STUDIA EKONOMICZNE ● ECONOMIC STUDIES
NR 1 (XCII) 2017

Martyna Kobus*, Olga Półchłopek**

TWO FACTOR COPULA MODEL IN HEALTH MEASUREMENT

(Artykuł nadesłany: 15.02.2017 r.; Zaakceptowany: 25.06.2017 r.)

ABSTRACT

Researchers emphasize the need to include broad information on health in health analyses (Conti et al. 2010, Heckman et al. 2011). This is mostly because single health indicators are not efficient in describing a person's health, and detailed information on health is increasingly available via e.g. ageing surveys. This, however, calls for a joint analysis of health indicators, taking into account the dependencies between various health variables. It is not self-evident how one should model such a multidimensional distribution, and in our previous paper (Kobus and Półchłopek, 2016) we offer a method for ordinal health data that models them in a flexible and computationally efficient way, the so-called factor copula models (Nikoloulopoulos and Joe 2015). Here we continue research in this area. We use 2-factor models to estimate 24-dimensional health distribution and show that they provide a highly accurate description of health data and perform better than standard analyses based on multivariate normal distribution. We find that a 2-factor model which is a combination of the $t(5)+t(4)$ copulas gives the highest likelihood. Based on this, we identify two major factors which govern the 24 health variable distributions and based on the estimated copula parameters we give them interpretation. Factor one relates to the general mood and attitude in life, whereas factor two describes physical health.

Keywords: multiple health indicators; interdependence; factor copulas.

JEL Classification: I31, D63

* Institute of Economics Polish Academy of Sciences, Warsaw; email:mkobus@inepan.waw.pl.
** Vistula University, Warsaw; email: olga.polchlopek@gmail.com.

INTRODUCTION

Socio-economic inequalities have received soaring attention in the last decades. Inequalities of health and well-being have attracted particular interest, especially as detailed health information is available through ageing surveys (e.g. SHARE in Europe). This includes information on not only physical health, but also mental and cognitive. Many of these indicators are in the form of ordinal data e.g. for which there are no numbers, but only the underlying ordering. One of most widely used health indicators, namely self-reported health status, is one example of an ordinal indicator. Here we have ordered categories of health statuses from very bad to very good. Health measurements studies are often restricted to this one variable (van Doorslaer et al., 1997, van Doorslaer and Koolman, 2004, Kunst et al., 2004, Cutler et al., 2015) or even if they use several health indicators, they either analyze them separately or aggregate them into a single index of health (e.g. Makdisi and Yazbeck, 2014). All this ignores the dependence structure in health data which is a distinctive feature of multidimensional versus unidimensional distributions. As it is widely acknowledged (Atkinson and Bourguignon, 1982), dependencies in well-being components increase inequality, as there is higher likelihood for individuals to suffer from multiple deprivations. This group is the proper target of any policy aimed at reducing health inequalities, especially for chronic diseases which plague modern societies and constitute an increasing drag on public finance, such as diabetes, hypertension, arthritis etc. That is to say, modelling health in a joint framework is a necessity for obtaining reliable health measurements, but it has been rare in the literature so far, also because of methodological and computational difficulties.

In our previous paper (Kobus and Półchłopek, 2016) on this topic we shed some light on how this can be done in the context of ordinal health indicators. We show that so called 1-factor copulas remarkably improve the modelling of health distributions, and detect complex non-linear behavior. Large part of the paper was devoted to the introduction of the concept. This paper is a continuation of this line of work, in which we apply more complex factor models i.e. 2-factor copulas. They are a more efficient tool for the joint treatment of ordinal health data which we describe in detail below.

A regular approach to multivariate modelling is based on normal (Gaussian) distribution. It is at the same time the most widely used and most criticized. The multivariate Gaussian distribution assumes that all margins, as well as the joint probability, are normal. However, in reality this is often difficult to observe, e.g. in finance income dispersion is often uneven or there may be share prices which are dependent only when they reach high values. This should be taken into account for a successful portfolio diversification. Especially when inequalities are concerned, it is highly inconvenient to assume normal distribution. The main drawbacks are the light tails and the lack of negative dependence. Tail dependence is a property that indicates how the 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. In a multivariate Gaussian distribution majority of mass (about 90%) is clustered in the center, while the tails are symmetrical and carry little mass. Negative dependence corresponds to negative correlation between variables – a situation also not covered by the normal distribution. These problems occur both on univariate and joint levels, so the mismatch caused by applying this method is compounded when the number of dimensions increases. Henceforth, it is important to study non-Gaussian models, as they have potential to be more accurate in atypical cases witnessed in risk analysis, insurance, finance and economics, and as our previous paper shows, in health analyses too.

A promising proposition of resolving the above mentioned problems is implementing copulas. A copula is a multivariate probability distribution function with uniform margins. Such function combined with marginal distribution characterize the joint distribution. For continuous data such representation is unique, which is stated in Sklar's theorem (Sklar, 1959). It is therefore a convenient method to construct joint distributions from marginal distributions. Moreover, copulas allow for a two-step estimation where the steps are independent. First, univariate margins are chosen, based on initial diagnostics of the dataset, and just any existing distributions can be used. Each item can be fit with a different distribution and different support – univariate Gaussian and *t* for retaining symmetry, gamma for exponential tails, Pareto for heavy tails, Poisson for integer-valued data etc. After the margins are picked, copula models can be considered based on the joint behavior of the data. Since there are numerous copulas to choose from, diverse in their properties, the multivariate model is highly flexible. There are also unimodal parametric copula families which introduce a continuous factor (a matrix or a vector) that can be fixed for each pair of variables for the best fit. So called extreme value copulas are characterized by uneven tails and can be rotated to obtain negative dependence. Such copulas are a good fit for responses derived from best-case or worst-case scenarios. For example, when asked about mobility limitation, the respondent takes into account only the 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, they take the minimum value of all the events relevant to the question (Nikoloulopoulos and Joe, 2015). In addition, there are many ways to combine copulas in order to obtain a joint distribution. Copulas usually have closed forms which are convenient in terms of computation because they allow analytical differentiation and integration instead of numerical. Although models that designate only one multivariate copula for all data (like the multivariate Gaussian model does) are in line with intuition, building high-dimensional copulas is considered a difficult computing problem. This implies the necessity of new solutions in numerical methods and algorithms for inference and simulation. This is why concepts of vine pair copula constructions (Aas et al., 2009) and factor copulas (Nikoloulopoulos and Joe, 2015) were proposed. They are based on linking different bivariate copulas to connect the variables separately in some order instead of all at once. The idea is to decompose a multivariate distribution into a series of bivariate copulas applied on original variables and on their conditional distribution functions.

This paper is a continuation of our research on English Longitudinal Study of Ageing (ELSA), waves 1 and 6. The previous paper (Kobus and Półchłopek, 2016) applies 1-factor copula models, where 1-factor indicates that there is a single unobservable factor that governs the behavior of all observed indicators. We showed that there are substantial dependencies in health data which cannot be easily neglected. In more detail, we found that factor copula models based on $t(4)$ and $t(5)$ copulas provide better fit than standard multivariate normal model. This suggests that health distributions are generated as a mixture of discretized means which is typical for a sample which is a mixture of heterogeneous groups. These interdependencies were present in all analyzed population subgroup. We observed the strongest dependence for items that express general optimism and such was the interpretation of the underlying factor. The paper closes with pointing to three directions for further research: (i) estimation of multi-factor models, (ii) estimation of structured factor copula models where a group structure of the data is implied (iii) applying different linking bivariate copulas. Hereby we address the first idea. We find that a 2-factor model which is a combination of the $t(5)+t(4)$ copulas gives the highest likelihood. Based on this, we can identify two major factors which summarize 24-dimensional distribution. Based on the estimated copula parameters we give the following interpretation of the two factors. Factor one relates to general mood and attitude in life, whereas factor two describes physical health. This differentiation is stable among population groups with most noticeable changes between men and women. Namely, for women first factor has a broader meaning and apart from general optimism it is also related to the sense of freedom in life.

The paper is organized as follows. In Section 1 we describe only the essential part of the methodology and estimation referring more interested readers to Kobus and Półchłopek (2016) which contains a more thorough description of the concepts used here. In Section 2 we describe the data and analyze descriptive statistics. Section 3 contains the results. Finally, we conclude in Section 4.

1. 2-FACTOR COPULA MODELS

The theory of factor copulas was developed by Nikoloulopoulos and Joe (2015) and the exact model used in this study is described in Kobus and Półchłopek (2016). Here we expand 1-factor copula model presented there. Please refer to the notation described in Kobus and Półchłopek (2016).

Latent factor models are models of high-dimensional data in which many items overlap or are correlated by design. Typically, the underlying concept (factor) is low-dimensional, but it is difficult to measure directly and thus proxied by many indicators that are easier to measure. Typical examples of include concepts such as quality of life, general intelligence or health. The general factor copula model is the following. Let $Y_i = (Y_{i1}, ..., Y_{id})$ be a vector of d ordinal variables each measured on a scale $\{0, ..., K-1\}$. These are observable indicators; here in empirical application $d = 24$ and each Y_j denotes a given health condition. The p-factor model assumes conditional independence of $Y_1, ..., Y_d$ given latent variables $X_1, ..., X_p$ (so called factors). The joint probability mass function (pmf) is therefore

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

where $F_{X_1, ..., X_p}$ is the joint distribution of latent factors. Copulas appear in how $P(Y_j = y_j | X_1 = x_1, ..., X_p = x_p)$ is modelled.

