INE PAN Working Paper Series

Paper number 41

Modelling health indicators in a joint framework via factor copula models

Martyna Kobus, Olga Półchłopek

Warsaw, June 2016

wp@inepan.waw.pl

Modelling health indicators in a joint framework via factor copula models

Martyna Kobus[*][†], Olga Półchłopek[‡]

20th June 2016

Abstract

The problems of ageing societies in advanced countries have recently put emphasis on the evaluation of health of the elderly. Health is likely to determine job market activity of the increasing parts of the society. Appropriate modelling of health conditions is therefore key for policymaking, in particular given that detailed health data are now available via ageing surveys. Thus there has been recently interest in modelling multiple ordinal health data. Makdisi and Yazbeck (2014) utilise the counting approach, which requires the transformation of multiple category health indicators into binary and lead to the loss of information, but also changes the dependence structure. We offer a different approach which does not have these limitations and is feasible for high-dimensional data (e.g. stochastic dominance methods are inconclusive for many dimensions (Duclos and Echevin 2012). We use recently developed methods based on so called vine pair-copula constructions (PCC) (Aas et al. 2009). We estimate a 1-factor copula model (Nikoloulopoulos and Joe 2015) for 24 health indicators taken from English Longitudinal Study of Ageing (ELSA) such as self-reported health status, mobility, eyesight, hearing and pain rating and questions related to emotional health. We show that there are substantial interdependencies in health data which cannot be neglected by dichotomisation and aggregation, nor can they be detected by the standard multivariate probit model. $t(4)-$ and $t(5)-$ factor copula model provides the best fit, and items that measure general optimism are most informative of the underlying factor. Groups are most heterogeneous along the employment status, with retired and disabled groups showing significantly more dependence than other groups in items related to mobility and general health status.

Keywords: multiple health indicators; interdependence; factor copulas; vine copulas
JEL codes: I31; D63

Introduction

Nowadays there are increasingly many health indicators available in health and ageing surveys. Such surveys include detailed information on functional limitations, cognitive, emotional and mental health, and health measurements such as hypertension, biomarkers etc. However, there are two major measurement problems in using this rich information. Firstly, researchers often restrict to self-assessed health only (van Doorslaer et al. 1997, van Doorslaer and Koolman 2004, Kunst et al. 2004, Cutler et al. 2015) or they aggregate various health indicators into a single index of health (e.g. Makdisi and Yazbeck 2014) or analyse various health indicators separately. This means that the dependence structure is ignored which means the loss of substantial part of the information on the health distribution. With fixed margins, generally, the more association between dimensions of wellbeing, the more inequality (Atkinson and Bourguignon 1982), because of higher likelihood for individuals to suffer from multiple deprivations. Interdependencies thus help to identify opportunities for policy intervention with the largest efficiency gains. This is particularly relevant for patients suffering from multiple chronic diseases, which given ageing population will become even more of a focus for public health policy and spending.

[*]Institute of Economics Polish Academy of Sciences, Nowy Swiat 72, 00-330 Warszawa; email: mkobus@inepan.waw.pl
[†]This work was supported by National Science Centre grant number 2013/09/D/HS4/01029, and grant MNiD carried out at INE PAN.
[‡]Vistula University, Stoklosy 3, 02-787 Warsaw, olga.polchlopek@gmail.com

Secondly, aggregation imposes some form of cardinalization of an often ordinal indicator (e.g. Mackenbach et al. 2008). Many health indicators are ordinal e.g. widely used self-reported health status, problems with vision, hearing, communication, speech, cognition, the feeling of pain, anxiety, depression etc. Cardinalization of ordinal variables has been criticised by Allison and Foster (2004). The choice of a particular numerical transformation (i.e. scale) of an ordinal indicator is arbitrary, on the other hand, the mean and the variance are not invariant to monotone transformations. Therefore the results may be often easily reversed when a different scale is chosen as examples by Allison and Foster (2004), Apouey and Silber (2013), Bond and Lang (2014) and Kobus (2015) show. This is the main drawback of the standard approach to measuring socioeconomic inequalities in health, namely, concentration curves, which is developed for ratio-scale variables and should not be used directly with ordinal data, as rightly criticised by Makdisi and Yazbeck (2014). Inequality measurement theory has been extended to account for ordinality of the data, but so far mostly for one health indicator (Allison and Foster 2004, Apouey 2007, Abul Naga and Yalcin 2008, Kobus and Miłoś 2012, Apouey and Silber 2013, Lazar and Silber 2013, Abul Naga and Stapenhurst 2015, Cowell and Flachaire 2015, Gravel et al. 2015, Lv et al. 2015, Kobus 2015). There is very few contributions up to date for multivariate health data: Sonne Schmidt et al. (2015) propose criteria for comparing distributions which are only tractable in the case of two binary indicators, so their applicability is very restricted. Duclos and Echevin (2012) propose a robust method for measuring health-income gradient based on dominance conditions. However, it is unlikely that dominance conditions are conclusive for 40 health indicators which are available for example in ageing surveys. Therefore, while elegant, these methods may provide little help for applied health economists.

The most comprehensive treatment of multiple health categories was offered recently by Makdisi and Yazbeck (2014). They use the insight from the counting approach to multidimensional poverty to measure the width of health problems i.e. the number of health problems. This further requires transforming each health indicator into a binary variable. They analyse US and Canadian data on vision, hearing, speech, emotion and other health indicators and transform them into 0-1 variables. For example, vision problems are recorded as 1 (the person has vision problems) if he or she has difficulty seeing with glasses. Dichotomization leads to a substantial loss of information. Moreover, it obscures the dependence structure. The goal of this paper is to show that interdependencies should not be easily neglected in case of categorical health data. To this end, we use recently developed factor copula models (Nikoloulopoulos and Joe 2015). These models omit some of the difficulties we mentioned, because they allow for a flexible modelling of multivariate distributions and are appropriate for modelling high-dimensional data. Although we cannot fully account for the ordinality of health data (i.e. in the margins), copula approach enables scale-free modelling of the dependence structure. We utilise 24 health indicators from Wave 1 and 6 of English Longitudinal Study of Ageing (ELSA) and provide evidence for stronger dependence of health indicators for both lack of health problems (lower tail dependence) and severe health problems (upper tail dependence). Thus we obtain a complex picture of health among the elderly, which cannot be fully reflected by the counting approach, and is enough to justify a truly multivariate approach to modelling health data.

In more detail, factor copula models we utilise are inspired by the modelling approach based on pair-copula constructions (PCC). PCC allows to model the dependence between each two dimensions via bivariate copulas which are building blocks of a multivariate copula. Such decomposition of a multivariate distribution into bivariate elements lowers substantially the level of integration[1] and makes maximum likelihood estimation feasible for high-dimensional data. Furthermore, distributions with tail dependence and asymmetric tail dependence can be accommodated using PCC, and such detailed picture of various interdependencies present in the data provides key information for efficient targeting of most vulnerable groups (multiply deprived). As Atkinson (2011, p. 326) states: "The copula diagram is helpful in thinking about policies to moderate the social gradient of health (...)." For example, let us assume that it turns out that for low socioeconomic status (SES) correlation between BMI and diabetes is stronger than for people with high SES. Yet the true picture may be that in fact for low BMI, SES is not important and dependence between BMI and diabetes is the same independent of SES, whereas it is for high BMI that dependence between BMI and diabetes is much higher for low SES than for high SES, so intervention should target only this group. That is to say, dependence in lower tail (low BMI) is different from the dependence in upper tail (high BMI); methods based on multivariate normal distribution and correlation cannot detect this. As another example, it is likely that

[1]Denoting by d the number of dimensions, instead of d-dimensional integral, with PCC one deals with d unidimensional integrals.

dependence between cognitive and mental health indicators might be different from the dependence between physical and mental health indicators; a modelling approach should allow for such heterogeneity.

The methods used here can be thought of as a generalisation of a standard approach to multivariate modelling based on multivariate normal distribution (MVN). MVN is a good premise, as many phenomena in nature evince Gaussian distribution. It arises as a limit of a scaled sum of weakly dependent random variables with no variable dominating i.e. Central Limit Theorem. It is constructed using the property that any linear combination of independent Gaussian random variables is Gaussian. As such it is simple and tractable. However, its limitations such as a requirement of normal margins, lack of negative dependence and tail dependence, are important from the point of view of applications. Overuse of Gaussian distribution in risk modelling is criticised, especially in the last few years e.g. it was even called "The formula that killed Wall Street" in a famous article by Felix Salmon published in Wired magazine. Particularly in finance, income dispersion is typically uneven or there may exist stock market shares which are dependent only when they reach high values. To model tail dependence appropriately is key for effective portfolio diversification. Tail dependence is a property that indicates how the joint probability behaves in extreme, low or high, values. Upper(lower) tail dependence means that large(low) values of two or more variables occur together more often. Multivariate Gaussian distribution does not have tail dependence, mass is clustered in the centre, tails are symmetrical and carry little mass, so they may underestimate the risk. These problems occur both on univariate and joint levels, so the mismatch caused by applying MVN is compounded when the number of dimensions increases. Therefore non-Gaussian models might be more appropriate in risk analysis, insurance, finance and economics. The study of copulas is one way to enter a non-Gaussian world.

A copula is a multivariate probability distribution function with uniform marginals. It is popular due to the celebrated Sklar's theorem (Sklar 1959) which states that a copula and marginal distribution characterise the joint distribution. For continuous distributions such representation is unique. Thus copula can be thought of as a method to construct joint distributions from marginal distributions. This modelling approach has the advantage of having univariate margins of different types (e.g. with potentially different scale/shape parameters) and the dependence structure can be modelled separately from univariate margins which leads to more efficient estimation. It is indeed much more flexible than well-known families of multivariate distributions such as normal, t, Pareto and other. Thus copulas allow for two-step estimation where the two steps are independent. First, univariate margins are chosen and they can be any e.g. univariate Gaussian and t for retaining symmetry, gamma for exponential tails, Pareto for heavy tails, Poisson for integer-valued data etc. Uniform margins are obtained via probability integral transform and in the second step copula is fit to the joint data. It can be chosen flexibly to model different types of dependence structure (e.g. exchangeable, conditional independence, both positive and negative dependence) and different types of joint tail behaviour (e.g. upper tail dependence, lower tail dependence, asymmetric tail dependence).

Despite theoretical elegance, building high-dimensional copulas is considered a difficult problem (Aas et al. 2009). However, this problem has been addressed recently by proposing the called vine pair copula constructions (Aas et al. 2009). The principle is to decompose a multivariate distribution into a series of bivariate copulas applied to original variables and to their conditional distribution functions. An example will help to present this concept. Let $\mathbb{X} = (X_1, X_2, X_3)$. Its density $f(x_1, x_2, x_3)$ can be factorised i.e. $f(x_1, x_2, x_3) = f_1(x_1)f(x_1|x_2)f(x_1|x_2, x_3)$. On the other hand, given that $F(x_1, x_2, x_3) = C(F_1(x_1), F_2(x_2), F_3(x_3))$ we can write joint density as

$$f(x_1, x_2, x_3) = c_{123}(F_1(x_1), F_2(x_2), F_3(x_3))f_3(x_3)f_2(x_2)f_1(x_1),$$

where $c_{123}(F_1(x_1), F_2(x_2), F_3(x_3))$ is a 3-variate copula density. Further, we get $f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2)$ and $f(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)} = c_{12}(F_1(x_1), F_2(x_2))f_1(x_1)$. We can also express $f(x_1|x_2, x_3)$ as

$$f(x_1|x_2, x_3) = c_{13|2}(F(x_1|x_2), F(x_3|x_2))f(x_1|x_2) =$$

$$c_{13|2}(F(x_1|x_2), F(x_3|x_2))c_{12}(F_1(x_1), F_2(x_2))f_1(x_1).$$

Continuing this we arrive at

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)c_{12}(F_1(x_1), F_2(x_2))c_{23}(F_2(x_2), F_3(x_3))\times$$

$$\times c_{13|2}(F(x_1|x_2), F(x_3|x_2)).$$

Altogether, this means that we can model multivariate distribution by having bivariate copulas as building blocks. Dependence between each two dimensions can be modelled flexibly and computation is fast. Such decomposition can be presented graphically (Bedford 2002) and is called the *(regular)* vine. Our vine has 3 variables and two trees: first tree in which variables 1 and 2 as well as 2 and 3 are connected via bivariate linking copulas, and second tree in which variables 1 and 3 are connected via a bivariate linking copula conditional on 2. If at some tree bivariate linking copulas become independence copula we say that the vine is truncated. Factor copulas (Nikoloulopoulos and Joe 2015) presented here are such truncated vines, however (Nikoloulopoulos and Joe 2015) develop them as conditional independence model and we follow this presentation.

We find that factor copula models based on $t(4)$ and $t(5)$ copulas provide better fit than standard multivariate normal model, which is a special case of models developed here. This suggest that our variables are generated as a mixture of discreticised means which is typical for a sample which is a mixture of heterogeneous groups. Furthermore, this suggests that there is enough dependence structure in categorical health data that cannot be easily neglected. Such interdependencies cannot be detected with the counting approach and we show a simple example of how such an approach obscures the dependence structure. These interdependencies are present in all analysed group distributions which supports a joint modelling framework for health data. The strongest dependence is with the items that express general optimism and such is the interpretation of the underlying factor. The most heterogenous are groups defined along the employment status dimension. For retired and disabled people the strongest dependence with the underlying factor is observed for self-reported health status and mobility, whereas for the employed group items seem to be linked to the factor more equally. Altogether this paper shows that there are substantial dependencies in health data which cannot be easily neglected. Furthermore, the modelling approach offered here is feasible for high-dimensional data.

The paper is organised as follows. Section 1 presents copula theory, with particular emphasis on ordinal data and factor copula models. In Section 2 we present the dataset and the results of estimating 1-factor copula models. Finally, we conclude (Section 4) by pointing to future research on improving modelling of high-dimensional health distributions..

