{-1}(v))$ | $\frac{2}{\pi}\arcsin\rho$ | $\frac{6}{\pi}\arcsin\big(\frac{\rho}{2}\big)$ |
| t-copula | $T_{\rho z}(T_z^{-1}(u), T_z^{-1}(v))$ | $\frac{2}{\pi}\arcsin\rho$ | $\frac{6}{\pi}\arcsin\big(\frac{\rho}{2}\big)$ |
| Frank[1] | $-\frac{1}{\rho}ln\bigg(1+\frac{(e^{-\rho u}-1)(e^{-\rho v}-1)}{e^{-\rho}-1}\bigg)$ | $1-\frac{4}{\rho}\Big(1-D_1(\rho)\Big)$ | $1-\frac{12}{\rho}\Big(D_2(-\rho)-D_1(-\rho)\Big)$ |
| Gumbel[2] | $exp\bigg\{-\Big[(-ln\,u^\rho)+(-ln\,v^\rho)\Big]^{1/\rho}\bigg\}$ | $1-\rho^{-1}$ | $12\int_0^1[1+A(w,1-w;\rho)]^{-2}\,dw-3$ |

*Note.* $\Phi_\rho$ denotes a bivariate standard normal cumulative distribution function with correlation $\rho$; $T_{pz}$ is a bivariate *t*-distribution with correlation $\rho$ and degrees of freedom $z$.
[1] Both Kendall's and Spearman's correlation coefficients based on Frank copula rely on *Debye* function given by $D_k(\rho) = \frac{k}{\rho^t}\int_0^\rho \frac{t^k}{e^t-1}\,dt$, where $k = 1, 2$ (see Genest (1987) and Nelsen (2006) for details).
[2] $A(w, 1-w; \rho)$ is the Pickands dependence function. For details on the function and its properties see Gudendorf and Segers (2010).
*Sources*: Frees and Valdez (1998); Huard et al. (2006); Nikoloulopoulos and Karlis (2008); Joe (2015).

### 3.2.3 Multidimensional rank correlation measures

Several scholars proposed multivariate extensions of Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients. This subsection adopts a $d$-dimensional generalization proposed by

Blumentritt and Schmid (2014) that are defined as follows

$$\rho_S = \frac{d+1}{2^d-(d+1)} \cdot \left\{ 2^d \int_{[0,1]^d} C(\mathbf{u})\, d\mathbf{u} - 1 \right\} \tag{3.10}$$

$$\tau_K = \frac{1}{2^{d-1}-1} \cdot \left\{ 2^d \int_{[0,1]} C(\mathbf{u})\, dC(\mathbf{u}) - 1 \right\} \tag{3.11}$$

with $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$. For further details and other multivariate extensions see Schmid and Schmidt (2007a) 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 the lower bound. This is due to the fact that perfect negative dependence is not possible if the number of dimensions $d > 2$. Therefore, both multivariate coefficients have the following lower bounds if $d \geq 3$ (Nelsen, 1996):

$$\begin{aligned} &\frac{2^d-(d+1)!}{d!\{2^d-(d+1)\}} \leq \rho_S \leq 1 \\ &\frac{-1}{2^{d-1}-1} \leq \tau_K \leq 1 \end{aligned} \tag{3.12}$$

Multivariate extensions of Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients from equations (3.10) and (3.11) are estimated nonparametrically by the *empirical copula* proposed by Deheuvels (2009). Let us consider the $d$-dimensional random vector $\mathbf{X}$ with joint distribution $F$, marginal distributions $F_j$ for $j = 1, \ldots, d$ and copula $C$ that are unknown. Let $X_1, \ldots, X_n$ be a random sample from $\mathbf{X}$. In the first step marginal distribution functions are estimated nonparametrically by

$$\hat{F}_{j,n}(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{\{X_{ij} \leq x\}} \tag{3.13}$$

where $\mathbb{1}$ is an indicator function that is equal to 1 when the underlying condition is satisfied. Further we define pseudo-observations $\hat{U}_{ij,n} := \hat{F}_{j,n}(X_{ij})$ and $\hat{\mathbf{U}}_{i,n} = (\hat{U}_{i1,n}, \ldots, \hat{U}_{id,n})$. Finally, the empirical copula is the empirical distribution that is defined as follows

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbb{1}_{\{\hat{U}_{ij,n} \leq u_i\}} \tag{3.14}$$

Estimators of multivariate Spearman's $\rho$ and Kendall's $\tau$ measures based on empirical copula are investigated in Schmid and Schmidt (2007a) and Blumentritt and Schmid (2014). Nonparametric estimators of these coefficients are obtained by plugging-in the definition of empirical copula $\hat{C}_n$ into equations (3.10) and (3.11) and are given by

$$\rho_S(\hat{C}_n) = \frac{d+1}{2^d-(d+1)} \cdot \left\{ \frac{2^d}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} (1-\hat{U}_{ij,n}) - 1 \right\} \tag{3.15}$$

$$\tau_K(\hat{C}_n) = \frac{1}{2^{d-1}-1} \cdot \left\{ \frac{2^d}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \prod_{j=1}^{d} \mathbb{1}_{\{\hat{U}_{ij,n} \leq \hat{U}_{kj,n}\}} - 1 \right\} \tag{3.16}$$

The interpretation of multivariate versions of Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients in the well-being context is the following. Both measures compare the specified society with a reference society. For Spearman's $\rho_S$ coefficient a society with independent dimensions serves as the reference point, while Kendall's $\tau_K$ measure considers a society with the same level of dependence as a reference. Thus, both measures can be interpreted as a probability that two randomly chosen individuals from each society outperform each other.

After having selected multivariate extensions of Spearman's $\rho_S$ and Kendall's $\tau_K$, we aim at developing properties, which copula-based dependence approach should satisfy. Chosen correlation coefficients assess the dependence using relative positions (or ranks) of each representative in the society. Consequently, we expect that selected association measures remain unaffected by certain alterations appearing in the distribution of ranks. Let $\hat{\mathbf{U}} = \big(\hat{U}_{11}, \ldots, \hat{U}_{ij}\big)$ with $i = 1, \ldots, n$ and $j = 1, \ldots, d$ be a matrix of normalized individual ranks of $n$ citizens in $d$ dimensions of well-being. Let also $\rho_S$ define a multivariate copula-based correlation coefficient. Then we expect $\rho_S$ to satisfy the following properties:

1. **Row symmetry**: if items of matrix of normalized ranks are rearranged, so that $\hat{\mathbf{U}}'$ is obtained from $\hat{\mathbf{U}}$ by reallocation of its rows, then $\rho_S\big(\hat{\mathbf{U}}'\big) = \rho_S\big(\hat{\mathbf{U}}\big)$. This property insures that the order, in which individuals are set in the matrix, does not influence the level of dependence within a society.

2. **Column symmetry**: the dependence among $d$ dimensions of well-being does not change if columns of matrix $\hat{\mathbf{U}}$ are put in a particular order. For instance, let the original vector of ranks for individual $i$ be $\hat{\mathbf{U}}_i = \big(\hat{U}_{i1}, \ldots, \hat{U}_{id}\big)$, while a modified vector is $\hat{\mathbf{U}}_i^\star = \big(\hat{U}_{id}, \ldots, \hat{U}_{i1}\big)$ and analogously for the rest of population. Then $\rho_S\big(\hat{\mathbf{U}}^\star\big) = \rho_S\big(\hat{\mathbf{U}}\big)$.

3. **Continuity**: the dependence measure is continuous over all individual ranks.

4. **Independence benchmark**: if positions of individuals in different dimensions are independent, then $\rho_S(\hat{\mathbf{U}}) = 0$.

5. **Normalization**: if individual normalized ranks are perfectly positively dependent, then $\rho_S(\hat{\mathbf{U}}) = 1$.

## 3.3 A simulation study

For a comprehensive understanding of the multivariate copula-based versions of Kendall's $\tau_K$ and Spearman's $\rho_S$ coefficients discussed earlier, we conduct a simulation study to compare their performance. Our aim is to explore how the dependence among well-being dimensions changes if individual ranks are distributed in a certain manner. Several hypothetical societies are created for this purpose and the details on each society are given later in this section.

In each hypothetical society individuals have certain achievements in three dimensions of well-being, which may correspond, but not necessarily, to income, education and health. Each individual achievement is transformed into a rank so that 1 means "the poorest" performance in the specified dimension in the whole society, while rank $n$ means "the richest" one respectively. The procedure is applied for all achievements across considered dimensions.

Let $S$ be the $n \times d$ matrix of individual dimension-specific normalized ranks, where $n$ is the size of population and $d$ indicates the number of dimensions. Note that normalized ranks (or pseudo-observations) exist on the interval $(0, 1)$ and are computed as follows:

$$\hat{U}_{ij} = \frac{1}{n+1} \cdot (\text{rank of } x_{ij} \text{ in } x_{1j}, \ldots, x_{nj}) \tag{3.17}$$

where $i = 1, \ldots, n$, $j = 1, \ldots, d$ and $x_{ij}$ is the typical element of matrix $S$. The normalization term $n + 1$ is applied for uniform transformation of individual ranks so that each rank lies inside the interval $(0, 1)$, which is necessary for applying copula-based dependence measures thereafter.

We simulate seven matrices whose entries are normalized ranks in three well-being dimensions. Each matrix corresponds to a hypothetical society of 1000 individuals (equation 3.18). The first community $S_1$ has the most unequal allocation of individual achievements: while the first representative is the poorest in all attributes, the last one has the highest positions in all dimensions. Each successive individual, who appears after the poorest one, performs better in all dimensions. Shifting from the first to the last person corresponds to a step-by-step simultaneous increase of achievements in each dimension. Hence, the poorest individual is followed by the one less poor, until the best performing person is reached. Consequently, in these scenarios the orderings among individuals are unambiguous.

$$
\begin{gathered}
S_1 = \begin{pmatrix} \text{I} & \text{E} & \text{H} \\ 0.0038 & 0.0038 & 0.0038 \\ 0.0040 & 0.0040 & 0.0040 \\ \vdots & \vdots & \vdots \\ 0.9972 & 0.9972 & 0.9972 \\ 0.9977 & 0.9977 & 0.9977 \end{pmatrix} \qquad S_2 = \begin{pmatrix} \text{I} & \text{E} & \text{H} \\ 0.9977 & 0.9977 & 0.9977 \\ 0.9972 & 0.9972 & 0.9972 \\ \vdots & \vdots & \vdots \\ 0.0040 & 0.0040 & 0.0040 \\ 0.0038 & 0.0038 & 0.0038 \end{pmatrix} \\
S_3 = \begin{pmatrix} 0.0038 & 0.0038 & 0.9977 \\ 0.0040 & 0.0040 & 0.9972 \\ \vdots & \vdots & \vdots \\ 0.9972 & 0.9972 & 0.0040 \\ 0.9977 & 0.9977 & 0.0038 \end{pmatrix} \qquad S_4 = \begin{pmatrix} 0.0038 & 0.9977 & 0.9977 \\ 0.0040 & 0.9972 & 0.9972 \\ \vdots & \vdots & \vdots \\ 0.9972 & 0.0040 & 0.0040 \\ 0.9977 & 0.0038 & 0.0038 \end{pmatrix} \\
S_5 = \begin{pmatrix} 0.9977 & 0.0038 & 0.0038 \\ 0.9972 & 0.0040 & 0.0040 \\ \vdots & \vdots & \vdots \\ 0.0040 & 0.9972 & 0.9972 \\ 0.0038 & 0.9977 & 0.9977 \end{pmatrix} \qquad S_6 = \begin{pmatrix} 0.4721 & 0.0011 & 0.9991 \\ 0.4892 & 0.0014 & 0.9987 \\ \vdots & \vdots & \vdots \\ 0.9025 & 0.9386 & 0.0023 \\ 0.7522 & 0.9904 & 0.0013 \end{pmatrix} \\
S_7 = \begin{pmatrix} 0.9558 & 0.0466 & 0.9997 \\ 0.1549 & 0.5402 & 0.9987 \\ \vdots & \vdots & \vdots \\ 0.7628 & 0.8205 & 0.0039 \\ 0.7201 & 0.0940 & 0.0025 \end{pmatrix}
\end{gathered} \tag{3.18}
$$

The scenario, when each following individual has better achievements in all dimensions, is an example of extreme distribution of benefits within a society and may be regarded as unrealistic. However, this hypothetical model is useful for visualizing the most unequal state, for example the one based on the caste system. Here the most privileged social class has an access to all resources, while the bottom social group is deprived in all dimensions of well-being. We expect that the dependence measured by copula-based coefficients is the highest for this scenario and the estimate should approach its upper bound of 1. Therefore, matrix $S_1$ is an essential benchmark of the highest possible multivariate dependence.

The next hypothetical society is given in matrix $S_2$. Obviously, matrices $S_1$ and $S_2$ have identical ranks. However, mentioned societies differ according to the column-wise order of pseudo-observations. Matrix $S_1$ presents an ascending order of dimensional ranks such that the first individual of the distribution is bottom-ranked in all dimensions and is outranked by all the successive individuals. The last person of the distribution is the richest across all dimensions and is not outperformed by anyone from the sample. By contrast, matrix $S_2$ presents a descending column-wise order of normalized ranks so that the first individual is the richest and the last one is the poorest.

By introducing this transformation of ordering among individuals we aim at proving that the level of dependence among dimensions is equal in the two societies $S_1$ and $S_2$ and, therefore, copula-based measures of association satisfy the property of row symmetry. It is a fundamental axiom of the dependence measure to be unaffected by the order in which individuals appear. Thus, our expectation is that the first two societies are equal with respect to their dependence and have the maximum possible value of correlation.

Starting from a hypothetical society described by matrix $S_3$, some discrepancy in the row-wise distribution of performances is allowed. Here achievements of the same individual are no longer identical across dimensions. For instance, the first person of matrix $S_3$ has the lowest achievements in the first two dimensions, although he outperforms the rest of the population in the last attribute. By downward move the successive individuals experience both a relative improvement - in the first and second columns compared to the first representative - and a decline if focusing on the last dimension. Hence, the second individual from matrix $S_3$ is richer in the first two dimensions and poorer in the last attribute than his predecessor and so on. Finally, the last representative of the society has the highest achievements in the first two columns compared to the rest of the population, although he suffers from the insufficient performance in the last dimension.

The design of the hypothetical society $S_3$ is motivated by the following idea: we assume that a relatively higher performance in one dimension can compensate for a lower achievement in the other attribute. The objective is to test whether the relationship between dimensions is of a compensable nature. Moreover, the society is constructed in such a manner that an improvement in one dimension is accompanied by a deterioration in the rest of them and visa versa. Therefore, our expectation is that rank correlation coefficients will reflect this process by producing a negative value of dependence.

The following two populations, namely $S_4$ and $S_5$, develop the same pattern of $S_3$ further. Thus, in $S_4$ the first person still has the lowest normalized rank in the first dimension, but now his achievements in the rest of attributes are the highest. As before the ranks are given in ascending order in the first column, followed by descending sequence in other columns. This structure is anew justified by the fact the dependence measure should be independent of row-wise order of items. Matrix $S_5$ aims to exploit the other fundamental property, namely symmetry, and reverses the sequence of columns.

Therefore, our expectation is that societies described by matrices $S_3$, $S_4$ and $S_5$ are indistinguishable with respect to the dependence among dimensions, whichever though is expected to be negative. Populations from these three scenarios differ from the caste system societies developed in $S_1$ and $S_2$, since the privileged group is able to control the majority of resources, but not all of them. Indeed, in the real-world examples there are individuals comparatively better-off in certain dimension(s), who are at the same time worse-off in other well-being attributes. However, in our model the distribution of performances is more extreme.

The penultimate society $S_6$ illustrates another case, when the multivariate dependence is expected to be highly negative. Ranks in the last column of $S_6$ are illustrated in a descending order. Within this population every high performance in either dimension is accompanied by low and middle achievements in the other two. Unlike previous populations, where simultaneous high performances in two attributes were allowed, this feature is eliminated in $S_6$. We believe that both correlation coefficients will let us discriminate between $S_6$ and the other "negative correlation" group of societies discussed above.

The last society $S_7$ contains normalized ranks distributed in unsystematic way. We do not impose either ascending or descending column-wise order. All individual achievements are allocated randomly. This last benchmark is assumed to show a fully independent society, where, for instance, earnings do not correlate with education and better health is not accompanied by neither high nor low educational attainment. Figure 3.20 in Appendix illustrates randomly generated populations of 1000 individuals, whose achievements correlate positively, negatively or are independently allocated.

In the described simulation we develop different distributions of rank-transformed well-being indicators. All scenarios incorporate a general idea of multidimensionality of well-being and highlight an essential role of each attribute for a society. Following leading scholars, who commonly acknowledge that income-based measures do not exhaustively assess multidimensional well-being, a majority of empirical studies model well-being using key indicators, namely income, education and health. It justifies a trivariate model of well-being adopted in the simulation. However, a specific empirical application may require a more advanced well-being modelling. As a result, a simulation with more than three dimensions is required to test how the dependence is captured by copula-based measures.

$$R_1 = \begin{pmatrix} \text{I} & \text{E} & \text{H} & \text{SW} \\ 0.0007 & 0.0007 & 0.0007 & 0.0007 \\ 0.0008 & 0.0008 & 0.0008 & 0.0008 \\ \vdots & \vdots & \vdots & \vdots \\ 0.9983 & 0.9983 & 0.9983 & 0.9983 \\ 0.9987 & 0.9987 & 0.9987 & 0.9987 \end{pmatrix} \quad R_2 = \begin{pmatrix} \text{I} & \text{E} & \text{H} & \text{SW} \\ 0.9987 & 0.9987 & 0.9987 & 0.9987 \\ 0.9983 & 0.9983 & 0.9983 & 0.9983 \\ \vdots & \vdots & \vdots & \vdots \\ 0.0008 & 0.0008 & 0.0008 & 0.0008 \\ 0.0007 & 0.0007 & 0.0007 & 0.0007 \end{pmatrix}$$

$$R3 = \begin{pmatrix} 0.0007 & 0.0007 & 0.0007 & 0.9987 \\ 0.0008 & 0.0008 & 0.0008 & 0.9983 \\ \vdots & \vdots & \vdots & \vdots \\ 0.9983 & 0.9983 & 0.9983 & 0.0008 \\ 0.9987 & 0.9987 & 0.9987 & 0.0007 \end{pmatrix} \quad R_4 = \begin{pmatrix} 0.0007 & 0.0007 & 0.9987 & 0.9987 \\ 0.0008 & 0.0008 & 0.9983 & 0.9983 \\ \vdots & \vdots & \vdots & \vdots \\ 0.9983 & 0.9983 & 0.0008 & 0.0008 \\ 0.9987 & 0.9987 & 0.0007 & 0.0007 \end{pmatrix} \tag{3.19}$$

$$R_5 = \begin{pmatrix} 0.0007 & 0.9987 & 0.9987 & 0.9987 \\ 0.0008 & 0.9983 & 0.9983 & 0.9983 \\ \vdots & \vdots & \vdots & \vdots \\ 0.9983 & 0.0008 & 0.0008 & 0.0008 \\ 0.9987 & 0.0007 & 0.0007 & 0.0007 \end{pmatrix} \quad R_6 = \begin{pmatrix} 0.9718 & 0.4761 & 0.8849 & 0.0012 \\ 0.7493 & 0.8403 & 0.8904 & 0.0018 \\ \vdots & \vdots & \vdots & \vdots \\ 0.0734 & 0.0010 & 0.9221 & 0.9991 \\ 0.7414 & 0.0376 & 0.0215 & 0.9992 \end{pmatrix}$$

$$R_7 = \begin{pmatrix} 0.2876 & 0.4090 & 0.9405 & 0.5281 \\ 0.5515 & 0.9568 & 0.6776 & 0.1029 \\ \vdots & \vdots & \vdots & \vdots \\ 0.0106 & 0.3274 & 0.6176 & 0.0766 \\ 0.8975 & 0.8877 & 0.2499 & 0.8689 \end{pmatrix}$$

We perform a similar computation in case of four dimensions contained in equation

(3.19) and consider seven scenarios, each containing normalized ranks of 1000 individuals. For the purpose of this study the fourth dimension is assumed to be subjective well-being. However, a choice of other attribute relevant for a specific empirical applications is valid.

New scenarios are extensions of three-dimensional examples in four-dimensional framework. Hence, representatives of society $R_1$ ($R_2$) are allocated in ascending (descending) order according to their ranked achievements. Similarly to previous simulation, the following four societies, $R_3$, $R_4$, $R_5$ and $R_6$, are expected to be negatively dependent since each individual has a mismatch of achievements across dimensions. For instance, the first representative of $R_3$ is "the poorest" in three dimensions and "the richest" in one attribute. The pronounced discrepancy of each representative across dimensions is the reason why negative association is expected. The last population $R_7$ has normalized ranks that are distributed at random. As a result, copula-based correlation coefficients should be equal to zero.

The results of estimation are given in Table 3.2. We report a mean of 1000 bootstrap replicates for each coefficient. The maximum positive dependence in three-dimensional scenario is observed in societies $S_1$ and $S_2$, while their four-dimensional counterparts $R_1$ and $R_2$ are maximally positively dependent as well. Since both copula-based measures attain their upper bound, this result proves row symmetry property.

Table 3.2: Dependence among dimensions estimated by multivariate copula-based Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients

| | Dimensions $d = 3$ | | | Dimensions $d = 4$ | |
|---|---|---|---|---|---|
| Scenario | $\rho_S$ | $\tau_K$ | Scenario | $\rho_S$ | $\tau_K$ |
| $S_1$ | 0.9968 | 1.0000 | $R_1$ | 0.9991 | 1.0000 |
| $S_2$ | 0.9991 | 0.9992 | $R_2$ | 0.9995 | 1.0000 |
| $S_3$ | -0.3299 | -0.2363 | $R_3$ | -0.0895 | -0.0280 |
| $S_4$ | -0.3310 | -0.2367 | $R_4$ | -0.2112 | -0.1417 |
| $S_5$ | -0.3308 | -0.2364 | $R_5$ | -0.0904 | -0.0279 |
| $S_6$ | -0.4824 | -0.3332 | $R_6$ | -0.2705 | -0.1428 |
| $S_7$ | 0.0001 | 0.0005 | $R_7$ | 0.0004 | 0.0003 |

*Note.* Mean of 1000 bootstrap replicates

As expected, populations $S_3 - S_6$ and $R_3 - R_6$ have a negative association among attributes. We emphasize that the result of identical dependence in $S_3$ and $S_5$ proves the proposed column symmetry axiom. In other words, the order of attributes in matrix of ranks does not influence the level of multivariate dependence. Remarkably, while scenarios $R_3$ and $R_5$ have slightly negative correlation, negative association in $S_3$ and $S_5$

is comparatively lower. A plausible explanation for this result is related to the construction of matrices. In the former populations a mismatch in the distribution of ranks appears in one out of four attributes, while in the latter it is in one out of three dimensions. A positive correlation between harmonized columns, for instance, as in society $R_3$, mitigates a discrepancy introduced by one attribute. In a three-dimensional scenario the role of mitigation of harmonized ranks is lower, since only two columns contain coherent ranks.

Societies $S_6$ and $R_6$ have the lowest negative association among three and four attributes respectively. Considering the case of $S_6$, each individual has contradictory ranks in two dimensions, while in the third dimension his performance is somewhat of a "middle ground" between the two. For instance, the the first individual is "the poorest" and "the richest" in the second and third dimensions respectively, while his achievement in the first one is the midpoint. Consequently, neither ranks are perfectly harmonized, which eliminates an option for mitigation.

A similar scenario is developed in four-dimensional case. However, unlike three-dimensional framework the dependence in societies $R_4$ and $R_6$ is sufficiently similar. As a result, if two pairs of attributes are negatively correlated, then the overall dependence is expected to be strongly negative too.

## 3.4 Results from empirical application

### 3.4.1 Data description

This analysis aims to understand the degree of dependence among some well-being dimensions. In this study, we focus on the following dimensions: income, education and health. These attributes are chosen in this study because they are the ones considered in the Human Development Index (Foster et al., 2005) and in the Multidimensional Poverty Index (UNDP, 2016). We will investigate the correlation between each pair of dimensions as well as their overall dependence.

To measure the dependence between dimensions, we use the data from the EU-SILC referred to the year 2015. In our study we consider the following European countries: Italy, Germany, Sweden, France and Poland abbreviated as IT, DE, SE, FR and PL respectively. These countries are selected to have representatives of different welfare state systems[1].

[1]Here we rely on *regime approach* to the classification of welfare states proposed by Esping-Andersen (1990). This approach classifies countries according to their political traditions and the extent of public responsibility for population's welfare. Ferrera (1996) extended this classification by adding the "Southern type" of welfare regime that characterizes Italy, Portugal, Spain and Greece. For other approaches to classification of welfare states see Bergqvist et al. (2013).