Here we present a 2-factor model ($p = 2$) in which X_1, X_2 are independent and uniformly distributed i.e. $X_1, X_2 \sim U(0, 1)$. From Sklar's theorem (Sklar, 1959) there exists a copula function $C_{X_1 j}$ such that $\mathrm{P}(X_1 \leq x_1, Y_j \leq y_j) = C_{X_1 j}(x_1, F_j(y_j))$, where F_j is the cdf of Y_j. It is a step function with jumps at $0, ..., K-1$. Let a_{jk} be the cutpoints for the j-th ordinal variable in the uniform scale. Thus $F_j(y_j) = a_{j, y_j + 1}$. Then, the conditional cdf is the following

$$F_{j|X_1}(y_j \mid x_1) := P(Y_j \leq y_j \mid X_1 = x_1) = \frac{\partial C_{X_1 j}(x_1, F_j(y_j))}{\partial x_1}. \tag{2}$$

Denoting copula density function $\frac{\partial C_{X_1 j}(x_1, F_j(y_j))}{\partial x_1}$ by $C_{j|X_1}(a_{j, y_j + 1} | x_1)$, we have that $C_{j|X_1}(a_{j, y_j + 1} | x_1) - C_{j|X_1}(a_{j, y_j} | x_1)$ gives the probability of $Y_j = y_j$ conditional on $X_1 = x_1$.

Let $C_{X_2 j}$ be a bivariate copula such that $\mathrm{P}(X_2 \leq x_2, Y_j \leq y_j | X_1 = x_1) = C_{X_2 j}(x_2, F_{j|X_1}(y_j | x_1))$ and $F_{j|X_1}(y_j | x_1)$ is given by (2). It is worth noting that a simplifying assumption is used, namely, the conditional copula for the univariate distributions $F_{Y_j|X_1}$ and $F_{X_2|X_1} = F_{X_2}$ does not depend on x_1. Denoting copula density function for factor X_2 by $C_{j|X_2}(F_{j|x_1}(y_j \mid x_1) \mid x_2) = \frac{\partial C_{X_2 j}(x_2, F_{j|x_1}(y_j \mid x_1))}{\partial x_2}$ we can write the joint pmf for the 2-factor model $P(Y_1 = y_1, ..., Y_d = y_d) =$

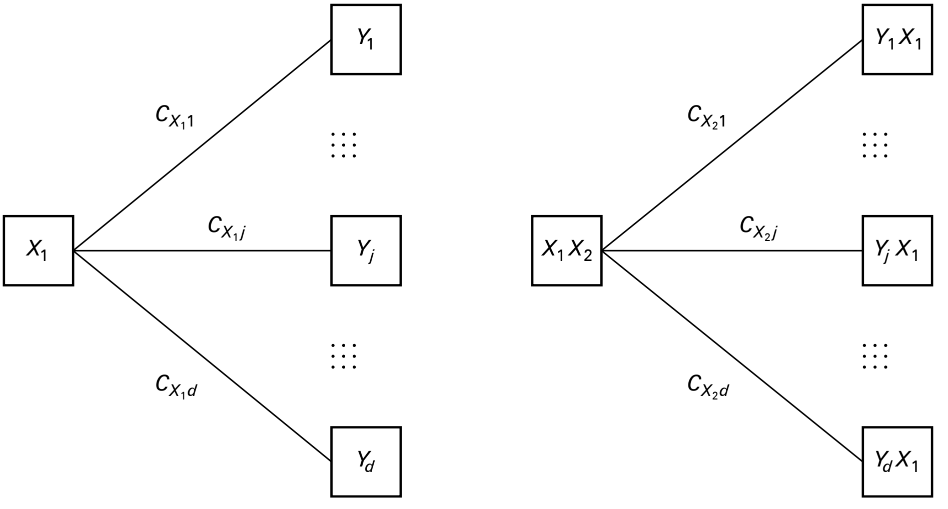
$$= \int_0^1 \int_0^1 \prod_{j=1}^{d} C_{j|X_2}(F_{j|X_1}(y_j \mid x_1) \mid x_2) - C_{j|X_2}(F_{j|X_1}(y_j - 1 \mid x_1) \mid x_2) dx_1 dx_2.$$

Less formally speaking, d items are linked to factor X_1 (1-factor model) and then conditional on X_1 another d items are linked to factor X_2 as pictured in Figure 1.

In theory, copulas $C_{X_1 1}, ..., C_{X_1 d}, C_{X_2 1}, ..., C_{X_2 d}$ can all come from different parametric families which compounds flexibility of the model. A special case is the discretized multivariate normal model for which all the copulas are bivariate Gaussian. A presentation of all copula families used in this paper is contained in

Kobus and Półchłopek (2016). Estimation follows two-step procedure described in detail in Kobus and Półchłopek (2016) and conducted using R package “CopulaModel”.

Figure 1. Graphic representation of the 2-factor copula model



Source: Nikoloulopoulos and Joe (2015).

We compare 2- and 1-factor models. For 1-factor models, the following copulas are tested: Gumbel, survival Gumbel, Gaussian, as well as copulas with 2, 3, 4, 5, 7 and 9 degrees of freedom. For 2-factor models 17 pairs of copulas are considered: Gumbel+Gumbel, s.Gumbel+s.Gumbel, Gaussian+Gaussian, $t(2)+t(3)$, $t(3)+t(2)$, $t(2)+t(4)$, $t(4)+t(2)$, $t(4)+t(5)$, $t(5)+t(4)$, $t(2)$+Gumbel, $t(3)$+Gumbel, $t(4)$+Gumbel, $t(5)$+Gumbel, Gumbel+$t(2)$, Gumbel+$t(3)$, Gumbel+$t(4)$ and Gumbel+$t(5)$. It is not so common to mix diverse parametric families for each factor because of the tail dependence inheritance. Moreover, the theory of factor copulas is relatively recent and such intricate models would require tools that have not yet been developed for ordinal data. For these reasons, only one family was chosen for each factor. The order of copulas in a single pair does matter, hence the use of symmetrical mixes such as $t(2)+t(4)$ and $t(4)+t(2)$.

2. DATA AND DESCRIPTIVE STATISTICS

ELSA (English Longitudinal Study of Ageing)[1] is a survey of quality of life among elderly people in the UK. Waves 1 and 6 (2002, 2012) were downloaded from the UK Data Archive. Although the dataset allows longitudinal analyses,

[1] 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

only cross-sectional studies were conducted to show if and how the conditions of the elderly have changed over time.

The multivariate model comprises of 24 variables: 19 items describing control ($C1$–$C4$), autonomy ($A1$–$A5$), pleasure ($P1$–$P5$) and self-realization ($S1$–$S5$), 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 analyze 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 (behavioral risk). This gives us 10 groups to analyze 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.

To compare distributions we use Wilcoxon rank-sum test (Wilcoxon, 1945, Mann and Whitney 1947). Rank-sum tests the hypothesis that two independent samples are from populations with the same distribution. Unlike Student's *t*-test, it does not require the assumption of normal distribution. The test is a standard choice for ordinal data. We choose three variables that later on prove relevant for copula analysis. They reflect both physical and mental health. With respect to ability to walk, the considered groups differ significantly (Table 3) i.e. men are different than women (sex), those who smoke are different from those who do not (smoking) etc. With respect to $P3$, namely, whether a respondent enjoys things he or she does, there are differences between all groups except for men and women and employed and unemployed. Here we cannot reject the null hypothesis that the samples come from the same distribution. Similarly for *S4* i.e. whether a respondent feels that life is full of opportunities. Typically groups differ except for men and women in Wave 6 and employed and unemployed in Wave 1.

Table 3. Wilcoxon test

	Wave 1			Wave 6		
	sex	age	smoking	sex	age	smoking
walk	2.8 (0.01)*	15.26 (0.00)*	3.31 (0.00)*	3.58 (0.00)*	17.69 (0.00)*	6.03 (0.00)*
*P*3	0.22 (0.83)	–3.22 (0.00)*	6.79 (0.00)*	–0.05 (0.96)	–7.24 (0.00)*	9.43 (0.00)*
*S*4	–2.03 (0.04)*	4.84 (0.00)*	6.11 (0.00)*	0.45 (0.65)	5.66 (0.00)*	7.81 (0.00)*

	Wave 1			Wave 6		
	employed	unemployed	disabled	employed	unemployed	disabled
walk	–17.88 (0.00)*	–3.36 (0.00)*	17.23 (0.00)*	–21.45 (0.00)*	–2.52 (0.01)*	20.63 (0.00)*
*P*3	1.25 (0.21)	3.97 (0.00)*	9.36 (0.00)*	5.33 (0.00)*	5.9 (0.00)*	12.81 (0.00)*
*S*4	–5.36 (0.00)*	0.36 (0.72)	8.16 (0.00)*	–5.9 (0.00)*	4.35 (0.00)*	11.79 (0.00)*

3. RESULTS

Most general result is such that in both waves, the combination of copulas with various degrees of freedom triumphs other models. In wave 6, all groups show the highest values of log-likelihood function when estimated with the pair of copulas $t(5)+t(4)$, where the numbers indicate degrees of freedom. In this model, the X_1 factor is linked to each of the 24 items using a $t(5)$ copula and then these distributions are combined with the X_2 factor using 24 $t(4)$ copulas. In wave 1 the groups go best with $t(5)+t(4)$ copulas as well, except for women, retired and unemployed people who are most accurately described with $t(4)+t(5)$ copulas. It is in line with the expectations, as the 1-factor models have also proved t copulas with 4 and 5 degrees of freedom to be the best choice.

Table 4 illustrates bivariate count distributions of variables *health* and *A4* simulated using copulas pair that gives best log likelihood ($t(5)+t(4)$ combination), as well as observed distribution. Because *A4* refers to restraints imposed by health condition, and so it 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 203 men (upper left corner) who rate their health as poor and simultaneously reported that it often stops them from doing what they want. On the other hand, 833 men (bottom right corner) say that their health was very good and that they do not feel constrained.