1 Theoretical model and its estimation

Before we present factor copula models that are suited for ordinal data we start with basic definitions related to copula theory. We start with bivariate copulas, but generalisations to multivariate copulas are straightforward (Nelsen 2006). The presentation of copula theory is based on (Nelsen 2006) and (Joe 2015).

1.1 Definitions

Let F, G be cumulative distribution functions. Let $X \sim F$ and $Y \sim G$. It is a well-known fact that if $X \sim F$ is a continuous random variable, then $F(X) \sim U(0, 1)$. If $U \sim U(0, 1)$ and F is a univariate cdf and F^{-1} is its generalised inverse, then $X = F^{-1}(U) \sim F$. Here X can be continuous or discrete. Moreover, for $U \sim U(0, 1)$ and $V \sim U(0, 1)$ we obtain by the inversion two independent samples from F and G, which are $X = F^{-1}(U)$ and $Y = G^{-1}(V)$ respectively.

Definition 1. *Let $U \sim U(0, 1)$, $V \sim U(0, 1)$, $X \sim F$ and $Y \sim G$. Additionally, let $(X, Y) \sim H$. If H is continuous, then $(U, V) = (F(X), G(Y)) \sim C$ where C is a copula.*

By definition, there are two marginal variables with cdfs and a function that combines these cdfs to return bivariate distribution. A copula is therefore a bivariate, or, in general, a multivariate distribution function whose margins are uniform. By Sklar's theorem, such function exists and is unique for continuous variables:

Theorem 1. *Let H be a bivariate cdf with univariate marginal cdfs, F and G. Then, there exists a bivariate copula C such that $H(x, y) = C(F(x), G(y))$ for all $(x, y) \in \mathbb{R}$. And conversely, if F and G are univariate continuous random variables, then $C(F(x), G(x))$ is a bivariate distribution for (X, Y) with marginal distributions F and G, respectively.*

Theorem 1 ensures uniqueness of copula only for continuous distributions. For discrete distributions there is a whole set of copulas which agree with Sklar's representation $H(x,y) = C(F(x),G(y))$. The problems related to copulas for discrete distributions are described in (Genest, Nešlehová 2007). Despite numerous problems, these authors do conclude that "copula modelling remains a valid option for constructing multivariate distributions with discrete margins" (pp. 507) and their article in fact shows that rank-based inference for copula parameters is not recommended in the discrete case. Here we deal with ordinal data that have mostly discrete distributions.[2] To be precise, we deal with data that are ordinal i.e. invariant to monotone transformations, and discrete i.e. probability mass is concentrated on a finite number of points. In the discrete case copula function is determined uniquely on intersections of $F(x), G(y)$ where $x = 0, ..., K-1$ and $y = 0, ..., L-1$ with K and L being the number of categories of, respectively, x and y. We have:

$$F(y_1, ..., y_d) = C(F_1(y_1), ..., F_d(y_d)) \text{ for } y_j \in \{0, ..., K-1\} \tag{1}$$

which leads to

$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)) \text{ for } u_j \in [0,1] \tag{2}$$

Copulas are particularly well-suited to study the dependence of ordinal data. As (Schweizer 1981) note: "it is precisely the copula which captures those properties of the joint distribution which are invariant under (...) strictly increasing transformations". With ordinal variables the only relevant information we have is about the ordering, therefore it is a very desirable property of a dependence measure to not change when the variable is transformed by a monotone transformation. Not surprisingly, copula which takes marginal cdf values has this property i.e. if C_{XY} is a copula function corresponding to the bivariate distribution of random variables X, Y and f, g are strictly monotonic continuous functions, then $C_{f(X)g(Y)} = C_{XY}$. For other monotonic transformations of X, Y the copula's behaviour is also known: for f increasing and g decreasing $C_{f(X)g(Y)}(u,v) = u - C_{XY}(u, 1-v)$ and for both f and g strictly decreasing it holds that $C_{f(X)g(Y)}(u,v) = u + v - 1 + C_{XY}(1-u, 1-v)$.

As mentioned, copulas enable appropriate modelling of tail dependence. We will now define these concepts formally. A bivariate copula C is reflection symmetric if for all $0 \leq u_1, u_2 \leq 1$ and density $c(u_1, u_2) = \frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2}$ it holds that $c(u_1, u_2) = c(1-u_1, 1-u_2)$. Otherwise, we say that C is reflection asymmetric and has more probability in the upper or lower tail. Upper tail dependence focuses on the upper quadrant of the distribution. It is measured by the upper tail dependence coefficient $\lambda_U \in [0,1]$:

$$\lambda_U = lim_{u \to 1^-} \left(\frac{C(u,u)}{1-u} \right) \tag{3}$$

If the limit exists and equals 0, then C does not have upper tail dependence. If $\lambda_U \in (0,1]$, then C has upper tail dependence. Similarly, if the limit given by:

$$\lambda_L = lim_{u \to 0^+} \left(\frac{C(u,u)}{u} \right) \tag{4}$$

exists, then C has lower tail dependence for $\lambda_L \in (0,1]$ and no lower tail dependence for $\lambda_L = 0$.

Only scale-invariant measures of association are suitable for ordinal data. These are, for example, widely known association coefficients such as Kendall's tau or Spearman'r rho. They are in fact measures based on the copula. Here we use Kendall's tau.

Definition 2. *(Nelsen 2006) Let X and Y be continuous random variables with a copula C. Then the population version of Kendall's τ for X and Y is given by*

$$\tau(X,Y) = -1 + 4 \int_0^1 \int_0^1 C(u,v) dC(u,v)$$

The sample version of Kendall's tau is defined in terms of concordance. Each pair of observations from the sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ is either concordant (i.e. $x_i > x_j, y_i > y_j$ or $x_i < x_j, y_i < y_j$) or discordant (i.e. $x_i > x_j, y_i < y_j$ or $x_i < x_j, y_i > y_j$). Kendall's tau is then the difference in the number of concordant

[2]This is not always the case e.g. BMI is an ordinal variable, test scores too, and they both have continuous distributions.

and discordant pairs divided by the number of all pairs. Here we use Kendall's tau because copula models with different parameters are not directly comparable, therefore parameter estimates and standard errors need to be compared on a Kendall's tau scale.

The dependence parameters θ vary across copula families. In order to compare different models, it is necessary to calculate Kendall's τ. The following transformation of copula parameters applies to Gaussian t copulas with $\theta \in (-1, 1)$ (Hult 2002)

$$\tau = \frac{2}{\pi} arcsin\theta, \tag{5}$$

whereas the following transformation applies to Gumbel copula with $\theta \in [1, \infty)$ (Genest 1986)

$$\tau = 1 - \frac{1}{\theta}. \tag{6}$$

Therefore, Kendall's τ for Gaussian and t copulas range from -1 to 1 and as for Gumbel, $\tau \in (0, 1)$, because this copula models only positive dependence.

1.2 Parametric copula families

Parametric copula families are most commonly used for their convenient properties (Joe 2015) i.e. they are easy to implement numerically as they have closed forms that do not contain integrals. Three copula families are considered due to their distinctive attributes: multivariate normal copulas, t copulas and Gumbel copulas, as well as related copulas, such as survival Gumbel (i.e. Gumbel copula rotated by 180 degrees to model lower tail dependence instead of upper tail dependence characteristic for Gumbel). Normal and t copulas are called elliptical copulas and Gumbel is a special case of an Archimedean copula. In describing various data generating processes that lead to particular families of copulas we follow intuition given by (Nikoloulopoulos and Joe 2015).

Let d be the number of (ordinal) variables whose multivariate distribution we would like to model. The first family, further referred to as Gaussian, presents reflection symmetry and no tail dependence which means that it is suitable when the data has been generating through some process of averaging (Nikoloulopoulos and Joe 2015) e.g. in answering a question a respondent is taking an average experience from all the events in his life relevant to this particular question. Gaussian copula is a good modelling choice when we expect the majority of respondents to fall into the middle categories with little fluctuation. Gaussian copula is defined in the following way

$$C(\vec{u}; \Sigma) = \Phi_d(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d); \Sigma) \tag{7}$$

where $\vec{u} \in [0, 1]^d$ and Σ is a $d \times d$ correlation matrix that serves as a parameter. For a bivariate copula, there is only one parameter $\theta \in (-1, 1)$:

$$C(u, v; \theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta) \tag{8}$$

There is always one parameter needed for each pair of variables. For Gaussian copula, both $\lambda_L = \lambda_U = 0$.

t copulas, as opposed to Gaussian copulas, cluster more probabilistic mass on the tails, simultaneously keeping the focus in the centre. Not only is it possible to set different degrees of freedom ν for the margins, but also to control skewness. t copulas model answers which are generated as combinations of means. This happens when respondents are of mixed populations, e.g. differ in sex or locations. Let T be univariate Student's t distribution cdf with ν degrees of freedom. Then, a multivariate t copula has the following form:

$$C(\vec{u}; \Sigma, \nu) = T_{d,\nu}(T_\nu^{-1}(u_1), ..., T_\nu^{-1}(u_d); \Sigma) \tag{9}$$

where Σ is positive definite parametric matrix. With $\nu \to \infty$, a Gaussian copula is obtained and for small values of ν there is more probability in the joint upper and lower tails. For the bivariate case, $\Sigma = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$ where $-1 < \theta < 1$ and the copula cdf is:

$$C(u, v; \theta, \nu) = T_{2,\nu}(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta) \tag{10}$$

Tails are symmetrical and the dependence coefficients can be written as:

$$\lambda_U = \lambda_L = 2T_{\nu+1}\left(-\frac{\sqrt{\nu+1}\sqrt{1-\theta}}{\sqrt{1+\theta}}\right) \tag{11}$$

where T stands for t Student cdf.

Gumbel copula family, on the other hand, is characterised by larger dependence in the upper tail, meaning that probability mass of the joint distribution is shifted towards the extreme values. These copulas capture only positive dependence. If two margins of K categories display negative dependence, one of them should be recoded from $K-1$ to 0 in order to use Gumbel copula. Such extreme value copulas are a good fit for responses derived from best-case or worst-case scenarios. E.g. when asked about mobility limitation, the respondent takes into account only events when his or her disability prevented them from performing some actions and based on this he or she chooses lower categories. In other words, he or she takes the minimum value of all the events relevant to the question.

Bivariate Gumbel copula with parameter θ is the following:

$$C(u,v;\theta) = e^{-((-lnu)^\theta+(-lnv)^\theta)^{\frac{1}{\theta}}} \tag{12}$$

where $1 \leq \theta < \infty$.

We can rotate a Gumbel copula by 180° and thus obtain lower tail dependence; such transformation is called survival Gumbel or reflected Gumbel copula. If $(U,V) \sim C$ for bivariate C, then $(1-U, 1-V) \sim C_r$ and $C_r(u,v) = u + v - 1 + C(1-u, 1-v)$. Therefore, the survival Gumbel has the following form:

$$C(u,v;\theta) = u + v - 1 + e^{-((-ln(1-u))^\theta+(-ln(1-v))^\theta)^{\frac{1}{\theta}}} \tag{13}$$

Other possibilities include rotations by 90° and 270° to model negative dependence.

$$C_{90^\circ}(u,v) = v - C(1-u, v) \tag{14}$$

$$C_{270^\circ}(u,v) = u - C(u, 1-v) \tag{15}$$

Figure 1: Contour plots of bivariate copulas with different behaviour in the tails

(a) Gaussian (b) $t(5)$ (c) Gumbel

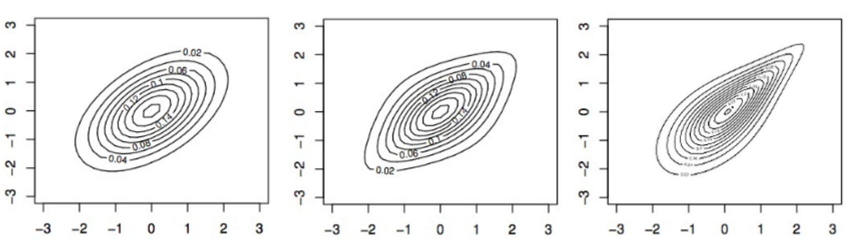


Figure 1 [22, ?] illustrates the support of different copulas. The isolines are sets of points with the same probability. There is more dependence in the tails if the mass is more concentrated around the diagonal. Gaussian copula does not have this property and Gumbel copula has it for higher values i.e. in the upper tail. For t copula tail dependence is symmetric.

1.3 Factor copulas. 1-factor copula model.

Inspired by the vine copula approach, (Nikoloulopoulos and Joe 2015) propose so called factor copula models for multivariate ordinal data. These are latent variable models for the analysis of high-dimensional item response data in which dependence comes from latent (unobservable) factors. This happens, for example, when there are 40-50 items in a questionnaire but some of them are associated as they aim to model the same concept (e.g. anxiety, depression). The theoretical concept is low-dimensional so latent (factor) models are suitable. The most well-known is the standard multivariate normal model. It is a special case of models proposed by (Nikoloulopoulos and Joe 2015) where all bivariate linking copulas are Gaussian copulas. Factor copula models can be interpreted as latent maxima/minima (in comparison to latent means) and as such they place more probability mass in the tails than a model based on multivariate normal distribution. For the first factor there are bivariate copulas that link each observed item (ordinal variable) with the latent factor, and for the second factor there are bivariate copulas that link each observed item with the second factor *conditional* on the first factor. As mentioned, the connection with vine copulas is that such factor models are in fact truncated vine copulas, however, following (Nikoloulopoulos and Joe 2015) we will motivate them as conditional independence models. An important computational advantage of factor copula models and the reason why (Joe 2015) describes them as the best modelling choice for multivariate data, is the need to estimate $\mathcal{O}(d)$ parameters instead of $\mathcal{O}(d^2)$.