While France is classified as a state with a corporatist regime, Sweden belongs to a social democratic welfare state (see Aspalter et al. (2009) for the classification). It is argued that Germany, traditionally included into a corporatist welfare regime, has adopted a liberal regime due to labour market reforms taking place in the last two decades (Siegel et al., 2014). Italy is a representative of Southern welfare regime combines corporatist traditions with universal benefits (Ferrera, 1996). Poland has undergone an economic transformation and has adopted a corporatist welfare regime (Aspalter et al., 2009).

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 may have not achieved their highest educational level yet.

The income dimension is represented by equivalised disposable income adjusted for a household size by OECD-modified equivalence scale[2]. Regarding income distribution several groups can be identified, for instance, one can distinguish between low, middle and high income earners. The proposed classification suggests that individual earnings are rather heterogeneous and, while income of top earners is affected by capital accumulation, the rest of the distribution is typically associated with employee's wage (Roine et al., 2009). Moreover, in the top earners group the intergenerational transmission of employers is remarkably higher than in other income groups (Corak, 2013). For the purpose of investigating correlation between income and educational dimension, we believe that effects of higher capital shares and the intergenerational transmission of employers should be disentangled from the role of educational attainment. Thus, we focus on the "core" of the income distribution that is mostly associated with wage earners and exclude those individuals, whose earnings belong to the top 1% of the distribution[3]. In addition, our analysis covers only positive values of disposable income excluding negative and zero values from the indicator. Hence, the distribution is both right- and left-truncated.

Figure 3.1 plots the distribution of equivalised disposable income and the corresponding kernel densities across selected countries. Whereas for all countries the distribution is right-skewed, in case of Sweden it tends to be symmetric and approximates the normal distribution. This tendency is justified by the fact Sweden is a

[2]For the construction of disposable income of each individual in the household the following values are assigned to its members according to the modified equivalence scale: a value of 1 is assigned to the head of household, 0.5 is given to other adults and 0.3 for children.

[3]If, instead, these top earners are included in the sample, the "direction" of dependence does not change, but its the magnitude is slightly affected compared to the truncated distribution (see Table 3.5 in Appendix).

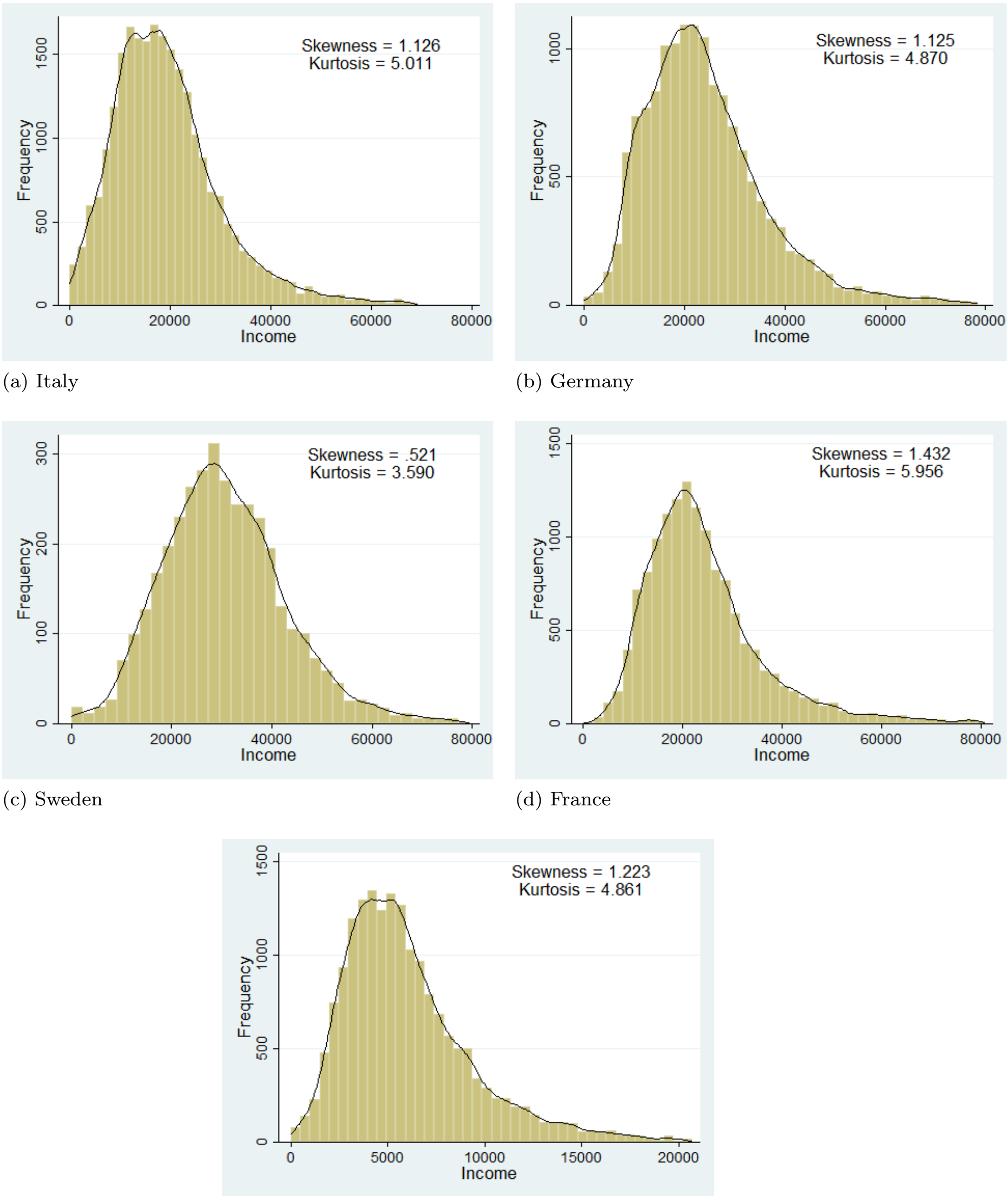
Table 3.3: Descriptive statistics of the well-being dimensions

| Country | Indicator[1] | M | SD | Min | Max |
|---|---|---|---|---|---|
| Italy (N = 24,099) | EDI (euro) | 19336 | 10735 | 4 | 69288 |
| | Years of education | 11.622 | 3.143 | 0 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.144 | 0.351 | 0 | 1 |
| | Good | 0.633 | 0.481 | 0 | 1 |
| | Fair | 0.165 | 0.371 | 0 | 1 |
| | Poor | 0.047 | 0.212 | 0 | 1 |
| | Very poor | 0.009 | 0.097 | 0 | 1 |
| Germany (N = 14,538) | EDI (euro) | 24310 | 12123 | 10 | 78579 |
| | Years of education | 13.922 | 2.169 | 4 | 16 |
| | Self-perceived general health | | | | |
| | Very good | 0.170 | 0.376 | 0 | 1 |
| | Good | 0.512 | 0.499 | 0 | 1 |
| | Fair | 0.240 | 0.427 | 0 | 1 |
| | Poor | 0.063 | 0.244 | 0 | 1 |
| | Very poor | 0.012 | 0.110 | 0 | 1 |
| Sweden (N = 3,654) | EDI (euro) | 30623 | 12270 | 14 | 79531 |
| | Years of education | 12.964 | 2.138 | 6 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.345 | 0.475 | 0 | 1 |
| | Good | 0.484 | 0.499 | 0 | 1 |
| | Fair | 0.134 | 0.341 | 0 | 1 |
| | Poor | 0.028 | 0.167 | 0 | 1 |
| | Very poor | 0.006 | 0.080 | 0 | 1 |
| France (N = 13,929) | EDI (euro) | 24396 | 11818 | 300 | 80900 |
| | Years of education | 12.005 | 3.009 | 0 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.236 | 0.424 | 0 | 1 |
| | Good | 0.485 | 0.499 | 0 | 1 |
| | Fair | 0.214 | 0.410 | 0 | 1 |
| | Poor | 0.056 | 0.230 | 0 | 1 |
| | Very poor | 0.006 | 0.081 | 0 | 1 |
| Poland (N = 17,747) | EDI (euro) | 6053 | 3330 | 17 | 20724 |
| | Years of education | 12.402 | 2.794 | 0 | 16 |
| | Self-perceived general health | | | | |
| | Very good | 0.142 | 0.349 | 0 | 1 |
| | Good | 0.476 | 0.499 | 0 | 1 |
| | Fair | 0.277 | 0.448 | 0 | 1 |
| | Poor | 0.089 | 0.285 | 0 | 1 |
| | Very poor | 0.012 | 0.113 | 0 | 1 |
| EU-28 (N = 270,108) | EDI (euro) | 15614 | 12868 | 3 | 118238 |
| | Years of education | 11.855 | 3.470 | 0 | 17 |
| | Self-perceived general health | | | | |
| | Very good | 0.233 | 0.423 | 0 | 1 |
| | Good | 0.481 | 0.499 | 0 | 1 |
| | Fair | 0.211 | 0.408 | 0 | 1 |
| | Poor | 0.060 | 0.238 | 0 | 1 |
| | Very poor | 0.012 | 0.112 | 0 | 1 |

*Note.* EDI = equivalised disposable income.

[1] The EU-SILC original variables involved are Total disposable household income (HY020), Highest ISCED level attained (PE040) and General health (PH010).

representative of a social democratic welfare state characterized by a pronounced redistributive role of taxes.




(a) Italy (b) Germany

(c) Sweden (d) France

(e) Poland

Figure 3.1: Distribution of equivalised disposable income by country. Black solid line on each graph shows kernel density

*Source:* authors' calculations based on data from the EU-SILC, 2015

The indicator of education is given by years of schooling (see Table 3.6 in Appendix), constructed in the following way: each individual is assigned a certain number of schooling years required for obtaining his highest educational level. Only those qualifications fall into a category of 'successfully completed' if an individual received a certificate after attending a program. Years of schooling associated with the same qualification may vary across countries.

Figure 3.2 displays the highest level of education in the countries under consideration. All certificates are classified into five groups to harmonize with the ISCED 2011. The first group comprises individuals with pre-primary and/or primary training. The nest group of educational attainments covers the lower secondary certificates holders. Upper secondary education (if further specified) includes general and vocational programmes, while last two groups distinguish between post-secondary further education and tertiary degrees.

The vast majority of respondents (from 42.1% to 62.7%) possess upper secondary qualification; Sweden is an exception with tertiary educational attainment as the largest category (41.1% of the sample). Germany (34%) has the next largest proportion of respondents with tertiary education level, followed by France (33.4%). The percentage of respondents, whose highest educational certificate corresponds to lower secondary programme, is the highest in Italy (28.5%), which is above the EU-28.

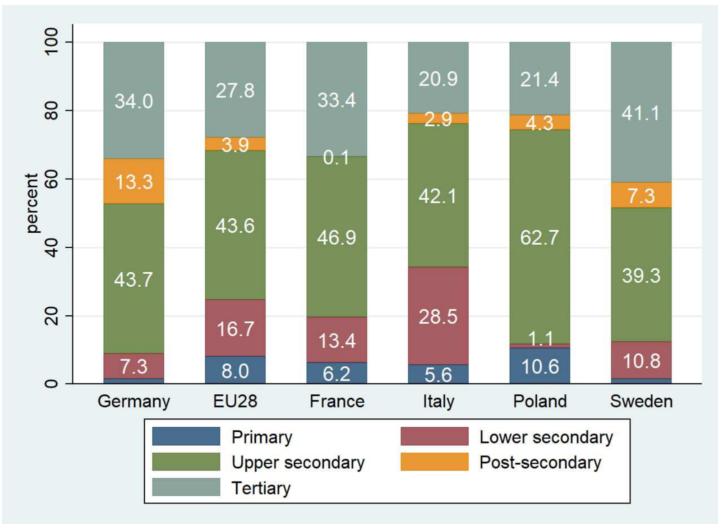



Figure 3.2: Stacked bar chart of the highest educational attainment by country, in percent
*Source:* authors' calculations based on data from the EU-SILC, 2015

The EU-SILC contains a set of questions on both subjective and objective measures of health. We limit our attention on the former, since the self-assessed health captures the general health status of the respondent and correlates with the objective measures of health (see, for example, Wu et al. (2013)). Our subjective health indicator includes five categories from very bad to very good. This subjective measure of health is intended to describe the respondent's health state in general regardless of current health problems or one's age.

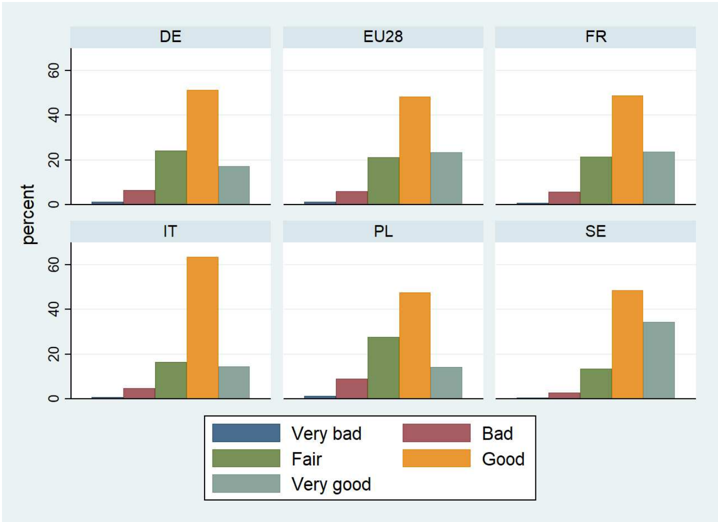



Figure 3.3: Bar chart of the self-perceived general health by country, in percent
*Source:* authors' calculations based on data from the EU-SILC, 2015

Figure 3.3 demonstrates the distribution of individuals over five subjective health categories. In the considered countries the largest share of respondents report their health status as good: from 47.6% in Poland to in 63.3% Italy. More than one third of sampled individuals in Sweden rate their health as excellent, while in the rest of the countries very good health is ranged between 14.4% and 23.6% of interviewees.

To give an initial insight into the interrelation between income and education variables, we compare the distribution of educational attainments over income groups across countries from the sample (see Figure 3.4). We allocate individual incomes across 5 quintiles, where the first quintile includes 20 percent of the population with the lowest income and the fifth quintile incorporates 20 percent of those with the highest income.

Figure 3.4 suggests that, as income goes up, the share of primary and secondary educational attainments decreases and they are substituted by post-secondary and tertiary education. The described pattern is observed across all countries indicating a certain degree of association between income and educational dimensions.

The distribution of self-assessed health status over income quintiles sheds light on the interrelation between subjective measure of health and the income variable. Figure

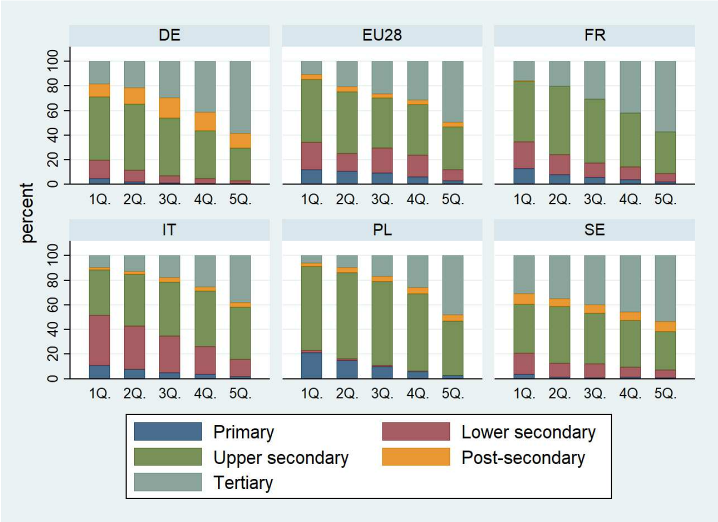



Figure 3.4: Stacked bar chart of educational attainment over income quintiles by country. 1Q. denotes the lowest quintiles, while 5Q. stays for the highest one
*Source:* authors' calculations based on data from the EU-SILC, 2015

3.5 illustrates how the respondents with different income levels assess their health. Our sample is again stratified by countries.

Remarkably, division of the respondents into income quintile groups leads to roughly identical distribution of subjective health in Italy. However, in the rest of the countries those with higher incomes report their health to be good or excellent more frequently.


Figure 3.5: Stacked bar chart of self-assessed health over income quintiles by country
*Source:* authors' calculations based on data from the EU-SILC, 2015

The relationship between self-assessed health and the highest education attainment is shown in Figure 3.6. In the European countries there is an inequality in the distribution of self-perceived health over educational levels. The largest share of citizens, who report good or excellent health, have obtained tertiary education. The most significant difference regarding subjective health is observed between primary and tertiary educational groups. This pattern is common for the selected countries as well as the EU as a whole.


Figure 3.6: Stacked bar chart of self-perceived general health over educational attainment by country. Each education level is represented by three-digit code according to the ISCED 2011
*Source:* authors' calculations based on data from the EU-SILC, 2015

### 3.4.2 Pairwise dependence estimation based on bivariate copulas

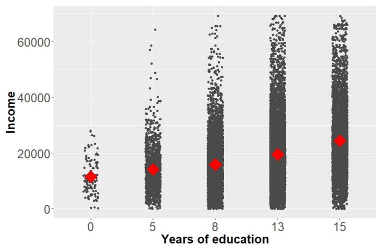
#### Income and Education

There is an emerging body of literature that examines the outcomes of schooling on both macro and micro levels. The role of education has been studied from different prospectives and economists commonly acknowledge monetary and non-monetary benefits of schooling. The level of education is usually approximated by the highest diploma or the total number of years of schooling. Although this approach does not shed light on the skills and more research is needed on it, the information on years of schooling is easily assessable and is still useful in the empirical studies.

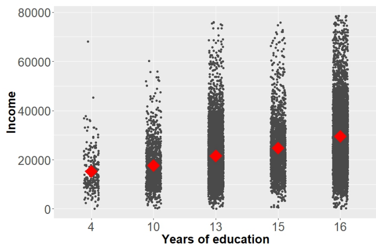
Education is commonly viewed as an investment into human capital done by individuals striving to receive higher earnings in the future. Thus, an essential benefit of education is a monetary one. However, the case of education-job mismatch is usually associated with an income penalty, but this problem lies beyond the scope of this paper.

As it was stressed previously, eduction is thought of as a sort of investment and the question of returns to education is of interest for both researchers and policy makers.

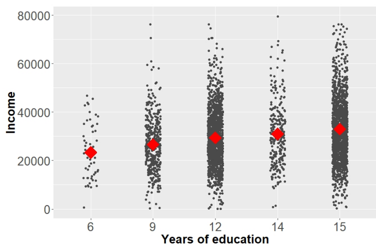
Therefore, the first pair of well-being dimensions we focus on is income-education. The highest education attainment considered in our sample is the tertiary one corresponding to 15-16 years of total duration of schooling. The concordance between income and years of education across countries is shown in Figure 3.7. 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 3.7 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 ($M$ = 24402.94 and $M$ = 19528.35 of tertiary and post-secondary educational groups respectively against $M$ = 11543.64 of group without any educational attainment in Italy). Sweden is an exception since the association between the income and education variables is the lowest.



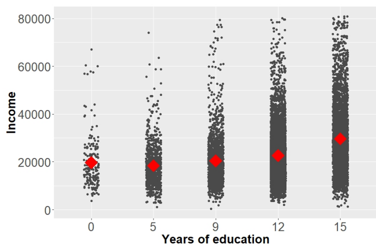

(a) Italy



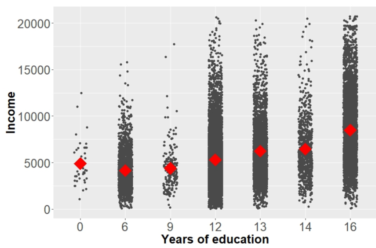

(b) Germany




(c) Sweden



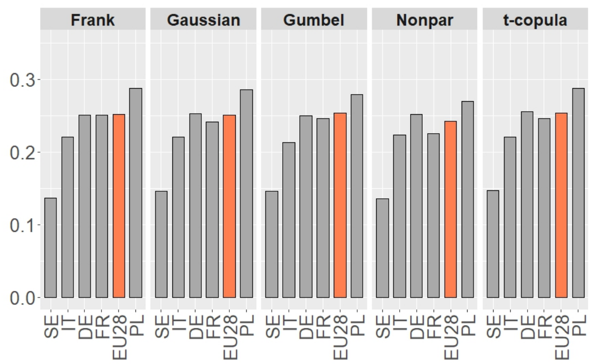

(d) France



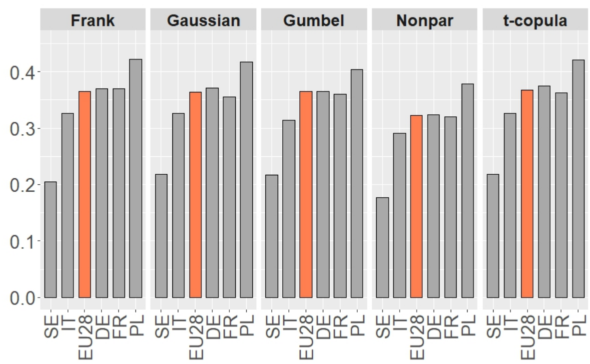

(e) Poland

Figure 3.7: Scatterplot of income versus years of education. 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. The resulting estimates of both coefficients are given in Figure 3.8 and are grouped according to the underlying copula function: four parametric copula families and an empirical copula. All correlation coefficients between income and education are statistically significant (see Table 3.7 in Appendix). The average dependence across 28 countries of the European Union (red bar in the figure) serves as a benchmark. To obtain this benchmark we compute the dependence in each country of the European Union and report a mean of the obtained correlation coefficients.




(a) Kendall's $\tau_K$




(b) Spearman's $\rho_S$

Figure 3.8: Copula-based correlation coefficients for income and education dimensions
*Source:* authors' calculations based on the EU-SILC, 2015.

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 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.

The ranking of countries is not preserved across five copula functions. These fluctuations stem from each copula's underlying assumptions regarding the form of distribution and the type of dependence captured. While for Sweden both Gaussian and Student's t-copula are better reflecting the symmetry of distribution and symmetric dependence in tails, for the rest of the countries Frank copula better captures the existing dependence concentrated in the middle of the distribution.

Two correlation coefficients do not report the same magnitude of dependence. It originates from the construction of each coefficient and the corresponding reference point used. Thus, Spearman's $\rho_S$ coefficient is interpreted as the dependence in a given society with respect to the independent one, while the reference of Kendall's $\tau_K$ is a society with the same level of dependence as in the given one. Consequently, the correlation reported by Spearman's $\rho_S$ is higher than the one estimated by Kendall's $\tau_K$.

As it was highlighted previously education is associated with both pecuniary and nonpecuniary gains. We will contribute to the discussion of nonpecuniary benefits when analysing the interrelationship between education and health. Here, instead, we focused on pecuniary gains created by education.

It is interesting to compare our results with the existing empirical studies. One of the stylized facts on education is that returns to schooling are positive and tend to be higher in developing countries compared to the developed ones (see Carneiro et al., 2011; Montenegro and Patrinos, 2014; Psacharopoulos and Patrinos, 2018 for other stylized facts). The research on causal effects of education on wages started from the fundamental work of Mincer (1974). Recent studies report significant positive causal effect of education on income (see, for example, Blundell et al., 2005; Oreopoulos and Petronijevic, 2013; Heckman et al., 2018). Additionally, there is a growing evidence in literature that income inequality and educational inequality positively correlate (Rodríguez-Pose and Tselios, 2008; Coady and Dizioli, 2017). Synthesizing the empirical evidence, income and education are positively associated dimensions of well-being. Hence, our results on positive correlation between educational attainment and income across European countries are in line with the findings of recent literature.