Table 4. Bivariate count distributions of the models with the best log-likelihood

	empirical distribution				1-factor				2-factor			
males	203	54	8	8	102	80	55	39	185	65	16	11
	212	293	87	50	150	217	131	140	193	285	123	48
	99	396	378	276	167	335	340	332	121	423	399	223
	28	208	456	833	118	290	404	689	46	196	424	831
females	190	56	6	4	95	94	39	40	175	57	8	10
	255	404	86	37	155	290	146	178	254	360	104	53
	111	528	402	331	190	427	360	416	123	560	430	265
	43	234	539	1100	120	424	461	891	45	260	463	1159
age 50–64	156	44	6	5	61	54	53	59	148	51	20	11
	157	248	86	46	104	140	121	138	151	212	126	36
	64	338	380	333	114	292	344	405	83	328	424	270
	28	160	497	1186	113	340	458	938	22	168	448	1236
age 65+	237	66	8	7	155	113	35	25	248	65	10	8
	310	449	87	41	223	332	191	137	316	372	126	42
	146	586	400	274	220	519	374	318	145	637	439	213
	43	282	498	747	127	398	401	613	54	302	412	792
non-smo-king	312	94	11	9	157	117	71	60	307	71	20	14
	396	586	151	71	266	448	266	260	369	567	218	90
	189	827	696	513	278	700	656	610	208	871	747	418
	66	403	946	1750	237	656	811	1427	60	431	816	1813
smoking	203	54	8	8	102	80	55	39	185	65	16	11
	212	293	87	50	150	217	131	140	193	285	123	48
	99	396	378	276	167	335	340	332	121	423	399	223
	28	208	456	833	118	290	404	689	46	196	424	831
retired	225	62	8	8	140	81	31	30	217	44	14	9
	331	479	96	43	253	379	195	147	312	456	151	55
	148	611	441	311	217	566	433	334	160	660	435	259
	44	284	556	911	140	405	452	755	52	284	505	945
employed	20	18	4	2	9	7	6	16	17	13	5	5
	52	148	59	33	25	102	78	115	65	120	52	47
	32	251	284	263	31	182	233	352	25	273	339	229
	13	129	380	914	47	277	419	703	11	148	358	895
unem-ployed	18	9	2	1	7	5	13	5	11	9	4	1
	26	48	15	8	21	25	22	26	23	44	18	16
	12	45	51	32	17	45	43	38	15	67	52	24
	10	28	57	106	19	56	55	71	8	24	63	89
disabled	130	21	0	1	109	27	6	4	127	12	1	3
	58	22	3	3	61	24	4	5	69	24	3	0
	18	17	4	1	23	10	1	2	17	15	5	2
	4	1	2	2	10	1	0	0	5	0	3	1

Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 5. Bivariate count distributions of items *A5* and *S5* modelled with copulas

	males				females					males				females			
empirical	63	176	188	63	65	166	308	89	5+4	70	136	177	96	87	163	260	131
	36	201	566	300	37	180	686	413		49	224	554	270	46	220	673	365
	24	123	499	440	34	162	549	526		15	152	522	423	30	153	601	514
	31	89	307	483	49	95	376	591		20	82	306	493	29	82	377	595
Gumbel	44	109	221	92	47	132	297	157	2+Gumbel	66	125	198	92	67	130	275	142
	54	233	543	263	80	217	655	382		37	245	543	262	47	218	707	357
	31	173	541	378	44	171	573	478		27	138	513	384	32	125	622	487
	13	67	297	530	27	87	389	590		27	82	288	562	42	92	384	599
s.Gumbel	42	115	223	95	47	132	297	157	3+Gumbel	64	120	195	93	65	126	289	148
	55	226	548	279	80	217	655	382		48	239	529	278	53	221	679	357
	29	168	516	376	44	171	573	478		25	152	490	394	40	141	606	491
	12	67	310	528	27	87	389	590		24	81	314	543	30	80	403	597
Gaussian	54	137	201	88	72	155	290	113	4+Gumbel	61	119	200	88	77	139	270	137
	55	214	537	301	69	238	653	387		50	243	520	285	52	231	681	388
	25	165	482	402	28	141	551	517		28	155	489	396	34	140	626	458
	12	76	369	471	19	73	416	604		21	77	317	540	28	84	383	598
2+3	79	160	133	107	94	164	201	184	5+Gumbel	60	117	204	87	52	119	271	186
	24	199	592	280	31	199	724	370		51	244	517	290	59	200	656	407
	18	137	567	377	14	136	657	479		28	155	487	395	40	161	607	472
	37	98	290	491	45	97	358	573		18	81	319	536	36	110	432	518
3+2	92	143	147	110	99	158	224	156	Gumbel+2	63	140	174	107	69	123	288	136
	25	207	589	250	39	210	690	370		36	226	560	270	68	232	675	356
	16	143	541	397	22	123	658	474		29	162	523	382	38	149	619	477
	27	99	278	525	41	99	339	624		21	66	295	535	29	83	321	663
2+4	64	149	159	97	85	136	256	146	Gumbel+3	62	140	181	86	69	124	299	129
	43	214	581	286	38	215	701	385		38	231	551	288	69	234	672	355
	21	140	514	406	27	131	586	503		33	154	505	395	37	145	601	487
	25	81	330	479	41	83	443	550		15	73	304	533	28	78	338	661

	males				females					males				females			
4+2	72	136	173	97	86	159	259	141	Gumbel+4	63	138	183	82	69	124	311	122
	45	214	556	275	48	213	665	370		40	232	552	285	65	240	661	359
	18	163	504	420	29	159	594	519		35	151	510	389	42	143	595	485
	20	87	307	502	33	79	383	589		13	71	309	536	27	78	345	660
4+5	69	139	179	96	90	138	246	152	Gumbel+5	61	138	184	80	69	124	313	118
	46	222	549	279	37	232	686	386		41	231	555	284	67	232	662	365
	18	148	539	403	26	134	609	488		33	153	506	395	39	145	598	485
	22	80	317	483	35	85	420	562		13	74	309	532	25	78	351	655

Source: Own calculations based on the ELSA wave 1 and wave 6.

Firstly, it can be seen already from Table 4 that the 2-factor models are better approximations than 1-factor models e.g. 185 of people in the first category as estimated by a 2-factor model is closer to 203 than 102 as estimated by a 1-factor model. The same goes for all population groups. In general, 1-factor model tends to spread mass more uniformly among categories than a 2-factor model, simply because it tries to model jointly the behavior of what seems to be two independent factors. Secondly, as shown in Table 5 which contains bivariate count distributions of items $A5$ and $S5$ used in Kobus and Półchłopek (2016), copulas estimate the probabilities far better than multivariate Gaussian models. Empirically $A5$, which refers to shortage of money, has more probability in the centre, it can be thus interpreted as a discretized mean, while $S5$ which indicates if the future looks good to the interviewee, has more mass in maximum values. This is why the joint probability is asymmetric and upper tail dependence is observed.

In each model, the dependence parameter vector $\vec{\theta}$ consists of 48 real values – the first 24 corresponding to copulas fitted for X_1 and the latter for X_2. The most accurate models use t copulas, hence the values range between –1 and 1.[2] Tables 8, 9, 10 and 11 in Appendix contain transformations of the parameters to Kendall's τ. The higher the copula parameter, the more dependence between a given item and the factor.[3] Copula models with different parameters are not

[2] For example, if Gumbel copula is used, then because it models positive dependence only, Kendall's tau values are between 0 and 1.

[3] As to the interpretation of factors, Nikoloulopoulos and Joe (2015) state that 2-factor models with t_ν copulas are clearly interpretable $\nu \leq 3$ when and that there is more variability when $\nu \geq 5$ which may affect identification of the factors. The varimax transformation (Kaiser, 1958) is necessary when such problems occur. Interpretation can also be distorted if variables have discrete distribution and if the sample is big, as standard errors decrease when (the number of records) or the number of categories increase. Large SEs (exceeding 0.9 in Kendall's scale) indicate that the model is non-identifiable. This is not the case in this study though, as SEs remain between 0.01 and 0.04, so the varimax rotation is not needed to explain the factors with their loadings.

directly comparable, therefore parameter estimates and standard errors need to be compared on a Kendall's tau scale (Genest and MacKay, 1986), (Hult and Lindskog, 2002). The following transformation of copula parameters applies to Gaussian t copulas with $\theta \in (-1,1)$ (Hult and Lindskog, 2002)

$$\tau = \frac{2}{\pi}\arcsin\theta, \tag{3}$$

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

$$\tau = 1 - \frac{1}{\theta}. \tag{4}$$

Therefore, Kendall's τ for Gaussian and copulas range from –1 to 1 and as for Gumbel, $\tau \in (0,1)$, because this copula models only positive dependence.

Let us first analyze the group of non-smokers (Appendix, Figure 2) which is the most numerous. The model has managed to reduce 24-dimensional health indicator to two major unobserved factors. Variables $C3, A1, P1, P2, P3, P5, S3$, $S4$ and $S5$ seem to be linked to the factor X_1, whereas X_2 is clearly linked to $A4$ and also to *health*, *walk*, *pain* and $C1$. Interpreting the items, X_1, variables strongly connected to it cover the topics of freedom ($C3, A1$), joy ($P3$), optimism ($P1, P2$, $S4$, $S5$), happiness and satisfaction ($P5$, $S1$–$S5$). It can be therefore interpreted as life attitude or general mood. The second factor is even more clear. It is mostly connected to physical health indicators such as self-reported health, mobility and pain rating, as well as two CASP-19 items, $C1$ and $A4$. The latter are about whether health or age prevents a respondent from doing something, so they have a clear association with physical health. Therefore, X_2 can be described as physical health/limitations. For the group of smokers the situation is very much similar. Variables related to X_1 ($P1, P2, P3, P4, P5, A2, A3$ account for optimism and the feeling of freedom, while the loadings of X_2 are the same as for the group of non-smokers. It seems that smoking does not affect perception of health, however, sex does (Appendix, Figure 3). For men, we still observe the pattern that X_1 is linked to $P1$, $P2$, $P3$, $S3$, $S4$, $S5$ which are all the feelings of optimism and enjoyment, and X_2 is linked to physical health (*health*, *walk*, *pain*, $A4$ and $C1$). The situation changes for women, for whom the first factor has a much broader meaning. Namely, apart from the variables we just mentioned in the men's case, X_1 is also linked to answers $C3$, $A1$, $A4$ and $S1$. They all relate to the sense of freedom, which apparently for women does not mean only freedom from the limitations of physical health, but a general ability to do what one wishes to do in life.