Definition 3. *Let $Y_i = (Y_{i1}, ..., Y_{id})$ be a vector of d ordinal variables each measured on a scale $\{0, ..., K-1\}$. The p-factor model assumes conditional independence of $Y_1, ..., Y_d$ given latent variables $X_1, ..., X_p$ (factors). The joint probability mass function (pmf) is therefore:*

$$P(Y_1 = y_1, ..., Y_d = y_d) = \int \prod_{j=1}^{d} P(Y_j = y_j | X_1 = x_1, ..., X_p = x_p) dF_{X_1,...,X_p}(x_1, ..., x_p)$$

where $F_{X_1,...,X_p}$ is the joint distribution of latent factors.

The conditional independence assumption allows us to replace d-dimensional integrals with the multiplication of one-dimensional integrals. Copulas appear in how $P(Y_j = y_j | X_1 = x_1, ..., X_p = x_p)$ is modelled. Here we present and estimate 1-factor copula model. Further research will focus on estimating 2-factor models for health data as well as more complex structured factor copula models (Krupskii and Joe 2015) which preserve the group structure e.g. some items are by design more associated than others. As in a standard model for ordinal data, factors gain their interpretation from the items they connect to the most i.e. via higher values of copula dependence parameters. For example, in case of health data one might expect items such as physical health and mobility to fall in one category (i.e. relate to one underlying factor), whereas items such as depression and anxiety to fall in a different category (i.e. relate to the other underlying factor). The factors are assumed to be independent to ease identifiability.[3]. Factor copula models do not have closed form cdfs, however, each of the d items is connected to each factor with a bivariate parametric copula, and typically no more than 2 factors are used. Moreover, the model allows mixtures of copula families, so the dependence between each two items is modelled flexibly.

A significant feature of factor copulas is that they inherit tail dependence (Hua 2014). If, for example, bivariate distributions of items j, k and factor X, namely, bivariate distributions of (Y_j, X) and (Y_k, X) are characterised by more probability in the upper tail, then (Y_j, Y_k) also has upper tail dependence. Therefore, one can infer from bivariate margins of the distribution of $Y = (Y_1, ..., Y_d)$ which copula families should be used with the data. If upper tail dependence is observed in (Y_j, Y_k), then a Gumbel copula should be used both with (Y_j, X) and (Y_k, X).[4]

Let the cutpoints in the $U(0, 1)$ scale be $a_{jk} = \Phi(\zeta_{jk})$ where ζ_{jk} are corresponding cutpoints in the $N(0, 1)$ scale for $j = 1, ..., d$. Let X be a latent variable, $X \sim U(0, 1)$. Given Theorem 1 there exists a copula C_j such that $F_{(X,Y_j)}(x, y_j) = C_j(x, F_{Y_j}(y_j))$. Then, the conditional cdf is the following

$$F_{Y_j|X}(y_j|x) = P(Y_j = y_j|X = x) = \frac{\partial C_j(x, F_{Y_j}(y_j))}{\partial x} := C_{j|X}(a_{j,y_j+1}|x) \tag{16}$$

and we have that

$$P(Y_1 = y_1, ..., Y_d = y_d) = \int_0^1 \prod_{j=1}^{d} P(Y_j = y_j|X = x)dx = \int_0^1 \prod_{j=1}^{d} (C_{j|X}(a_{j,y_j+1}|x) - C_{j|X}(a_{j,y_j}|x))dx \tag{17}$$

Clearly, $C_{j|X}(a_{j,y_j+1}|x) - C_{j|X}(a_{j,y_j}|x)$ gives the probability of $Y_j = y_j$ conditional on $X = x$. We have d items so d bivariate linking copulas, so d parameters to estimate.

[3] They do not have to be independent for more complex structured factor copula models (Krupskii and Joe 2015)

[4] Or other copulas with properties similar to Gumbel like Galambos or Clayton.

1.4 Two-step estimation

First, for a random vector $\vec{Y_i} = (Y_{i1}, ..., Y_{id})$ with $i = 1, ..., n$ we estimate univariate cutpoints a_{jk} for $j = 1, ..., d$ using empirical distribution:

$$\hat{a}_{j0} = 0,\ \hat{a}_{j1} = p_0,\ \hat{a}_{j2} = p_0 + p_1, ..., \hat{a}_{jK_j} = p_{j0} + ... + p_{j,K_j-1} = 1 \tag{18}$$

with p_k being a sample proportion, $p_k = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{y_i = k\}$. Note that K_j depends on j; the variables do not need to have the same number of categories. Uniform cutpoints $a_{jk} = \Phi(\zeta_{jk})$ are then converted to corresponding cutpoints ζ_{jk} in $N(0,1)$ with Φ^{-1} i.e. inverse standard normal cdf. With these held fixed, dependence parameters $\theta \in M_{d\times d}$ are estimated using MLE approach and maximising log-likelihood function:

$$\ell(\theta) = \sum_{i=1}^{n} \ln P(Y_{ij} = y_{ij}, j = 1, ..., d; \theta) = \sum_{i=1}^{n} lnP(\vec{Y_i} = (y_1, ..., y_d), \theta) \tag{19}$$

The probability under the sum can be evaluated as (Panagiotelis 2012):

$$P(\vec{Y_i} = (y_1, ..., y_d); \theta) = \sum_{k_1=0,1} ... \sum_{k_d=0,1} (-1)^{k_1+...+k_d} P(Y_1 \leq y_1 - k_1, ..., Y_d \leq y_d - k_d); \theta) \tag{20}$$

which for $d = 2$ and $\theta \in \mathbb{R}$ is in line with inclusion-exclusion principle:

$$\begin{aligned} P(\vec{Y_i} = (y_1, y_2); \theta) = P(Y_1 \leq y_1, Y_2 \leq y_2; \theta) - P(Y_1 \leq y_1, Y_2 \leq y_2 - 1; \theta) + \\ -P(Y_1 \leq y_1 - 1, Y_2 \leq y_2; \theta) + P(Y_1 \leq y_1 - 1, Y_2 \leq y_2 - 1; \theta) = C_\theta(a_{1,y_1+1}, a_{2,y_2+1}) + \\ -C_\theta(a_{1,y_1+1}, a_{2,y_2}) - C_\theta(a_{1,y_1}, a_{2,y_2+1}) + C_\theta(a_{1,y_1}, a_{2,y_2}) \end{aligned} \tag{21}$$

where $y_1 = 0, ..., K_1 - 1$, $y_2 = 0, ..., K_2 - 1$. One chooses copula family with the highest likelihood.

1.5 Simulation in R

Estimation was conducted using R package "CopulaModel" developed by H. Joe and P. Krupskii. For 1-factor models, the following copulas were tested in the process, regardless of the initial diagnostics of the dataset: Gumbel, survival Gumbel, Gaussian, as well as t copulas with 2, 3, 4, 5, 7 and 9 degrees of freedom.

The algorithm for copula simulation is as follows. For each group of respondents univariate cutpoints are estimated based on the sample as in (18), and for each copula or a set of copulas d parameter values are estimated in line with the formula 19 for the log-likelihood function. It is possible to maximise the function numerically using the Newton-Raphson method but for large d it is extremely time-consuming. Inference Function of Margins (IFM) described by H. Joe (Joe 2005) was used, since $d = 24$. This method is efficient regarding both computing time and asymptotic variance. Using the vector of parameters, a simulation is conducted to generate a twin dataset with the same number of observations whose multivariate distribution is given by the copula. One should notice that since the forms of bivariate copulas in 8, 10 and 12 denote continuous distributions, some figures have to be rounded up or down randomly. The obtained database depends on seed settings of a random number generator. Afterwards, the same diagnostics can be run on both original dataset and a simulated one for comparison.

2 Health modelling via 1-factor copula model

ELSA (English Longitudinal Study of Ageing[5] is a survey of quality of life among older people in the UK. Waves 1 and 6 (2002, 2012) were downloaded from the UK Data Archive. Although the dataset allows longitudinal analyses, only cross-sectional were conducted to show if and how the conditions of the elderly have changed over time.

[5]Marmot M., Oldfield Z., Clemens S., Blake M., Phelps A., Nazroo J., Steptoe A., Rogers N., Banks J. (2016), *English Longitudinal Study of Ageing*, Waves 0-7, 1998-2015 [data collection], 24th Edition, UK Data Service, SN: 5050, http://dx.doi.org/10.5255/UKDA-SN-5050-11

2.1 Data

The multivariate model comprises of 24 variables: 19 items describing control, autonomy, self-realisation and pleasure, each rated on a scale from 1 to 4, self-reported health status, mobility, eyesight, hearing and pain rating.

Table 1: CASP-19 variables

C1	How often feels age prevents them from doing things they like
C2	How often feels what happens to them is out of their control
C3	How often feels free to plan for the future
C4	How often feels left out of things
A1	How often can do the things they want to do
A2	How often family responsibilities prevent them from doing things they want to do
A3	How often feels they can please themselves with what they do
A4	How often feels their health stops them from doing what they want to do
A5	How often shortage of money stops them doing things
P1	How often looks forward to each day
P2	How often feels that their life has meaning
P3	How often enjoys the things they do
P4	How often enjoys being in the company of others
P5	How often looks back on their life with a sense of happiness
S1	How often feels full of energy these days
S2	How often chooses to do things they have never done before
S3	How often feels satisfied with the way their life has turned out
S4	How often feels that life is full of opportunities
S5	How often feels the future looks good to them

The data has been adjusted to fit the assumptions in the model. The number of categories differed across variables and therefore was reduced to 4 by collapsing higher responses (pairing "very good" with "excellent" as one category). Therefore, some items, such as health, vision and hearing, are self-rated on the following scale: 1-"poor", 2-"fair", 3-"good", 4-"excellent". Both eyesight and hearing questions assume reporting the senses using everyday correcting devices such as glasses, contact lenses or hearing aid. Blind people fall into the first category of "poor" eyesight. Pain rating is derived from two separate questions: "Are you bothered by pain?" and "How much does it hurt?" by adding the fourth, "no pain", category. Therefore, the responses are: 1-"severe", 2-"moderate", 3-"mild", 4-"no pain". Mobility was measured by asking the respondents about how much difficulty they associate with walking for a quarter of a mile. The possible answers were: 1-"unable to do it", 2-"much difficulty", 3-"some difficulty", 4-"no difficulty". Ordering of responses was reversed in some cases to assure positive dependence. After all necessary adjustments, the waves 1 and 6 consist of 4650 and 7915 observations, respectively, as shown in Table 2.

We analyse the distribution of the groups defined based on sex, age (50-64 years old group and 65+), employment (retired, employed, unemployed, disabled), as well as smoking (behavioural risk). This gives us 10 groups to analyse in each wave.

Table 2: Sample sizes by population groups

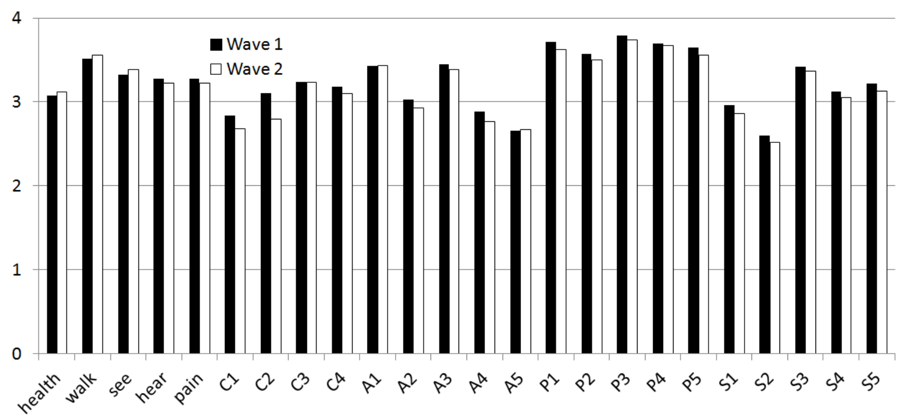
	size (wave 1)	% (wave 1)	size (wave 6)	% (wave 6)
males	2174	44,49%	3589	45,34%
females	2476	50,67%	4326	54,66%
age 50-64	2542	52,02%	3734	47,18%
age 65+	2108	43,13%	4181	52,82%
non-smoking	3848	78,74%	7020	88,69%
smoking	802	16,41%	895	11,31%
retired	2329	47,66%	4558	57,59%
employed	1588	32,49%	2602	32,87%
unemployed	475	9,72%	468	5,91%
disabled	258	5,28%	287	3,63%
total	4650		7915	

Source: Own calculations based on the ELSA Wave 1 and Wave 6

2.2 Descriptive statistics

As a first approximation, Tables 14 and 15 and Tables 11 and 12 contain information on, respectively, group means and differences if scale 1-4 of ordinal variables is used. Figure 2 presents it graphically.

Figure 2: Mean values of variables in waves 1 and 6



Source: Own calculations based on the ELSA Wave 1 and Wave 6

It can be easily inferred that, on average, younger people tend to feel better physically than older people, which does not, however, translate to better mental health. On the other hand, smoking can be seen related to both worse physical and mental condition. The people who still work are of better general health but not with respect to general happiness; here retired are better off. The unemployed are less satisfied than both mentioned groups and the disabled present the worst condition of all. There is no visible pattern to how men differ from women - each group dominates the other in nearly equal number of items and the differences are often insignificant.

Comparing between waves (Table 13) what stands out is a considerable decline (of more than 0.1) among the unemployed respondents which means that they feel worse than ten years ago. Items *C1, C2, A4, S1* show noticeable decline in all groups which indicates that people may feel more tired and physically constrained by their conditions in wave 6 comparing to wave 1. In 2012 people of all ages over 50 reported significantly better health and mobility than in 2002. People older than 65 have bettered in eyesight as well, however, both hearing worsened in both age groups. Control domain consisting of variables *C1, C2, C3 and C4* experienced dramatic decrease regardless of the age of respondents; respondents feel less in control of their actions now. Two variables concerning autonomy, *A1* and *A5*, improved for 65+ age group. The first indicator refers to a general power to do the things one wants to do and the latter indicates whether a respondent feels financial constraints. Indicators measuring pleasure experienced a slight decline over the period.