#### Income and Health

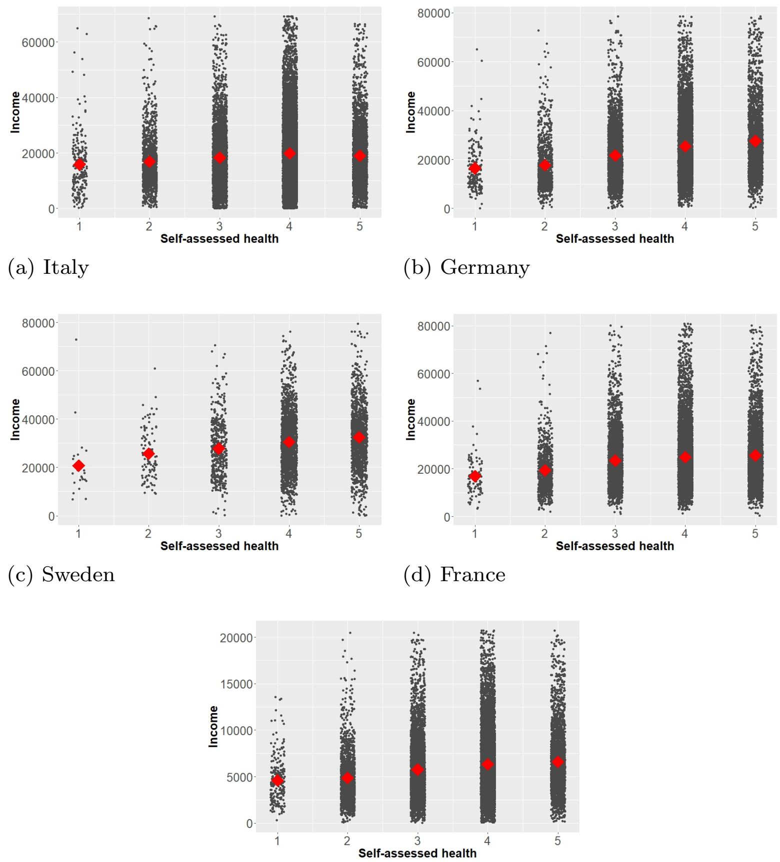
Individuals with higher income on average are healthier than their peers with lower earnings. The presence of inequalities in health caused by socio-economic determinants, with income as one of the most important factors, is called the social gradient in health. Albeit intuitive, the relationship between income and health is complex and the causality between the two variables can have either direction. However, the role of health selection, i.e. the situation when individuals with poor health are trapped into lower income quintiles, is suggested to be limited, since it reduces the association between income and health but does not eliminate the last (Benzeval and Judge, 2001). In our analysis the presence of causality between well-being dimensions is not essential. Instead, we are focused on the association between the dimensions and its magnitude.

It is worth noting that the EU-SILC is particularly rich in collecting different sources of income on household's and individual levels. Different measures of income have similar effects on health, validating our choice of a household's size adjusted disposable income as the indicator (for the comparison of different income measures in the context of health see Geyer (2011)).

The scatterplot displayed in Figure 3.9 illustrates the concordance between equivalised disposable income and self-assessed health. In Italy and Poland the association is low, 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 3.10. Correlation significance tests for Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients are given in Table 3.8 in Appendix. This pairwise dependence is generally lower than the one between income and education for all countries. In Germany, where correlation is the highest, 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. This result may be interpreted in the light of welfare system in health prevailing in each country. Thus, in Germany welfare state retrenchment that occurred in the last decades has increased health inequalities (Siegel et al., 2014), while in Italy the existing welfare regime may have a protective effect on health of the citizens.

The estimated magnitude of the dependence between income and health is higher for Spearman's $\rho_S$ than for Kendall's $\tau_K$ regardless of the underlying copula. Taking the example of Gaussian copula function, in Germany the interdependence between income and subjectively assessed health is around 0.26 as measured by Spearman's $\rho_S$ and is



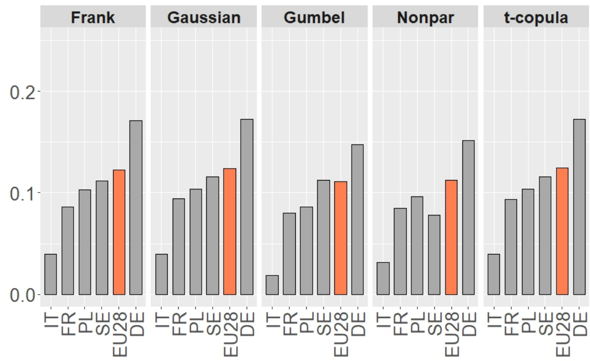

(a) Italy (b) Germany (c) Sweden (d) France (e) Poland

Figure 3.9: Scatterplot of income versus health status. Red diamonds represent mean of income for each health status

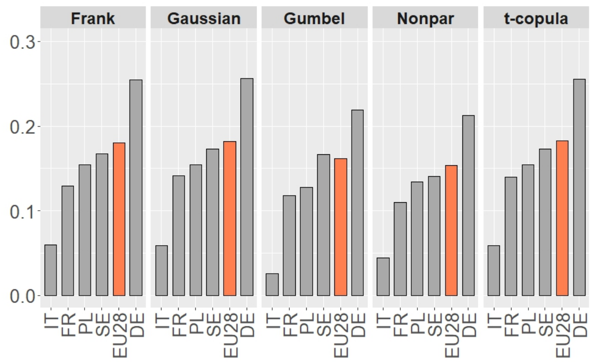
approximately 0.17 for Kendall's $\tau_K$. These copula-based correlation coefficients can differ in magnitude because they reflect different patterns of dependence.

Although the estimated correlation does not entirely overlap across copulas, the ranking of countries given by both correlation coefficients is sufficiently similar. The dissimilarities occurred in case of Gumbel copula suggest that the correlation between two dimensions does not reflect extreme values. Instead, the association in both tails is symmetric as captured by Gaussian and Student's copula functions. Some discrepancy is observed in the ranking based on empirical copula, especially for Kendall's tau correlation coefficient. A possible explanation for this fluctuation in ranking is an additional source of uncertainty, since the original population is compared with the reference one whose ranks are generated randomly.

This subsection contributes to the understanding of the relationship between income and health and complements previous studies on the positive association between these




(a) Kendall's $\tau_K$




(b) Spearman's $\rho_S$

Figure 3.10: Copula-based correlation coefficients for income and self-assessed health
*Source:* authors' calculations based on the EU-SILC, 2015.

dimensions of well-being (Benzeval and Judge, 2001; Furnée et al., 2011; Karlsdotter et al., 2012; Detollenaere et al., 2018). Although positive relationship between income and health can be partially explained by the employment status due to health selection (Stronks et al., 1997), this intermediate link between employment and health does not fully capture the association between income and health. The evidence from other studies suggests that income inequality is negatively associated with population health (Pickett and Wilkinson, 2015; Baeten et al., 2013; Rözer and Volker, 2016).

The evidence of positive association between income and health naturally brings the discussion to the next step, namely the relation between the lack of income and health. Intuitively, if better-off individuals report better self-perceived health, then citizens whose income falls below a poverty threshold are more vulnerable in terms of their health. As suggested by the report of European Commission (2013), there is a strong negative association between material deprivation[4] and health in the European countries. In light

[4]Material deprivation is an indicator in EU-SILC that measures a share of population not able to

of the facts emphasized previously, we are able to conclude that material factors matter for one's health status. A possible channel that links insufficient income with health is unmet medical need, i.e. the situation when individual has to undergo a medical examination but refuses to receive the treatment because he is not able to afford it. Undoubtedly, the effect of income poverty on health can be mitigated by the universal access to a national healthcare system, but the research of the mechanism lies beyond the scope of present Chapter.

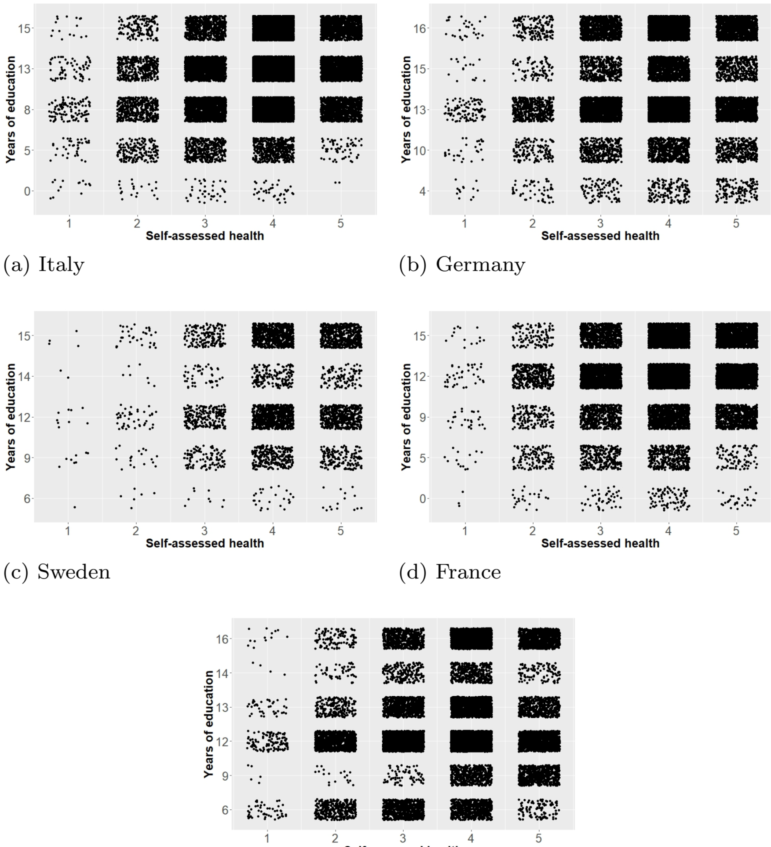
**Education and Health**

Education as an investment into human capital is associated with consequent returns in the form of higher earnings. This view of financial benefits accompanying education was discussed before in the section about the dependence between income and education. However, schooling does not bring only monetary benefits, it also has nonpecuniary effects on one's life including higher job and life satisfaction, lower risk of being unemployed as well as better health status and longevity (Oreopoulos and Salvanes, 2011). The positive relation between education and health is a well-established fact in economic literature. Economists widely acknowledge that not only healthcare coverage but also schooling plays a role for health. Yet there is no consensus in literature regarding a causal role of education in reducing the probability of adverse health behaviours and some studies establish a link between schooling and participation in health-related behaviours (Cutler and Lleras-Muney, 2010; Böckerman and Maczulskij, 2016), whereas others do not report a causal relationship between the two (Tenn et al., 2010; Kemptner et al., 2011).

The relationship between education and health across the European countries is shown in Figure 3.11. The subjective health status is plotted against 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 especially in Poland, Italy, France and Germany. In Sweden respondents tend to report their health status as excellent or very good regardless of their educational attainment.

Correlations significance tests for both correlation coefficients are provided in Table 3.9 in Appendix. 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

afford three out of nine items. For a more detailed description of indicator and a full list of items see European Commission (2013).

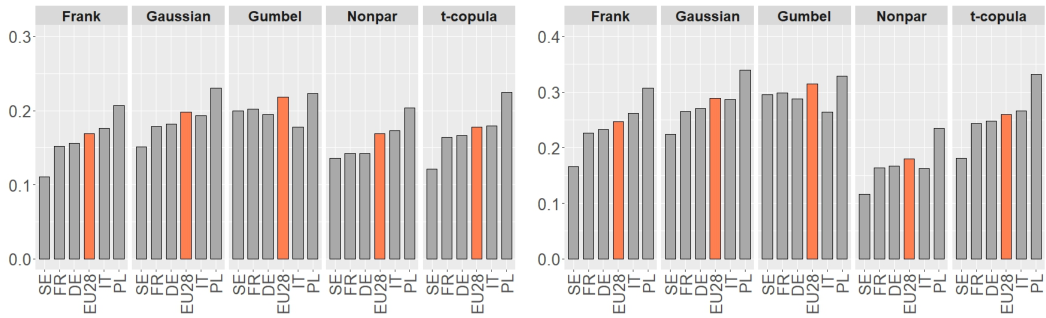



(a) Italy (b) Germany

(c) Sweden (d) France

(e) Poland

Figure 3.11: Scatterplot of education versus health status




(a) Kendall's $\tau_K$ (b) Spearman's $\rho_S$

Figure 3.12: Copula-based correlation coefficients for education and self-assessed health
*Source:* authors' calculations based on the EU-SILC, 2015.

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.

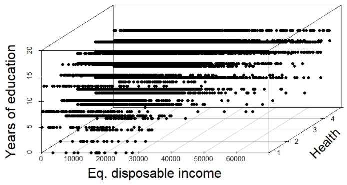
Albeit the majority of copula families allow establishing a stable ranking with Poland and Italy above the EU average benchmark and Germany, France and Sweden below it, some disparities in ranking occur. The strongest disparity is observed for Gumbel copula. A similar trend is observed also for previous pairs of well-being dimensions. A plausible explanation of this mismatch of rankings is a sensitivity of Gumbel copula to extreme values in the upper tail of the distribution. Thus, a country with the highest dependence (described by a symmetric correlation in tails) does not necessarily have the strongest association in the upper tail as captured by Gumbel copula.

In this subsection we investigate the relationship between education and health using copula-based measures of dependence. The results suggest that years of schooling and self-perceived health are positively associated and the magnitude of this association varies across countries. A number of studies emphases that not only cognitive skills contributed to this association, but also personality traits tend to play an important role. Yet the exact mechanism behind the education-health gradient is beyond the scope of present paper, because for our purposes only the magnitude of the correlation is crucial. Our results are consistent with the empirical evidence found in recent literature on the positive association between education and health for the UK (Conti et al., 2010; Conti and Hansman, 2013), the Netherlands (Bijwaard et al., 2015), Finland (Böckerman and Maczulskij, 2016) and the US data (Cutler and Lleras-Muney, 2010).

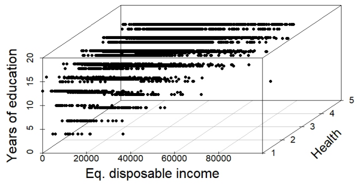
### 3.4.3 Overall dependence estimation using empirical copula

Previous subsections have done an extensive analysis of the dependence between all pairs of well-being attributes. However, bivariate aspects of dependency do not apply directly to the overall interrelation among variables. Moreover, bivariate and multivariate dependence patterns may be contradictory (Pèrez and Prieto-Alaiz, 2016). Therefore, multivariate extensions of copula-based measures are necessary for giving a comprehensive picture of interrelation among individual positions across attributes.

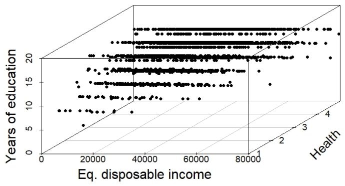
We aim at understanding the overall interdependence among income, education and health. Figure 3.13 illustrates achievements of all respondents in the selected dimensions. As it appears in Figure, the association between these indicators is positive in all countries. In other words, individuals with higher earnings are also healthier and better educated, occupying higher rank positions in all indicators than their peers with lower income. Performances in three attributes are more concentrated in Germany and Poland, whereas in France and Sweden the multivariate distribution of well-being achievements is more dispersed.



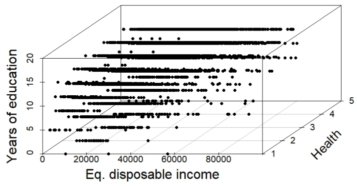

(a) Italy



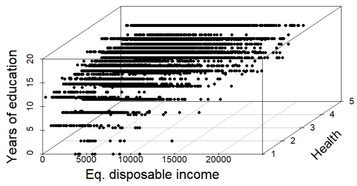

(b) Germany




(c) Sweden




(d) France




(e) Poland

Figure 3.13: The distribution of individual achievements across three dimensions of well-being by country

Copula-based correlation coefficients are flexible dependence measures primarily because they assess the dependency among rank-transformed variables. It enables modelling more general types of dependence and excludes oversensitivity to outliers as in case of Pearson linear correlation. Furthermore, there exist multivariate extensions of Spearman's $\rho_S$ and Kendall's $\tau_K$ making them particularly useful in multidimensional well-being context.

We use multivariate versions of these copula-based measures from (3.10)-(3.11) and estimate the global dependence among "core" dimensions of well-being, namely income, education and health, in 2015. The estimator is based on the empirical copula function. We perform 500 bootstrap replicates of nonparametric estimation of both indices and report mean of these replicates as well as 95% confidence intervals in Table 3.4.

Table 3.4: The overall dependence among dimensions measured by multidimensional Spearman's $\rho_S$ and Kendall's $\tau_K$ correlation coefficients

| | Spearman's $\rho_S$[1] | 95% CI (bootstrap) | Kendall's $\tau_K$[1] | 95% CI (bootstrap) |
|---|---|---|---|---|
| Italy | 0.168 | [0.153, 0.186] | 0.141 | [0.135, 0.149] |
| Germany | 0.239 | [0.214, 0.264] | 0.193 | [0.177, 0.197] |
| Sweden | 0.149 | [0.108, 0.194] | 0.127 | [0.105, 0.144] |
| France | 0.200 | [0.178, 0.224] | 0.162 | [0.142, 0.163] |
| Poland | 0.248 | [0.226, 0.268] | 0.196 | [0.194, 0.212] |

*Note.* CI = confidence interval.
[1] Mean of 500 nonparametric bootstrap replicates.

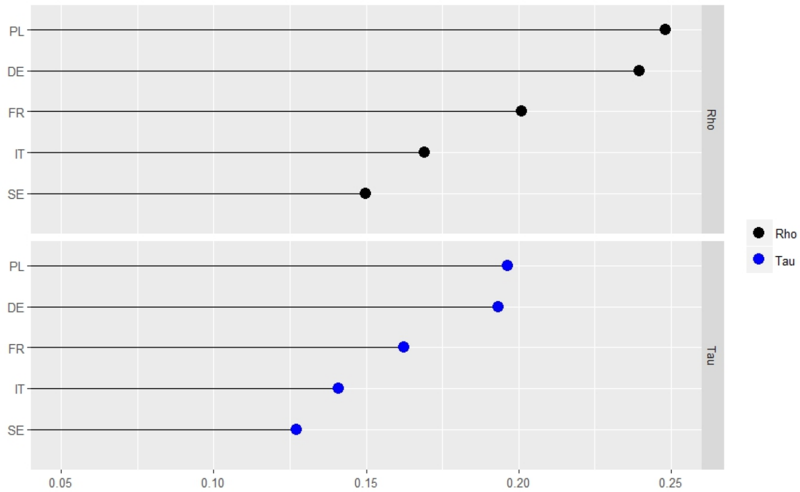



Figure 3.14: The overall dependence among dimensions estimated by Spearman's $\rho_S$ and Kendall's $\tau_K$ correlation coefficients
*Source:* authors' calculations based on the EU-SILC, 2015.

The resulting ranking of countries according to the overall dependence among key attributes is shown in Figure 3.14. With respect to Spearman's $\rho_S$ coefficient, the highest overall dependence is observed in Poland with coefficient at around 0.25, followed by Germany, where global interdependence is approximately 0.24. France appears in the middle of the ranking, while Italy and Sweden have the lowest dependence in the sample - around 0.17 and 0.15 respectively. A comparable ranking is obtained from multivariate Kendall's $\tau_K$ coefficient even though these measures are grasping non-identical aspects of dependency. According to the results, Poland and Germany are again top-ranked with coefficients slightly below 0.20. Finally, France outperforms Sweden and Italy in term of the overall dependence with the coefficient around 0.16.

## 3.5 The evolution of dependence among well-being attributes

To complement the results on bivariate and multivariate dependency patterns in the European countries obtained in previous sections, we enhance the analysis by covering a longer time interval. The application of copula-based measures using recent data from the EU-SILC revealed that individual performances in key well-being dimensions are associated. For instance, high-income earners enjoy higher educational level and better subjectively assessed health.

These findings are well-documented in the literature. Yet, to the best of our knowledge, the time evolution of interdependence among well-being indicators has not been extensively considered. A work by Pèrez and Prieto-Alaiz (2016) is a prominent example of copula application to the multidimensional well-being with an extensive study of the evolution of dependence among welfare components in the world. Mentioned paper investigates the evolution of multivariate dependence in the world economy, meaning that the overall pattern of dimensional association is build upon a heterogeneous sample of countries. Naturally each country's institutions and public sector policy (i.e. related to education coverage, health care services etc.) shape the interrelation among well-being domains. Therefore, it is worth identifying country-specific features of dependence evolution on the example of European countries.

Along with a scarce contribution on the evolution of dependence among dimensions, another motivation for considering an extended time interval is the recent economic and financial crisis. Its consequences adversely affected the global economy causing a downward trend in the economic growth. Together with unemployment and the increased sovereign debt as immediate consequences of the financial crisis of 2008, we believe that its negative impact for well-being is linked to an increased interdependence among key attributes.

To disentangle the effect of financial crisis it is necessary to evaluate the dimensional dependency in pre- and post-crisis periods. The beginning of this financial crisis is commonly associated with the bankruptcy of Lehman Brothers in September 2008. Consequently, to evaluate its welfare effects we need to consider a period prior to the collapse. In particular, we choose year 2006 as a pre-crisis reference point for dependence. Our post-crisis indicator is year 2010, when individual achievements should have adjusted to new conditions.

We begin with the evolution of the HDI in the selected countries. The HDI is a composite indicator that aggregates information in three well-being dimensions: living standard, knowledge and health. Each dimension is represented by its indicator(s): gross national income per capita is a proxy for a standard of living; education is reflected by two sub-pillars, namely by mean of years of schooling for adults aged 25 years or older and by expected years of education for children; finally, life expectancy at birth is an indicator of health.

The unit of analysis is specified at country level. Each dimensional index is normalized on the interval $[0, 1]$, according to the following equation (UNDP, 2016):

$$\text{Dimensional indicator} = \frac{\text{actual value} - \text{minimum value}}{\text{maximum value} - \text{minimum value}} \tag{3.20}$$

where minimum and maximum values are the so called "natural zeros" and "aspirational goals" respectively. For instance, according to the methodology of the United Nations Development Programme, a minimum value of life expectancy is fixed at 20 years, while its "aspirational target" is 85 years of age. Since education is approximated by two sub-indices, an arithmetic mean of the two is taken, which is a proxy for knowledge. Three indicators are aggregated using a multiplicative method, in particular, the geometric mean of three pillars.

The evolution of the HDI in five European countries over almost a decade is demonstrated in Figure 3.15, where red dot indicates mean, while a black solid line shows median.

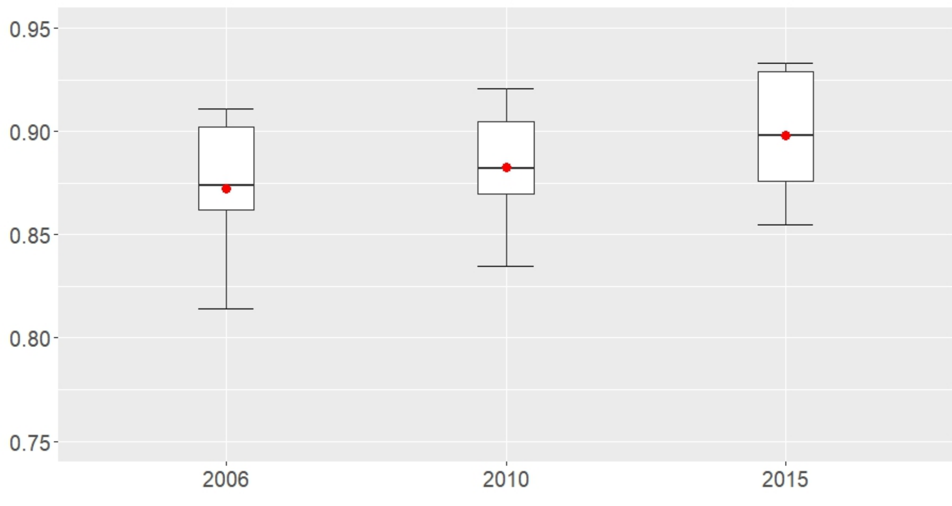



Figure 3.15: The evolution of the Human Development Indicator in 2006, 2010 and 2015
*Source:* authors' calculations based on the Human Development Data from UNDP, 2006-2015.

As it appears from the graph, the level of human development is very high in the selected countries. In pre-crisis period all countries were above 0.80 in terms of the HDI. In the following years a mean value gradually increased, achieving its maximum in 2015. Although the overall well-being improved during the period under consideration as summarized by the composite indicator, its dimensions might have followed diverse development trajectories across countries.

We are interested in the interdependence among the underlying dimensions of the HDI and its evolution during the comparable period. In this section we extend our application of copula-based measures of dependence by prolonging a considered time span. Since EU-SILC data is being produced annually, it grants an opportunity to monitor the time change of individual achievements. The variables we are considering in our analysis, i.e. total disposable household income, the highest educational attainment and self-perceived health, are collected yearly providing a basis for comparing the evolution of association among welfare variables over time. The descriptive statistics of three well-being indicators in 2010 and 2006 is summarized in Tables 3.10-3.11 in Appendix.

Figure 3.16 illustrates the evolution of the overall dependence between income, education and health as measured by multivariate Spearman's $\rho_S$ coefficient. Notably, each country followed a unique path of the global dependence among attributes. In the selected states the overall interrelation remained positive during the observed time interval.