With age, walking and the feeling that health prevents one from doing things they want ($A4$) emerges as two indicators driving factor X_2 (Appendix, Figure 4). When it comes to factor X_1 there are many indicators that link to X_1 with a similar loading for both age groups. These are a few traditional ones $P1$, $P2$, $P3$, $S3$, $S4$, $S5$ but also $P5$ and $S1$. $P5$ refers to looking back at one's life with a sense of happiness and $S1$ to feeling full of energy. The importance of these two for the general positive feeling about life increases with age. The last group we analyze is related to employment status. Here too clear interpretation for factors X_1 and X_2 can be drawn. What changes between groups is the importance of indicators $P5$ and $S1$. $P5$ is more linked to X_1 for employed than for other groups, whereas for the retired $S1$ is linked to both factor first and second.

The 2-factor model has not only proved to be better than multivariate normal approach for modelling data with intermediate tail dependence, but has also provided highly accurate approximation. Moreover, it enables successful identification of unobservable determinants and reveals important links between various health indicators. The novelty of this method is that it detects complex dependencies present in the data. This provides not only for the desired dimensional reduction but also for the natural interpretation of underlying factors.

CONCLUSION

The analysis identified two main factors governing 24-dimensional health distribution. For Wave 1 the single factor was interpreted as positive life attitude, as variables $S1$ and $S5$ appeared to have the highest linking parameters, and for Wave 6 also $S3$, $P1$ and $P3$ were relevant (enjoying things in life). It is in line with the description of X_1 in this study. The other factor turned out to represent physical constraints. It seems that since the first factor appeared in both 1- and 2-factor models, it is the main determinant of well-being, or more accurately, people's perception of it. The fact that the copula models evinced the highest values of the log-likelihood function implies that the dataset describes a population which is a mixture of subpopulations and the latent variables can be interpreted as compilations of means that vary across these subgroups. This paper shows that 2-factor copula models provide a better fit than 1-factor copula models presented in Kobus and Półchłopek (2016) and also better than standard tools (based on MVN), and should be used with health data when joint analysis is necessary. The output of these estimations can now be used for further analyses.

From the policy perspective, the added value of factor copula models in health measurement is the following. Firstly, it allows for the computationally efficient estimation of the joint health distribution (impossible with ordinal probit model with high-dimensional data) which naturally can then be used for further models and purposes. The necessity to consider joint health distributions comes from the fact that single health indicators are typically poor proxies of general health. Here not only many health indicators can be considered but also the dependencies between them. Such dependencies, known in the health literature as comorbidity pose a fiscal challenge to healthcare policy. Secondly, the model allows for a very flexible modelling of these dependencies, hence nonlinearities in the relationship between health conditions are taken into account (as opposed to ordinal probit model). Thirdly, it detects independent dimensions and allows for their interpretation. This reduces the dimensionality of the problem and shapes policymakers' thinking about most relevant health differences between various groups. Finally, the detected nonlinearities (i.e. upper and lower tail dependence) help to target most vulnerable groups most effectively. In particular, if the copula between two health dimensions in a given group is upper tail dependent one may conclude that the problem concentrates mostly for the high values of the considered health conditions and therefore targeting this group may prove to be the most efficient, given limited budget.

Although only the most relevant results were shown here, this study was very comprehensive (we estimated 17 different models). Still, it could be developed further. Firstly, one could fit margins and add covariates in copula estimations (Nikoloulopoulos and Karlis, 2010). Secondly, even more copula combinations (models) can be tested. Thirdly, the structure of the data (division into C, A, S and P questions) can be a priori taken into account via the so-called structured factor copula models (Krupskii and Joe, 2015).[4] CASP-19 indeed has a group structure (control, autonomy, pleasure, self-realization). In these models there are many zeros in vectors of loadings (the factors are loaded on separate subsets of variables) which means fewer dependence parameters. Furthermore, the factors can be dependent. In this model, on the other hand, each factor can be linked to any other item, and factors are independent.

REFERENCES

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

Abul Naga R.H., Yalcin T. (2008), *Inequality measurement for ordered response health data*, "Journal of Health Economics", 27(6): 1614–1625.

Abul Naga R.H., Stapenhurst Ch. (2015), *Estimation of inequality indices of the cumulative distribution function*, "Economics Letters", 130: 109–112.

Allison R.A., Foster J.E. (2004), *Measuring health inequality using qualitative data*, "Journal of Health Economics", 23(3): 505–524.

Apouey B. (2007), *Measuring health polarization with self-assessed health data*, "Health Economics", 16: 875–894.

Apouey B., Silber J. (2013), *Inequality and bi-polarization in socioeconomic status and health: ordinal approaches*, "SSRN Electronic Journal".

Atkinson A.B. (1970), *On the measurement of inequality*, "Journal of Economic Theory", 2(3): 244–263.

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

[4] Structured factor copulas are extensions to the Gaussian bi-factor model covered by Gibbons and Hedecker (1992), as well as Holzinger and Swineford (1937). However, structured copulas have only been developed for continuous data Krupskii and Joe (2015).

Atkinson A.B., Bourguignon F. (1982), *The Comparison of Multi Dimensioned Distributions of Economic Status*, "Review of Economic Studies", 49(2):183–201.

Bedford T., Cooke R.M. (2002), *Vines – a new graphical model for dependent random variables*, "Annals of Statistics", 30: 1031–1068.

Duclos J.Y., Echevin D. (2012), *Health and income: A robust comparison of Canada and the US*, "Journal of Health Economics", 30: 293–302.

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

Cowell F., Flachaire E. (2015), *Inequality with ordinal data*, STICERD 130226 Working Paper.

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.

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

Genest C., MacKay J. (1986), *The joy of copulas: bivariate distributions with uniform marginals*, "The American Statistician", 40: 280–283.

Genest C., Nešlehová J. (2007), *A Primer on Copulas for Count Data*, "ASTIN Bulletin", 37: 475–515.

Gibbons R., Hedeker D. (1992), *Full-information item bi-factor analysis*, "Psychometrika", 57: 423–436.

Gravel N., Magdalou B., Moyes P. (2015), *Ranking distributions of an ordinal attribute*, <halshs-01082996v2>

Heckman J., Humphries J.E., Urzua S., Veramendi G. (2011), *The effects of educational choices on labor market, health, and social outcomes*, University of Chicago, unpublished manuscript.

Holzinger K.J., Swineford F. (1937), *The bi-factor method*, "Psychometrika", 2: 41–54.

Hua L., Joe H. (2014), *Strength of tail dependence based on conditional tail expectation*, "Journal of Multivariate Analysis", 123: 143–159.

Hult H., Lindskog F. (2002), *Multivariate extremes, aggregation and dependence in elliptical distributions*, "Advances in Applied Probability", 34: 587–608.

Joe H. (2005), *Asymptotic efficiency of the two-stage estimation method for copula--based models*, "Journal of Multivariate Analysis", 94: 401–419.

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

Kaiser H.F. (1958), *The varimax criterion for analytic rotation in factor analysis*, "Psychometrika", 23: 187–200.

Krupskii P., Joe H. (2013), *Factor copula models for multivariate data*, "Journal of Multivariate Analysis", 120: 85–101.

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

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 countrie*s, "International Journal of Epidemiology", 34: 295–305.

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.

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.

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.

Makdisi P., Yazbeck M. (2014), *Measuring socioeconomic health inequalities in the presence of multiple categorical information*, "Journal of Health Economics", 34: 84–95.

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

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

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

Nikoloulopoulos A.K., Karlis D. (2010), *Regression in a copula model for bivariate count data*, "Journal of Applied Statistics", 37(9): 1555–1568.

Panagiotelis A., Czado C., Joe H. (2012), *Pair Copula Constructions for Multivariate Discrete Data*, "Journal of the American Statistical Association", 107–499.

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.

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

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

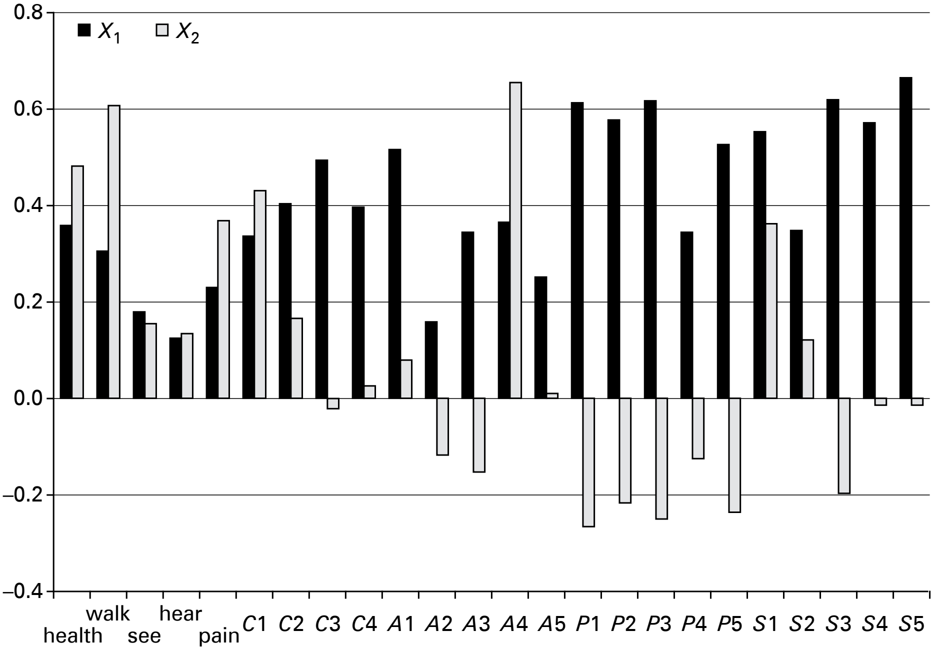
van Doorslaer E., Koolman X. (2004), *Explaining the differences in income-related health inequalities across European countries*, "Health Economics", 13: 609–628.

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.