3 Results

Table 3 presents bivariate count distributions of CASP-19 variables *A5* and *S5* simulated with various copulas, as well as observed distribution. *A5* refers to financial constraints, whereas *S5* refers to prospects for the future. Because item *A5* is linked to negative events, the order of categories was reversed. The table illustrates how many respondents are estimated to fall into each bivariate category, e.g. in the original study there were 63 men who reported their money restrictions to occur "often" and who, simultaneously, never feel that their future looks good. On the other hand, there were 483 men who were optimistic about their future and did not feel constrained by money. From this initial diagnostics it seems that empirically *A5* has more probability in the centre, it is therefore a discretised mean variable, whereas *S5* witnesses more mass in extreme (maximum) values. This is why the joint probability is asymmetric and upper tail dependence is observed. Simulations with Gumbel copulas have more mass in the upper quadrant but the least in the lower quadrant of all estimated distributions. Gaussian copulas, on the other hand, underestimate the upper tails and shift the mass towards the middle categories, leaving little negative dependence. t copulas give the closest approximation compared to the other families - $t(5)$ and $t(7)$ seem to be the closest fits for males and $t(4)$ or $t(5)$ for females. The initial diagnostics show that the dependence of financial constraints and one's view about the future is not linear and is stronger for both those financially constrained and financially unconstrained. These non-linearities cannot be detected with the Gaussian model and cannot be seen if the data were dichotomised. With two four-categories variables there is nine different ways of where to put a cutoff (three different ways for one item given that they are in an increasing order so that 1 in the lower categories implies 1 in higher categories). In Table 4 we present the distribution of *A5* and *S5* among males after the following dichotomisations: (i) a respondent has health problems if he reports categories 2, 3, or 4 for both indicators (ii) a respondent has health problems if he reports categories 3, or 4 for both indicators (iii) a respondent has health problems if he reports category 4 for both indicators. Therefore, we observe three 2×2 distributions. Marginal distributions change, but the dependence structure changes too and it changes fundamentally. If cases (i) and (ii) are considered then we conclude that *A5* and *S5* are strongly related in the upper tail, on the other hand this is reversed for (iii). Clearly, dichotomisation obscures the dependence structure. Furthermore, from policy point of view, the existence of tail dependence implies that, with marginals unchanged, there is a certain level of comorbidity in emotional health. Emotional health conditions tend to occur together and tend to *not* occur together. In line with inequality measurement theory axioms, such property of a health distribution implies that it is more unequal.

Table 3: Comparison of bivariate count distributions for items A5 and S5

	males				females			
empirical distribution	63	176	188	63	65	166	308	89
	36	201	566	300	37	180	686	413
	24	123	499	440	34	162	549	526
	31	89	307	483	49	95	376	591
Gumbel	49	145	246	79	46	152	292	132
	57	220	544	290	63	243	691	362
	31	147	477	404	46	150	570	485
	18	59	298	525	25	75	354	640
s.Gumbel	41	137	244	90	44	128	306	147
	55	212	535	313	62	246	662	387
	32	151	470	395	52	163	551	480
	18	72	323	501	33	73	394	598
Gaussian	59	144	235	72	56	156	281	129
	52	221	526	326	60	240	676	388
	26	146	458	423	39	151	537	528
	15	63	351	472	22	76	422	565
t(2)	75	158	160	104	75	179	220	166
	32	218	579	300	35	214	715	377
	9	118	508	416	20	131	617	488
	30	95	303	484	46	105	354	584
t(3)	75	152	180	94	67	177	238	152
	37	217	561	305	40	224	699	380
	13	129	490	426	29	138	590	498
	26	84	318	482	36	94	382	582
t(4)	72	149	191	88	64	173	248	141
	41	225	550	308	44	235	691	383
	19	128	481	425	34	138	572	510
	22	80	325	485	30	87	398	578
t(5)	71	148	201	88	64	167	253	135
	45	226	539	310	45	240	693	381
	18	128	481	424	34	142	568	517
	22	76	328	484	28	86	397	576
t(7)	69	148	207	81	62	167	262	133
	43	227	541	312	52	235	690	382
	24	135	474	424	32	146	555	519
	16	72	334	482	27	85	400	579
t(9)	70	148	217	75	61	162	269	133
	45	220	538	319	56	236	683	380
	25	138	468	424	33	145	553	526
	17	66	342	477	27	83	405	574

Source: Own calculations based on the ELSA Wave 1 and Wave 6

Table 4: Males joint *A5, S5* distribution after dichotomisation with different cut-offs

	Categories 2, 3, 4 = 1		Categories 3, 4 = 1		Categories 4 = 1	
empirical	63	427	476	1117	1876	803
distribution	91	3008	267	1729	427	483

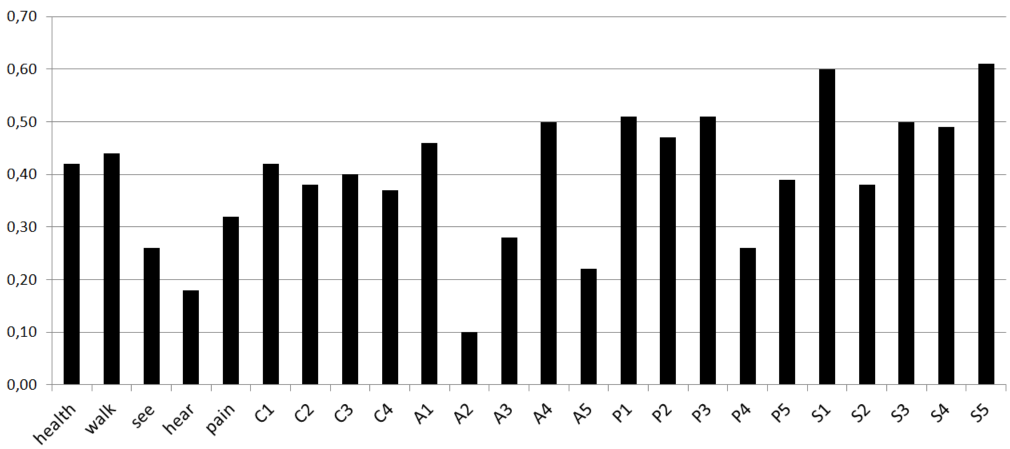
Source: Own calculations based on the ELSA Wave 1 and Wave 6

Tables 5 and 6 display absolute values of the log-likelihood function for all tested copulas and univariate distributions in both waves of the data. The smallest figures are emphasised in bold. Both waves are dominated by t copulas with 4 degrees of freedom; in wave 1 only the groups of employed, unemployed and disabled people can be estimated most accurately with a $t(5)$ copula and in wave 6 there are two such groups, the females and the unemployed. All tested distributions are characterised with intermediate dependence in both upper and lower tail. Such interdependence is impossible to detect if the data are dichotomised and then aggregated. Furthermore, the results imply that the responses are the mixture of discreticised means which is typical for mixed populations (Nikoloulopoulos and Joe 2015). We thus find evidence that factor copula model based on t copulas works better for our health data than purely Gaussian model; the improvement is more evident for Wave 6 than Wave 1.

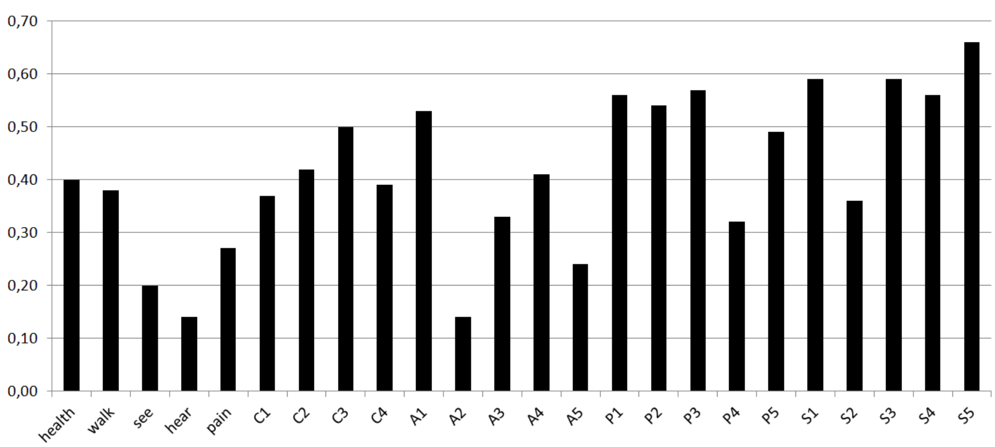
As to the interpretation of the factor we analyse the values of copula parameters corresponding to the model with the maximum likelihood (Tables 16 and 17). The interpretation is available without the need for a varimax rotation typical of factor models. Parameters that can be compared between models are given on a Kendall's tau scale (Tables 18 and 19). The higher the copula parameter, the more dependent a given item on a factor, which provides for a factor's interpretation. Figure 3 illustrates the parameters for the groups of non-smokers (as they are most numerous). In wave 1 items *S1* and *S5* appear to be most linked with the factor. They describe a person's level of energy and views for the future, thus the factor can be interpreted as general optimism in life. In wave 6 there are more items with dependence parameters close to 0.6; these are *S1, S3, S5, P1* and *P3*. The interpretation of factor one as measuring general optimism is now extended to cover life satisfaction and the feeling of enjoyment about things one does.

Figure 3: Parameters on a Kendall's tau scale

(a) Wave 1



(b) Wave 6



Source: Own calculations based on the ELSA Wave 1 and Wave 6

Table 5: Absolute values of log-likelihood function for wave 1

	Gumbel	s.Gumbel	Gaussian	t(2)	t(3)	t(4)	t(5)	t(7)	t(9)
males	49620,53	49505,16	49527,02	49506,85	49380,29	**49362,82**	49368,40	49389,42	49406,95
females	55201,25	55112,61	55166,12	55071,06	54924,82	**54905,49**	54913,19	54944,38	54974,52
age 50-64	55049,87	54887,60	54979,59	54829,26	54702,98	**54695,07**	54709,00	54746,49	54779,05
age 65+	48757,13	48706,06	48713,25	48672,09	48544,95	**48528,84**	48537,17	48562,03	48582,68
non-smoking	85566,36	85428,67	85460,71	85408,05	85173,08	**85141,29**	85152,18	85191,21	85226,21
smoking	49620,53	49505,16	49527,02	49506,85	49380,29	**49362,82**	49368,40	49389,42	49406,95
retired	52828,69	52829,01	52817,13	52741,65	52604,01	**52589,33**	52600,26	52630,72	52656,79
employed	32308,87	32349,51	32303,31	32329,87	32224,60	32204,03	**32202,70**	32212,44	32222,93
unemployed	10884,02	10890,54	10893,65	10882,91	10847,59	10839,54	**10838,83**	10843,70	10849,61
disabled	6748,956	6754,429	6749,422	6769,521	6740,441	6732,230	**6730,039**	6730,551	6732,406

Table 6: Absolute values of log-likelihood function for wave 6

	Gumbel	s.Gumbel	Gaussian	t(2)	t(3)	t(4)	t(5)	t(7)	t(9)
males	80854,65	80717,91	80846,76	80457,84	80259,61	**80254,10**	80291,59	80378,07	80449,67
females	96224,81	96084,29	96137,74	96065,23	95740,58	95670,10	**95665,24**	95705,32	95754,27
age 50-64	80293,86	80103,56	80259,77	79918,39	79692,60	**79673,95**	79701,84	79775,39	79838,67
age >65	95291,40	95158,20	95225,41	95066,25	94764,68	**94711,52**	94721,97	94783,42	94844,76
non-smoking	155769,7	155532,7	155668,2	155360,9	154870,6	**154785,7**	154802,2	154902,5	155004,2
smoking	80854,65	80717,91	80846,76	80457,84	80259,61	**80254,10**	80291,59	80378,07	80449,67
retired	102116,5	102098,0	102094,7	101980,7	101635,0	**101567,6**	101572,6	101631,5	101693,6
employed	53573,58	53645,68	53640,69	53546,96	53322,45	**53281,89**	53287,39	53325,96	53365,15
unemployed	11008,83	10991,50	10993,84	11002,42	10959,10	10947,77	**10945,50**	10948,37	10953,16
disabled	7481,391	7476,571	7478,870	7469,904	7447,853	**7444,286**	7445,429	7450,359	7454,723

Source: Own calculations based on the ELSA Wave 1 and Wave 6

The most heterogeneity is found among groups based on employment status. In Wave 1 (Table 18) there is substantial dependence of general health status and mobility for both retired and disabled and much less for the employed group. In other words, among retired and disabled whether the person is mobile or not is related stronger to his/her subjective health status than in the employed group. Furthermore, these two are one of key health indicators for these two groups. The similar difference is observed for indicators *C1* and *C2* which measure how much one feels constrained by age and feels lack of control. These two are more interdependent among retired and disabled than among employed group. In Wave 1 for the unemployed group items such as *P1* and *P3* appear as the most informative of the underlying factor, whereas for other groups all items in P (pleasure) category are equally important. *P1* and *P3* measure the enjoyment one has about every day and things in life. In Wave 6, *P1* becomes the key item for the employed group too. All in all, the employed group seems different in terms of emotional health comparing to other groups. The dependence parameters seem more evenly distributed in this group, suggesting that the underlying factor may point to a general level of well-being, whereas for other groups, factor's interpretation may be more specific and focus mostly on physical limitations and the indicators for depression e.g. energy levels.

Tables 7 and 8 present univariate distributions of all items for two groups: males and females. There are only small differences in the original and estimated margins. Margins were estimated with $t(4)$ and $t(5)$ copulas for men and women respectively, with parameters given in table 17.