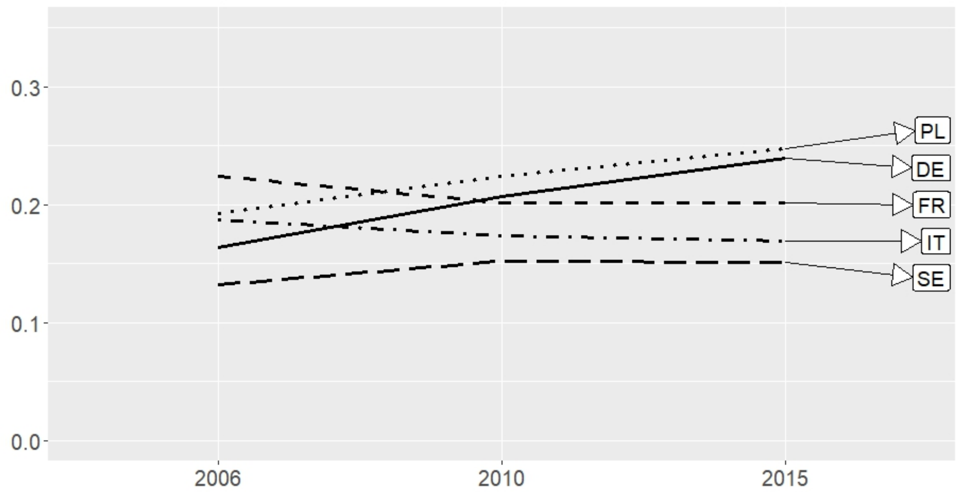



Figure 3.16: The evolution of global dependence between education and health measured by multivariate copula-based Spearman's $\rho_S$ in Italy (dot-dash line), Germany (solid line), Sweden (long-dash line), France (dashed line) and Poland (dotted line)
*Source:* authors' calculations based on the EU-SILC, 2006-2015.

Although in the pre-crisis period the highest overall dependency was observed in France, in the post-crisis period the association among well-being indicators had a decreasing trend there and stabilized in 2015 at the level of 0.20. Unlike France, Germany, Poland and Sweden underwent a raising interconnectedness among key attributes. In other words, in the mentioned states the post-crisis period is characterized by a higher coordination of individual positions in three dimensions. The evolution of the overall dependence brings us to an important conclusion. Despite a clear increase of the welfare as measured by the HDI, the achievements in the underlying attributes remain positively dependent in all countries. Moreover, the financial crisis facilitated this dependence, especially in Poland and Germany.

To extend these results, we additionally focus on the bivariate dependence among all pairs of attributes. Figure 3.17 depicts the development of pairwise association between income and education from 2006 until 2015. Besides France, earnings and years of schooling became more dependent in all countries. The association between these indicators grew sharply in Poland after 2010 and reached the value of almost 0.40. Overall, the interdependence between income and knowledge, the latter approximated by years of schooling, remains the highest compared to the other pairs.

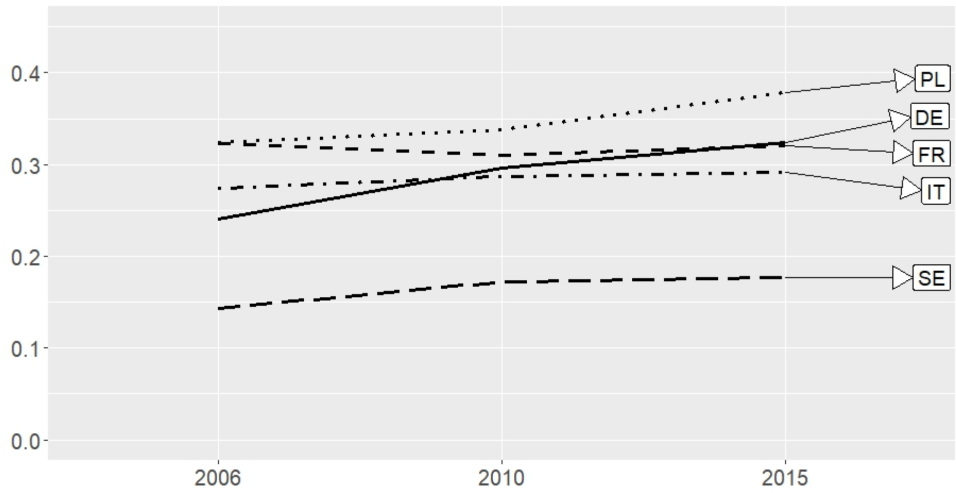



Figure 3.17: The evolution of dependence between income and education measured by Spearman's $\rho_S$ in Italy (dot-dash line), Germany (solid line), Sweden (long-dash line), France (dashed line) and Poland (dotted line)
*Source:* authors' calculations based on the EU-SILC, 2006-2015.

The evolution of dependency between the next pair of attributes, namely income and health, is shown in Figure 3.18. These dimensions were slightly positively associated in Italy, whereas in the rest of countries they became more interrelated in 2015. This finding signalizes that citizens, who occupy higher ranks in income, are expected to report better health status.

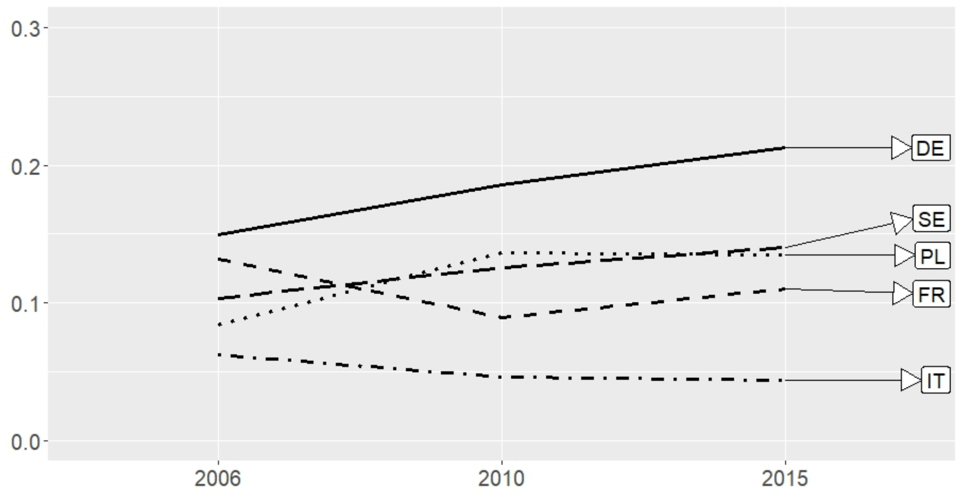



Figure 3.18: The evolution of dependence between income and health measured by copula-based Spearman's $\rho_S$ in Italy (dot-dash line), Germany (solid line), Sweden (long-dash line), France (dashed line) and Poland (dotted line)
*Source:* authors' calculations based on the EU-SILC, 2006-2015.

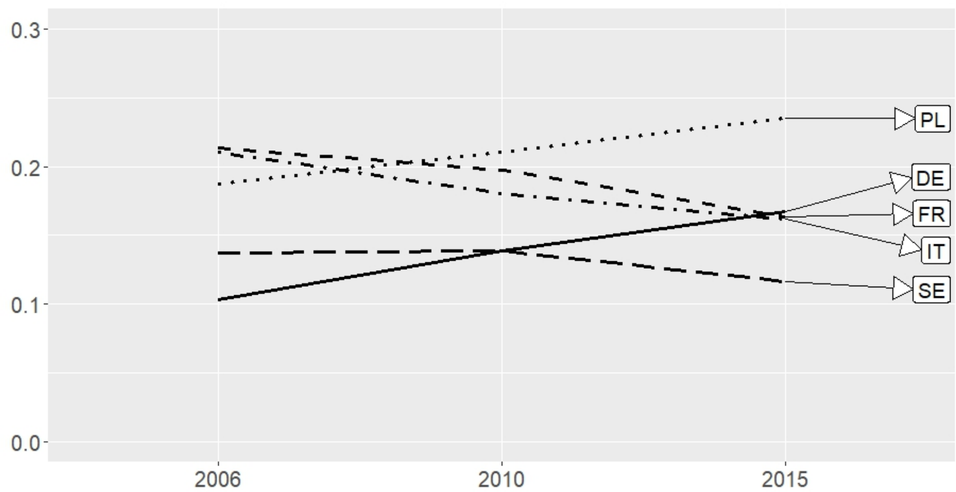



Figure 3.19: The evolution of dependence between education and health measured by copula-based Spearman's $\rho_S$ in Italy (dot-dash line), Germany (solid line), Sweden (long-dash line), France (dashed line) and Poland (dotted line)
*Source:* authors' calculations based on the EU-SILC, 2006-2015.

The last pair of attributes is the one of education and health. As estimated by Spearman's $\rho_S$, during the specified time interval performances in these attributes remained being harmonized in a sense that higher achievements in both dimensions tend to occur simultaneously. However, each country had a specific evolution pattern: the interrelation between years of schooling and self-perceived health strengthened in Poland and Germany, while in the remaining nations this pairwise dependence lessened.

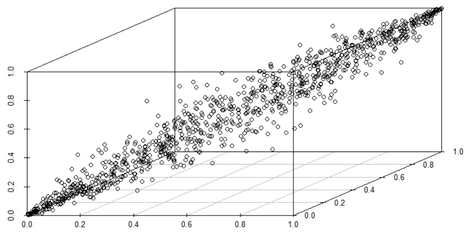
We conclude this section with a brief summary on how the bivariate and multivariate dependency evolved over nearly a decade. The financial crisis had a clear-cut effect on the relationship between earnings and educational level, causing a stronger linkage between the two. The rest of pairs remained positively dependent, although in each country fluctuations of dimensional dependency were observed.

## 3.6 Concluding remarks

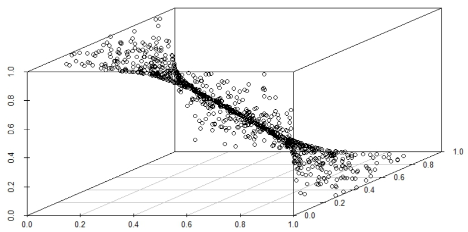
Accepting the common notion that well-being is a multidimensional phenomenon, in this paper we study the interdependence among its "core" attributes in the European countries. Since well-being data is often represented by ordinal variables rather than continuous ones, we explore the dependency patterns by applying flexible copula-based statistical tools. In particular, we use Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients that apply rank transformation to the original variables allowing more general types of association.

Our empirical study contains an application of bivariate and multivariate copula-based dependence measures to the selected European countries using the data from the EU-SILC. Main findings on the reference year 2015 are complemented by the additional estimations for pre- and post-crisis time intervals. The results suggest that key well-being dimensions are positively interconnected as estimated by bivariate and multivariate copula-based coefficients. Moreover, various evolution patterns were identified, indicating that the interconnectedness among attributes fluctuated unequally due to the financial crisis and is country-specific. Finally, despite a pronounced increase in human development during 2006-2015, income, education and health continue being positively interdependent. This result provides an evidence that human capabilities might be unequally distributed within societies, so that high-income earners tend to be better-off in both education and health.

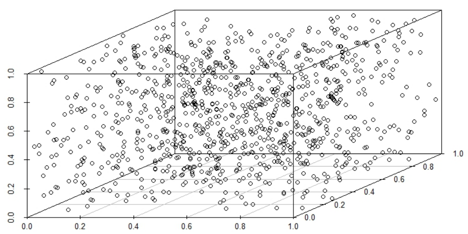
## 3.7 Appendix



(a)



(b)



(c)

Figure 3.20: A simulated population of 1000 individuals from a society with positive (a), negative (b) and independent (c) distribution of ranks across three well-being dimensions

Table 3.5: The Dependence Between Income and Education Measured by Spearman's $\rho_S$ Coefficient: the Role of Top Income Earners

| Country | Gaussian | t-copula | Frank | Gumbel | Empirical |
|---|---|---|---|---|---|
| Italy | | | | | |
| Without (N = 24,099) | 0.33 | 0.33 | 0.33 | 0.31 | 0.29 |
| With (N = 24,340) | 0.33 | 0.33 | 0.33 | 0.32 | 0.30 |
| Germany | | | | | |
| Without (N = 14,538) | 0.37 | 0.38 | 0.37 | 0.37 | 0.32 |
| With (N = 14,684) | 0.38 | 0.38 | 0.38 | 0.37 | 0.33 |
| Sweden | | | | | |
| Without (N = 3,654) | 0.22 | 0.22 | 0.20 | 0.22 | 0.18 |
| With (N = 3,690) | 0.23 | 0.23 | 0.21 | 0.22 | 0.18 |
| France | | | | | |
| Without (N = 13,929) | 0.36 | 0.36 | 0.37 | 0.36 | 0.32 |
| With (N = 14,069) | 0.36 | 0.37 | 0.37 | 0.36 | 0.32 |
| Poland | | | | | |
| Without (N = 17,747) | 0.42 | 0.42 | 0.42 | 0.40 | 0.38 |
| With (N = 17,925) | 0.43 | 0.43 | 0.43 | 0.41 | 0.39 |

Table 3.6: The Number of Schooling Years Corresponding to Each Educational Attainment

| | | Italy | | Germany | | Sweden | | France | | Poland | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Level | ISCED | YRS | Cum. | YRS | Cum. | YRS | Cum. | YRS | Cum. | YRS | Cum. |
| Primary | 100 | 5 | 5 | 4 | 4 | 6 | 6 | 5 | 5 | 6 | 6 |
| Lower secondary | 200 | 3 | 8 | 5-6 | 9-10 | 3 | 9 | 4 | 9 | 3 | 9 |
| Upper secondary | 300 | 5 | 13 | 3 | 12-13 | 3 | 12 | 3 | 12 | 3 | 12 |
| Post-secondary | 400 | 2 | 15 | 2 | 14-15 | 1-2 | 13-14 | 1-2 | 13-14 | 1-2 | 13-14 |
| Tertiary | 500 | 2 | 15 | 3 | 16 | 2-3 | 14-15 | 2-3 | 14-15 | 3 | 15-16 |
| Beginning | 100 | 6 | | 6 | | 7 | | 6 | | 7 | |
| Compulsory | - | 10 | | 12 | | 9 | | 10 | | 9 | |

*Note.* ISCED = International Standard Classification of Education; YRS = total duration of educational level in years; Cum. = cumulative duration of schooling in years; Beginning - age at the start of compulsory education; Compulsory = duration of compulsory schooling in years.
*Sources:* UNESCO (2012a,b).

Table 3.7: Correlation Significance Tests for Spearman's $\rho_S$ and Kendall's $\tau_K$ Coefficients Estimated Between Income and Years of Education, 2015

| | Gaussian | | | t-copula | | | Frank | | | Gumbel | | | Empirical | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI |
| | | | | | | | Spearman's $\rho_S$ | | | | | | | | |
| Italy | 0.33 | 53.46 ($p$<.001) | 0.31-0.34 | 0.33 | 53.65 ($p$<.001) | 0.31-0.34 | 0.33 | 53.57 ($p$<.001) | 0.31-0.34 | 0.31 | 51.41 ($p$<.001) | 0.30-0.33 | 0.29 | 47.30 ($p$<.001) | 0.28-0.30 |
| Germany | 0.37 | 48.19 ($p$<.001) | 0.36-0.39 | 0.38 | 48.81 ($p$<.001) | 0.36-0.39 | 0.37 | 47.99 ($p$<.001) | 0.36-0.38 | 0.37 | 47.30 ($p$<.001) | 0.35-0.38 | 0.32 | 41.28 ($p$<.001) | 0.31-0.34 |
| Sweden | 0.22 | 13.50 ($p$<.001) | 0.19-0.25 | 0.22 | 13.53 ($p$<.001) | 0.19-0.25 | 0.20 | 12.63 ($p$<.001) | 0.17-0.24 | 0.22 | 13.41 ($p$<.001) | 0.19-0.25 | 0.18 | 10.86 ($p$<.001) | 0.15-0.21 |
| France | 0.36 | 44.85 ($p$<.001) | 0.34-0.37 | 0.36 | 45.94 ($p$<.001) | 0.35-0.38 | 0.37 | 47.01 ($p$<.001) | 0.36-0.38 | 0.36 | 45.60 ($p$<.001) | 0.35-0.37 | 0.32 | 39.92 ($p$<.001) | 0.31-0.34 |
| Poland | 0.42 | 61.19 ($p$<.001) | 0.40-0.43 | 0.42 | 61.72 ($p$<.001) | 0.41-0.43 | 0.42 | 61.97 ($p$<.001) | 0.41-0.43 | 0.40 | 58.76 ($p$<.001) | 0.39-0.42 | 0.38 | 54.45 ($p$<.001) | 0.37-0.39 |
| | | | | | | | Kendall's $\tau_K$ | | | | | | | | |
| Italy | 0.22 | 35.08 ($p$<.001) | 0.21-0.23 | 0.22 | 35.20 ($p$<.001) | 0.21-0.23 | 0.22 | 35.08 ($p$<.001) | 0.21-0.23 | 0.21 | 33.88 ($p$<.001) | 0.20-0.22 | 0.20 | 32.23 ($p$<.001) | 0.19-0.22 |
| Germany | 0.25 | 31.45 ($p$<.001) | 0.24-0.27 | 0.26 | 31.84 ($p$<.001) | 0.24-0.27 | 0.25 | 31.25 ($p$<.001) | 0.24-0.27 | 0.25 | 31.13 ($p$<.001) | 0.23-0.27 | 0.25 | 30.61 ($p$<.001) | 0.23-0.26 |
| Sweden | 0.15 | 8.94 ($p$<.001) | 0.11-0.18 | 0.15 | 8.96 ($p$<.001) | 0.12-0.18 | 0.14 | 8.36 ($p$<.001) | 0.11-0.17 | 0.15 | 8.92 ($p$<.001) | 0.11-0.18 | 0.12 | 7.39 ($p$<.001) | 0.09-0.15 |
| France | 0.24 | 29.33 ($p$<.001) | 0.23-0.26 | 0.25 | 30.02 ($p$<.001) | 0.23-0.26 | 0.25 | 30.61 ($p$<.001) | 0.24-0.27 | 0.25 | 30.01 ($p$<.001) | 0.23-0.26 | 0.24 | 29.63 ($p$<.001) | 0.23-0.26 |
| Poland | 0.29 | 39.69 ($p$<.001) | 0.27-0.30 | 0.29 | 40.01 ($p$<.001) | 0.27-0.30 | 0.29 | 40.04 ($p$<.001) | 0.27-0.30 | 0.28 | 38.68 ($p$<.001) | 0.26-0.29 | 0.28 | 38.71 ($p$<.001) | 0.26-0.29 |

*Note.* Coef. = correlation coefficient; CI = confidence interval.
$p$-values are reported in parentheses.

Table 3.8: Correlation Significance Tests for Spearman's $\rho_S$ and Kendall's $\tau_K$ Coefficients Estimated Between Income and Self-Assessed Health, 2015

| | Gaussian | | | t-copula | | | Frank | | | Gumbel | | | Empirical | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI |
| | | | | | | | Spearman's $\rho_S$ | | | | | | | | |
| Italy | 0.06 | 9.18 ($p$<.001) | 0.05-0.07 | 0.06 | 9.17 ($p$<.001) | 0.05-0.07 | 0.06 | 9.25 ($p$<.001) | 0.05-0.07 | 0.03 | 4.01 ($p$<.001) | 0.01-0.04 | 0.04 | 6.82 ($p$<.001) | 0.03-0.06 |
| Germany | 0.26 | 31.95 ($p$<.001) | 0.24-0.27 | 0.26 | 31.90 ($p$<.001) | 0.24-0.27 | 0.25 | 31.71 ($p$<.001) | 0.24-0.27 | 0.22 | 27.10 ($p$<.001) | 0.20-0.23 | 0.21 | 26.28 ($p$<.001) | 0.20-0.23 |
| Sweden | 0.17 | 10.59 ($p$<.001) | 0.14-0.20 | 0.17 | 10.63 ($p$<.001) | 0.14-0.20 | 0.17 | 10.26 ($p$<.001) | 0.14-0.20 | 0.17 | 10.21 ($p$<.001) | 0.13-0.19 | 0.14 | 8.57 ($p$<.001) | 0.11-0.17 |
| France | 0.14 | 16.86 ($p$<.001) | 0.13-0.16 | 0.14 | 16.62 ($p$<.001) | 0.12-0.16 | 0.13 | 15.37 ($p$<.001) | 0.11-0.15 | 0.12 | 14.01 ($p$<.001) | 0.10-0.13 | 0.11 | 13.05 ($p$<.001) | 0.09-0.13 |
| Poland | 0.15 | 20.84 ($p$<.001) | 0.14-0.17 | 0.15 | 20.84 ($p$<.001) | 0.14-0.17 | 0.15 | 20.78 ($p$<.001) | 0.14-0.17 | 0.13 | 17.11 ($p$<.001) | 0.11-0.14 | 0.13 | 18.08 ($p$<.001) | 0.12-0.15 |
| | | | | | | | Kendall's $\tau_K$ | | | | | | | | |
| Italy | 0.04 | 6.12 ($p$<.001) | 0.03-0.05 | 0.04 | 6.11 ($p$<.001) | 0.03-0.05 | 0.04 | 6.17 ($p$<.001) | 0.03-0.05 | 0.02 | 2.88 ($p$<.001) | 0.01-0.03 | 0.04 | 6.46 ($p$<.001) | 0.03-0.05 |
| Germany | 0.17 | 21.10 ($p$<.001) | 0.16-0.19 | 0.17 | 21.07 ($p$<.001) | 0.16-0.19 | 0.17 | 20.92 ($p$<.001) | 0.16-0.19 | 0.15 | 18.01 ($p$<.001) | 0.13-0.16 | 0.15 | 18.41 ($p$<.001) | 0.14-0.17 |
| Sweden | 0.12 | 7.03 ($p$<.001) | 0.08-0.15 | 0.12 | 7.06 ($p$<.001) | 0.08-0.15 | 0.11 | 6.81 ($p$<.001) | 0.08-0.14 | 0.11 | 6.84 ($p$<.001) | 0.08-0.14 | 0.13 | 8.14 ($p$<.001) | 0.10-0.17 |
| France | 0.09 | 11.21 ($p$<.001) | 0.08-0.11 | 0.09 | 11.05 ($p$<.001) | 0.08-0.11 | 0.09 | 10.22 ($p$<.001) | 0.07-0.10 | 0.08 | 9.48 ($p$<.001) | 0.06-0.10 | 0.08 | 9.85 ($p$<.001) | 0.06-0.10 |
| Poland | 0.10 | 13.84 ($p$<.001) | 0.09-0.12 | 0.10 | 13.85 ($p$<.001) | 0.09-0.12 | 0.10 | 13.80 ($p$<.001) | 0.09-0.12 | 0.09 | 11.56 ($p$<.001) | 0.07-0.10 | 0.09 | 12.13 ($p$<.001) | 0.08-0.11 |

*Note.* Coef. = correlation coefficient; CI = confidence interval.
$p$-values are reported in parentheses.

Table 3.9: Correlation Significance Tests for Spearman's $\rho_S$ and Kendall's $\tau_K$ Coefficients Estimated Between Education and Self-Assessed Health, 2015

| | Gaussian | | | t-copula | | | Frank | | | Gumbel | | | Empirical | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI | Coef. | $t$ | 95% CI |
| | | | | | | | Spearman's $\rho_S$ | | | | | | | | |
| Italy | 0.29 | 46.40 ($p<.001$) | 0.27-0.30 | 0.27 | 42.96 ($p<.001$) | 0.25-0.28 | 0.26 | 42.16 ($p<.001$) | 0.25-0.27 | 0.26 | 42.48 ($p<.001$) | 0.25-0.28 | 0.16 | 25.52 ($p<.001$) | 0.15-0.17 |
| Germany | 0.27 | 33.84 ($p<.001$) | 0.26-0.29 | 0.25 | 30.79 ($p<.001$) | 0.23-0.26 | 0.23 | 28.88 ($p<.001$) | 0.22-0.25 | 0.29 | 36.25 ($p<.001$) | 0.27-0.30 | 0.17 | 20.42 ($p<.001$) | 0.15-0.18 |
| Sweden | 0.22 | 13.93 ($p<.001$) | 0.19-0.26 | 0.18 | 11.13 ($p<.001$) | 0.15-0.21 | 0.17 | 10.17 ($p<.001$) | 0.13-0.20 | 0.29 | 18.65 ($p<.001$) | 0.26-0.32 | 0.12 | 7.10 ($p<.001$) | 0.08-0.15 |
| France | 0.26 | 32.40 ($p<.001$) | 0.25-0.28 | 0.24 | 29.68 ($p<.001$) | 0.23-0.26 | 0.23 | 27.45 ($p<.001$) | 0.21-0.24 | 0.30 | 36.91 ($p<.001$) | 0.28-0.31 | 0.16 | 19.55 ($p<.001$) | 0.15-0.18 |
| Poland | 0.34 | 48.12 ($p<.001$) | 0.33-0.35 | 0.33 | 46.85 ($p<.001$) | 0.32-0.34 | 0.31 | 42.96 ($p<.001$) | 0.29-0.32 | 0.33 | 46.31 ($p<.001$) | 0.32-0.34 | 0.23 | 32.19 ($p<.001$) | 0.22-0.25 |
| | | | | | | | Kendall's $\tau_K$ | | | | | | | | |
| Italy | 0.19 | 30.56 ($p<.001$) | 0.18-0.21 | 0.18 | 28.34 ($p<.001$) | 0.17-0.19 | 0.18 | 27.79 ($p<.001$) | 0.16-0.19 | 0.18 | 28.09 ($p<.001$) | 0.17-0.19 | 0.18 | 27.92 ($p<.001$) | 0.16-0.19 |
| Germany | 0.18 | 22.32 ($p<.001$) | 0.17-0.20 | 0.17 | 20.34 ($p<.001$) | 0.15-0.18 | 0.16 | 19.08 ($p<.001$) | 0.14-0.17 | 0.19 | 23.93 ($p<.001$) | 0.18-0.21 | 0.17 | 20.54 ($p<.001$) | 0.15-0.18 |
| Sweden | 0.15 | 9.22 ($p<.001$) | 0.12-0.18 | 0.12 | 7.38 ($p<.001$) | 0.09-0.15 | 0.11 | 6.75 ($p<.001$) | 0.08-0.14 | 0.20 | 12.30 ($p<.001$) | 0.17-0.23 | 0.13 | 8.09 ($p<.001$) | 0.10-0.16 |
| France | 0.18 | 21.38 ($p<.001$) | 0.16-0.19 | 0.16 | 19.61 ($p<.001$) | 0.15-0.18 | 0.15 | 18.15 ($p<.001$) | 0.14-0.17 | 0.20 | 24.34 ($p<.001$) | 0.19-0.22 | 0.15 | 18.45 ($p<.001$) | 0.14-0.17 |
| Poland | 0.23 | 31.53 ($p<.001$) | 0.22-0.24 | 0.22 | 30.72 ($p<.001$) | 0.21-0.24 | 0.21 | 28.20 ($p<.001$) | 0.19-0.22 | 0.22 | 30.50 ($p<.001$) | 0.21-0.24 | 0.20 | 27.12 ($p<.001$) | 0.19-0.21 |

*Note.* Coef. = correlation coefficient; CI = confidence interval.
$p$-values are reported in parentheses.