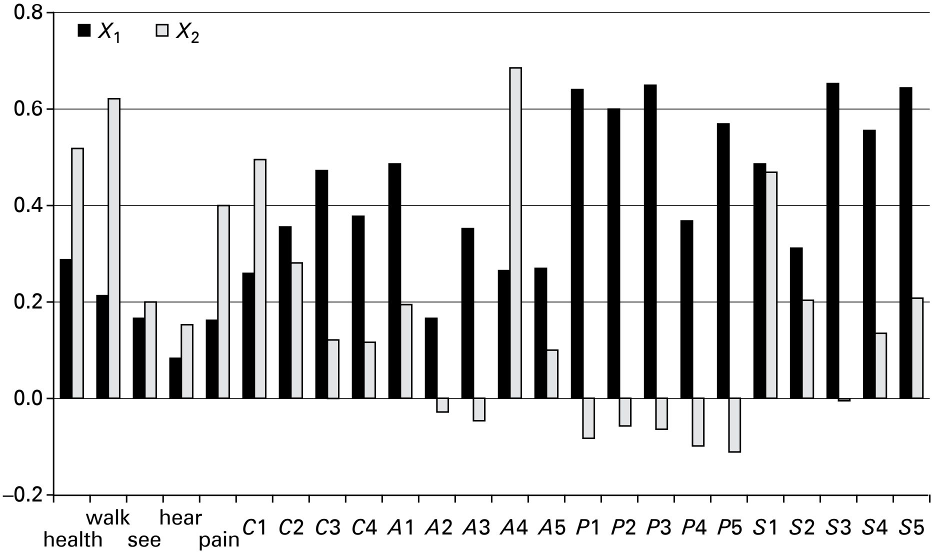
APPENDIX

Figure 2. Parameters *non-smoking* and *smoking* in Kendall's scale for the 2-factor models

(a) non-smoking

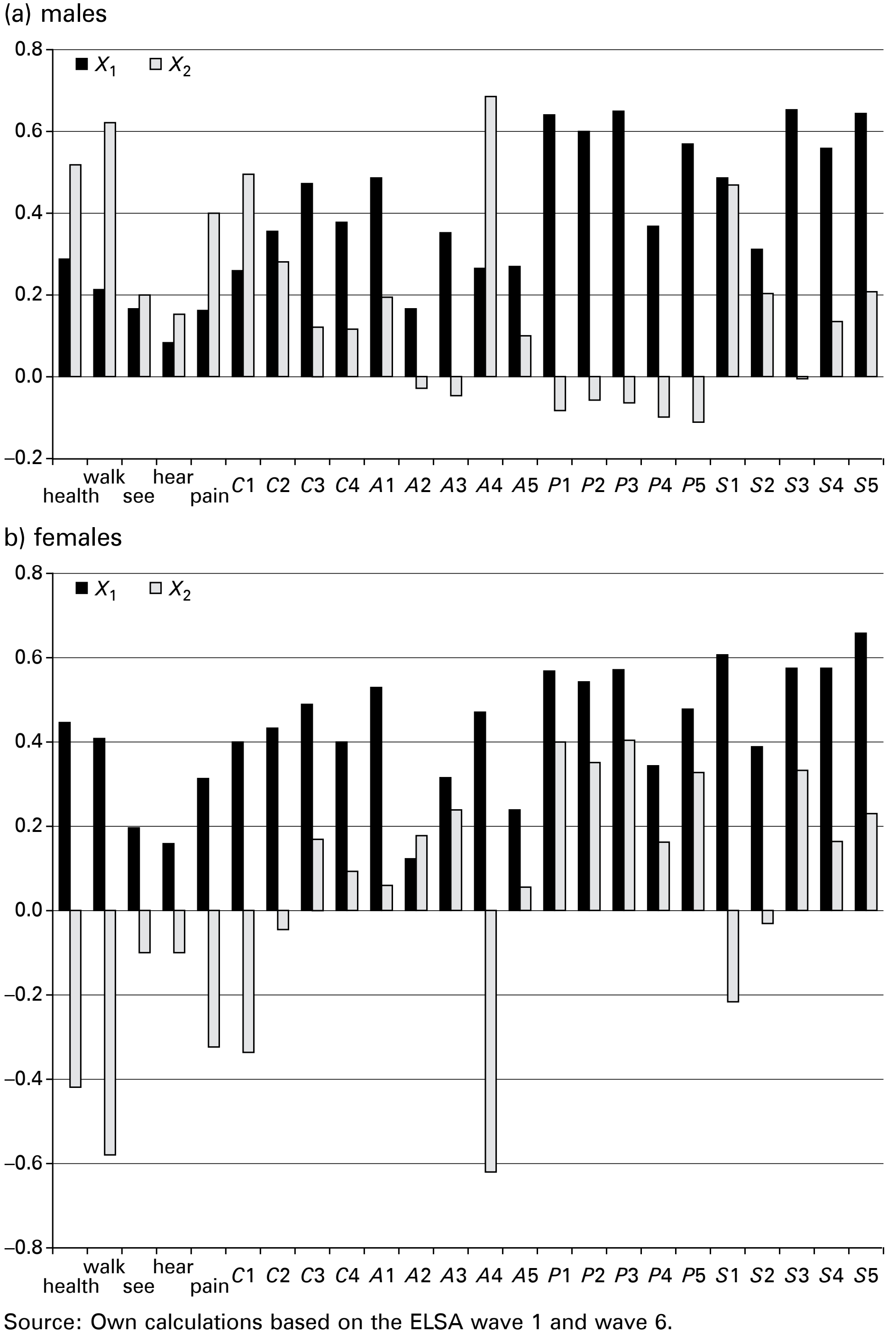


b) smoking



Source: Own calculations based on the ELSA wave 1 and wave 6.

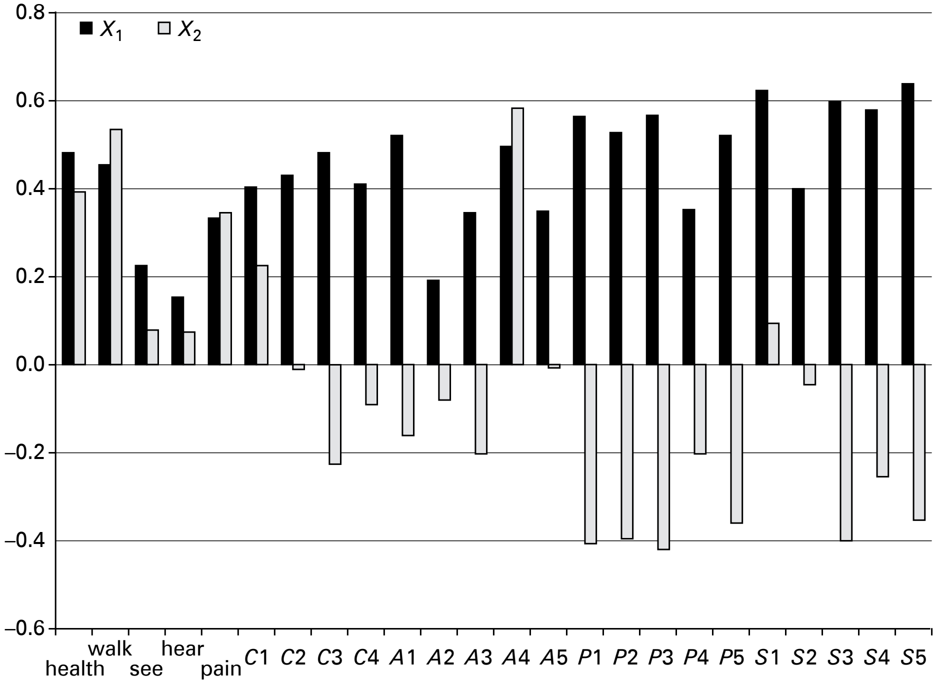
Figure 3. Parameters *males* and *females* in Kendall's scale for the 2-factor models



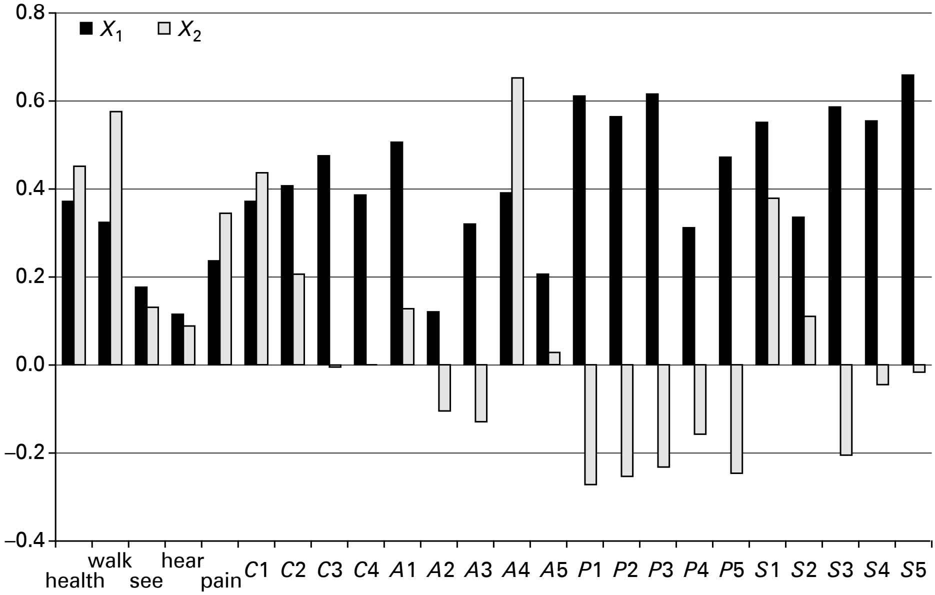
Source: Own calculations based on the ELSA wave 1 and wave 6.