Table 7: Univariate margins for the group of males in wave 6

	empirical distribution				estimated with t(4) copula			
	1	2	3	4	1	2	3	4
health	273	642	1149	1525	276	638	1174	1501
walk	220	187	414	2768	226	168	421	2774
see	76	279	1359	1875	70	271	1417	1831
hear	209	762	1241	1377	216	783	1183	1407
pain	239	630	464	2256	251	642	457	2239
C1	374	1380	1094	741	359	1342	1148	740
C2	230	1023	1614	722	244	1024	1611	710
C3	142	492	1226	1729	142	498	1228	1721
C4	124	669	1376	1420	117	699	1341	1432
A1	78	238	1274	1999	73	244	1302	1970
A2	138	1031	1329	1091	150	1006	1340	1093
A3	93	302	1521	1673	89	304	1539	1657
A4	542	951	929	1167	537	922	930	1200
A5	490	1103	1086	910	500	1124	1053	912
P1	44	146	945	2454	36	137	965	2451
P2	87	282	1093	2127	78	273	1123	2115
P3	18	49	784	2738	16	47	779	2747
P4	21	96	1151	2321	11	106	1163	2309
P5	40	199	1120	2230	31	203	1141	2214
S1	229	795	1792	773	207	845	1765	772
S2	359	1394	1498	338	359	1364	1545	321
S3	93	284	1387	1825	95	285	1370	1839
S4	156	674	1577	1182	151	674	1565	1199
S5	154	589	1560	1286	154	582	1547	1306

Table 8: Univariate margins for the group of females in wave 6

	empirical distribution				estimated with t(4) copula			
	1	2	3	4	1	2	3	4
health	256	782	1372	1916	268	769	1393	1896
walk	349	232	553	3192	370	222	540	3194
see	86	384	1681	2175	81	395	1685	2165
hear	109	522	1378	2317	104	532	1427	2263
pain	345	1038	571	2372	366	1002	562	2396
C1	375	1449	1433	1069	387	1492	1388	1059
C2	255	1305	1845	921	271	1333	1833	889
C3	219	612	1570	1925	244	627	1543	1912
C4	143	1031	1531	1621	142	1050	1517	1617
A1	94	348	1545	2339	96	345	1566	2319
A2	253	1268	1398	1407	235	1342	1317	1432
A3	104	299	1543	2380	90	311	1514	2411
A4	599	1222	1033	1472	560	1235	1006	1525
A5	628	1316	1271	1111	619	1359	1261	1087
P1	43	183	1139	2961	35	210	1145	2936
P2	69	267	1312	2678	70	287	1314	2655
P3	20	67	935	3304	17	66	957	3286
P4	18	66	988	3254	14	72	960	3280
P5	32	236	1332	2726	28	221	1310	2767
S1	280	960	2208	878	269	982	2221	854
S2	465	1574	1847	440	432	1557	1903	434
S3	123	374	1689	2140	132	374	1698	2122
S4	203	807	1910	1406	200	800	1921	1405
S5	185	603	1919	1619	171	635	1911	1609

Source: Own calculations based on the ELSA Wave 1 and Wave 6

There exists no coherent way to illustrate a 24-dimensional distribution, thus we present estimated bivariate margins for variables *health* and *A4* (Table 9) which measures the feeling of autonomy. The two variables are expected to be dependent, since lower levels of general health imply e.g. mobility limitations. Indeed, a vast majority of probability mass is gathered near the main diagonal as can be seen in Table 9. For each decomposed group, the table contains output from a copula with the highest value of log-likelihood function (cf. table 6). The maximum deviation in all presented values equals 8,11% (compared to the magnitude of each subsample) and average deviation is lower than 2%.

Table 9: Bivariate count distributions of the models with the best log-likelihood

	empirical				1-factor			
males	203	54	8	8	102	80	55	39
	212	293	87	50	150	217	131	140
	99	396	378	276	167	335	340	332
	28	208	456	833	118	290	404	689
females	190	56	6	4	95	94	39	40
	255	404	86	37	155	290	146	178
	111	528	402	331	190	427	360	416
	43	234	539	1100	120	424	461	891
age 50-64	156	44	6	5	61	54	53	59
	157	248	86	46	104	140	121	138
	64	338	380	333	114	292	344	405
	28	160	497	1186	113	340	458	938
age 65+	237	66	8	7	155	113	35	25
	310	449	87	41	223	332	191	137
	146	586	400	274	220	519	374	318
	43	282	498	747	127	398	401	613
non-smoking	312	94	11	9	157	117	71	60
	396	586	151	71	266	448	266	260
	189	827	696	513	278	700	656	610
	66	403	946	1750	237	656	811	1427
smoking	203	54	8	8	102	80	55	39
	212	293	87	50	150	217	131	140
	99	396	378	276	167	335	340	332
	28	208	456	833	118	290	404	689
retired	225	62	8	8	140	81	31	30
	331	479	96	43	253	379	195	147
	148	611	441	311	217	566	433	334
	44	284	556	911	140	405	452	755
employed	20	18	4	2	9	7	6	16
	52	148	59	33	25	102	78	115
	32	251	284	263	31	182	233	352
	13	129	380	914	47	277	419	703
unemployed	18	9	2	1	7	5	13	5
	26	48	15	8	21	25	22	26
	12	45	51	32	17	45	43	38
	10	28	57	106	19	56	55	71
disabled	130	21	0	1	109	27	6	4
	58	22	3	3	61	24	4	5
	18	17	4	1	23	10	1	2
	4	1	2	2	10	1	0	0

Source: Own calculations based on the ELSA Wave 1 and Wave 6

4 Conclusion

Considering causal links would add further layers of complexity to the model and is beyond the scope of this proposal. However, there is still a lot to be learnt about the joint distribution of health. This is an area of active research in the last decade too as detailed information on health has become available. A disaggregated view of health will inform researchers and policymakers on most vulnerable groups (in terms of both worse unidimensional health distributions and the incidence of multiple health deprivations) and on the potential areas of intervention.

Factor copulas are indeed powerful and flexible tools in modelling multivariate data, including difficult cases of high-dimensional and ordinal data that defy regular procedures (see e.g. Oh and Patton 2015). They have been so far applied in psychometrics (Nikoloulopoulos and Joe 2015) and continuous stock returns (Krupskii and Joe 2013), (Krupskii and Joe 2015), (Oh and Patton 2015). They are very flexible in modelling complex dependence with particular emphasis on tail dependence. In this paper we present their usefulness in modelling ordinal health data. The can detected complexities and nonlinearities in the data that are not detectable to the methods proposed in the health and health inequality measurement literature so far. We show that there is a potential in applying factor copula models to health data and provide evidence that richer models provide for a better fit. The presentation here should be developed in the following directions: (i) estimation of 2− and possibly 3-factor models (ii) estimation of structured factor copula models for ordinal data where the group structure (such as autonomy, control etc. in CASP-19) is taken into account in the statistical models (structured copulas have only been developed for continuous data (Krupskii and Joe 2015)). Structured factor copulas are extensions to the Gaussian bi-factor model covered by (Gibbons i Hedeker 1992) as well as (Holzinger i Swineford 1937). (iii) choosing different bivariate linking copulas in a given tree, which has to be done with caution as it obscures model comparisons. Concerning the latter, Table 10 illustrates bivariate distribution of variables *C3* and *S5*. Clearly, they are upper dependent; the probability mass is concentrated in the upper quadrant. Not surprisingly, the corresponding count figures delivered by the Gumbel copula provide the best fit.

Another important extension of the model is the addition of regression parameters, not only for marginals (i.e. via probit regressions), but also for the dependence parameters, as dependence may potentially depend on a different set of covariates.

Table 10: Comparison of modelling upper tail dependence of items *C3* and *S5*

empirical				Gumbel				$t(4)$				Gaussian			
39	43	77	95	28	59	127	32	47	68	103	36	32	62	123	32
41	87	118	30	16	77	140	52	19	63	125	64	21	64	130	60
9	98	**411**	**145**	27	92	**345**	**206**	16	78	**334**	**226**	25	91	**310**	**246**
8	47	**429**	**799**	20	65	**427**	**763**	18	67	**478**	**734**	16	69	**472**	**723**

Source: Own calculations based on the ELSA Wave 1 and Wave 6

5 Appendix

Table 11: t test for means

	Wave 1			Wave 6		
	sex	age	smoking	sex	age	smoking
health	0,042 (0,028)	-0,195 (0,028)*	-0,291 (0,037)*	0,05 (0,021)*	-0,233 (0,021)*	-0,326 (0,033)*
walk	-0,063 (0,027)*	-0,404 (0,027)*	-0,122 (0,036)*	-0,074 (0,02)*	-0,343 (0,02)*	-0,18 (0,031)*
see	-0,084 (0,023)*	-0,175 (0,023)*	-0,125 (0,03)*	-0,028 (0,016)	-0,142 (0,016)*	-0,142 (0,026)*
hear	0,284 (0,026)*	-0,25 (0,026)*	-0,059 (0,034)	0,31 (0,019)*	-0,282 (0,019)*	0,008 (0,031)
pain	-0,133 (0,03)*	-0,089 (0,03)*	-0,126 (0,04)*	-0,171 (0,023)*	-0,12 (0,023)*	-0,152 (0,036)*
C1	0,108 (0,028)*	-0,582 (0,027)*	0,016 (0,038)	0,125 (0,021)*	-0,556 (0,02)*	-0,06 (0,033)
C2	-0,003 (0,027)	-0,017 (0,027)	-0,102 (0,036)*	0,005 (0,019)	-0,094 (0,019)*	-0,123 (0,03)*
C3	-0,085 (0,029)*	-0,15 (0,029)*	-0,207 (0,038)*	-0,063 (0,019)*	-0,049 (0,019)*	-0,181 (0,03)*
C4	-0,032 (0,026)	0,019 (0,026)	-0,093 (0,035)*	-0,07 (0,019)*	0,005 (0,019)	-0,089 (0,03)*
A1	0,007 (0,024)	-0,107 (0,024)*	-0,178 (0,032)*	-0,03 (0,016)	-0,026 (0,016)	-0,178 (0,026)*
A2	0,051 (0,028)	0,445 (0,027)*	-0,006 (0,037)	-0,025 (0,02)	0,333 (0,02)*	0,079 (0,032)*
A3	0,136 (0,023)*	0,111 (0,023)*	-0,074 (0,031)*	0,103 (0,017)*	0,148 (0,016)*	-0,065 (0,026)*
A4	0,03 (0,032)	-0,418 (0,031)*	-0,114 (0,042)*	0,023 (0,024)	-0,419 (0,023)*	-0,148 (0,038)*
A5	0,068 (0,03)*	0,262 (0,029)*	-0,297 (0,039)*	-0,011 (0,023)	0,315 (0,022)*	-0,337 (0,036)*
P1	-0,017 (0,017)	0,048 (0,017)*	-0,191 (0,022)*	0,004 (0,014)	0,103 (0,014)*	-0,205 (0,022)*
P2	0,041 (0,021)*	-0,019 (0,021)	-0,192 (0,027)*	0,06 (0,016)*	-0,006 (0,016)	-0,194 (0,025)*
P3	-0,007 (0,014)	0,036 (0,014)*	-0,125 (0,018)*	0 (0,011)	0,082 (0,011)*	-0,181 (0,018)*
P4	0,118 (0,016)*	0,012 (0,016)	-0,076 (0,021)*	0,12 (0,012)*	0,018 (0,012)	-0,082 (0,019)*
P5	0,023 (0,018)	0,116 (0,018)*	-0,19 (0,024)*	0,017 (0,014)	0,078 (0,014)*	-0,178 (0,023)*
S1	0,035 (0,024)	-0,21 (0,024)*	-0,208 (0,032)*	-0,015 (0,018)	-0,15 (0,018)*	-0,211 (0,029)*
S2	0,06 (0,025)*	-0,308 (0,025)*	-0,118 (0,034)*	0,017 (0,018)	-0,212 (0,018)*	-0,156 (0,029)*
S3	0,014 (0,022)	0,066 (0,022)*	-0,191 (0,029)*	-0,026 (0,017)	0,032 (0,017)	-0,207 (0,026)*
S4	0,054 (0,025)*	-0,14 (0,025)*	-0,218 (0,033)*	-0,01 (0,019)	-0,107 (0,019)*	-0,254 (0,029)*
S5	0,056 (0,024)*	-0,163 (0,024)*	-0,221 (0,031)*	0,041 (0,019)*	-0,11 (0,018)*	-0,223 (0,029)*