Table 3.10: Descriptive statistics of three well-being dimensions in post-crises period (2010)

| Country | Indicator[1] | M | SD | Min | Max |
|---|---|---|---|---|---|
| Italy (N = 27,364) | EDI (euro) | 19089 | 10124 | 4 | 64805 |
| | Years of education | 10.860 | 3.530 | 0 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.165 | 0.372 | 0 | 1 |
| | Good | 0.604 | 0.489 | 0 | 1 |
| | Fair | 0.184 | 0.387 | 0 | 1 |
| | Poor | 0.038 | 0.191 | 0 | 1 |
| | Very poor | 0.007 | 0.088 | 0 | 1 |
| Germany (N = 16,044) | EDI (euro) | 22071 | 10744 | 83 | 73303 |
| | Years of education | 13.887 | 2.136 | 4 | 16 |
| | Self-perceived general health | | | | |
| | Very good | 0.175 | 0.380 | 0 | 1 |
| | Good | 0.530 | 0.499 | 0 | 1 |
| | Fair | 0.226 | 0.418 | 0 | 1 |
| | Poor | 0.055 | 0.229 | 0 | 1 |
| | Very poor | 0.011 | 0.105 | 0 | 1 |
| Sweden (N = 4,936) | EDI (euro) | 22044 | 8430 | 0.94 | 50964 |
| | Years of education | 12.731 | 2.155 | 6 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.394 | 0.488 | 0 | 1 |
| | Good | 0.440 | 0.496 | 0 | 1 |
| | Fair | 0.126 | 0.332 | 0 | 1 |
| | Poor | 0.031 | 0.174 | 0 | 1 |
| | Very poor | 0.007 | 0.088 | 0 | 1 |
| France (N = 14,655) | EDI (euro) | 23786 | 120458 | 180 | 87035 |
| | Years of education | 11.482 | 3.367 | 0 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.243 | 0.429 | 0 | 1 |
| | Good | 0.495 | 0.499 | 0 | 1 |
| | Fair | 0.204 | 0.403 | 0 | 1 |
| | Poor | 0.050 | 0.220 | 0 | 1 |
| | Very poor | 0.007 | 0.085 | 0 | 1 |
| Poland (N = 20,192 ) | EDI (euro) | 4910 | 2731 | 23 | 17584 |
| | Years of education | 11.995 | 2.822 | 0 | 16 |
| | Self-perceived general health | | | | |
| | Very good | 0.149 | 0.356 | 0 | 1 |
| | Good | 0.450 | 0.497 | 0 | 1 |
| | Fair | 0.289 | 0.454 | 0 | 1 |
| | Poor | 0.097 | 0.295 | 0 | 1 |
| | Very poor | 0.014 | 0.116 | 0 | 1 |

*Note.* EDI = equivalised disposable income.

[1] The EU-SILC original variables involved are Total disposable household income (HY020), Highest ISCED level attained (PE040) and General health (PH010).

Table 3.11: Descriptive statistics of three well-being dimensions in pre-crises period (2006)

| Country | Indicator[1] | M | SD | Min | Max |
|---|---|---|---|---|---|
| Italy (N = 32,103) | EDI (euro) | 17392 | 9167 | 16 | 59180 |
| | Years of education | 10.545 | 3.708 | 0 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.152 | 0.359 | 0 | 1 |
| | Good | 0.515 | 0.499 | 0 | 1 |
| | Fair | 0.282 | 0.450 | 0 | 1 |
| | Poor | 0.042 | 0.202 | 0 | 1 |
| | Very poor | 0.008 | 0.089 | 0 | 1 |
| Germany (N = 17,980) | EDI (euro) | 19913 | 9511 | 56 | 65918 |
| | Years of education | 13.932 | 2.105 | 4 | 16 |
| | Self-perceived general health | | | | |
| | Very good | 0.131 | 0.337 | 0 | 1 |
| | Good | 0.518 | 0.499 | 0 | 1 |
| | Fair | 0.274 | 0.446 | 0 | 1 |
| | Poor | 0.064 | 0.245 | 0 | 1 |
| | Very poor | 0.012 | 0.108 | 0 | 1 |
| Sweden (N = 4,752) | EDI (euro) | 20021 | 7561 | 0.11 | 50904 |
| | Years of education | 12.464 | 2.334 | 6 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.368 | 0.482 | 0 | 1 |
| | Good | 0.427 | 0.494 | 0 | 1 |
| | Fair | 0.151 | 0.358 | 0 | 1 |
| | Poor | 0.042 | 0.201 | 0 | 1 |
| | Very poor | 0.011 | 0.103 | 0 | 1 |
| France (N = 13,671) | EDI (euro) | 18860 | 8981 | 485 | 61678 |
| | Years of education | 11.308 | 3.337 | 0 | 15 |
| | Self-perceived general health | | | | |
| | Very good | 0.267 | 0.442 | 0 | 1 |
| | Good | 0.498 | 0.500 | 0 | 1 |
| | Fair | 0.174 | 0.379 | 0 | 1 |
| | Poor | 0.053 | 0.224 | 0 | 1 |
| | Very poor | 0.008 | 0.088 | 0 | 1 |
| Poland (N = 24,545) | EDI (euro) | 3505 | 2102 | 19 | 13142 |
| | Years of education | 11.608 | 2.953 | 0 | 16 |
| | Self-perceived general health | | | | |
| | Very good | 0.123 | 0.329 | 0 | 1 |
| | Good | 0.438 | 0.496 | 0 | 1 |
| | Fair | 0.305 | 0.460 | 0 | 1 |
| | Poor | 0.116 | 0.320 | 0 | 1 |
| | Very poor | 0.017 | 0.130 | 0 | 1 |

*Note.* EDI = equivalised disposable income.

[1] The EU-SILC original variables involved are Total disposable household income (HY020), Highest ISCED level attained (PE040) and General health (PH010).

# Chapter 4

# A new proposal of multidimensional poverty index based on copulas

## 4.1 Introduction

In the last decade composite indicators became widely applied in multidimensional well-being and poverty measurement. International organizations commonly use them in the annual reports to summarize the performance of countries in economic spheres. Consequently, economists, policy-makers and non-academic stakeholders entered the discussion on the use of composite indicators as well as their advantages and pitfalls.

Let us focus on multidimensional poverty measurement to illustrate an example. Poverty can be defined as a failure to achieve a sufficient level over a set of resources. This set usually includes income but it is not restricted by this dimension only. Income alone as a measure of poverty gives a primary understanding of one's deprivation, but at the same time it narrows down the complex phenomenon from multiple aspects to a single dimension. Therefore, poverty can be defined as unsatisfactory performance over several attributes.

While scholars widely acknowledge the complexity of welfare and poverty, there is a lack of consensus among them regarding the social evaluation of these notions. In literature there exists two opposite approaches: already mentioned synthetic index method and a dashboard of indicators (Aaberge and Brandolini, 2015). Each approach has its pros and cons. Although a dashboard presents a variety of socio-economic indicators giving a detailed picture on how a given society succeeds in various well-being dimensions (Ciommi et al., 2017), it overlooks relevant interrelations among attributes. In the context of dashboard method the researcher may investigate literacy rate and income per capita of a society, but potential interconnection between the two variables is left beyond the scope. Due to a variety of available indicators dashboard does not offer simple comparisons over time and across regions. A composite indicator synthesizes information from several dimensions that are reflecting a certain aspect of a multivariate phenomenon to produce a single value. By contrast, this method allows a complete ordering of countries according to their performance by summarizing information over several dimensions into a single number (Ferreira and Lugo, 2013). Despite this advantage composite indicator may include arbitrary choices in the identification and the aggregation procedures influencing the eventual ordering of countries. Moreover, an information loss occurs due to the aggregation step.

Empirical applications of multidimensional poverty indices face several challenges such as the selection of relevant attributes to represent multidimensional poverty, establishment of deprivation cut-offs in each dimension, as well as the identification criterion and suitable weighting scheme (Greco et al., 2019; Maricic et al., 2019). The latter establishes a target importance of the underlying indicator for the multidimensional phenomenon reflected by the composite index. We highlight that dimensional weights do not measure the importance of variables; instead they govern the contribution of each dimension to the overall outcome of the composite indicator by affecting individual scores (see Paruolo et al. (2013), Schlossarek et al. (2019) and the references therein for the discussion of the importance of variables and nominal weights). Technically, in weighting step dimensional weights are computed and attached to pillars and sub-pillars of a composite indicator (Maggino, 2017). As a result, the information from underlying indicators is synthesized into a numeric value to return the overall level of multidimensional poverty or well-being. Composite indicator's outcome and the resulting ranking can strongly vary if dimensional weights are modified (Decancq et al., 2013b). However, there is no consensus in literature on how to select an optimal weighting method (Santos and Villatoro, 2018).

Decancq and Lugo (2013) classify weights of the composite indices into three groups: data-driven, normative and hybrid. As suggested by the term data-driven weights do not originate from the experts' opinion on the optimal trade-off between dimensions (Maricic et al., 2019). The advantage of this class of weights is their distributional predetermination; it allows data to "speak for itself". However, constructing dimensional weights using pure statistical tools is associated with certain drawbacks. In particular, a (possible) lack of economic interpretation is the weak spot of this approach (Schlossarek et al., 2019). Moreover, the relationship between dimensions as measured by statistical tools might not necessarily reflect the true interconnection among them (Saisana and Tarantola, 2002).

Unlike data-driven group, normative weights do not origin from certain characteristics of distribution. By contrast, this approach establishes dimensional weights according to experts' judgement on the best trade-off between attributes. Stakeholders reallocate the fixed number of points among the selected well-being dimensions; as a result, dimensional weights are defined as an average of experts' opinion about the trade-off (Saisana and Tarantola, 2002). All variables can be weighted equally as well, emphasizing an equivalent importance of each indicator; this weighting scheme is included in the normative class of weights.

Hybrid weights combine characteristics of the two approaches: they are data-driven, while a normative judgement on a trade-off is included likewise. Preferences-based weights illustrate an example of this duality, since representatives of the analysed society are interviewed about their view on an optimal trade-off among dimensions.

Table 4.1 summarizes weighting approaches used in well-being, inequality and poverty composite indices. The majority of reviewed measures use equal weights for the respective indicators and subindicators (if any) due to the interpretation simplicity. When several dimensions are chosen to give a more complete snapshot of welfare, equal weighting is a straightforward method to synthesize information from several pillars. Despite its simplicity and intuitive meaning this method has its drawbacks. Although equal weighting approach interprets each attribute as evenly important, it proposes to apply a fixed set of weights across countries causing a bottleneck on "fair" comparison. There might be different views on optimal weighting across countries; consequently, some of them may not accept a resulting ranking in terms of multidimensional poverty due to "unfair" weighting scheme. Moreover, equal weight of each dimension may not be desirable if some indicators are interdependent.

Statistical weights, which are based on such methods as Principle Component Analysis (PCA) or Factor Analysis (FA), belong to a group of data-driven weighting methods. In brief, PCA explains the total variability of the original data with several *principle components*, which are linear combinations of the original variables. Hence, the main purpose of PCA is the reduction of dimensionality. Weights of indicators can be

Table 4.1: Weighting schemes of composite indicators in some works on well-being, poverty and inequality measurement literature

| Authors | Composite indicator | Dimensions | Weights |
|---|---|---|---|
| Alkire and Jahan (2018) | Multidimensional Poverty Index | Living standards, education, health | EW |
| Bossert et al. (2013) | Material Deprivation Index | Household-level variables related to the ability to face unexpected expenses, pay mortgage, bills etc. | SPW |
| Bourguignon and Chakravarty (2003) | Multidimensional Poverty Index | Income and education | EW |
| Decancq and Lugo (2012) | Multidimensional Inequality Index | Household expenditures, health, education, housing quality | EW |
| Maasoumi and Lugo (2008) | Multiattribute Poverty Measure | Expenditures, level of hemoglobin, education | EW |
| Mitra and Brucker (2017) | Multiple Deprivations Index | Income, education, health, employment status, health insurance | EW |
| Nilsson (2010) | Multidimensional Inequality Index | Expenditures, education, health, land holdings | EW |
| Somarriba and Pena (2009) | Quality of Life Indicator | Employment and job satisfaction, dwelling, education, income, leisure, health, social life, safety of environment, life satisfaction | SW |
| UNDP (2016) | Human Development Index | Living standard, education, health | EW |

*Note.* EW = equal weights; AW = arbitrary weights; SW = statistical weights; SPW = stated-preference weights.

computed using the linear combination able to explain the largest share of variance (Greco et al., 2019). Using PCA to weight dimensions is suggested to be an objective weighting approach if any subjective beliefs of a researcher about the trade-off are not embedded in the procedure (Maggino, 2017). Despite attractive statistical properties these weights should be interpreted with care, since they are not based on the theoretical framework on well-being.

Stated-preferences weights offer the middle ground between data-driven and normative classes. Briefly, the trade-off among dimensions is defined through a survey by asking an opinion of respondents about their perception of different attributes. Therefore, citizens are involved in the decision-making process on welfare or poverty evaluation. For instance, the Eurobarometer survey in 2007 collected opinions of the EU citizens about their perception of poverty (Guio et al., 2009). The questions from this survey were harmonized with the data collected by the EU-SILC, what allows weights to be computed by directly incorporating respondents' preferences.

Selection of an appropriate weighting scheme is not the only debate in the context of multidimensional poverty measurement. There is a growing evidence in the literature that well-being dimensions are interrelated. For example, better self-perceived health is often reported by those citizens, whose earnings belong to higher income quintiles. Moreover, education has both monetary and non-monetary benefits: possessing higher educational attainment usually means earning higher income (Montenegro and Patrinos, 2014), while better educated individuals usually enjoy better health status and longevity (Oreopoulos and Salvanes, 2011).

This dependence among dimensions matters for multidimensional poverty measurement, since income is no longer the only attribute under consideration. Societies with highly dependent well-being attributes may be poorer than expected. However, the described interrelation is generally overlooked by poverty indices. Moreover, there is a current lack of research on how indicators' interdependence should be handled in multidimensional poverty measures.

In this paper we focus on the interdependence among dimensions by overcoming the limitations of equal weighting approach. Therefore, the contribution of this paper is twofold. Firstly, we propose a new copula-based multidimensional poverty index, which explores the interrelation among the underlying well-being attributes. An application of copula function in welfare data is beneficial due to its useful statistical properties: it allows capturing a broader class of dependence structures than linear correlation coefficient. Secondly, we innovate over the weighting scheme by directly incorporating the dependence among attributes estimated with copula into the dimensional weights. We offer a flexible country-specific weighting approach, which includes several controls to be chosen by the researcher. In particular, beliefs-adjusting parameter adapts the weights according to the opinion of the practitioner about how the dependence among dimensions should be regulated. In this regard, the proposed approach reconciles the opposite views on the interdependence among dimensions, i.e. if they should be associated with higher weights to account for the dependence or with lower weights to eliminate the redundancy of the composite indicator.

The rest of the paper is structured as follows. In Section 2 we develop methodological framework for the proposed copula-based multidimensional poverty measure discussing also main properties. In Section 3 we apply the proposed indicator to estimate multidimensional poverty in European countries using EU-SILC dataset and compare the new index with other approaches. Finally, Section 4 contains concluding remarks.

## 4.2 A class of copula-based multidimensional poverty indices

### 4.2.1 Identification step

Let population of $n$ individuals have achievements in a fixed number $d$ of dimensions. The distribution of all attributes across individuals is summarized in $n \times d$ matrix $X$. A typical element of the achievements' matrix, $x_{ij}$, defines the performance of individual $i$ in well-being dimension $j$. Unlike univariate poverty, in multidimensional poverty measurement there is a specific cut-off for all attributes. Therefore, let the vector of dimension-specific thresholds be given by $z = (z_1, z_2, z_3, \ldots, z_d)$. If individual achievement in dimension $j$ is below its corresponding cut-off $z_j$, then he is deprived in this attribute.

After establishing dimension-specific cut-offs to identify deprivations in each attribute we need to specify a criterion under which respondents are multidimensionally poor. Let $k$ be such poverty identification criterion that distinguishes multidimensionally poor from non-poor. The across-dimension cut-off $k$ is essential because concepts of deprivation and poverty do not coincide in multidimensional context. In general, the identification criterion ranges between 1 and $d$, the former corresponds to union approach, while the latter gives the intersection method.

### 4.2.2 Weighting and aggregation step

Our proposed weighting scheme is based on copula functions, which are able to separate the dependence structure from marginal distributions. Since copula uses rank-transformed data rather than original values, it captures more general types of dependence than linear correlation does. Well-being data rarely belongs to the class of continuous variables, therefore, copula-based measures are more suitable in welfare context (Pèrez and Prieto-Alaiz, 2016).

As already discussed in the previous Chapters, a $d$-dimensional copula $C$ from $[0,1]^d$ to $[0,1]$ grasps the dependence among $d$ random variables and links their joint distribution with respective margins:

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d))\,, \tag{4.1}$$

where $F_1(x_1), F_2(x_2), \ldots, F_d(x_d)$ are marginal distributions and $F(x_1, x_2, \ldots, x_d)$ is the joint distribution function. Copula is a flexible statistical tool that allows us describing the joint distribution of $d$ random variables with their marginal distributions and a copula function.

As already discussed in Chapter 3, Spearman's $\rho_S$ and Kendall's $\tau_K$ are measures of association that depend only on copula and, consequently, can be formulated in terms of it. The bivariate versions of mentioned correlation coefficients are given by:

$$\begin{aligned} \tau_K &= 4\int_0^1\int_0^1 C(u,v)\,dC(u,v) - 1 \\ \rho_S &= 12\int_0^1\int_0^1 uv\,dC(u,v) - 3\,, \end{aligned} \tag{4.2}$$

where $(u, v)$ are respective uniform margins. We recall that the two measures from 4.2 range between -1 and 1, corresponding to maximum positive and maximum negative dependence respectively (see Chapters 2 and 3 for more details). Since copula function transforms the original values into ranks, Spearman's $\rho_S$ and Kendall's $\tau_K$ are more flexible measures of dependence than common linear correlation coefficient.

The details on estimation of both measures were provided in Chapter 3. In brief, copula-based versions of Spearman's $\rho_S$ and Kendall's $\tau_K$ from equation (4.2) can be estimated assuming that the underlying copula takes a certain parametric form, e.g. Gaussian. Alternatively, both coefficients can be estimated nonparametrically using the empirical copula function. Corresponding nonparametric estimators are given by

$$\begin{aligned} \rho_S &= \frac{d+1}{2^d-(d+1)}\cdot\left\{\frac{2^d}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}(1-\hat{U}_{ij,n})-1\right\} \\ \tau_K &= \frac{1}{2^{d-1}-1}\cdot\left\{\frac{2^d}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\prod_{j=1}^{d}\mathbb{1}_{\{\hat{U}_{ij,n}\le\hat{U}_{kj,n}\}}-1\right\} \end{aligned} \tag{4.3}$$

where $\hat{U}_{ij,n}$ defines the rank of individual $i$ in dimension $j$ with respect to achievements of the rest of population in this dimension.

After having provided the definitions of copula-based measures of dependence we can define dimensional weights. Let $w^t = (w_1^t, w_2^t, \ldots, w_d^t)$ be a vector of weights with $\sum_{j=1}^{d} w_j^t = 1$, where $w_j^t > 0$ is a positive weight assigned to well-being dimension $j$. The weighting scheme we propose incorporates the interdependence between pairs of dimensions estimated by copula-based dependence measures given in equation (4.2). The weight of each dimension is governed by the strength of its interconnectedness with other attributes as measured by copula-based coefficients. In particular, dimensional weight is defined as the ratio of dependence due to the considered dimension and the sum of bivariate dependences existing between all pairs of attributes:

$$w_j^t = \frac{\sum\limits_{j \neq k}^{d} \left| \delta_{jk}^t \right|^{\theta_{jk} \times \beta}}{\sum\limits_{j=1}^{d} \sum\limits_{k=1}^{d} \left| \delta_{jk}^t \right|^{\theta_{jk} \times \beta}} \tag{4.4}$$

with $\delta_{jk}^t$ is a bivariate dependence between dimensions $j$ and $k$ at time $t$ as estimated by copula-based coefficient. Thereafter we skip the superscript $t$ in the notation of weights for brevity keeping in mind that dimensional weights evolve over time. The numerator in formula (4.4) measures the dependence that is driven by dimension $j$, while the denominator measures the total pairwise dependence existing among dimensions under consideration. With $\theta_{jk} \geq 1$ we denote a positive parameter that models the elasticity of substitution between each pair of dimensions. In principle, one can assume the same elasticity of substitution between all pairs of attributes for simplicity, albeit it is not compulsory. The higher is the value of $\theta$, the lower is the level of substitution.

The elasticity of substitution is driven by the value judgement of the practitioner. We indicate two special cases relevant for this parameter. Firstly, if $\theta = 1$, then dimensions are assumed to be perfect substitutes. In this case a better performance in one dimension offsets a lower achievement in the other one. The normative choice of the elasticity of substitution has an important policy implication: if dimensions are perfect substitutes, then policy-makers can focus on those attributes, which are able to improve welfare at lower cost or require less effort (Pinar, 2019). Consequently, individuals are likely to keep their unbalanced performances across dimensions, while the overall welfare (poverty) in the state is expected to increase (decrease). Secondly, dimensions are assumed to be perfect complements if $\theta$ goes to infinity. In case dimensional weights are modelled as somewhat complementary policy implications are different from perfect substitutability scenario: an optimal improvement of welfare (or reduction of multidimensional poverty) is accomplished if achievements in all attributes simultaneously improve (Pinar, 2019).

An important normative control involved in the proposed weighting procedure is a "belief-adjusting" parameter $\beta$, which mirrors an opinion of the practitioner on how the interdependence among indicators should affect a trade-off between them. Some scholars may argue that higher dependence should be penalized by a lower dimensional weight; by contrast, others might support a view that a higher weight should be assigned to highly interconnected dimensions. Therefore, a certain degree of flexibility is required to handle the dependence among underlying indicators of multidimensional poverty measures.

We highlight three special cases of the proposed weights from (4.4). The practitioner may choose to assign higher weights to highly dependent dimensions and fix $\beta$ equal to 1. In such case the more dependence is created by a certain dimension, the higher weight will be attached to it. Alternatively, if the practitioner is convinced that the highest weight should be allocated to the least associated dimensions, then he chooses $\beta = -1$ and changes the relation between the dependence and weighting to opposite. This normative control implies that the weight of indicator is inversely proportional to its interrelation with the rest of indicators. Finally, a special case of equal weighting is obtained if $\beta$ is equal to zero.

Dimensional weights specified in equation (4.4) belong to a hybrid group according to the classification suggested by Decancq and Lugo (2013). The intuition of the proposed weighting is the following: the importance of each well-being dimension is measured as the share of dependence that the dimension has with the others, with respect to the total pairwise dependence among all dimensions. Thereby the weight of each attribute is governed by its contribution to the total pairwise dependence. According to the proposed method a higher correlation is reflected in a higher (lower) dimensional weight if the opinion-adjusting parameter $\beta$ is positive (negative).