Figure 4. Parameters *age 50–64* and *age 65+* in Kendall's scale for the 2-factor models

(a) age 50–64



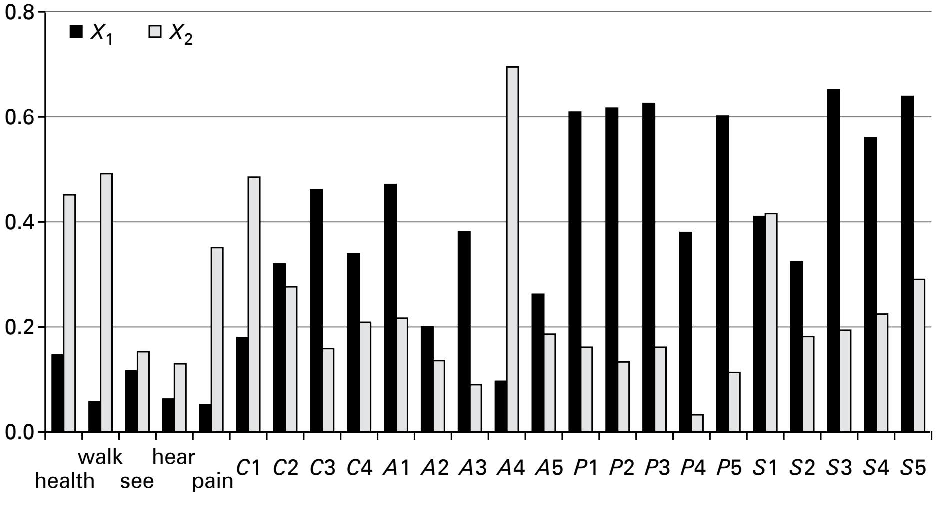
(b) age 65+



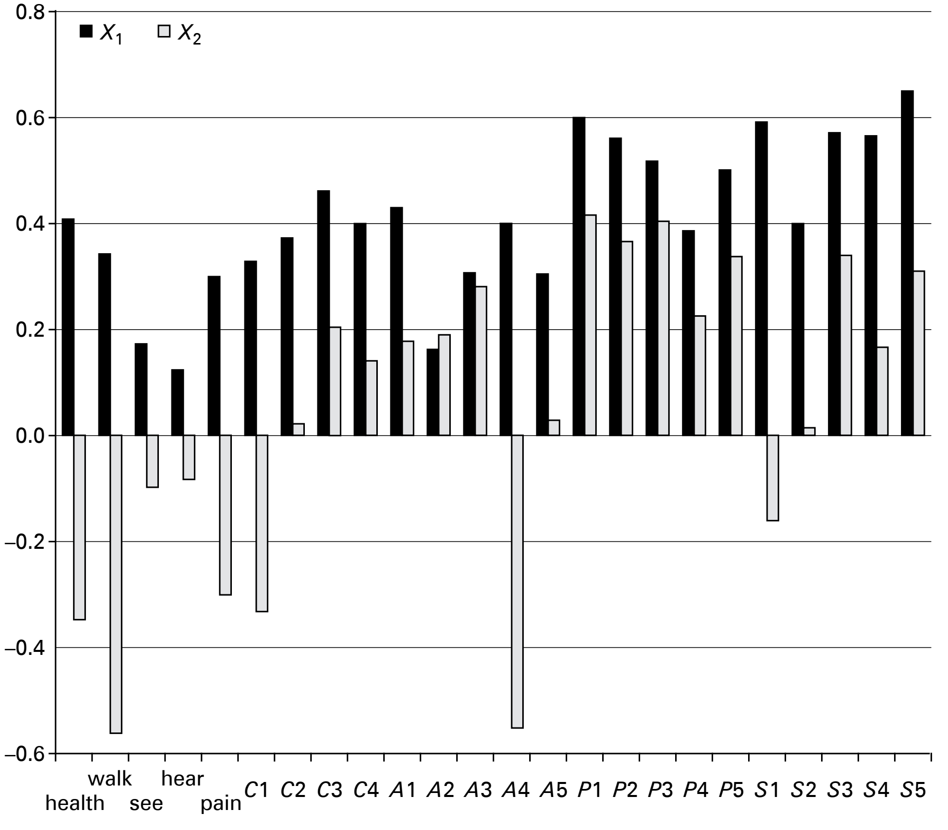
Source: Own calculations based on the ELSA wave 1 and wave 6.

Figure 5. Parameters *employed*, *unemployed*, *retired* and *disabled* in Kendall's scale for the 2-factor models

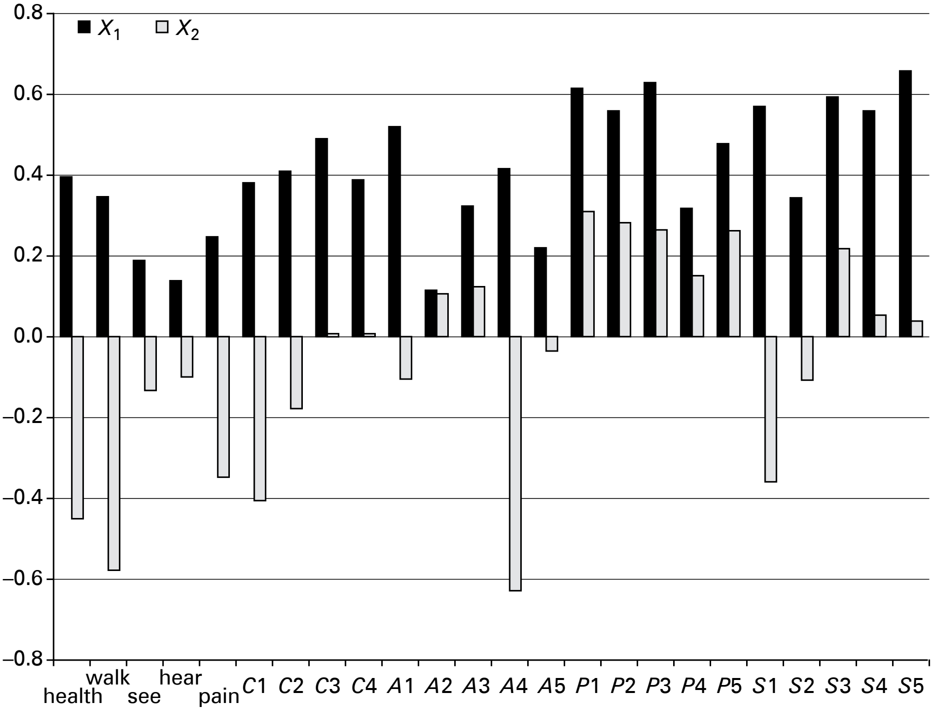
(a) employed



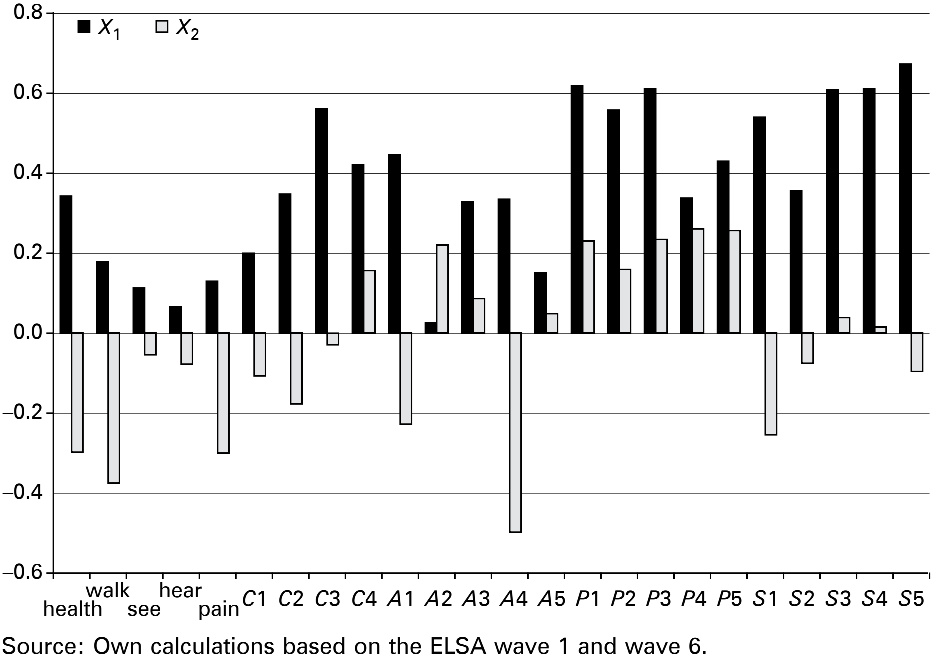
(b) unemployed



(c) retired



(d) disabled



Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 6. Means of variables according to the group in wave 1

	all	males	females	age 50–64	age 65+	non--smoking	smoking	retired	employed	unem-ployed	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
*C*1	2.8406	2.7833	2.8910	3.1046	2.5223	2.8378	2.8541	2.6458	3.2242	2.8084	2.2985
*C*2	3.1022	3.1040	3.1006	3.1098	3.0930	3.1198	3.0175	3.1181	3.2229	3.0358	2.3372
*C*3	3.2415	3.2870	3.2015	3.3096	3.1594	3.2773	3.0698	3.2194	3.3942	3.1474	2.6744
*C*4	3.1791	3.1960	3.1644	3.1703	3.1898	3.1952	3.1022	3.2194	3.2487	3.0821	2.5659
*A*1	3.4254	3.4218	3.4285	3.4740	3.3667	3.4561	3.2781	3.3985	3.5693	3.4274	2.7791
*A*2	3.0297	3.0023	3.0537	2.8281	3.2728	3.0307	3.0249	3.2181	2.8161	2.8526	2.9690
*A*3	3.4475	3.3753	3.5109	3.3973	3.5081	3.4602	3.3865	3.5298	3.4011	3.3705	3.1318
*A*4	2.8837	2.8675	2.8978	3.0732	2.6551	2.9033	2.7893	2.7389	3.3369	2.8632	1.4380
*A*5	2.6600	2.6237	2.6918	2.5413	2.8031	2.7113	2.4140	2.7935	2.5749	2.6716	1.9574
*P*1	3.7181	3.7272	3.7100	3.6963	3.7443	3.7510	3.5599	3.7540	3.7330	3.6632	3.4031
*P*2	3.5757	3.5538	3.5949	3.5842	3.5655	3.6089	3.4165	3.5848	3.6272	3.5200	3.2791
*P*3	3.7869	3.7907	3.7835	3.7707	3.8065	3.8085	3.6833	3.8162	3.8010	3.7305	3.5388
*P*4	3.6963	3.6334	3.7516	3.6908	3.7030	3.7095	3.6334	3.7050	3.7173	3.6674	3.5426
*P*5	3.6432	3.6311	3.6539	3.5905	3.7068	3.6759	3.4863	3.6844	3.6272	3.6400	3.3760
*S*1	2.9628	2.9443	2.9790	3.0578	2.8482	2.9987	2.7905	2.9081	3.1958	2.9621	2.0233
*S*2	2.6011	2.5690	2.6292	2.7408	2.4326	2.6214	2.5037	2.5281	2.8319	2.5011	2.0233
*S*3	3.4189	3.4117	3.4253	3.3891	3.4549	3.4519	3.2606	3.4564	3.4742	3.4316	2.7171
*S*4	3.1265	3.0975	3.1519	3.1900	3.0498	3.1640	2.9464	3.0957	3.2702	3.0884	2.5891
*S*5	3.2163	3.1868	3.2423	3.2903	3.1271	3.2544	3.0337	3.1782	3.3785	3.2400	2.5194

Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 7. Means of variables according to the group in wave 6

	all	males	females	age 50–64	age 65+	non--smoking	smoking	retired	employed	unem-ployed	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
*C*1	2.6820	2.6135	2.7388	2.9756	2.4198	2.6887	2.6291	2.5211	3.0042	2.7179	2.2578
*C*2	2.7909	2.7880	2.7933	2.8404	2.7467	2.8048	2.6816	2.7905	2.8935	2.6645	2.0732
*C*3	3.2310	3.2655	3.2023	3.2571	3.2076	3.2514	3.0704	3.2587	3.2875	3.0192	2.6237
*C*4	3.1020	3.1402	3.0703	3.0994	3.1043	3.1120	3.0235	3.1270	3.1499	2.9509	2.5157
*A*1	3.4306	3.4472	3.4168	3.4446	3.4181	3.4507	3.2726	3.4491	3.5019	3.3034	2.6969
*A*2	2.9263	2.9398	2.9152	2.7507	3.0832	2.9174	2.9966	3.0564	2.7529	2.6090	2.9512
*A*3	3.3864	3.3302	3.4330	3.3082	3.4561	3.3937	3.3285	3.4717	3.3044	3.1944	3.0871
*A*4	2.7706	2.7582	2.7809	2.9920	2.5728	2.7873	2.6391	2.6360	3.1660	2.7543	1.3484
*A*5	2.6672	2.6732	2.6623	2.5011	2.8156	2.7053	2.3687	2.8350	2.5104	2.3462	1.9477
*P*1	3.6206	3.6186	3.6223	3.5664	3.6690	3.6437	3.4391	3.6731	3.5992	3.5491	3.0976
*P*2	3.4983	3.4656	3.5254	3.5016	3.4953	3.5202	3.3263	3.5110	3.5392	3.4466	3.0105
*P*3	3.7391	3.7392	3.7390	3.6960	3.7776	3.7595	3.5788	3.7815	3.7244	3.6453	3.3519
*P*4	3.6740	3.6082	3.7286	3.6647	3.6824	3.6833	3.6011	3.6863	3.6891	3.6368	3.4042
*P*5	3.5530	3.5436	3.5608	3.5118	3.5898	3.5731	3.3955	3.5915	3.5615	3.4316	3.0627
*S*1	2.8582	2.8663	2.8516	2.9373	2.7876	2.8821	2.6715	2.8293	3.0281	2.8077	1.8606
*S*2	2.5151	2.5057	2.5229	2.6272	2.4150	2.5328	2.3765	2.4667	2.6849	2.4103	1.9164
*S*3	3.3632	3.3775	3.3514	3.3463	3.3784	3.3866	3.1799	3.4059	3.3939	3.2030	2.6690
*S*4	3.0491	3.0546	3.0446	3.1058	2.9986	3.0779	2.8235	3.0410	3.1714	2.8675	2.3659
*S*5	3.1308	3.1084	3.1493	3.1888	3.0789	3.1560	2.9330	3.1270	3.2440	3.0235	2.3380

Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 8. Parameters in Kendall's scale for factor X_1 in wave 1

	males	females	age 50–64	age 65+	non-smo-king	smoking	retired	employed	unem-ployed	disabled
	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$
health	0.4492	0.5478	0.4449	0.3654	0.3878	0.4492	0.5081	0.2078	0.4777	0.3212
walk	0.4495	0.6340	0.4301	0.3531	0.3744	0.4495	0.6146	0.1550	0.5492	0.1984
see	0.2582	0.2660	0.2609	0.2080	0.2357	0.2582	0.2443	0.1711	0.2761	0.1696
hear	0.1930	0.1906	0.1864	0.1255	0.1630	0.1930	0.1873	0.1573	0.2046	0.0112
pain	0.3280	0.4077	0.3214	0.2670	0.2866	0.3280	0.3944	0.1293	0.3361	0.1533
$C1$	0.4140	0.5103	0.3945	0.3971	0.3907	0.4140	0.5325	0.2453	0.4338	0.2173
$C2$	0.3888	0.2981	0.3906	0.3648	0.3708	0.3888	0.3439	0.2649	0.1474	0.3347
$C3$	0.3788	0.1906	0.4020	0.3888	0.3966	0.3788	0.2071	0.3915	0.0550	0.4795
$C4$	0.3547	0.2268	0.4054	0.3496	0.3773	0.3547	0.2612	0.3055	0.1029	0.3482
$A1$	0.4154	0.3325	0.4491	0.4404	0.4376	0.4154	0.3358	0.3893	0.1748	0.4572
$A2$	0.1057	–0.0296	0.1892	0.0820	0.1252	0.1057	0.0261	0.2002	–0.0838	0.0391
$A3$	0.2672	0.0564	0.3233	0.2784	0.3009	0.2672	0.1103	0.3527	–0.0627	0.2829
$A4$	0.5027	0.7034	0.4886	0.4387	0.4507	0.5027	0.6988	0.2307	0.5322	0.4353
$A5$	0.2511	0.1028	0.3194	0.1648	0.2307	0.2511	0.1339	0.2434	–0.0238	0.2090
$P1$	0.4883	0.1488	0.5473	0.5771	0.5476	0.4883	0.1606	0.5656	0.0644	0.5275
$P2$	0.4617	0.1288	0.5139	0.5166	0.5057	0.4617	0.1369	0.5697	0.0202	0.5384
$P3$	0.5290	0.1244	0.5562	0.5813	0.5508	0.5290	0.1781	0.5971	0.0596	0.5916
$P4$	0.2802	0.0705	0.3210	0.2704	0.2864	0.2802	0.0468	0.3554	0.0608	0.2743
$P5$	0.3881	0.0488	0.4697	0.4060	0.4302	0.3881	0.0498	0.5088	–0.0762	0.4172
$S1$	0.6117	0.4836	0.5993	0.5769	0.5738	0.6117	0.4986	0.4668	0.3422	0.5672
$S2$	0.4074	0.2724	0.3810	0.4056	0.3753	0.4074	0.2509	0.3177	0.1816	0.4301
$S3$	0.5188	0.1542	0.5794	0.5324	0.5464	0.5188	0.1582	0.5878	0.0098	0.5291
$S4$	0.5139	0.2239	0.5320	0.5338	0.5211	0.5139	0.2007	0.5345	0.0634	0.5536
$S5$	0.6240	0.2681	0.6393	0.6351	0.6351	0.6240	0.2787	0.6215	0.0808	0.6178

Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 9: Parameters in Kendall's scale for factor X_2 in wave 1

	males	females	age 50–64	age 65+	non-smo-king	smoking	retired	employed	unem-ployed	disabled
	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$
health	–0.3946	0.1895	0.4733	–0.3929	–0.4287	–0.3946	0.1811	–0.4368	0.3437	–0.1997
walk	–0.5601	0.1087	0.5756	–0.5231	–0.5535	–0.5601	0.1326	–0.4900	0.1798	–0.5393
see	–0.1143	0.1129	0.1117	–0.1504	–0.1565	–0.1143	0.1038	–0.1280	0.1833	–0.0092
hear	–0.0944	0.0496	0.0562	–0.1217	–0.1179	–0.0944	0.0513	–0.1031	0.0946	0.0468
pain	–0.2966	0.1043	0.3371	–0.3044	–0.3074	–0.2966	0.0892	–0.3138	0.1717	–0.1885
$C1$	–0.3795	0.2024	0.3067	–0.4340	–0.4174	–0.3795	0.1908	–0.4521	0.3302	–0.0415
$C2$	–0.1104	0.2757	0.0962	–0.2224	–0.1261	–0.1104	0.2387	–0.2213	0.4031	–0.0807
$C3$	0.0963	0.3767	–0.1354	0.0227	0.0543	0.0963	0.3222	0.0534	0.4300	0.0181
$C4$	0.0062	0.3422	–0.0188	–0.1036	–0.0164	0.0062	0.2750	–0.1446	0.4605	0.1705
$A1$	0.0010	0.3931	–0.0315	–0.1669	–0.0806	0.0010	0.3310	–0.0386	0.4719	–0.1740
$A2$	0.0606	0.1772	–0.0629	–0.0059	0.0945	0.0606	0.0780	–0.1198	0.3027	0.2175
$A3$	0.1352	0.3422	–0.1610	0.0469	0.1384	0.1352	0.2497	–0.0129	0.4355	0.0851
$A4$	–0.6094	0.2947	0.6321	–0.6437	–0.6504	–0.6094	0.2555	–0.7002	0.4108	–0.3415
$A5$	0.0128	0.2039	0.0237	–0.0254	0.0165	0.0128	0.1318	–0.1333	0.3333	0.1829
$P1$	0.3673	0.6234	–0.3118	0.2723	0.3008	0.3673	0.6259	0.0705	0.6358	0.2256
$P2$	0.3452	0.5899	–0.3271	0.2645	0.2892	0.3452	0.5662	0.1390	0.5352	0.2264
$P3$	0.3120	0.6243	–0.3347	0.2020	0.3030	0.3120	0.5798	0.0574	0.6681	0.3590
$P4$	0.2146	0.3119	–0.1740	0.1470	0.1590	0.2146	0.3097	0.0867	0.3202	0.2013
$P5$	0.3551	0.5336	–0.3229	0.2765	0.3102	0.3551	0.5094	0.0868	0.5566	0.4168
$S1$	–0.1971	0.4825	0.1805	–0.3490	–0.3084	–0.1971	0.4840	–0.2546	0.5293	–0.1519
$S2$	0.0185	0.2864	–0.0232	–0.0541	–0.0760	0.0185	0.3300	–0.0523	0.2847	–0.0174
$S3$	0.3009	0.5974	–0.2851	0.1853	0.2508	0.3009	0.5604	0.0697	0.6379	0.1460
$S4$	0.2264	0.5046	–0.1939	0.1330	0.1241	0.2264	0.5166	–0.0219	0.5039	0.0922
$S5$	0.2140	0.6121	–0.2130	0.0964	0.1172	0.2140	0.6129	–0.0373	0.6357	0.0557

Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 10. Parameters in Kendall's scale for factor X_1 in wave 6

	males	females	age 50–64	age 65+	non-smo-king	smoking	retired	employed	unem-ployed	disabled
	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$
health	0.2886	0.4461	0.4832	0.3704	0.3595	0.2886	0.3986	0.1474	0.4091	0.3427
walk	0.2144	0.4096	0.4542	0.3240	0.3063	0.2144	0.3485	0.0582	0.3418	0.1801
see	0.1665	0.1982	0.2264	0.1767	0.1792	0.1665	0.1906	0.1171	0.1724	0.1129
hear	0.0818	0.1597	0.1538	0.1158	0.1252	0.0818	0.1387	0.0634	0.1237	0.0667
pain	0.1599	0.3125	0.3330	0.2356	0.2304	0.1599	0.2482	0.0527	0.2987	0.1313
*C*1	0.2604	0.4008	0.4054	0.3706	0.3374	0.2604	0.3815	0.1802	0.3295	0.2011
*C*2	0.3555	0.4335	0.4309	0.4067	0.4046	0.3555	0.4117	0.3208	0.3729	0.3497
*C*3	0.4716	0.4890	0.4828	0.4763	0.4942	0.4716	0.4903	0.4626	0.4616	0.5608
*C*4	0.3776	0.4032	0.4115	0.3856	0.3988	0.3776	0.3895	0.3400	0.3994	0.4201
*A*1	0.4857	0.5293	0.5197	0.5063	0.5171	0.4857	0.5207	0.4716	0.4300	0.4488
*A*2	0.1639	0.1242	0.1926	0.1201	0.1593	0.1639	0.1164	0.2017	0.1636	0.0276
*A*3	0.3529	0.3159	0.3458	0.3219	0.3453	0.3529	0.3249	0.3821	0.3066	0.3288
*A*4	0.2642	0.4725	0.4968	0.3918	0.3659	0.2642	0.4186	0.0982	0.4011	0.3355
*A*5	0.2691	0.2399	0.3485	0.2063	0.2525	0.2691	0.2213	0.2635	0.3049	0.1521
*P*1	0.6398	0.5703	0.5657	0.6120	0.6134	0.6398	0.6154	0.6095	0.6025	0.6206
*P*2	0.6017	0.5428	0.5281	0.5652	0.5773	0.6017	0.5597	0.6166	0.5601	0.5594
*P*3	0.6492	0.5724	0.5661	0.6167	0.6176	0.6492	0.6305	0.6260	0.5172	0.6117
*P*4	0.3677	0.3450	0.3531	0.3125	0.3442	0.3677	0.3184	0.3808	0.3858	0.3399
*P*5	0.5676	0.4785	0.5219	0.4730	0.5260	0.5676	0.4802	0.6025	0.5034	0.4304
*S*1	0.4842	0.6072	0.6232	0.5525	0.5530	0.4842	0.5721	0.4113	0.5926	0.5420
*S*2	0.3100	0.3882	0.4001	0.3358	0.3482	0.3100	0.3437	0.3241	0.3710	0.3570
*S*3	0.6522	0.5748	0.5974	0.5857	0.6199	0.6522	0.5935	0.6517	0.5733	0.6101
*S*4	0.5575	0.5755	0.5793	0.5548	0.5727	0.5575	0.5611	0.5613	0.5678	0.6150
*S*5	0.6437	0.6579	0.6398	0.6581	0.6653	0.6437	0.6597	0.6405	0.6512	0.6742

Source: Own calculations based on the ELSA wave 1 and wave 6.

Table 11: Parameters in Kendall's scale for factor X_2 in wave 6

	males	females	age 50–64	age 65+	non-smoking	smoking	retired	employed	unemployed	disabled
	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$	$t(4)+t(5)$	$t(5)+t(4)$
health	0.5188	–0.4183	0.3921	0.4522	0.4817	0.5188	–0.4501	0.4514	–0.3490	–0.2982
walk	0.6219	–0.5798	0.5353	0.5760	0.6069	0.6219	–0.5766	0.4931	–0.5625	–0.3730
see	0.1992	–0.1001	0.0793	0.1309	0.1559	0.1992	–0.1334	0.1534	–0.0987	–0.0546
hear	0.1540	–0.1011	0.0747	0.0877	0.1356	0.1540	–0.1013	0.1302	–0.0835	–0.0764
pain	0.4000	–0.3234	0.3461	0.3454	0.3677	0.4000	–0.3477	0.3517	–0.3014	–0.2990
*C*1	0.4953	–0.3372	0.2248	0.4380	0.4309	0.4953	–0.4064	0.4860	–0.3333	–0.1057
*C*2	0.2814	–0.0454	–0.0112	0.2042	0.1678	0.2814	–0.1791	0.2779	0.0224	–0.1773
*C*3	0.1208	0.1694	–0.2256	–0.0042	–0.0211	0.1208	0.0099	0.1591	0.2038	–0.0277
*C*4	0.1162	0.0946	–0.0912	0.0015	0.0078	0.1162	0.0086	0.2088	0.1414	0.1575
*A*1	0.1940	0.0596	–0.1609	0.1281	0.0788	0.1940	–0.1042	0.2177	0.1769	–0.2278
*A*2	–0.0297	0.1789	–0.0791	–0.1048	–0.1155	–0.0297	0.1058	0.1363	0.1901	0.2218
*A*3	–0.0449	0.2390	–0.2020	–0.1298	–0.1511	–0.0449	0.1221	0.0905	0.2805	0.0862
*A*4	0.6842	–0.6184	0.5844	0.6522	0.6544	0.6842	–0.6291	0.6952	–0.5533	–0.4956
*A*5	0.1021	0.0559	–0.0072	0.0300	0.0119	0.1021	–0.0357	0.1876	0.0274	0.0489
*P*1	–0.0834	0.4012	–0.4063	–0.2713	–0.2646	–0.0834	0.3089	0.1621	0.4166	0.2305
*P*2	–0.0583	0.3517	–0.3957	–0.2533	–0.2158	–0.0583	0.2818	0.1331	0.3656	0.1614
*P*3	–0.0633	0.4031	–0.4216	–0.2317	–0.2488	–0.0633	0.2649	0.1615	0.4040	0.2356
*P*4	–0.0991	0.1627	–0.2032	–0.1572	–0.1239	–0.0991	0.1504	0.0336	0.2255	0.2623
*P*5	–0.1120	0.3279	–0.3604	–0.2453	–0.2350	–0.1120	0.2623	0.1138	0.3378	0.2591
*S*1	0.4704	–0.2168	0.0932	0.3788	0.3632	0.4704	–0.3584	0.4163	–0.1630	–0.2548
*S*2	0.2028	–0.0288	–0.0457	0.1111	0.1213	0.2028	–0.1062	0.1825	0.0134	–0.0754
*S*3	–0.0032	0.3332	–0.3993	–0.2052	–0.1970	–0.0032	0.2187	0.1940	0.3402	0.0404
*S*4	0.1359	0.1662	–0.2542	–0.0437	–0.0145	0.1359	0.0526	0.2247	0.1658	0.0157
*S*5	0.2061	0.2299	–0.3548	–0.0153	–0.0145	0.2061	0.0410	0.2913	0.3112	–0.0944

Source: Own calculations based on the ELSA wave 1 and wave 6.

DWUWSKAŹNIKOWY MODEL KOPUŁ CZYNNIKOWYCH W OCENIE STANU ZDROWIA

STRESZCZENIE

Analiza łącznych rozkładów wskaźników zdrowia jest konieczna do zrozumienia stanu zdrowia starzejących się społeczeństw i efektywnej polityki zdrowotnej (Conti et al., 2010; Heckman et al., 2011). W poprzednim artykule (Kobus and Półchłopek, 2016) zaproponowałyśmy metodę modelowania porządkowych zmiennych zdrowotnych, tj. tak zwane kopuły czynnikowe (Nikoloulopoulos and Joe, 2015). W niniejszym artykule rozwijamy nasze badania w tym zakresie, wykorzystując modele dwuczynnikowe do estymacji 24-wymiarowych rozkładów i pokazując, że pozwalają one na bardzo precyzyjny opis rozkładów zdrowia. Wykazujemy również, że kopuły czynnikowe oparte na kombinacjach rozkładów i spisują się lepiej niż standardowe podejście oparte na wielowymiarowych rozkładach gaussowskich. Dodatkowo identyfikujemy dwa główne czynniki, które wpływają na rozkład 24 zmiennych, jak również interpretujemy je na podstawie uzyskanych parametrów. Pierwszy czynnik związany jest z usposobieniem i podejściem do życia, a drugi ze zdrowiem fizycznym.

Słowa kluczowe: dane porządkowe, zdrowie, nierówności, kopuły czynnikowe.

Klasyfikacja JEL: I31; D63