Table 12: t test for means of employment groups

	Wave 1			Wave 6		
	employed	unemployed	disabled	employed	unemployed	disabled
health	0,378 (0,029)*	0,027 (0,048)	-1,254 (0,063)*	0,353 (0,021)*	0,041 (0,045)*	-1,38 (0,056)*
walk	0,483 (0,027)*	0,204 (0,049)*	-1,257 (0,067)*	0,436 (0,02)*	0,144 (0,046)*	-1,323 (0,059)*
see	0,196 (0,024)*	-0,051 (0,04)*	-0,413 (0,053)*	0,151 (0,017)*	-0,016 (0,036)	-0,427 (0,046)*
hear	0,242 (0,028)*	0,127 (0,046)*	-0,149 (0,061)*	0,233 (0,021)*	0,193 (0,043)*	-0,109 (0,054)*
pain	0,251 (0,032)*	-0,013 (0,052)	-1,166 (0,068)*	0,251 (0,024)*	-0,071 (0,05)*	-1,177 (0,063)*
C1	0,578 (0,03)*	0,163 (0,049)*	-0,347 (0,064)*	0,483 (0,022)*	0,197 (0,045)*	-0,263 (0,057)*
C2	0,105 (0,029)*	-0,082 (0,047)*	-0,781 (0,062)*	0,103 (0,02)*	-0,126 (0,041)*	-0,717 (0,051)*
C3	0,175 (0,031)*	-0,072 (0,05)*	-0,545 (0,065)*	0,029 (0,021)	-0,239 (0,041)*	-0,635 (0,052)*
C4	0,029 (0,028)	-0,137 (0,045)*	-0,654 (0,06)*	0,023 (0,02)	-0,176 (0,041)*	-0,611 (0,052)*
A1	0,171 (0,026)*	0,029 (0,042)*	-0,619 (0,056)*	0,053 (0,017)*	-0,146 (0,035)*	-0,752 (0,045)*
A2	-0,402 (0,029)*	-0,365 (0,047)*	-0,249 (0,06)*	-0,304 (0,021)*	-0,447 (0,043)*	-0,105 (0,053)*
A3	-0,129 (0,025)*	-0,159 (0,04)*	-0,398 (0,052)*	-0,167 (0,017)*	-0,277 (0,035)*	-0,385 (0,044)*
A4	0,598 (0,032)*	0,124 (0,054)*	-1,301 (0,069)*	0,53 (0,025)*	0,118 (0,051)*	-1,288 (0,063)*
A5	-0,219 (0,032)*	-0,122 (0,052)*	-0,836 (0,066)*	-0,325 (0,024)*	-0,489 (0,048)*	-0,887 (0,06)*
P1	-0,021 (0,018)	-0,091 (0,029)*	-0,351 (0,038)*	-0,074 (0,015)*	-0,124 (0,029)*	-0,576 (0,037)*
P2	0,042 (0,022)	-0,065 (0,036)*	-0,306 (0,047)*	0,028 (0,017)	-0,064 (0,034)*	-0,501 (0,044)*
P3	-0,015 (0,015)	-0,086 (0,023)*	-0,277 (0,031)*	-0,057 (0,012)*	-0,136 (0,023)*	-0,43 (0,029)*
P4	0,012 (0,017)	-0,038 (0,027)*	-0,162 (0,036)*	0,003 (0,013)	-0,05 (0,026)*	-0,282 (0,033)*
P5	-0,057 (0,019)*	-0,044 (0,03)*	-0,308 (0,04)*	-0,03 (0,015)*	-0,16 (0,03)*	-0,529 (0,038)*
S1	0,288 (0,025)*	0,054 (0,042)*	-0,885 (0,055)*	0,199 (0,019)*	-0,022 (0,04)	-0,969 (0,05)*
S2	0,304 (0,027)*	-0,027 (0,045)	-0,505 (0,058)*	0,218 (0,019)*	-0,056 (0,039)*	-0,55 (0,049)*
S3	0,018 (0,023)	-0,025 (0,038)	-0,739 (0,051)*	-0,012 (0,018)	-0,203 (0,036)*	-0,737 (0,045)*
S4	0,174 (0,027)*	-0,007 (0,044)	-0,507 (0,058)*	0,13 (0,02)*	-0,174 (0,041)*	-0,675 (0,051)*
S5	0,2 (0,025)*	0,062 (0,041)*	-0,659 (0,055)*	0,117 (0,019)*	-0,104 (0,04)*	-0,789 (0,051)*

Table 13: t test for means across waves

	males	females	age 50-64	age 65+	non-smoking	smoking
health	0,036 (0,023)*	0,044 (0,022)	0,076 (0,021)*	0,038 (0,023)*	0,028 (0,018)	-0,007 (0,035)*
walk	0,046 (0,022)	0,036 (0,021)*	0,037 (0,019)*	0,099 (0,023)*	0,039 (0,018)	-0,019 (0,033)*
see	0,034 (0,018)	0,09 (0,017)*	0,059 (0,017)	0,092 (0,018)*	0,058 (0,015)*	0,041 (0,027)*
hear	-0,066 (0,021)*	-0,04 (0,02)*	-0,012 (0,019)*	-0,044 (0,021)*	-0,058 (0,017)*	0,008 (0,032)
pain	-0,03 (0,025)*	-0,068 (0,024)	-0,029 (0,023)*	-0,06 (0,025)	-0,057 (0,02)*	-0,083 (0,038)
C1	-0,17 (0,023)*	-0,152 (0,022)*	-0,129 (0,021)*	-0,103 (0,023)*	-0,149 (0,018)*	-0,225 (0,035)*
C2	-0,316 (0,021)*	-0,307 (0,02)*	-0,269 (0,019)*	-0,346 (0,021)*	-0,315 (0,017)*	-0,336 (0,032)*
C3	-0,021 (0,021)*	0,001 (0,021)	-0,053 (0,02)*	0,048 (0,022)*	-0,026 (0,017)*	0,001 (0,032)*
C4	-0,056 (0,021)*	-0,094 (0,02)*	-0,071 (0,02)*	-0,085 (0,021)*	-0,083 (0,017)*	-0,079 (0,032)
A1	0,025 (0,018)	-0,012 (0,017)	-0,029 (0,017)*	0,051 (0,018)*	-0,005 (0,015)	-0,005 (0,027)*
A2	-0,062 (0,022)*	-0,139 (0,021)*	-0,077 (0,021)*	-0,19 (0,022)*	-0,113 (0,018)*	-0,028 (0,033)*
A3	-0,045 (0,018)	-0,078 (0,017)*	-0,089 (0,017)	-0,052 (0,018)*	-0,067 (0,015)*	-0,058 (0,028)
A4	-0,109 (0,026)*	-0,117 (0,025)*	-0,081 (0,024)*	-0,082 (0,026)*	-0,116 (0,021)*	-0,15 (0,04)
A5	0,049 (0,024)	-0,03 (0,023)	-0,04 (0,023)*	0,012 (0,025)*	-0,006 (0,02)*	-0,045 (0,038)*
P1	-0,109 (0,015)*	-0,088 (0,014)*	-0,13 (0,014)*	-0,075 (0,015)*	-0,107 (0,012)*	-0,121 (0,023)*
P2	-0,088 (0,017)*	-0,069 (0,016)*	-0,083 (0,016)*	-0,07 (0,018)*	-0,089 (0,014)*	-0,09 (0,027)*
P3	-0,052 (0,012)*	-0,045 (0,011)*	-0,075 (0,011)*	-0,029 (0,012)*	-0,049 (0,01)*	-0,105 (0,019)*
P4	-0,025 (0,013)*	-0,023 (0,012)*	-0,026 (0,012)	-0,021 (0,013)*	-0,026 (0,01)*	-0,032 (0,02)*
P5	-0,087 (0,015)*	-0,093 (0,015)*	-0,079 (0,015)*	-0,117 (0,015)*	-0,103 (0,012)*	-0,091 (0,024)*
S1	-0,078 (0,02)*	-0,127 (0,019)*	-0,12 (0,019)*	-0,061 (0,02)	-0,117 (0,016)*	-0,119 (0,031)*
S2	-0,063 (0,02)*	-0,106 (0,019)*	-0,114 (0,018)*	-0,018 (0,02)*	-0,089 (0,016)*	-0,127 (0,03)
S3	-0,034 (0,018)*	-0,074 (0,017)*	-0,043 (0,017)	-0,077 (0,018)*	-0,065 (0,015)*	-0,081 (0,028)*
S4	-0,043 (0,02)*	-0,107 (0,019)*	-0,084 (0,019)*	-0,051 (0,021)	-0,086 (0,016)*	-0,123 (0,031)*
S5	-0,078 (0,02)*	-0,093 (0,019)*	-0,102 (0,018)*	-0,048 (0,02)	-0,098 (0,016)*	-0,101 (0,031)*

	retired	employed	unemployed	disabled
health	0,035 (0,022)*	0,01 (0,025)*	0,05 (0,044)	-0,091 (0,059)*
walk	0,052 (0,022)*	0,005 (0,023)*	-0,008 (0,042)	-0,014 (0,057)*
see	0,069 (0,017)*	0,024 (0,02)*	0,104 (0,035)*	0,055 (0,046)*
hear	-0,044 (0,021)	-0,054 (0,023)*	0,022 (0,041)*	-0,004 (0,055)*
pain	-0,068 (0,024)	-0,068 (0,027)*	-0,127 (0,048)	-0,079 (0,065)*
C1	-0,125 (0,022)	-0,22 (0,025)*	-0,09 (0,044)*	-0,041 (0,059)*
C2	-0,328 (0,02)*	-0,329 (0,023)*	-0,371 (0,04)*	-0,264 (0,054)*
C3	0,039 (0,021)	-0,107 (0,024)*	-0,128 (0,041)*	-0,051 (0,054)*
C4	-0,092 (0,02)*	-0,099 (0,023)*	-0,131 (0,04)	-0,05 (0,054)*
A1	0,051 (0,018)	-0,067 (0,02)*	-0,124 (0,034)	-0,082 (0,046)*
A2	-0,162 (0,021)*	-0,063 (0,025)*	-0,244 (0,043)	-0,018 (0,057)
A3	-0,058 (0,018)*	-0,097 (0,02)	-0,176 (0,035)	-0,045 (0,047)*
A4	-0,103 (0,025)	-0,171 (0,028)*	-0,109 (0,05)	-0,09 (0,067)*
A5	0,042 (0,024)*	-0,065 (0,027)*	-0,325 (0,048)	-0,01 (0,064)*
P1	-0,081 (0,014)*	-0,134 (0,017)*	-0,114 (0,029)	-0,306 (0,04)*
P2	-0,074 (0,017)*	-0,088 (0,019)*	-0,073 (0,034)	-0,269 (0,045)*
P3	-0,035 (0,012)*	-0,077 (0,014)*	-0,085 (0,024)	-0,187 (0,032)*
P4	-0,019 (0,013)*	-0,028 (0,015)*	-0,031 (0,026)	-0,138 (0,034)*
P5	-0,093 (0,015)*	-0,066 (0,017)*	-0,208 (0,03)*	-0,313 (0,041)*
S1	-0,079 (0,019)*	-0,168 (0,022)*	-0,154 (0,039)*	-0,163 (0,052)*
S2	-0,061 (0,019)	-0,147 (0,022)*	-0,091 (0,038)	-0,107 (0,051)*
S3	-0,051 (0,018)*	-0,08 (0,02)*	-0,229 (0,035)	-0,048 (0,048)*
S4	-0,055 (0,02)*	-0,099 (0,022)*	-0,221 (0,039)	-0,223 (0,053)*
S5	-0,051 (0,019)*	-0,134 (0,022)*	-0,216 (0,039)*	-0,181 (0,052)*

Table 14: Means of variables according to the group in wave 1

	all	males	females	age 50-64	age 65+	non-smoking	smoking	retired	employed	unemployed	disabled
health	3,0798	3,0575	3,0994	3,1684	2,9730	3,1299	2,8392	3,0176	3,3955	3,0442	1,7636
walk	3,5166	3,5501	3,4871	3,6998	3,2955	3,5377	3,4152	3,4006	3,8835	3,6042	2,1434
see	3,3239	3,3684	3,2847	3,4032	3,2282	3,3454	3,2207	3,2851	3,4811	3,2337	2,8721
hear	3,2716	3,1205	3,4043	3,3851	3,1347	3,2817	3,2232	3,1842	3,4263	3,3116	3,0349
pain	3,2787	3,3496	3,2165	3,3190	3,2301	3,3004	3,1746	3,2589	3,5101	3,2463	2,0930
C1	2,8406	2,7833	2,8910	3,1046	2,5223	2,8378	2,8541	2,6458	3,2242	2,8084	2,2985
C2	3,1022	3,1040	3,1006	3,1098	3,0930	3,1198	3,0175	3,1181	3,2229	3,0358	2,3372
C3	3,2415	3,2870	3,2015	3,3096	3,1594	3,2773	3,0698	3,2194	3,3942	3,1474	2,6744
C4	3,1791	3,1960	3,1644	3,1703	3,1898	3,1952	3,1022	3,2194	3,2487	3,0821	2,5659
A1	3,4254	3,4218	3,4285	3,4740	3,3667	3,4561	3,2781	3,3985	3,5693	3,4274	2,7791
A2	3,0297	3,0023	3,0537	2,8281	3,2728	3,0307	3,0249	3,2181	2,8161	2,8526	2,9690
A3	3,4475	3,3753	3,5109	3,3973	3,5081	3,4602	3,3865	3,5298	3,4011	3,3705	3,1318
A4	2,8837	2,8675	2,8978	3,0732	2,6551	2,9033	2,7893	2,7389	3,3369	2,8632	1,4380
A5	2,6600	2,6237	2,6918	2,5413	2,8031	2,7113	2,4140	2,7935	2,5749	2,6716	1,9574
P1	3,7181	3,7272	3,7100	3,6963	3,7443	3,7510	3,5599	3,7540	3,7330	3,6632	3,4031
P2	3,5757	3,5538	3,5949	3,5842	3,5655	3,6089	3,4165	3,5848	3,6272	3,5200	3,2791
P3	3,7869	3,7907	3,7835	3,7707	3,8065	3,8085	3,6833	3,8162	3,8010	3,7305	3,5388
P4	3,6963	3,6334	3,7516	3,6908	3,7030	3,7095	3,6334	3,7050	3,7173	3,6674	3,5426
P5	3,6432	3,6311	3,6539	3,5905	3,7068	3,6759	3,4863	3,6844	3,6272	3,6400	3,3760
S1	2,9628	2,9443	2,9790	3,0578	2,8482	2,9987	2,7905	2,9081	3,1958	2,9621	2,0233
S2	2,6011	2,5690	2,6292	2,7408	2,4326	2,6214	2,5037	2,5281	2,8319	2,5011	2,0233
S3	3,4189	3,4117	3,4253	3,3891	3,4549	3,4519	3,2606	3,4564	3,4742	3,4316	2,7171
S4	3,1265	3,0975	3,1519	3,1900	3,0498	3,1640	2,9464	3,0957	3,2702	3,0884	2,5891
S5	3,2163	3,1868	3,2423	3,2903	3,1271	3,2544	3,0337	3,1782	3,3785	3,2400	2,5194