We provide a simple example to illustrate the proposed weights. Let us suppose that well-being is modelled by three dimensions: income, education and health. The part of dependence due to income stems from its interconnectedness with other attributes, namely education and health (Figure 4.1). Consequently, the weight for income is obtained from income-education and income-health relation, whereas the linkage between education and health is not considered.

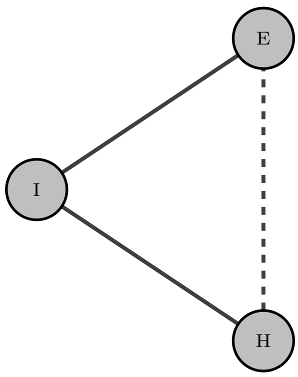



Figure 4.1: Example: bivariate dependencies relevant for the weight of income indicator

The weighting scheme we propose is not purely defined by the interrelation among dimensions as measured by copula; necessary normative controls are contained as well.

Therefore, the copula-based weighting method integrates the features of data-driven and normative groups, which is typical for hybrid class of weights. In particular, we suggest belief-adjusting parameter that offers necessary flexibility regarding how the dependence is handled. In other words, this normative parameter can take different values according to the objectives of each application and the preferences of stakeholders: for instance, the researcher can rationalize that dimensional weight is directly proportional to the strength of its relation with other attributes, which is obtained when $\beta$ takes the value of 1. Alternatively, an application may require that the importance of strongly associated dimensions is reduced accordingly; in this case the researcher keeps $\beta$ equal to -1, so that the weight is inversely proportional to the dependence associated with certain dimension. Finally, if either research question or policy priority require all dimensions to be equally weighted, this particular case is secured for $\beta = 0$.

The second normative control incorporated in the proposed copula-based weighting scheme is the degree of substitutability between the underlying attributes. Whether dimensions are assumed to be perfect substitutes is expected to have an impact on the overall multidimensional poverty. Intuitively, a higher degree of complementarity would return a higher multidimensional poverty compared to perfect substitutability case. In turn, it would promote policy to target a harmonized improvement in well-being dimensions rather than allocate resources in a single attribute.

We then suggest an additive aggregation procedure for the proposed multidimensional poverty index. Inspired by Foster-Greer-Thorbecke (FGT) family of poverty measures we defined an individual weighted deprivation score as follows

$$\lambda_i = \sum_{j=1}^{d} w_j \big(\frac{z_j - x_{ij}}{z_j}\big)^{\alpha} \mathbb{1}_{\{\zeta_i \geq k\}} \tag{4.5}$$

Then, the copula-based multidimensional poverty measure that we propose is defined as

$$\begin{aligned} P(X; z) &= \frac{1}{n} \sum_{i=1}^{n} \lambda_i \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d} w_j \big(\frac{z_j - x_{ij}}{z_j}\big)^{\alpha} \mathbb{1}_{\{\zeta_i \geq k\}} \end{aligned} \tag{4.6}$$

Parameter $\alpha \geq 0$ is interpreted as the aversion towards poverty: the higher is $\alpha$, the more sensitive is the poverty index to extreme poverty. If $\alpha = 0$, the proposed index (4.6) takes the form of headcount ratio, while for $\alpha = 1$ it is a poverty gap index. Finally, $\alpha = 2$

indicates a squared poverty gap.

Multidimensional poverty measure proposed in (4.6) is based on a double (row-first) aggregation procedure. Firstly, the aggregation is done across $j$ dimensions for each individual and can be interpreted as a weighted deprivation score of each representative of a given society. In the second step $n$ deprivation scores are aggregated to get an overall multidimensional poverty. The row-first type of aggregation requires that individual achievements in different attributes are collected from the same data source. If the data on joint distribution of attributes is available, it allows tracking the performance of the same respondents across dimensions. This information is useful for identifying a dependence pattern among dimensions of well-being. Additionally, if a longitudinal component is available, the evolution of association among dimensions can be traced over time.

### 4.2.3 Properties

We discuss a non-exhaustive list of axioms relevant for multidimensional poverty indicators. The properties mirror normative judgements on a desirable performance of a composite indicator under a certain distributional profile.

**Property 1.** Symmetry (SYM). Let $S$ be $n \times n$ permutation matrix; then for any $(X; z)$ the following holds: $P(SX; z) = P(X; z)$.

**Property 2.** Replication invariance (RI). If $Y$ is derived from $X$ by replicating its rows a finite number times, then $P(Y; z) = P(X; z)$.

**Property 3.** Scale invariance (SI). If $\Omega = diag(\omega_1, \dots, \omega_n)$ is a positive diagonal matrix, then for any $(X; z)$ the following is true: $P(X\Omega; z\Omega) = P(X; z)$.

Symmetry is an essential and widely advocated axiom since it requires a poverty measure to be anonymous: individual achievements contained in matrix $X$ should be treated equally for any $i = 1, \dots, n$. As a result, multidimensional poverty is driven by attributes' distribution in a society, while other information is irrelevant. If symmetry is satisfied, multidimensional poverty index is not a subject to gender, age, racial or any other type of discrimination. We highlight that symmetry is related to rows (individuals) of achievements' matrix, whereas for its columns (attributes) trade-off decisions are acceptable.

Under replication invariance a poverty measure does not change if each member of society is cloned a finite number of times. This property defines poverty as a per-capita phenomenon: between two societies $A$ and $B$ poverty indices usually define that one poorer, which has a higher number of poor over the population size. This property is useful for comparing poverty rates across countries with different number of citizens. Additionally, it allows unbiased poverty tracking over time given fluctuations in population size.

The principle of scale invariance requires a poverty measure to be unaffected by a modification of indicator's unit of measurement if the corresponding cut-off is adjusted. For instance, if the duration of education is reported in years rather than months should not affect the overall poverty level.

**Property 4.** Monotonicity (MON). If individual $i$ experiences an improvement in his deprived dimension $j$, such that $x_{ij} < x'_{ij} < z_j$, while the outcome of the rest of individuals is fixed, then $P(X'; z) < P(X; z)$.

**Property 5.** Dimensional monotonicity (DMON). If an individual $i$ improves his performance so that $x_{ij} < z_j \leq x'_{ij}$ keeping unchanged achievements of the rest of population, then $P(X'; z) < P(X; z)$.

Monotonicity and dimensional monotonicity allow establishing a dominance relationship between two societies having identical outcomes for all citizens apart from individual $i$. Two types of improvement are introduced: the first one, which raises the achievement keeping it below the dimension-specific cut-off, and the second one that moves corresponding outcome above the threshold. According to monotonicity the improved profile is socially preferred compared to the initial distribution, emphasizing the importance of each individual performance. In turn, dimensional monotonicity highlights each attribute for the overall poverty: if individual $i$ does not experience a shortfall in dimension $j$ that was formerly deprived, then poverty indicator should correspondingly decrease.

**Property 6.** Weak poverty focus (WFOC). When a non-poor individual $i$ improves his outcome in dimension $j$ such that $x_{ij} < x'_{ij}$, then $P(X; z) = P(X'; z)$.

**Property 7.** Strong poverty focus (SFOC). An improvement in a non-deprived outcome such that $z_j < x_{ij} < x'_{ij}$ does not alter the overall poverty: $P(X; z) = P(X'; z)$.

Whereas inequality and well-being indicators consider total distribution of attributes, by construction multidimensional poverty indices are sensitive to the lowest part of the distribution. Weak poverty focus ensures that an improvement in any outcome of non-poor individual does not affect a poverty measure, while a strong version of this property requires that any improvement in a non-deprived dimension does not influence the overall poverty evaluation. Intuitively, a poverty measure must be blind to the achievements of non-poor citizens. The motivation to introduce strong poverty focus is the following: if an increase in a non-deprived attribute was associated with a lower multidimensional poverty, then policy-makers could allocate resources in this attribute to reduce the number of poor. However, an anti-poverty strategy should grant resources to attributes, in which multidimensionally poor experience a deficit.

**Property 8.** Normalization (NORM). If $x_{ij} \geq z_j$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, d$, then $P(X; z) = 0$.

The copula-based multidimensional poverty measure that we propose in this Chapter satisfies SYM, RI, SI, WFOC, SFOC, NORM. The property of NORM is satisfied by the proposed index since it produces the value of zero if achievements of all individuals in the society exceed dimension-specific cut-offs. The axioms of invariance group, namely SYM, RI and SI, are verified as well. In particular, the proposed index is invariant under any rearrangement of the order, in which individuals are contained in matrix $X$. By construction, the proposed index considers solely individual achievements in all dimensions, whereas other details regarding matrix $X$ are not relevant. Similarly, if all rows of the original achievements matrix are replicated a certain number of times, then the overall multidimensional poverty is not affected since it is defined as per-capita phenomenon.

Modification of measurement unit of any dimension and the corresponding cut-off does not influence the copula-based index, since the weighted individual deprivation from equation (4.5) is not affected by this change. Firstly, regardless of measurement unit a proper threshold is supplied, so that the presence (or absence) of deprivation is correctly identified. Secondly, dimensional weights are computed using rank-transformed data, which does not depend on measurement unit of attributes.

In general, MON and DMON are not satisfied by the proposed index since dimensional weights change as well: due to suggested improvement in deprived dimension a new dependence structure occurs, so that dimensional weights change accordingly. A weak version of both properties, namely $P(X'; z) \leq P(X; z)$, holds if the weights are kept constant.

Finally, the proposed measure satisfies WFOC since any improvement of non-poor individuals does not influence the overall multidimensional poverty, while SFOC is fulfilled if the weights are constant.

## 4.3 An empirical application of the new multidimensional poverty index

In this section we apply the multidimensional poverty measure proposed using the data from the EU Survey on Income and Living Conditions (EU-SILC). The EU-SILC is an annual survey on income distribution and social exclusion that encompasses households and individuals of the EU member states. Our empirical illustration covers a selection of European countries – Italy, Germany, Sweden, France and Poland – and is referred to cross-sectional component of the survey at three time points: 2006, 2010 and 2015. We first present the results on the copula-based weighting and the proposed index in 2015, while the time evolution of multidimensional poverty is discussed at the end of this section.

Although researchers agree on the fact that welfare is multivariate, a consensus on which attributes should be selected has not been reached yet. We follow the majority of empirical studies and assume that well-being is represented by three attributes, namely income, education and health, which form a so-called "core" of well-being. Each attribute is mapped by an indicator: equivalised disposable household income, years of schooling and self-perceived health respectively. A summary of each indicator's construction and the corresponding EU-SILC variable involved is provided in Table 4.2. Our unit of analysis is individual, that is we investigate shortfalls of each adult in the household. For the purpose of this study we include male and female respondents aged 26-65 years, excluding young adults between 16 and 25 years of age, since they can be involved in education and might not achieved their highest educational level at the time of interview.

Before proceeding with the proposed multidimensional poverty index dimension-specific cut-offs need to be established. Poverty threshold for income indicator is fixed at 60% of median income. Since countries in our sample are roughly homogeneous in terms of their educational systems, the cut-off for education is set at 12 years of schooling corresponding to upper-secondary educational level according to ISCED 2011 (UNESCO, 2012a). In our context it is a reasonable threshold due to the fact that our sample consists of European countries. Otherwise, if the sample was constructed from the respondents of developing countries, a lower educational threshold – commonly specified at primary school diploma – would be required (Bourguignon and Chakravarty, 2003; Alkire and Foster, 2011). Consequently, in present framework an individual is identified as deprived in education if his years of training do not reach the specified threshold corresponding to upper secondary education.

Table 4.2: Well-being dimensions and the corresponding indicators

| Dimension | Indicator | Construction |
|---|---|---|
| Income | Equivalised disposable household income | Total disposable household income (HY020) adjusted for a household size by OECD-modified equivalence scale[1] |
| Education | Years of schooling | Total duration of education in years according to the highest educational attainment (PE040) |
| Health | Self-perceived health | Subjectively perceived general health status (PH010) ranged from very bad to excellent |

*Note.* The EU-SILC code of the original variables, which are used for constructing the indicators, is given in parenthesis.

[1] Total disposable income is adjusted to household needs considering the number of its members and their age. According to the OECD-modified equivalence scale a value of 1 is assigned to household head, 0.5 is given to other adults, while 0.3 is assigned to children.

General health status is a subjective measure of health assessed by individuals themselves. Respondents are able to categorize their health in one out of five categories from very poor to excellent. Dimension-specific cut-off for health status is fixed at the level "fair", meaning that individuals with either very poor or poor health are deprived (Alkire and Foster, 2011).

We have established dimension-specific thresholds that allow identifying individual shortfalls in each attribute. The snapshot of attribute-wise deprivations as well as the overlap between them is reported in Table 4.3. In Italy a little over one fifth of the respondents are deprived in income, whereas due to education criterion one third of interviewed citizens experience a shortfall. According to both criteria the share of deprived individuals over the population size is the largest in Italy. Conversely, if the criterion is health, then Poland has the highest percentage of respondents with a shortfall.

Looking at a deficit in one dimension independently from the others provides a fragmented understanding of ill-being patterns and overlooks a (possible) interplay among attributes. Indeed, the same individual may accumulate monetary and non-monetary deprivations and analysing the dashboard of indicators does not enable researcher to establish these links. The degree of overlap between shortfalls is of importance for practitioners and policy-makers since it uncovers crucial aspect of multifaceted phenomenon and sheds light on the "concentration" of deprivations.

The degree of interdependence among attributes, which should not be neglected in multidimensional poverty measurement, can be demonstrated by the percentage of respondents who are deprived in pairs of indicators (Atkinson and Lugo, 2010; Ferreira and Lugo, 2013). Remarkably, the largest overlap between income and education is observed in Italy, suggesting that its citizens are inclined to accumulate deprivations in these attributes. Consequently, in Italy some below-threshold income earners tend to experience a shortfall in education, which proves an importance of non-monetary attributes for poverty tracking. Similarly, the rest of countries demonstrate the same pattern of income-education overlap.

Table 4.3: Deprivations in income, years of schooling and self-perceived health in 2015

| Percentage of individuals | Italy | Germany | Sweden | France | Poland |
|---|---|---|---|---|---|
| deprived in income indicator | 20.13 | 17.65 | 13.98 | 13.43 | 18.41 |
| deprived in education indicator | 33.05 | 8.94 | 12.28 | 20.65 | 11.88 |
| deprived in health indicator | 5.72 | 7.61 | 3.53 | 6.31 | 10.26 |
| deprived in income and education | 9.96 | 3.50 | 3.03 | 5.19 | 4.33 |
| deprived in income and health | 1.59 | 3.34 | 1.01 | 1.86 | 3.08 |
| deprived in education and health | 3.26 | 1.56 | 0.82 | 2.40 | 2.64 |
| deprived in all dimensions | 1.12 | 0.96 | 0.38 | 0.91 | 1.12 |

The dependence among all deprivations across countries is illustrated by three-dimensional Venn diagrams (see Figure 4.3 in Appendix). Figure offers some intuition about how deprivations are interdependent in each country and illustrates the same pattern of the joint distribution of deprivations reported in Table 4.3. Each circle denotes the percentage of deprived respondents in each dimension, while an intersection between two circles shows the percent of deprived in both dimensions. The larger the size of a circle or an intersection area, the more individuals experience a shortfall in respective attribute(s).

The mismatch between monetary and non-monetary deprivations occurs: in Italy the highest percent of deprived have a shortfall in education. In turn, the intersection between education-deprived and income-deprived is captured at 10%. Obviously, the overlap between shortfalls is country-specific: while in Germany the intersection between income-education and income-health pairs of deprivations are almost equal, in Italy the former pair overweights the latter.

In this paper we aim at addressing the issue of interrelated individual performances in the context of multidimensional poverty measurement. For this purpose we propose a new class of copula-based multidimensional poverty indices (4.6) and apply it to selected European countries. We do not adopt a certain value of poverty identification criterion; instead, we allow the parameter to vary so that union (i.e. $k = 1$), intermediate (i.e. $1 < k < d$) and intersection (i.e. $k = d$) criteria are obtained. We choose three distinctive values of poverty aversion parameter and let the elasticity of substitution vary likewise. In the initial specification we assume that attributes are *perfect substitutes* and derive dimensional weights according to this assumption. The results suggest that copula-based approach gives more weight to education than to other dimensions in Italy, France and Poland (Table 4.4). By contrast, in Germany and Sweden education and income are weighted approximately equally, which is in line with the overlap between these attributes.

It is worth noting that in case of Sweden the derived copula-based weights are close to the equal weighting benchmark. A plausible explanation for this result is the extent of pairwise association between dimensions. For instance, the correlation between income and education does not exceed 0.22 and similarly for other pairs (we refer the reader to Chapter 3 for further details). Furthermore, this result may also be motivated by the role of government in well-being of its citizens. The proximity between copula-based approach and the equal-weighting benchmark in case of Sweden brings us to the following conclusion: equal weighting is a reasonable choice if the dependence between dimensions is low and does not go beyond a certain threshold.

In the second and third specifications we relax the assumption of perfect substitutability and obtain weights dimensions that are *complementary* to a certain extent. To model some degree of complementarity of attributes we fix $\theta = 5$ and $\theta = 10$. Under this assumption education anew is assigned a higher weight than other indicators. However, the weight of income has raised due to the assumed complementarity. Obviously, when dimensions are somewhat complementary, a deficit in one indicator can not be compensated by a surplus in the other one, confirming a modified weighting structure. Since the highest pairwise dependence is observed between income and education, which are now complements to some degree, higher weight is given to both of them.

For the sake of brevity in the subsequent analysis we attach weights based on empirical copula to attributes, since different copulas produce similar results. Furthermore, empirical copula function does not make any assumptions regarding the form of marginal distributions. The poverty identification criterion takes three values that correspond to union, intermediate and intersection approaches.

The application of the proposed poverty index is summarized in Table 4.6. The results are stratified by the elasticity of substitution between well-being dimensions. Table contains the level of poverty as the headcount ratio, the poverty gap and the squared poverty gap. The first column provides the list of countries, while the successive ones contain the absolute value of the proposed multidimensional poverty measure. The

Table 4.4: Copula-based weights using Spearman's $\rho_S$ and Kendall's $\tau_K$ referred to the year 2015, in percent

| | | Spearman's rho | | | | | Kendall's tau | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | Indicator | Normal | t-copula | Frank | Gumbel | Empirical | Normal | t-copula | Frank | Gumbel | Empirical |
| | | | | | | $\theta = 1$ | | | | | |
| Italy | Income | 29.65 | 30.44 | 30.70 | 28.98 | **34.41** | 29.66 | 30.47 | 30.73 | 29.11 | 31.01 |
| | Education | **44.02** | **43.89** | **43.65** | **46.30** | **43.95** | **44.10** | **43.96** | **43.72** | **46.20** | **44.05** |
| | Health | 26.33 | 25.68 | 25.65 | 24.72 | 21.64 | 26.24 | 25.57 | 25.55 | 24.69 | 24.95 |
| Germany | Income | **33.99** | **34.78** | **35.53** | 32.87 | **37.28** | **34.03** | **34.84** | **35.59** | **32.92** | **35.07** |
| | Education | **35.91** | **35.66** | **35.23** | **37.32** | **34.92** | **35.99** | **35.74** | **35.29** | **37.44** | **35.68** |
| | Health | 30.09 | 29.55 | 29.24 | 29.82 | 27.80 | 29.98 | 29.42 | 29.12 | 29.64 | 29.25 |
| Sweden | Income | 31.42 | **33.65** | **34.05** | 27.99 | **36.06** | 31.40 | **33.66** | **34.06** | 27.93 | 32.78 |
| | Education | **35.90** | **34.87** | **34.40** | **37.73** | **33.77** | **35.93** | **34.89** | **34.41** | **37.75** | **35.28** |
| | Health | 32.67 | 31.48 | 31.56 | **34.28** | 30.16 | 32.67 | 31.46 | 31.54 | **34.33** | 31.93 |
| France | Income | 32.69 | **33.58** | **34.31** | 30.95 | **36.15** | 32.73 | **33.64** | **34.38** | 31.04 | **33.81** |
| | Education | **39.95** | **39.88** | **40.12** | **41.67** | **39.74** | **40.05** | **39.97** | **40.21** | **41.71** | **40.06** |
| | Health | 27.36 | 26.54 | 25.57 | 27.38 | 24.11 | 27.23 | 26.38 | 25.41 | 27.24 | 26.13 |
| Poland | Income | 31.47 | 31.76 | 32.64 | 30.99 | **34.29** | 31.49 | 31.80 | 32.70 | 31.10 | 32.37 |
| | Education | **40.74** | **40.71** | **40.41** | **41.71** | **40.15** | **40.89** | **40.85** | **40.54** | **41.84** | **40.72** |
| | Health | 27.79 | 27.53 | 26.95 | 27.30 | 25.56 | 27.62 | 27.35 | 26.76 | 27.07 | 26.92 |
| | | | | | | $\theta = 5$ | | | | | |
| Italy | Income | **34.02** | **37.42** | **38.10** | **35.78** | **47.61** | **34.27** | **37.71** | **38.36** | **36.13** | **39.99** |
| | Education | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** |
| | Health | 16.01 | 12.62 | 11.94 | 14.23 | 2.41 | 15.76 | 12.33 | 11.68 | 13.87 | 10.04 |
| Germany | Income | **38.89** | **41.84** | **43.90** | **35.69** | **47.44** | **39.16** | **42.13** | **44.13** | **36.18** | **42.85** |
| | Education | **44.14** | **43.94** | **43.24** | **46.81** | **44.01** | **44.37** | **44.19** | **43.47** | **47.03** | **44.20** |
| | Health | 16.97 | 14.22 | 12.86 | 17.50 | 8.55 | 16.47 | 13.68 | 12.40 | 16.79 | 12.95 |
| Sweden | Income | 24.47 | **35.89** | **37.38** | 9.68 | **44.43** | 24.36 | **35.94** | **37.42** | 9.45 | 31.67 |
| | Education | **43.57** | **40.42** | **38.70** | **47.83** | **38.09** | **43.66** | **40.49** | **38.75** | **47.86** | **41.64** |
| | Health | 31.96 | 23.69 | 23.92 | **42.50** | 17.48 | 31.98 | 23.57 | 23.83 | **42.69** | 26.69 |
| France | Income | **39.77** | **42.90** | **45.13** | **35.20** | **47.92** | **40.09** | **43.20** | **45.36** | **35.75** | **43.79** |
| | Education | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** |
| | Health | 10.90 | 7.74 | 5.36 | 15.04 | 2.54 | 10.54 | 7.38 | 5.09 | 14.47 | 6.75 |
| Poland | Income | **35.72** | **37.08** | **40.34** | **35.46** | **45.11** | **36.17** | **37.56** | **40.78** | **36.28** | **39.77** |
| | Education | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** |
| | Health | 14.72 | 13.37 | 10.16 | 14.77 | 5.36 | 14.23 | 12.84 | 9.67 | 13.93 | 10.65 |

*Note.* Weights > 33% are listed in boldface.

Table 4.4: (continued)

| Country | Indicator | Spearman's rho | | | | | Kendall's tau | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Normal | t-copula | Frank | Gumbel | Empirical | Normal | t-copula | Frank | Gumbel | Empirical |
| | | | | | | $\theta = 10$ | | | | | |
| Italy | Income | **40.95** | **44.91** | **45.55** | **43.18** | **49.00** | **41.29** | **45.19** | **45.78** | **43.58** | **47.05** |
| | Education | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** |
| | Health | 9.05 | 5.09 | 4.45 | 6.82 | 1.00 | 8.71 | 4.81 | 4.22 | 6.42 | 2.95 |
| Germany | Income | **45.06** | **47.59** | **48.73** | **41.94** | **49.00** | **45.39** | **47.81** | **48.84** | **42.67** | **48.25** |
| | Education | **48.62** | **48.67** | **48.43** | **49.00** | **48.98** | **48.75** | **48.81** | **48.57** | **49.00** | **48.85** |
| | Health | 6.32 | 3.73 | 2.84 | 8.46 | 1.20 | 5.86 | 3.38 | 2.59 | 7.67 | 2.91 |
| Sweden | Income | 18.00 | **39.88** | **41.77** | 1.81 | **48.74** | 17.84 | **39.98** | **41.84** | 1.71 | **32.31** |
| | Education | **47.97** | **45.34** | **43.39** | **49.00** | **44.23** | **48.03** | **45.42** | **43.47** | **49.00** | **46.32** |
| | Health | **34.03** | 14.79 | 14.84 | **48.33** | 7.03 | 34.13 | 14.60 | 14.69 | **48.43** | 21.37 |
| France | Income | **46.80** | **48.63** | **49.00** | **42.40** | **49.00** | **47.04** | **48.76** | **49.00** | **43.07** | **48.99** |
| | Education | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** |
| | Health | 3.21 | 1.38 | 1.00 | 7.60 | 1.00 | 2.98 | 1.25 | 1.00 | 6.93 | 1.02 |
| Poland | Income | **42.96** | **44.47** | **47.22** | **42.73** | **49.00** | **43.49** | **44.96** | **47.51** | **43.68** | **46.83** |
| | Education | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** | **49.00** |
| | Health | 7.05 | 5.54 | 2.78 | 7.27 | 1.00 | 6.51 | 5.05 | 2.49 | 6.32 | 3.17 |

*Note.* Weights > 33% are listed in boldface.

estimation is made with copula-based versions of Spearman's rho and Kendall's tau coefficients.