Table 15: Means of variables according to the group in wave 6

	all	male	female	age 50-64	age 65+	non-smoking	smoking	retired	employed	unemployed	disabled
health	3,1212	3,0939	3,1438	3,2442	3,0112	3,1580	2,8324	3,0527	3,4058	3,0940	1,6725
walk	3,5563	3,5965	3,5229	3,7373	3,3946	3,5766	3,3966	3,4524	3,8885	3,5962	2,1289
see	3,3870	3,4023	3,3742	3,4620	3,3200	3,4030	3,2615	3,3539	3,5046	3,3376	2,9268
hear	3,2241	3,0549	3,3645	3,3733	3,0909	3,2232	3,2313	3,1402	3,3728	3,3333	3,0314
pain	3,2264	3,3199	3,1489	3,2898	3,1698	3,2436	3,0916	3,1909	3,4416	3,1197	2,0139
C1	2,6820	2,6135	2,7388	2,9756	2,4198	2,6887	2,6291	2,5211	3,0042	2,7179	2,2578
C2	2,7909	2,7880	2,7933	2,8404	2,7467	2,8048	2,6816	2,7905	2,8935	2,6645	2,0732
C3	3,2310	3,2655	3,2023	3,2571	3,2076	3,2514	3,0704	3,2587	3,2875	3,0192	2,6237
C4	3,1020	3,1402	3,0703	3,0994	3,1043	3,1120	3,0235	3,1270	3,1499	2,9509	2,5157
A1	3,4306	3,4472	3,4168	3,4446	3,4181	3,4507	3,2726	3,4491	3,5019	3,3034	2,6969
A2	2,9263	2,9398	2,9152	2,7507	3,0832	2,9174	2,9966	3,0564	2,7529	2,6090	2,9512
A3	3,3864	3,3302	3,4330	3,3082	3,4561	3,3937	3,3285	3,4717	3,3044	3,1944	3,0871
A4	2,7706	2,7582	2,7809	2,9920	2,5728	2,7873	2,6391	2,6360	3,1660	2,7543	1,3484
A5	2,6672	2,6732	2,6623	2,5011	2,8156	2,7053	2,3687	2,8350	2,5104	2,3462	1,9477
P1	3,6206	3,6186	3,6223	3,5664	3,6690	3,6437	3,4391	3,6731	3,5992	3,5491	3,0976
P2	3,4983	3,4656	3,5254	3,5016	3,4953	3,5202	3,3263	3,5110	3,5392	3,4466	3,0105
P3	3,7391	3,7392	3,7390	3,6960	3,7776	3,7595	3,5788	3,7815	3,7244	3,6453	3,3519
P4	3,6740	3,6082	3,7286	3,6647	3,6824	3,6833	3,6011	3,6863	3,6891	3,6368	3,4042
P5	3,5530	3,5436	3,5608	3,5118	3,5898	3,5731	3,3955	3,5915	3,5615	3,4316	3,0627
S1	2,8582	2,8663	2,8516	2,9373	2,7876	2,8821	2,6715	2,8293	3,0281	2,8077	1,8606
S2	2,5151	2,5057	2,5229	2,6272	2,4150	2,5328	2,3765	2,4667	2,6849	2,4103	1,9164
S3	3,3632	3,3775	3,3514	3,3463	3,3784	3,3866	3,1799	3,4059	3,3939	3,2030	2,6690
S4	3,0491	3,0546	3,0446	3,1058	2,9986	3,0779	2,8235	3,0410	3,1714	2,8675	2,3659
S5	3,1308	3,1084	3,1493	3,1888	3,0789	3,1560	2,9330	3,1270	3,2440	3,0235	2,3380

Table 16: Estimated parameter values in wave 1

	males	females	age 50-64	age 65+	non-smoking	smoking	retired	employed	unemployed	disabled
	t(4)	t(4)	t(4)	t(4)	t(4)	t(4)	t(4)	t(5)	t(5)	t(5)
health	0,638489	0,618545	0,616641	0,619140	0,616763	0,638489	0,621972	0,428571	0,562755	0,623471
walk	0,666895	0,618333	0,641397	0,642934	0,640686	0,666895	0,651505	0,408877	0,434152	0,652900
see	0,400040	0,390414	0,396874	0,364010	0,393926	0,400040	0,366974	0,304262	0,382715	0,368165
hear	0,304209	0,253068	0,280417	0,245259	0,278719	0,304209	0,262664	0,266243	0,248011	0,263002
pain	0,501910	0,480925	0,475994	0,485375	0,483800	0,501910	0,480367	0,306052	0,374916	0,481262
C1	0,589991	0,604302	0,553078	0,660110	0,613261	0,589991	0,634332	0,465542	0,564669	0,631641
C2	0,566821	0,568973	0,567401	0,592367	0,560357	0,566821	0,577424	0,445203	0,615086	0,576308
C3	0,575493	0,605720	0,610242	0,553095	0,587694	0,575493	0,554153	0,558955	0,614907	0,553274
C4	0,514719	0,580114	0,586444	0,532692	0,545591	0,514719	0,537517	0,484345	0,670280	0,540063
A1	0,624819	0,713086	0,674314	0,672719	0,664823	0,624819	0,663205	0,586526	0,696640	0,655694
A2	0,145931	0,163318	0,279864	0,121497	0,155154	0,145931	0,106807	0,321818	0,367361	0,108658
A3	0,405906	0,447385	0,493764	0,403530	0,427844	0,405906	0,390020	0,528175	0,556350	0,390222
A4	0,694227	0,705304	0,660865	0,739996	0,704415	0,694227	0,722342	0,482594	0,612516	0,720922
A5	0,373205	0,318981	0,472674	0,243698	0,331448	0,373205	0,277935	0,395249	0,433873	0,277887
P1	0,685422	0,763638	0,759082	0,714527	0,718829	0,685422	0,742271	0,744460	0,844839	0,745627
P2	0,652261	0,723138	0,727265	0,642272	0,672280	0,652261	0,678317	0,734659	0,728173	0,679694
P3	0,729117	0,748215	0,767331	0,730604	0,719293	0,729117	0,729373	0,777349	0,861692	0,732399
P4	0,412188	0,431320	0,479361	0,353159	0,401193	0,412188	0,384606	0,484582	0,483417	0,389872
P5	0,552050	0,604592	0,665918	0,507250	0,569616	0,552050	0,557212	0,680376	0,688234	0,560937
S1	0,820864	0,819788	0,802104	0,836542	0,812994	0,820864	0,837109	0,708079	0,757317	0,836317
S2	0,594480	0,569175	0,561669	0,590850	0,563814	0,594480	0,585937	0,483176	0,497644	0,585969
S3	0,705427	0,733827	0,786230	0,665477	0,702127	0,705427	0,684634	0,754382	0,794450	0,689418
S4	0,702399	0,722167	0,738591	0,674406	0,695805	0,702399	0,688080	0,728640	0,697294	0,691641
S5	0,823726	0,824272	0,846251	0,786144	0,813817	0,823726	0,811507	0,815682	0,821737	0,813856

Table 17: Estimated parameter values in wave 6

	males	females	age 50-64	age 65+	non-smoking	smoking	retired	employed	unemployed	disabled
	t(4)	t(5)	t(4)	t(4)	t(4)	t(4)	t(4)	t(4)	t(5)	t(4)
health	0,608663	0,586518	0,575703	0,585917	0,585917	0,608663	0,630738	0,430701	0,474505	0,492401
walk	0,578920	0,554954	0,554796	0,561349	0,561349	0,578920	0,600923	0,357690	0,357802	0,236009
see	0,339404	0,294043	0,314884	0,301783	0,301783	0,339404	0,317395	0,250769	0,225212	0,164937
hear	0,194497	0,230162	0,202160	0,212953	0,212953	0,194497	0,224368	0,164188	0,165883	0,114333
pain	0,414018	0,423225	0,400376	0,406440	0,406440	0,414018	0,427703	0,276674	0,338515	0,160350
C1	0,564708	0,532135	0,513461	0,548661	0,548661	0,564708	0,606239	0,481531	0,372208	0,307150
C2	0,622587	0,619960	0,616648	0,617843	0,617843	0,622587	0,623196	0,584637	0,541036	0,479528
C3	0,697956	0,718181	0,732160	0,708017	0,708017	0,697956	0,701769	0,691205	0,694221	0,769373
C4	0,571490	0,590763	0,604582	0,578758	0,578758	0,571490	0,561521	0,566662	0,607879	0,595062
A1	0,738376	0,748061	0,753629	0,742198	0,742198	0,738376	0,748021	0,719171	0,648209	0,626871
A2	0,213759	0,196221	0,296298	0,211134	0,211134	0,213759	0,146415	0,352950	0,298583	0,035620
A3	0,483542	0,501475	0,552190	0,492302	0,492302	0,483542	0,470769	0,568293	0,513411	0,449009
A4	0,612093	0,594194	0,557102	0,597555	0,597555	0,612093	0,661456	0,429092	0,417724	0,435900
A5	0,424207	0,354568	0,508298	0,371124	0,371124	0,424207	0,327802	0,471850	0,431861	0,199670
P1	0,758542	0,802206	0,826511	0,774502	0,774502	0,758542	0,767598	0,809716	0,865471	0,820609
P2	0,740991	0,775637	0,795095	0,753016	0,753016	0,740991	0,718266	0,809552	0,832286	0,762607
P3	0,775143	0,806743	0,826562	0,783892	0,783892	0,775143	0,791724	0,820805	0,800013	0,805697
P4	0,466179	0,533401	0,562943	0,487098	0,487098	0,466179	0,442603	0,540112	0,614476	0,496068
P5	0,686467	0,704773	0,779560	0,690834	0,690834	0,686467	0,631081	0,783847	0,764942	0,593487
S1	0,807442	0,790903	0,791543	0,795820	0,795820	0,807442	0,815515	0,727198	0,734224	0,734985
S2	0,537785	0,572122	0,580831	0,539887	0,539887	0,537785	0,532144	0,536277	0,543257	0,521961
S3	0,799775	0,806679	0,856571	0,796686	0,796686	0,799775	0,764971	0,844296	0,834617	0,812446
S4	0,767805	0,799043	0,815862	0,775245	0,775245	0,767805	0,755663	0,791024	0,790305	0,815782
S5	0,856856	0,875377	0,884462	0,861027	0,861027	0,856856	0,848646	0,871775	0,886881	0,863954

Table 18: Parameter values on a Kendall's tau scale for wave 1

	males		females		age 50-64		age 65+		non-smoking		smoking		retired		employed		unemployed		disabled	
copula	t(4)		t(4)		t(4)		t(4)		t(4)		t(4)		t(4)		t(5)		t(5)		t(5)	
	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE
health	0,44	0,01	0,42	0,01	0,42	0,01	0,43	0,01	0,42	0,01	0,44	0,01	0,43	0,01	0,28	0,02	0,38	0,03	0,43	0,01
walk	0,46	0,02	0,42	0,02	0,44	0,02	0,44	0,02	0,44	0,01	0,46	0,02	0,45	0,02	0,27	0,03	0,29	0,04	0,45	0,02
see	0,26	0,02	0,26	0,02	0,26	0,02	0,24	0,02	0,26	0,01	0,26	0,02	0,24	0,02	0,20	0,02	0,25	0,03	0,24	0,02
hear	0,20	0,02	0,16	0,02	0,18	0,02	0,16	0,02	0,18	0,01	0,20	0,02	0,17	0,02	0,17	0,02	0,16	0,04	0,17	0,02
pain	0,33	0,02	0,32	0,02	0,32	0,02	0,32	0,02	0,32	0,01	0,33	0,02	0,32	0,02	0,20	0,02	0,24	0,04	0,32	0,02
C1	0,40	0,01	0,41	0,01	0,37	0,01	0,46	0,01	0,42	0,01	0,40	0,01	0,44	0,01	0,31	0,02	0,38	0,03	0,44	0,01
C2	0,38	0,01	0,39	0,01	0,38	0,01	0,40	0,01	0,38	0,01	0,38	0,01	0,39	0,01	0,29	0,02	0,42	0,03	0,39	0,01
C3	0,39	0,02	0,41	0,01	0,42	0,01	0,37	0,02	0,40	0,01	0,39	0,02	0,37	0,02	0,38	0,02	0,42	0,03	0,37	0,01
C4	0,34	0,02	0,39	0,01	0,40	0,01	0,36	0,02	0,37	0,01	0,34	0,02	0,36	0,01	0,32	0,02	0,47	0,03	0,36	0,01
A1	0,43	0,02	0,51	0,01	0,47	0,01	0,47	0,02	0,46	0,01	0,43	0,02	0,46	0,01	0,40	0,02	0,49	0,03	0,46	0,01
A2	0,09	0,02	0,10	0,02	0,18	0,02	0,08	0,02	0,10	0,01	0,09	0,02	0,07	0,02	0,21	0,02	0,24	0,03	0,07	0,02
A3	0,27	0,02	0,30	0,02	0,33	0,02	0,26	0,02	0,28	0,01	0,27	0,02	0,26	0,02	0,35	0,02	0,38	0,03	0,26	0,02
A4	0,49	0,01	0,50	0,01	0,46	0,01	0,53	0,01	0,50	0,01	0,49	0,01	0,51	0,01	0,32	0,02	0,42	0,03	0,51	0,01
A5	0,24	0,02	0,21	0,02	0,31	0,01	0,16	0,02	0,22	0,01	0,24	0,02	0,18	0,02	0,26	0,02	0,29	0,03	0,18	0,02
P1	0,48	0,02	0,55	0,02	0,55	0,01	0,51	0,02	0,51	0,01	0,48	0,02	0,53	0,02	0,53	0,02	0,64	0,03	0,54	0,02
P2	0,45	0,02	0,51	0,01	0,52	0,01	0,44	0,02	0,47	0,01	0,45	0,02	0,47	0,02	0,53	0,02	0,52	0,03	0,48	0,02
P3	0,52	0,02	0,54	0,02	0,56	0,02	0,52	0,02	0,51	0,01	0,52	0,02	0,52	0,02	0,57	0,02	0,66	0,03	0,52	0,02
P4	0,27	0,02	0,28	0,02	0,32	0,02	0,23	0,02	0,26	0,01	0,27	0,02	0,25	0,02	0,32	0,02	0,32	0,04	0,25	0,02
P5	0,37	0,02	0,41	0,02	0,46	0,01	0,34	0,02	0,39	0,01	0,37	0,02	0,38	0,02	0,48	0,02	0,48	0,03	0,38	0,02
S1	0,61	0,01	0,61	0,01	0,59	0,01	0,63	0,01	0,60	0,01	0,61	0,01	0,63	0,01	0,50	0,02	0,55	0,03	0,63	0,01
S2	0,41	0,01	0,39	0,01	0,38	0,01	0,40	0,01	0,38	0,01	0,41	0,01	0,40	0,01	0,32	0,02	0,33	0,03	0,40	0,01
S3	0,50	0,01	0,52	0,01	0,58	0,01	0,46	0,02	0,50	0,01	0,50	0,01	0,48	0,01	0,54	0,02	0,58	0,03	0,48	0,01
S4	0,50	0,01	0,51	0,01	0,53	0,01	0,47	0,01	0,49	0,01	0,50	0,01	0,48	0,01	0,52	0,02	0,49	0,03	0,49	0,01
S5	0,62	0,01	0,62	0,01	0,64	0,01	0,58	0,01	0,61	0,01	0,62	0,01	0,60	0,01	0,61	0,02	0,61	0,03	0,61	0,01

Table 19: Parameter values on a Kendall's tau scale for wave 6

	males		females		age 50-64		age 65+		non-smoking		smoking		retired		employed		unemployed		disabled	
copula	t(4)		t(4)		t(4)		t(4)		t(4)		t(4)		t(4)		t(5)		t(5)		t(5)	
	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE	θ	SE
health	0,42	0,01	0,40	0,01	0,39	0,01	0,42	0,01	0,40	0,01	0,42	0,01	0,43	0,01	0,28	0,01	0,31	0,03	0,43	0,01
walk	0,39	0,01	0,37	0,01	0,37	0,02	0,41	0,01	0,38	0,01	0,39	0,01	0,41	0,01	0,23	0,03	0,23	0,04	0,41	0,01
see	0,22	0,01	0,19	0,01	0,20	0,01	0,19	0,01	0,20	0,01	0,22	0,01	0,21	0,01	0,16	0,02	0,14	0,04	0,21	0,01
hear	0,12	0,01	0,15	0,01	0,13	0,01	0,12	0,01	0,14	0,01	0,12	0,01	0,14	0,01	0,11	0,02	0,11	0,04	0,14	0,01
pain	0,27	0,01	0,28	0,01	0,26	0,01	0,28	0,01	0,27	0,01	0,27	0,01	0,28	0,01	0,18	0,02	0,22	0,04	0,28	0,01
C1	0,38	0,01	0,36	0,01	0,34	0,01	0,42	0,01	0,37	0,01	0,38	0,01	0,41	0,01	0,32	0,01	0,24	0,03	0,41	0,01
C2	0,43	0,01	0,43	0,01	0,42	0,01	0,43	0,01	0,42	0,01	0,43	0,01	0,43	0,01	0,40	0,01	0,36	0,03	0,43	0,01
C3	0,49	0,01	0,51	0,01	0,52	0,01	0,48	0,01	0,50	0,01	0,49	0,01	0,50	0,01	0,49	0,01	0,49	0,03	0,50	0,01
C4	0,39	0,01	0,40	0,01	0,41	0,01	0,38	0,01	0,39	0,01	0,39	0,01	0,38	0,01	0,38	0,01	0,42	0,03	0,38	0,01
A1	0,53	0,01	0,54	0,01	0,54	0,01	0,53	0,01	0,53	0,01	0,53	0,01	0,54	0,01	0,51	0,01	0,45	0,03	0,54	0,01
A2	0,14	0,01	0,13	0,01	0,19	0,01	0,10	0,01	0,14	0,01	0,14	0,01	0,09	0,01	0,23	0,01	0,19	0,03	0,09	0,01
A3	0,32	0,01	0,33	0,01	0,37	0,01	0,31	0,01	0,33	0,01	0,32	0,01	0,31	0,01	0,38	0,01	0,34	0,03	0,31	0,01
A4	0,42	0,01	0,41	0,01	0,38	0,01	0,45	0,01	0,41	0,01	0,42	0,01	0,46	0,01	0,28	0,01	0,27	0,03	0,46	0,01
A5	0,28	0,01	0,23	0,01	0,34	0,01	0,20	0,01	0,24	0,01	0,28	0,01	0,21	0,01	0,31	0,01	0,28	0,03	0,21	0,01
P1	0,55	0,01	0,59	0,01	0,62	0,01	0,55	0,01	0,56	0,01	0,55	0,01	0,56	0,01	0,60	0,01	0,67	0,03	0,56	0,01
P2	0,53	0,01	0,57	0,01	0,59	0,01	0,51	0,01	0,54	0,01	0,53	0,01	0,51	0,01	0,60	0,01	0,63	0,03	0,51	0,01
P3	0,56	0,01	0,60	0,01	0,62	0,01	0,56	0,01	0,57	0,01	0,56	0,01	0,58	0,01	0,61	0,01	0,59	0,03	0,58	0,01
P4	0,31	0,01	0,36	0,01	0,38	0,01	0,28	0,01	0,32	0,01	0,31	0,01	0,29	0,01	0,36	0,02	0,42	0,03	0,29	0,01
P5	0,48	0,01	0,50	0,01	0,57	0,01	0,42	0,01	0,49	0,01	0,48	0,01	0,43	0,01	0,57	0,01	0,55	0,03	0,43	0,01
S1	0,60	0,01	0,58	0,01	0,58	0,01	0,60	0,01	0,59	0,01	0,60	0,01	0,61	0,01	0,52	0,01	0,52	0,03	0,61	0,01
S2	0,36	0,01	0,39	0,01	0,39	0,01	0,35	0,01	0,36	0,01	0,36	0,01	0,36	0,01	0,36	0,01	0,37	0,03	0,36	0,01
S3	0,59	0,01	0,60	0,01	0,65	0,01	0,54	0,01	0,59	0,01	0,59	0,01	0,55	0,01	0,64	0,01	0,63	0,02	0,55	0,01
S4	0,56	0,01	0,59	0,01	0,61	0,01	0,54	0,01	0,56	0,01	0,56	0,01	0,55	0,01	0,58	0,01	0,58	0,02	0,55	0,01
S5	0,66	0,01	0,68	0,01	0,69	0,01	0,64	0,01	0,66	0,01	0,66	0,01	0,65	0,01	0,67	0,01	0,69	0,02	0,65	0,01

References

[1] Aas K., Czado C., Frigessi A., and Bakken H. (2009), *Pair-Copula Constructions of Multiple Dependence*, Insurance, Mathematics and Economics, Vol. 44, p. 182–198.

[2] Abul Naga R H, Yalcin T. 2008. Inequality measurement for ordered response health data, Journal of Health Economics; 27(6):1614-1625.

[3] Abul Naga R H, Stapenhurst Ch. 2015. Estimation of inequality indices of the cumulative distribution function, Economics Letters 130:109-112.

[4] Allison R A, Foster J E. 2004. Measuring health inequality using qualitative data, Journal of Health Economics; 23(3):505-524.

[5] Apouey B. 2007. Measuring health polarization with self-assessed health data, Health Economics, 16:875-894.

[6] Apouey B, Silber J. 2013, Inequality and bi-polarization in socioeconomic status and health: ordinal approaches, SSRN Electronic Journal.

[7] Atkinson A B. 1970. On the measurement of inequality, Journal of Economic Theory, 2(3):244-263.

[8] Atkinson A B. 2011. On lateral thinking, Journal of Economics Inequality 9: 319-328.

[9] Atkinson A B, Bourguignon F. 1982. The Comparison of MultiDimensioned Distributions of Economic Status, Review of Economic Studies, 49(2):183-201.

[10] Bedford T., Cooke R.M. (2002), Vines — a new graphical model for dependent random variables, Annals of Statistics, Vol. 30, p. 1031–1068.

[11] Duclos J Y, Echevin D. 2012. Health and income: A robust comparison of Canada and the US, Journal of Health Economics 30: 293–302.

[12] Cowell F, Flachaire E. 2015. Inequality with ordinal data, STICERD 130226 Working Paper.

[13] Cutler D M, Huang W, Lleras-Muney A. 2015. When does education matter? The protective effect of education for cohorts graduating in bad times, Social Science & Medicine 127:63-73.

[14] Cutler D, Lleras-Muney A, Vogl T. 2011. Socioeconomic status and health: dimensions and mechanisms, The Oxford Handbook of Health Economics.

[15] Genest C., MacKay J. (1986), *The joy of copulas: bivariate distributions with uniform marginals*, The American Statistician, Vol. 40, p. 280-283.

[16] Genest C., Nešlehová J. (2007), *A Primer on Copulas for Count Data*, ASTIN Bulletin, Vol. 37, p. 475-515.

[17] Gibbons R., Hedeker D. (1992), *Full-information item bi-factor analysis*, Psychometrika, Vol. 57, p. 423-436.

[18] Gravel N, Magdalou B, Moyes P. 2015, Ranking distributions of an ordinal attribute <halshs-01082996v2>

[19] Holzinger K.J., Swineford F. (1937), *The bi-factor method*, Psychometrika, Vol. 2, p. 41-54.

[20] Hua L., Joe H. (2014), *Strength of tail dependence based on conditional tail expectation*, Journal of Multivariate Analysis, Vol. 123, p. 143-159.

[21] Hult H., Lindskog F. (2002), *Multivariate extremes, aggregation and dependence in elliptical distributions*, Advances in Applied Probability, Vol. 34, p. 587-608.

[22] Joe H. (2005), *Asymptotic efficiency of the two-stage estimation method for copula-based models*, Journal of Multivariate Analysis, Vol. 94, p. 401-419.

[23] Joe H. (2015), *Dependence modelling with copulas*, Taylor & Francis Group.

[24] Krupskii P., Joe H. (2013), *Factor copula models for multivariate data*, Journal of Multivariate Analysis, Vol. 120, p. 85–101.

[25] Krupskii P., Joe H. (2015), *Structured factor copula models: Theory, inference and computation*, Journal of Multivariate Analysis, Vol. 138, p. 53–73.

[26] Kunst A E, Bos V, Lahelma E, Bartley M, Lissau I, Regidor E, et al.. 2004. Trends in socioeconomic inequalities in sefl-assessed health in 10 European countries, International Journal of Epidemiology 34: 295-305.

[27] Lazar A. Silber J. 2013. On the cardinal measurement of health inequality when only ordinal information is available on individual health status, Health Economics 22(1):106-113.

[28] Mackenbach J P, Stirbu I, Roskam A J R, Schaap M M, Menvielle G, Leinsalu M, Kunst A E. 2008. Socioeconomic inequalities in health in 22 European countries, New England Journal of Medicine 358:2468-2481.

[29] Lv G, Wang Y, Xu Y. 2015. On a new class of measures for health inequality based on ordinal data, Journal of Economic Inequality 13:465-477.

[30] Makdisi P., Yazbeck M. (2014), *Measuring socioeconomic health inequalities in the presence of multiple categorical information*, Journal of Health Economics, Vol. 34, p. 84-95.

[31] Marmot M., Oldfield Z., Clemens S., Blake M., Phelps A., Nazroo J., Steptoe A., Rogers N., Banks J. (2016), *English Longitudinal Study of Ageing*, Waves 0-7, 1998-2015 [data collection], 24th Edition, UK Data Service, SN: 5050, http://dx.doi.org/10.5255/UKDA-SN-5050-11

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

[33] Nikoloulopoulos A.K., Joe H. (2015), *Factor copula models for item response data*, Psychometrika.

[34] Panagiotelis A., Czado C., Joe H. (2012), *Pair Copula Constructions for Multivariate Discrete Data*, Journal of the American Statistical Association, p. 107-499.

[35] Oh D.H., Patton A.J. (2015), *Modelling Dependence in High Dimensions with Factor Copulas*, Finance and Economics Discussion Series 2015-051, Board of Governors of the Federal Reserve System.

[36] Sklar A. (1959), *Fonctions de répartition à n dimensions et leurs marges*, Publications de l'Institut de Statistique de l'Université de Paris, Vol. 8, p. 229–231.

[37] Schweizer B., Wolff E.F. (1981), *On nonparametric measures of dependence for random variables*, Annals of Statistics, 9:870885.

[38] van Doorslaer E, Koolman X. 2004. Explaining the differences in income-related health inequalities across European countries, Health Economics 13:609-628.

[39] van Doorslaer E, Wagstaff A, Bleichrodt H, Calonge S, Gerdtham U G, Gerfin M, Geurts J, Gross L, Hakkinen U, Leu R E, O'Donnell O, Propper C, Puffer F, Rodriguez M, Sundberg G, Winkelhake O. 1997. Income-related inequalities in health: some international comparisons, Journal of Health Economics 16:93-112.

Streszczenie

W ostatniej dekadzie mamy dostęp do szczegółowych danych zdrowotnych (np.: badanie SHARE). Modelowanie łącznego rozkładu wielu wskaźników zdrowotnych nie jest łatwym zadaniem. Literatura nie jest duża i nie adresuje właściwie np.: struktury zależności między wskaźnikami. Współwystępowanie chorób jest zaś kluczowe dla wydatków na opiekę zdrowotną. Celem artykułu jest pokazanie, że zależności te występują w stopniu, który nie może być ignorowany oraz rozszerzenie literatury o metody, które modelują łączny rozkład zdrowia elastycznie i wydajnie obliczeniowo. Są to dostępne od niedawna tzw. pair-copula constructions (PCC) (Aas et al. 2009). Mogą być one użyte przy wielu wymiarach zdrowia, gdzie inne metody są niekonkluzywne (Duclos i Echevin 2012). W oparciu o dane z English Longitudinal Study of Ageing (ELSA) nt. statusu zdrowotnego, mobilności, wzroku, słuchu czy zdrowia emocjonalnego, estymujemy jednoczynnikowy model (Nikoloulopoulos i Joe 2015). Wskaźniki zdrowia wykazują zależność w ogonach; kopule $t(4)$ i $t(5)$ wykazują najlepsze dopasowanie. Szczegółowe zależności są nie do wykrycia przez niedawno rozwinięte podejścia (Makdisi i Yazbeck 2014).