The highest multidimensional poverty is ascertained in Italy regardless of the identification criterion, the existence of substitutability between dimensions and the underlying correlation coefficient (Table 4.6). This result suggests that Italy dominates other countries in terms of multidimensional poverty estimated with the proposed index. If a perfect substitutability assumption is relaxed, headcount ratio increases as expected. However, the assumption of complementarity does not necessarily lead to a higher poverty if the aversion parameter $\alpha$ is positive and intermediate identification criterion is applied.

This result is motivated by the structure of deprivations in each society rather than the association among dimensions. The assumption of complementarity makes weights of income and education grow, while moving downwards weight of health. This modification results in a reallocation of deprivation scores, i.e. sum of weighted deprivations. The reallocation of weights and deprivation scores can be well-balanced - the case of Germany and Sweden - keeping the multidimensional poverty at the same level.

We now compare the performance of the proposed index with the developed in literature multidimensional poverty measures (Bourguignon and Chakravarty, 2003; Chakravarty, 2009; Alkire and Foster, 2011). Table 4.7 contains the results on multidimensional poverty estimation in the European countries in 2006, 2010 and 2015. An absolute value of each poverty measure is reported together with 95% bootstrap confidence interval. We adopt union identification criterion (i.e. $k = 1$) to make indices comparable and assume, where applicable, perfect substitutability between dimensions (i.e. $\theta = 1$). In addition, we fix poverty aversion parameter $\alpha = 0$ so that the headcount ratio is produced. The copula-based multidimensional poverty measure is computed with three alternative sets of weights: weighting structure that assigns higher weights to more interrelated dimensions ($\beta = 1$), set of equal weights ($\beta = 0$) and weighting scheme that distributes higher weights to less interconnected dimensions ($\beta = -1$). Finally, for the existing approaches we use equal weighting scheme.

The ranking of countries in terms of multidimensional poverty implemented by the proposed copula-based index is in line with the results from the existing approaches. However, the absolute values of poverty measures diverge: a kind of upper and lower "bounds" are established by Bourguignon and Chakravarty index and Watts index respectively, while the estimates of the approach we propose here is included into this range.

Table 4.6: Multidimensional poverty measured with the proposed copula-based index in 2015: substitutability versus complementarity

| | Substitutes ($\theta = 1$) | | | | | | Complements ($\theta = 10$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Spearman's rho | | | Kendall's tau | | | Spearman's rho | | | Kendall's tau | | |
| Country | $k = 1$ | $k = 2$ | $k = 3$ | $k = 1$ | $k = 2$ | $k = 3$ | $k = 1$ | $k = 2$ | $k = 3$ | $k = 1$ | $k = 2$ | $k = 3$ |
| | | | | | | Headcount ratio ($\alpha = 0$) | | | | | | |
| Italy | **0.228** | **0.100** | **0.012** | **0.224** | **0.098** | **0.012** | **0.262** | **0.114** | **0.012** | **0.259** | **0.112** | **0.012** |
| Germany | 0.115 | 0.045 | 0.009 | 0.113 | 0.045 | 0.009 | 0.127 | 0.046 | 0.009 | 0.126 | 0.047 | 0.009 |
| Sweden | 0.101 | 0.029 | 0.004 | 0.099 | 0.029 | 0.004 | 0.123 | 0.034 | 0.004 | 0.108 | 0.031 | 0.004 |
| France | 0.146 | 0.058 | 0.010 | 0.145 | 0.058 | 0.010 | 0.167 | 0.064 | 0.010 | 0.167 | 0.064 | 0.010 |
| Poland | 0.137 | 0.058 | **0.012** | 0.136 | 0.058 | **0.012** | 0.148 | 0.061 | **0.012** | 0.147 | 0.061 | **0.012** |
| | | | | | | Poverty gap ($\alpha = 1$) | | | | | | |
| Italy | **0.085** | **0.039** | **0.005** | **0.083** | **0.038** | **0.005** | **0.097** | **0.044** | **0.005** | **0.096** | **0.043** | **0.005** |
| Germany | 0.033 | 0.014 | 0.003 | 0.033 | 0.014 | 0.003 | 0.034 | 0.014 | 0.003 | 0.034 | 0.014 | 0.003 |
| Sweden | 0.030 | 0.009 | 0.001 | 0.029 | 0.009 | 0.001 | 0.035 | 0.010 | 0.001 | 0.031 | 0.009 | 0.001 |
| France | 0.050 | 0.020 | 0.004 | 0.050 | 0.020 | 0.004 | 0.056 | 0.022 | 0.004 | 0.056 | 0.022 | 0.004 |
| Poland | 0.052 | 0.023 | **0.005** | 0.052 | 0.023 | **0.005** | 0.055 | 0.024 | **0.005** | 0.055 | 0.024 | **0.005** |
| | | | | | | Squared gap ($\alpha = 2$) | | | | | | |
| Italy | **0.038** | **0.019** | **0.003** | **0.038** | **0.018** | **0.003** | **0.045** | **0.021** | **0.003** | **0.044** | **0.021** | **0.003** |
| Germany | 0.014 | 0.006 | 0.002 | 0.014 | 0.006 | 0.002 | 0.014 | 0.006 | 0.002 | 0.014 | 0.006 | 0.002 |
| Sweden | 0.012 | 0.004 | 0.001 | 0.011 | 0.004 | 0.001 | 0.013 | 0.004 | 0.001 | 0.012 | 0.004 | 0.001 |
| France | 0.023 | 0.010 | 0.002 | 0.023 | 0.010 | 0.002 | 0.026 | 0.011 | 0.002 | 0.026 | 0.011 | 0.002 |
| Poland | 0.025 | 0.011 | 0.002 | 0.025 | 0.011 | 0.002 | 0.027 | 0.012 | 0.002 | 0.027 | 0.012 | 0.002 |

*Note.* Identification criterion $k$ specifies minimum number of deprivations individual should have a shortfall in to be identified as multidimensionally poor. Parameter $\beta$ is assigned the value of 1. Top-ranked values of poverty index for each specification are given in boldface.

Table 4.7: The evolution of multidimensional poverty estimated by the proposed index and the existing approaches at three time points: 2006, 2010 and 2015

| | | Bourguignon and Chakravarty (2003) | | Alkire and Foster (2011) | | Multidimensional Watts Index[1] | | Copula-based index ($\beta = 1$) | | Copula-based index ($\beta = 0$) | | Copula-based index ($\beta = -1$) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country | Year | HR | 95% CI | HR | 95% CI | Index | 95% CI | HR | 95% CI | HR | 95% CI | HR | 95% CI |
| Italy | 2015 | 0.452 | [0.445, 0.458] | 0.198 | [0.194, 0.201] | 0.117 | [0.113, 0.121] | 0.228 | [0.225, 0.232] | 0.198 | [0.194, 0.201] | 0.164 | [0.161, 0.167] |
| | 2010 | 0.503 | [0.497, 0.509] | 0.218 | [0.215, 0.221] | 0.150 | [0.145, 0.155] | 0.262 | [0.258, 0.265] | 0.218 | [0.215, 0.221] | 0.159 | [0.156, 0.161] |
| | 2006 | 0.540 | [0.535, 0.546] | 0.234 | [0.231, 0.236] | 0.168 | [0.163, 0.172] | 0.275 | [0.272, 0.279] | 0.234 | [0.231, 0.236] | 0.176 | [0.173, 0.178] |
| Germany | 2015 | 0.263 | [0.255, 0.270] | 0.111 | [0.108, 0.115] | 0.045 | [0.043, 0.047] | 0.115 | [0.112, 0.119] | 0.111 | [0.108, 0.115] | 0.107 | [0.104, 0.111] |
| | 2010 | 0.257 | [0.250, 0.264] | 0.106 | [0.103, 0.109] | 0.039 | [0.038, 0.041] | 0.111 | [0.108, 0.115] | 0.106 | [0.103, 0.109] | 0.101 | [0.098, 0.104] |
| | 2006 | 0.247 | [0.240, 0.254] | 0.099 | [0.096, 0.102] | 0.040 | [0.038, 0.042] | 0.103 | [0.100, 0.106] | 0.099 | [0.097, 0.102] | 0.095 | [0.092, 0.098] |
| Sweden | 2015 | 0.249 | [0.235, 0.264] | 0.098 | [0.092, 0.104] | 0.040 | [0.036, 0.044] | 0.101 | [0.095, 0.108] | 0.098 | [0.092, 0.104] | 0.095 | [0.089, 0.101] |
| | 2010 | 0.226 | [0.214, 0.238] | 0.089 | [0.084, 0.094] | 0.040 | [0.036, 0.043] | 0.091 | [0.085, 0.096] | 0.089 | [0.084, 0.094] | 0.087 | [0.082, 0.092] |
| | 2006 | 0.264 | [0.251, 0.277] | 0.101 | [0.096, 0.107] | 0.045 | [0.042, 0.049] | 0.104 | [0.098, 0.109] | 0.101 | [0.096, 0.107] | 0.099 | [0.094. 0.104] |
| France | 2015 | 0.318 | [0.310, 0.326] | 0.135 | [0.132, 0.139] | 0.095 | [0.089, 0.101] | 0.146 | [0.142, 0.150] | 0.135 | [0.132, 0.139] | 0.126 | [0.122, 0.129] |
| | 2010 | 0.346 | [0.339, 0.355] | 0.146 | [0.142, 0.149] | 0.099 | [0.094, 0.104] | 0.163 | [0.159, 0.167] | 0.146 | [0.142, 0.149] | 0.125 | [0.122, 0.128] |
| | 2006 | 0.379 | [0.371, 0.387] | 0.161 | [0.157, 0.165] | 0.121 | [0.113, 0.128] | 0.176 | [0.172, 0.180] | 0.161 | [0.157, 0.165] | 0.144 | [0.140, 0.148] |
| Poland | 2015 | 0.315 | [0.308, 0.322] | 0.136 | [0.132, 0.139] | 0.077 | [0.074, 0.080] | 0.137 | [0.134, 0.141] | 0.136 | [0.132, 0.139] | 0.136 | [0.133, 0.139] |
| | 2010 | 0.334 | [0.327, 0.341] | 0.145 | [0.142, 0.149] | 0.081 | [0.078, 0.084] | 0.148 | [0.144, 0.151] | 0.145 | [0.142, 0.149] | 0.145 | [0.141, 0.148] |
| | 2006 | 0.382 | [0.376, 0.388] | 0.168 | [0.165, 0.172] | 0.107 | [0.103, 0.110] | 0.171 | [0.168, 0.175] | 0.168 | [0.165, 0.172] | 0.167 | [0.164, 0.170] |

*Note.* HR = headcount ratio; CI = confidence interval. We report nonparametric bootstrap confidence interval based on 1000 replicates. Where applicable well-being dimensions are assumed to be perfect substitutes. For all measures, except Watts index, the poverty aversion parameter $\alpha$ equals zero. We follow union criterion to poverty identification for comparability purpose. The weights of the poverty measure we propose are estimated with Spearman's $\rho_S$ based on empirical copula.

[1] The multidimensional extension of Watts index was defined by Chakravarty (2009).

The evolution of multidimensional poverty estimated with the proposed index is illustrated in Figure 4.2. As Figure suggests multidimensional poverty decreased in the considered time range in all countries, except Germany, where a slight increase is observed. We compare three special cases of belief-adjusting parameter $\beta$. Remarkably, the highest poverty is observed if dimensional weights are computed with $\beta = 1$, that is when weights are directly proportional to the interconnectedness among dimensions. If the role of interrelation is eliminated, i.e. $\beta$ is equal to zero, the obtained effects are twofold. Firstly, weighting dimensions equally reduces the level of multidimensional poverty in all countries. Secondly, the differences in terms of poverty between some of them are notably reduced. In particular, it is the case of France and Poland: for $\beta = 1$ the former is poorer than the latter, while in case of equal weighting the dominance relationship between them is not established.

The last special case considers $\beta = -1$; in other words, higher dependence among dimensions is adjusted by lower weights, while the least dependent dimension is attached the highest weight. The reverse link between the dependence and the corresponding weight redefines the trade-off among attributes and redistributes the contribution of each indicator to the overall poverty. As a result, lower multidimensional poverty is ascertained in the considered countries.

A plausible explanation for this is related to different aspects of multidimensional poverty captured by the underlying indicators given the trade-off between them. If the interdependence among dimensions is accounted for by assigning relatively higher weights, it means that the contribution of each indicator is decomposed into the existence of deprivation and the degree of interrelation among performances in monetary and non-monetary attributes. By contrast, if the reverse relation between the dependence and the weights is established, then the role of highly interconnected attributes is lowered, thereby reducing the overall multidimensional poverty. Consequently, the degree of dependence among attributes shapes the magnitude of multidimensional poverty.

## 4.4 Concluding remarks

In this paper we contributes to the discussion on the dependence among dimensions in the context of multidimensional poverty measurement and propose a new class of multidimensional poverty indices based on copula. We innovate over a weighting scheme by incorporating the dependence among attributes into it. More general types of

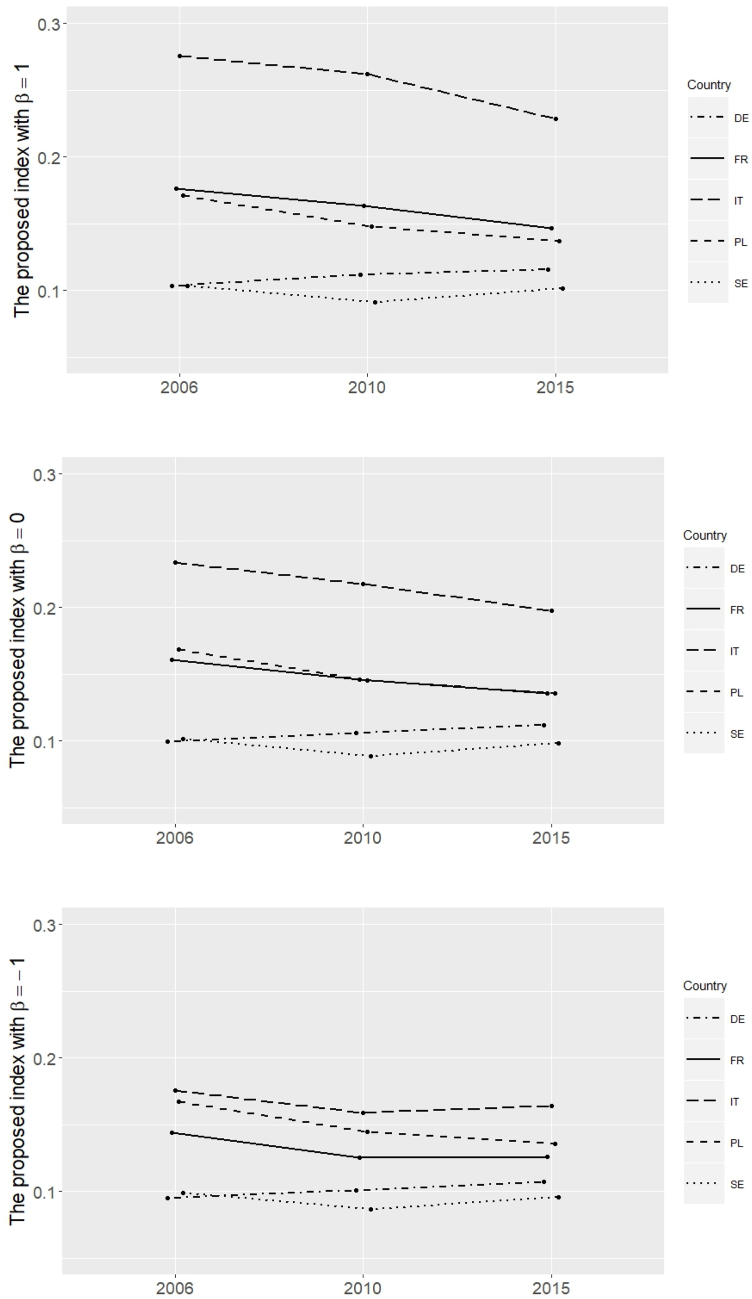



Figure 4.2: The evolution of multidimensional poverty estimated with the proposed copula-based index in Germany (dot-dash line), France (solid line), Italy (long-dash line), Poland (dashed line) and Sweden (dotted line) for $\beta = 1$, $\beta = 0$ and $\beta = -1$
*Source:* authors' calculations based on the EU-SILC referred to the years 2006, 2010 and 2015.

dependence among attributes are captured due to the application of copula function in the framework.

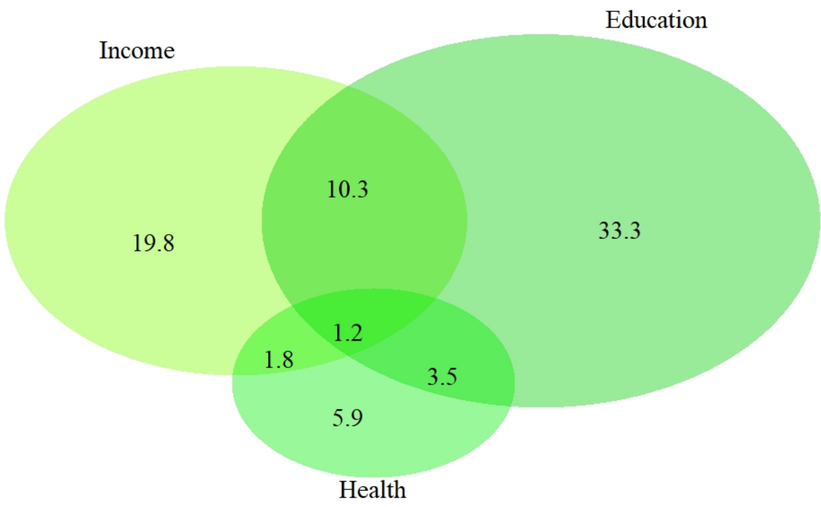
The proposed copula-based index offers a necessary flexibility due to the introduced normative controls. In particular, dimensions can be considered either perfect substitutes or complements to a certain extent depending on the choice of the practitioner and purposes of the empirical application. Moreover, the interdependence among attributes can have the twofold effect on the trade-off among dimensions: higher weights can be attached either to the highly interrelated indicators or to the least dependent ones. The proposed weighting scheme also includes equal weighting as a special case.

The results suggest that the degree of substitutability between dimensions is of importance for establishing the trade-off among them: the more complementary are the interrelated dimensions the more weight is assigned to them. Therefore, to increase the well-being of citizens an improvement in both attributes is essential, promoting a more harmonized enhancement of performances across dimensions rather than a focus on a single attribute.

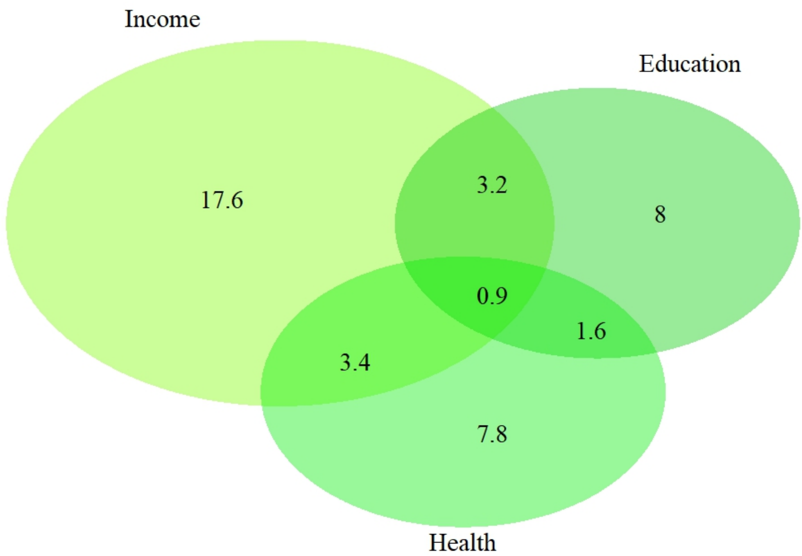
The proposed copula-based index is then applied to the European countries to measure the evolution of multidimensional poverty at three points in time, namely 2006, 2010 and 2015. During the considered time span multidimensional poverty in the European countries lowered. Assigning higher weights to more related attributes leads to a higher multidimensional poverty in all countries compared to the other cases (i.e. $\beta = 0$ and $\beta = -1$). This result is motivated by the fact that not only the below-threshold individual performances contribute to the overall poverty level, but also the degree of interdependence among attributes plays an important role.

Despite its useful statistical properties copula function is associated with a certain limitations in the welfare application. Although it goes beyond the common linear correlation and allows establishing broader types of dependence, copula function can be applied only with quantitative or ordinal variables. Categorical variables, which are frequently found in the multidimensional poverty framework, are excluded. Hence, the limitation of the proposed approach is that it can derive weights only for continuous or ordinal variables, excluding qualitative ones.

## 4.5 Appendix

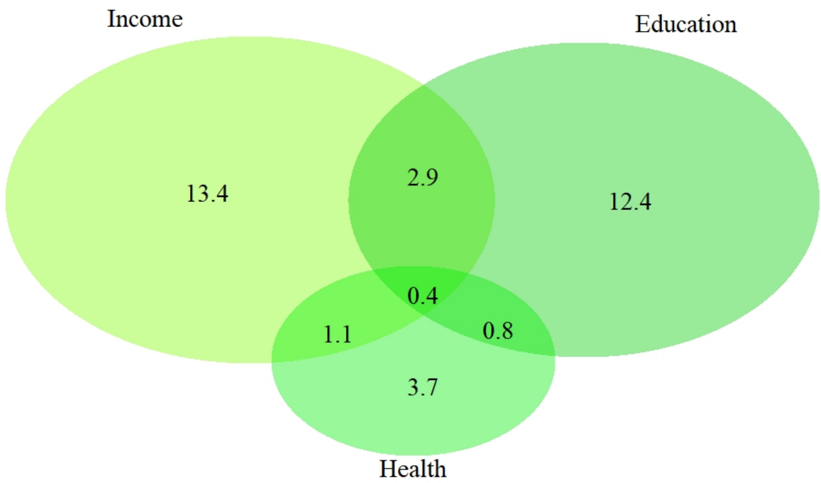



(a) Italy (N = 22,593)

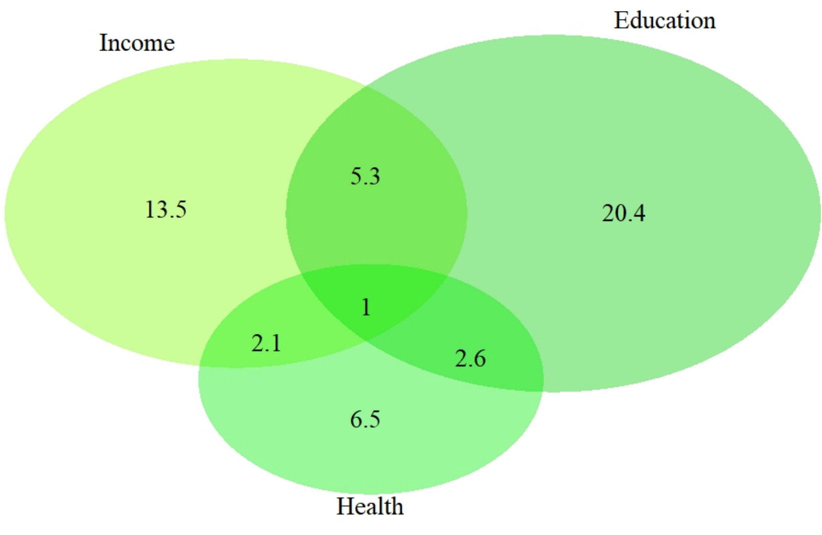



(b) Germany (N = 14,046)

Figure 4.3: The overlap among deprivations in income, education and health, in percent
*Source:* authors' calculations based on the EU-SILC, survey year 2015.

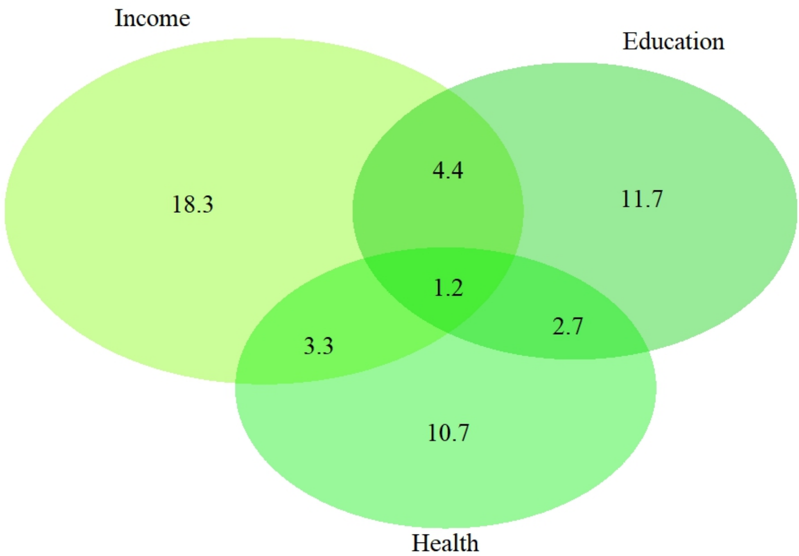



(c) Sweden (N = 3,415)




(d) France (N = 13,065)

Figure 4.3: The overlap among deprivations in income, education and health, in percent (cont.)




(e) Poland (N = 16,795)

Figure 4.3: The overlap among deprivations in income, education and health, in percent (cont.)

# General Conclusion

This dissertation contributes to the literature on the multidimensional poverty measurement and addresses the issues of interdependence among well-being dimensions. Copula function plays a central role in estimating the interdependence among attributes, which captures some general dependence structures existing in bivariate and multivariate cases. We provide results on the time evolution of pairwise and overall dependence in the European countries. A new proposal of multidimensional poverty index is done in this dissertation, which incorporates the interrelation among well-being variables.

In the first Chapter we give an overview of the approaches to poverty measurement, discussing their strengths and drawbacks. In addition, we formulate properties relevant for the univariate and the multivariate poverty indices. A special focus of the Chapter is the selection of an appropriate weighting scheme in the context of interrelation among attributes. In the second Chapter we introduce copula function and summarize main theorems and properties associated with it. Most recent applications of copula function in the well-being framework are outlined. Finally, we suggest to apply copula for welfare variables and propose a possible channel to include the estimated dependence in the composite indicators.

The third Chapter is methodologically linked to the first one and addresses the problem of interdependence among well-being dimensions further. Throughout the Chapter we assume that multivariate well-being is reflected in three dimensions, i.e. income, education and health. We apply copula-based dependence measures, i.e. Spearman's $\rho_S$ and Kendall's $\tau_K$ coefficients, to the selected European countries using the EU-SILC data. The results suggest that in 2015 the dependence was positive in all countries considering either pairs of attributes or the overall interrelation. Additionally, we monitor the time evolution of the interdependence using three time points, namely 2006, 2010 and 2015. The main findings suggest that in the post-crises period the relation among attributes strengthened in some countries. In particular, the income-education pair demonstrates higher coordination of the individual achievements after the financial crises. Comparing these findings with the evolution of the human development as measured by the HDI allows us to make the following conclusion. Although the overall human development kept being very high and even improved during the period under consideration, the dependence among attributes remained positive in all countries, while in some of them it augmented.

In the fourth Chapter we propose a new class of copula-based poverty indices by innovating over the weighting scheme. In particular, we propose weights that are derived from the dependence among dimensions measured by copula and contain necessary normative controls. Dimensional weights we propose offer an essential flexibility and can be adjusted to the specific objectives of the empirical application. The practitioner can model the dimensions as perfect substitutes or he may assume them being somewhat complementary. Moreover, an additional parameter is suggested that affects the trade-off among attributes in terms of the estimated dependence: the practitioner can choose to assign higher weights to more dependent indicators or he may reduce their contribution to the overall index by changing the value of the belief-adjusting parameter. An equal weighting is included in the proposed weighting scheme as a special case. The proposed copula-based multidimensional poverty index is then applied to five European countries using three points in time. The results suggest that during the period from 2006 to 2015 multidimensional poverty decreased. If the role of the dependence is eliminated by assigning equal weights to the attributes, the overall multidimensional poverty is lower compared to the model when the relation between weights and the dependence is directly proportional. Moreover, if the least dependent dimensions are attached the highest weights, then the lowest multidimensional poverty is obtained. Our conclusion is that the strength of the interdependence among attributes shapes the multidimensional poverty.

# Bibliography

Aaberge, R. and Brandolini, A. (2015). *Handbook of Income Distribution*, volume 2A, chapter Multidimensional Poverty and Inequality, pages 142–216. North Holland.

Alkire, S. (2007). *The Many Dimensions of Poverty*, chapter Choosing Dimensions: The Capability Approach and Multidimensional Poverty, pages 89–119. Palgrave Macmillan, London.

Alkire, S. and Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95:476–487.

Alkire, S., Foster, J., and Santos, M. E. (2011). Where did identification go? *The Journal of Economic Inequality*, 9(3):501–505.

Alkire, S., Foster, J., Seth, S., Santos, M. E., Roche, J. M., and Ballon, P. (2015). *Measuring Multidimensional Poverty: The Axiomatic Approach*. Oxford University Press.

Alkire, S. and Jahan, S. (2018). The new global mpi 2018: Aligning with the sustainable development goals'. Technical report, HDRO Occasional Paper, United Nations Development Programme (UNDP).

Aspalter, C., Jinsoo, K., and Sojeung, P. (2009). Analysing the welfare state in poland, the czech republic, hungary and slovenia: An ideal-typical perspective. *Social Policy Administration*, 43(2):170–185.

Athanassoglou, S. (2015). Multidimensional welfare rankings under weight imprecision: a social choice perspective. *Social Choice and Welfare*, 44(4):719–744.

Atkinson, A. and Bourguignon, F. (1982). The comparison of multidimensional distributions on economic status. *Review of Economic Studies*, 49:183–201.

Atkinson, A. and Lugo, M. A. (2010). Growth, poverty and distribution in tanzania. International Growth Centre Working Paper 10/0831.

Atkinson, A. B. (2003). Multidimensional deprivation: contrasting social welfare and counting approaches. *Journal of Economic Inequality*, 1:51–65.

Atkinson, A. B. (2011). On lateral thinking. *The Journal of Economic Inequality*, 9(3):319–328.

Baeten, S., Ourti, T. V., and van Doorslaer, E. (2013). Rising inequalities in income and health in china: Who is left behind? *Journal of Health Economics*, 32(6):1214–1229.

Balakrishnan, N. and Lai, C.-D. (2009). *Continuous Bivariate Distributions*. Springer-Verlag New York, second edition edition.

Böckerman, P. and Maczulskij, T. (2016). The education-health nexus: Fact and fiction. *Social Science and Medicine*, 150:112–116.

Benzeval, M. and Judge, K. (2001). Income and health: the time dimension. *Social Science and Medicine*, 52(9):1371–1390.

Bergqvist, K., Åberg Yngwe, M., and Lundberg, O. (2013). Understanding the role of welfare state characteristics for health and inequalities – an analytical review. *BMC Public Health*, 13:1–20.

Bijwaard, G. E., Kippersluis, H., and Veenman, J. (2015). Education and health: The role of cognitive ability. *Journal of Health Economics*, 42:29–43.

Blumentritt, T. and Schmid, F. (2014). Nonparametric estimation of copula-based measures of multivariate association from contingency tables. *Journal of Statistical Computation and Simulation*, 84(4):781–797.

Blundell, R., Dearden, L., and Sianesi, B. (2005). Evaluating the effect of education on earnings: models, methods and results from the national child development survey. *Journal of the Royal Statistical Society: Series A*, 168(3):473–512.

Bonhomm, S. and Robin, J.-M. (2009). Assessing the equalizing force of mobility using short panels: France, 1990–2000. *The Review of Economic Studies*, 79:63–92.

Bossert, W., Chakravarty, S. R., and D'Ambrosio, C. (2013). Multidimensional poverty and material deprivation with discrete data. *Review of Income and Wealth*, 59(1):29–43.

Bourguignon, F. and Chakravarty, S. R. (1999). *Advances in Econometrics, Income Distribution and Scientific Methodology. Essays in Honor of Camilo Dagum*, chapter A Family of Multidimensional Poverty Measures, pages 331–344. Physica-Verlag.

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

Bronshtein, I., Semendyayev, K., Musiol, G., and Mühlig, H. (2007). *Handbook of Mathematics*. Springer-Verlag Berlin Heidelberg, 5th edition.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating marginal returns to education. *The American Economic Review*, 101(6):2754–2781.

Castagnoli, E. and Muliere, P. (1989). *Income and Wealth Distribution, Inequality and Poverty*, conference proceedings A Note On Inequality Measures And The Pigou-Dalton Principle Of Transfers, pages 171–182.

Chakravarty, S. R. (2009). *Inequality, Polarization and Poverty. Advances in Distributional Analysis*. Economic Studies in Inequality, Social Exclusion and Well-Being. Springer-Verlag New York.

Chakravarty, S. R. and Silber, J. (2008). *Quantitative Approaches to Multidimensional Poverty Measurement*, chapter Measuring Multidimensional Poverty: The Axiomatic Approach, pages 192–209. Palgrave Macmillan UK.

Charpentier, A., Fermanian, J.-D., and Scaillet, O. (2007). *Copulas: From Theory to Application in Finance*, chapter The estimation of copulas : theory and practice, pages 35–64. London : Risk Books.

Cherchye, L., Ooghe, E., and Puyenbroeck, T. V. (2008). Robust human development rankings. *The Journal of Economic Inequality*, 6(4):287–321.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons Inc, Chichester, UK ; Hoboken, NJ.

Ciommi, M., Gigliarano, C., Emili, A., Taralli, S., and Chelli, F. M. (2017). A new class of composite indicators for measuring well-being at the local level: An application to the equitable and sustainable well-being (BES) of the italian provinces. *Ecological Indicators*, 76:281–296.

Coady, D. and Dizioli, A. (2017). Income inequality and education revisited: Persistence, endogeneity, and heterogeneity. Working Paper 17/126, International Monetary Fund.

Conti, G. and Hansman, C. (2013). Personality and the education-health gradient: a note on "understanding differences in health behaviors by education". *Journal of Health Economics*, 32:480–485.

Conti, G., Heckman, J., and Urzua, S. (2010). The education-health gradient. *The American Economic Review*, 100(2):234–238.

Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society*, 43(2):210–218.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29:1–28.

de la Vega, C. L. and Aristondo, O. (2012). Proposing indicators to measure achievement and shortfall inequality consistently. *Journal of Health Economics*, 31(4):578–583.

de la Vega, M. C. L. (2010). *Studies in Applied Welfare Analysis: Papers from the Third ECINEQ Meeting*, volume (Research on Economic Inequality, Volume 18), chapter Counting poverty orderings and deprivation curves, pages 153 – 172. Emerald Group Publishing Limited.

Dearden, L., Fitzsimons, E., Goodman, A., and Kaplan, G. (2008). Higher education funding reforms in england : the distributional effects and the shifting balance of costs. *The Economic Journal*, 118(526):F100–F125.

Decancq, K. (2014). Copula-based measurement of dependence between dimensions of well-being. *Oxford Economic Papers*, 66(3):681–701.

Decancq, K., Goedeme, T., den Bosch, K. V., and Vanhille, J. (2013a). *Reconciling Work and Poverty Reduction: How Successful Are European Welfare States?*, chapter The Evolution of Poverty in the European Union: Concepts, Measurement, and Data. Oxford University Press.

Decancq, K. and Lugo, M. A. (2012). Inequality of wellbeing: A multidimensional approach. *Economica*, 79(316):721–746.

Decancq, K. and Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, 32(1):7–34.

Decancq, K., Ootegem, L. V., and Verhofstadt, E. (2013b). What if we voted on the weights of a multidimensional well-being index? an illustration with flemish data. *Fiscal Studies*, 34(3):315–332.

Deheuvels, P. (2009). A multivariate bahadur–kiefer representation for the empirical copula process. *Zapiski Nauchnykh Seminarov POMI*, 364:120–147.

Detollenaere, J., Desmarest, A.-S., Boeckxstaens, P., and Willemsc, S. (2018). The link between income inequality and health in europe, adding strength dimensions of primary care to the equation. *Social Science and Medicine*, 201:103–110.

dos Santos Silva, R. and Lopes, H. F. (2008). Copula, marginal distributions and model selection: a bayesian note. *Statistics and Computing*, 18(3):313–320.

Duclos, J.-Y., Sahn, D. E., and Younger, S. D. (2006). Robust multidimensional poverty comparisons. *The Economic Journal*, 116(514):943–968.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). *Risk Management: Value at Risk and Beyond*, chapter Correlation and Dependence in Risk Management: Properties and Pitfalls, page 176–223. Cambridge University Press.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5).

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Cambridge, Policy Press.

European Commission (2013). Health inequalities in the eu - final report of a consortium. consortium lead: Sir michael marmot. Technical report, European Commission Directorate-General for Health and Consumers.

Fan, Y. and Patton, A. J. (2014). Copulas in econometrics. *The Annual Review of Economics*, 6:179–200.

Ferreira, F. H. G. (2011). Poverty is multidimensional. but what are we going to do about it? *The Journal of Economic Inequality*, 9(3):493–495.

Ferreira, F. H. G. and Lugo, M. A. (2013). Multidimensional poverty analysis: Looking for a middle ground. *The World Bank Research Observer*, 28(2):220–235.

Ferrera, M. (1996). The 'southern model' of welfare in social europe. *Journal of European Social Policy*, 6(1):17–37.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3):761–766.

Foster, J. E. (2006). *Poverty, Inequality and Development. Essays in Honor of Erik Thorbecke*, chapter Poverty Indices, pages 41–65. Springer US.

Foster, J. E., Lopez-Calva, L. F., and Szekely, M. (2005). Measuring the distribution of human development: methodology and an application to mexico. *Journal of Human Development*, 6(1):5–25.

Foster, J. E., McGillivray, M., and Seth, S. (2013). Composite indices: Rank robustness, statistical association, and redundancy. *Econometric Reviews*, 32(1):35–56.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25.

Furnée, C. A., Groot, W., and Pfann, G. A. (2011). Health and income: a meta-analysis to explore cross-country, gender and age differences. *European Journal of Public Health*, 21(6):775–780.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semi parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Genest, C., Nešlehová, J., and Ghorbal, N. B. (2011). Estimators based on kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2):157–177.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043.

Geyer, S. (2011). Income, income, or income? the effects of different income measures on health in a national survey. *Journal of Epidemiology and Community Health*, 65(6):491–496.

Gibbons, J. D. (1993). *Nonparametric measures of association.* Quantitative Applications in the Social Sciences. Sage Publications, Inc.

Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(1):61–94.

Gudendorf, G. and Segers, J. (2010). *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, chapter Extreme-Value Copulas, pages 127–145. Springer-Verlag Berlin Heidelberg.

Guio, A.-C., Fusco, A., and Marlier, E. (2009). A european union approach to material deprivation using eu-silc and eurobarometer data. *IRISS Working Paper 2009-19*.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126(S1):S197–S246.

Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes. Statistical Theory and Applications in Science, Engineering and Economics*, 3(3):279–290.

Hougaard, P. (1986). A class of multivanate failure time distributions. *Biometrika*, 73(3):671–678.

Huard, D., Évin, G., and Favre, A.-C. (2006). Bayesian copula selection. *Computational Statistics & Data Analysis*, 51(2):809–822.

ISTAT (2018). Rapporto bes 2018: Il benessere equo e sostenibile in italia. Technical report, L'Istituto nazionale di statistica.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Springer Science+Business Media.

Joe, H. (2015). *Dependence Modeling with Copulas*. CRS Press.

Kakwani, N. C. (1980). *Income inequality and poverty : methods of estimation and policy applications*. New York: Oxford University Press.

Karlsdotter, K., Martín, J. J. M., and del Amo González, M. P. L. (2012). Multilevel analysis of income, income inequalities and health in spain. *Social Science and Medicine*, 74(7):1099–1106.

Kemptner, D., Jürges, H., and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from germany. *Journal of Health Economics*, 30(2):340–354.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6):2836–2850.

Kobus, M. (2014). Multidimensional polarization for ordinal data. *ECINEQ Working Paper 326*, pages 1–21.

Kobus, M. and Kurek, R. (2017). Copula-based measurement of interdependence for discrete distributions. *ECINEQ Working Paper 431*.

Kojadinovic, I. and Yan, J. (2010). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, 47(1):52–63.

Kolm, S.-C. (1977). Multidimensional egalitarianisms. *The Quarterly Journal of Economics*, 91(1):1–13.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284).

Kurkowiak, B., Hametner, M., Dimitrova, A., Dimitrova, A., Pisano, U., Gschwend, E., Reutimann, J., Rüegge, B., Zandonella, R., Graf, A., Gerstetter, C., Knoblauch, D., and Umpfenbach, K. (2015). Sustainable development in the european union. 2015 monitoring report of the eu sustainable development strategy. Technical report, European Union.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5):1137–1153.

Maasoumi, E. and Lugo, M. A. (2008). *Quantitative Approaches to Multidimensional Poverty Measurement*, chapter The Information Basis of Multivariate Poverty Assessments, pages 1–29. Palgrave Macmillan UK.

Maggino, F. (2017). *Complexity in Society: From Indicators Construction to their Synthesis*, chapter Dealing with Syntheses in a System of Indicators, pages 115–137. Springer International Publishing.

Marcantoni, E. (2014). *Collateralized Debt Obligations. A Moment Matching Pricing Technique based on Copula Functions*. BestMasters. Gabler Verlag.

Mari, D. D. and Kotz, S. (2001). *Correlation and Dependence*. Imperial College Press.

Maricic, M., Egea, J. A., and Jeremic, V. (2019). A hybrid enhanced scatter search—composite i-distance indicator (ess-cidi) optimization approach for determining weights within composite indicators. *Social Indicators Research*, 144(2):497–537.

Marmot, M., Friel, S., Bell, R., Houweling, T. A., and Taylor., S. (2008). Closing the gap in a generation: health equity through action on the social determinants of health. Technical report, Commission on Social Determinants of Health. Geneva, World Health Organization.

McGillivray, M. (1991). The human development index: Yet another redundant composite development indicator? *World Development*, 19(10):1461–1468.

Mincer, J. (1974). *Schooling, Experience and Earnings*. Columbia University Press for National Bureau of Economic Research, New York.

Mitra, S. and Brucker, D. L. (2017). Income poverty and multiple deprivations in a high-income country: The case of the united states. *Social Science Quarterly*, 98(1):37–56.

Montenegro, C. E. and Patrinos, H. A. (2014). Comparable estimates of returns to schooling around the world. Policy Research Working Paper 7020, World Bank.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD publishing.

Nelsen, R. B. (1996). *Distribution with Fixed Marginals and Related Topics. IMS Lecture Notes – Monograph Series*, volume 28, chapter Nonparametric Measures of Multivariate Association, pages 223–232. Institute of Mathematical Statistics, Hayward, CA.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag New York, 2 edition.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Copula model evaluation based on parametric bootstrap. *Computational Statistics and Data Analysis*, 52:3342–3353.

Nilsson, T. (2010). Health, wealth and wisdom: Exploring multidimensional inequality in a developing country. *Social Indicators Research*, 95(2):299–323.

Nussbaum, M. C. (2011). *Creating Capabilities: The Human Development Approach*. Harvard University Press.

Oreopoulos, P. and Petronijevic, U. (2013). Making college worth it: A review of the returns to higher education. *Future of Children*, 23(1):41–65.

Oreopoulos, P. and Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, 25(1):159–184.

Paruolo, P., Saisana, M., and Saltelli, A. (2013). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society*, 176(3):609–634.

Pèrez, A. and Prieto-Alaiz, M. (2016). Measuring the dependence among dimensions of welfare: A study based on spearman's footrule and gini's gamma. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 24(1):87–105.

Permanyer, I. (2011). Assessing the robustness of composite indices rankings. *The Review of Income and Wealth*, 57(2):306–326.

Permanyer, I. (2012). Uncertainty and robustness in composite indices rankings. *Oxford Economic Papers*, 64(1):57–79.

Permanyer, I. (2014). Assessing individuals' deprivation in a multidimensional framework. *Journal of Development Economics*, 109:1–16.

Pickett, K. E. and Wilkinson, R. G. (2015). Income inequality and health: A causal review. *Social Science and Medicine*, 128:316–326.

Pinar, M. (2019). Multidimensional well-being and inequality across the european regions with alternative interactions between the well-being dimensions. *Social Indicators Research*, 144(1):31–72.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60(310):516–522.

Psacharopoulos, G. and Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, 26(5):445–458.

Quinn, C. (2007). Using copulas to measure association between ordinal measures of health and income. *HEDG Working Paper 07/24*. University of York.

Randall, C., Cochrane, A., Jones, R., and Manclossi, S. (2019). Measuring national well-being in the uk: international comparisons, 2019. Technical report, National Statistics Office of the UK.

Ravallion, M. (2011). On multidimensional indices of poverty. *The Journal of Economic Inequality*, 9(2):235–248.

Rippin, N. (2010). Poverty severity in a multidimensional framework: The issue of inequality between dimensions. *Courant Research Centre, Discussion Paper 47, University of Goettingen*, University of Goettingen. University of Goettingen.

Rodríguez-Pose, A. and Tselios, V. (2008). Education and income inequality in the regions of the european union. Discussion Paper 11, Spatial Economics Research Centre.

Roine, J., Vlachos, J., and Waldenström, D. (2009). The long-run determinants of inequality: What can we learn from top income data? *Journal of Public Economics*, 93(7-8):974–988.

Rözer, J. J. and Volker, B. (2016). Does income inequality have lasting effects on health and trust? *Social Science and Medicine*, 149:37–45.

Saaty, R. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling*, 9(3-5):161–176.

Saisana, M., Saltelli, A., and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):307–323.

Saisana, M. and Tarantola, S. (2002). State-of-the-art report on current methodologies and practices for composite indicator development. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/9253d939-b47b-4428-b792-619e6b6c8645.

Santos, M. E. and Villatoro, P. (2018). A multidimensional poverty index for latin america. *Review of Income and Wealth*, 64(1):52–82.

Schlossarek, M., Syrovátka, M., and Vencálek, O. (2019). The importance of variables in composite indices: A contribution to the methodology and application to development indices. *Social Indicators Research*, page 1–36.

Schmid, F. and Schmidt, R. (2007a). Multivariate extensions of spearman's rho and related statistics. *Statistics and Probability Letters*, 77(4):407–416.

Schmid, F. and Schmidt, R. (2007b). Nonparametric inference on multivariate versions of blomqvist's beta and related measures of tail dependence. *Metrika*, 66(3):323–354.

Schweizer, B. (1991). *Advances in Probability Distributions with Given Marginals*, chapter Thirty Years of Copulas, pages 13–50. Springer Netherlands.

Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879–885.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, 44(2):219–231.

Sen, A. (1999). *Commodities and Capabilities.* Oxford India Paperbacks. Oxford University Press.

Siegel, M., Vogt, V., and Sundmacher, L. (2014). From a conservative to a liberal welfare state: Decomposing changes in income-related health inequalities in germany, 1994–2011. *Social Science and Medicine*, 108:10–19.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leur marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231.

Somarriba, N. and Pena, B. (2009). Synthetic indicators of quality of life in europe. *Social Indicators Research*, 94(1):115–133.

Stiglitz, J., Sen, A., and Fitoussi, J. (2009). Report by the commission on the measurement of economic performance and social progress. Technical report.

Stronks, K., van de Mheen, H., van den Bos, J., and Mackenbach, J. (1997). The interrelationship between income, health and employment status. *International Journal of Epidemiology*, 26(3):592 – 600.

Tenn, S., A.Herman, D., and Wendling, B. (2010). The role of education in the production of health: An empirical analysis of smoking behavior. *Journal of Health Economics*, 29(3):404–417.

Trivedi, P. K. and Zimmer, D. M. (2007). Copula modeling: An introduction for practitioners. *Foundations and Trends® in Econometrics*, 1(1):1–111.

Tsui, K.-Y. (2002). Multidimensional poverty indices. *Social Choice and Welfare*, 19(1):69–93.

UNDP (2016). Human development report: Human development for everyone. Technical report, Human Development Report Office, New York: USA.

UNDP (2019). Human development data (1990-2017).

UNESCO (2012a). International standard classification of education isced 2011. Technical report, UNESCO Institute for Statistics.

UNESCO (2012b). Isced mappings. http://uis.unesco.org/en/isced-mappings. Accessed: 2018-03-10.

Wu, S., Wang, R., Zhao, Y., Ma, X., Wu, M., Yan, X., and He, J. (2013). The relationship between self-rated health and objective health status: a population-based study. *BMC Public Health*, 13(320).

Zheng, B. (1993). An axiomatic characterization of the watts poverty index. *Economics Letters*, 42:81–85.