MODELLING OF THE DISTRIBUTION OF THE MATERIAL DEPRIVATION INDEX BASED ON SHARE DATA

Ivana Malá

Abstract

A material deprivation index from the database of the Survey of Health, Ageing and Retirement in Europe (SHARE) exploring the European population aged 50 and over in 2013 is analysed. The probability distribution of the index is modelled for 15 participating countries. There are too many zero values (these values of the index mean the absence of material-related problems, frequencies are from 16% in Estonia to 74% in the Netherlands and Sweden and 77% in Denmark) in the data to model the distribution as a continuous one. A mixture model is applied, a logistic regression model is used to model the impact of explanatory variables (such as gender, age or country) on the probability of no deprivation. The continuous part of the mixture is modelled with the use of a finite mixture of normal components. The distributions of the index in analysed countries are similar in the shape but different in the scale.

Key words: SHARE, deprivation, logistic regression, finite mixture

JEL Code: I31, C25, C46

Introduction

Reducing deprivation and social exclusion of inhabitants is the target of the strategy for 2020. For this reason, it is essential to construct tools that enable to describe and quantify so subjective phenomenons such as a quality of life or deprivation and exclusion. For material deprivation, there are more composite indicators that can reflect the material situation in households reliably; these indicators try to include more characteristics than just the household income (Bellani, D´Ambrosio, 2011; Bellani, 2013; Hyde, 2003). However, no construct is generally accepted as the best and universally applicable. It is important to remember that such an effort is always just an attempt to quantify a very general situation involving a strong subjective element involving personal perception and feelings.

The population of European inhabitants above 50 is the target population of the extensive survey Survey of Health, Ageing and Retirement in Europe (SHARE Release Guide 7.0.0., 2019; Börsch-Supan, 2016), the analysed index of material deprivation *depmat* is

adapted to the situation of elderly respondents (Adena et al., 2015). The included questions in the questionnaire indicate material inconvenience or problems connected with the availability of services, medical care and social relations. In the 5th wave, 14 European countries and Israel took part, especially Austria, Belgium, the Czech Republic, Denmark, Estonia, France, Germany, Italy, Luxembourg, the Netherlands, Slovenia, Spain, Sweden and Switzerland. In the construction of the index, hedonic weights are used to build the composite index of deprivation on the 0-1 scale (Saisana, Saltelli, 2005). Higher values of the index mean a higher level of deprivation; value zero indicates no problems included in the questioned items.

This contribution aims to model the distribution of the material deprivation index with the use of a mixture of a discrete part (for values 0) and a continuous part. The differences between participating countries in the survey are investigated.

1 Model and Data

1.1 The composite index of material deprivation

In the 5th wave of the survey, two composite indicators are included for social and material deprivation. The index of material deprivation, analysed in this contribution, is an aggregate measure of material conditions of Europeans aged over 50 years, comprising a set of 11 criterion items that refer to two broad domains: the inability to meet basic needs and financial difficulties (Myck, 2015; Adena, 2015; Malá, 2018 for the Czech Republic). The value of this index is equal for all members of the household, as its value depends only on the situation of a household.

Alternative answers yes (in case of problems) or no (if there are no problems) are weighted to the composite indices using hedonic weights. The indicator is transformed into a <0,1> scale from no deprivation (*depmat*=0) to the highest degree of deprivation (*depmat*=1).

1.2 Finite mixture model

We will suppose that the distribution of the material deprivation index Y (Dalrymple et al., 2003) is a mixture of two distributions; a discrete part (for Y=0) and a continuous part (for positive values Y>0). The continuous part is modelled as a mixture of K component densities.

Let $\pi_0(\mathbf{x}) = P(Y=0|\mathbf{x})$ denote the probability of no deprivation, given the vector of $m \geq 1$ explanatory variables $\mathbf{x}=(x_1, x_2, ..., x_m)'$. The logistic regression model is applied in the form

$$\text{logit}\left(P(Y=0|\mathbf{x})\right) = \text{logit}(\pi_0(\mathbf{x})) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m, \qquad (1)$$

where $\boldsymbol{\beta}=(\beta_0, \beta_1, ..., \beta_m)'$ is an (m+1)-dimensional vector of unknown regression parameters.

We include explanatory variables for age (5 years groups, 8 categories, 50-54, 55-59, ..., 85+), gender (male, female), size of household (factor with three categories 2, 3, 4, 4+ members) and a country (15 categories) in the logistic regression model. All variables are coded with dummy variables; the basic category respondent is a woman from Austria, aged 50-54 and living alone. We have a logistic regression model with m=25 regression parameters.

Using this notation, the normal mixture model with K components is given by

$$f(y|\mathbf{x}) = \pi_0(\mathbf{x})\mathrm{I}_{\{y=0\}} + (1-\pi_0(\mathbf{x}))\sum_{j=1}^{K}\pi_j f_j(y), \qquad (2)$$

where $\mathrm{I}_{\{\}}$ is an indicator function, $0 \le \pi_j \le 1,\ j = 1, 2, ..., K$, $\sum_{j=1}^{K}\pi_j = 1$ and $f_j,\ j = 1, 2, ..., K$ are normal or lognormal component densities specified by the component parameters (μ_j, σ_j^2). In the model, there are $m + 3K$ unknown parameters to be estimated.

For the continuous part of the model, the component membership is not observable, and we use artificial components based on observed data estimated for the whole sample. Using the AIC criterion, two to four normal and lognormal components were treated K=2-4, and a mixture of 3 lognormal components was selected as a compromise between model quality and component identification. The problem with the normal mixture was a too small standard error in the component of small values.

All computations were performed in R (R Core Team, 2017). The normal component parameters μ and σ and component weights π in (2) were estimated with *mixtools* package (Benaglia et al., 2009) and lognormal mixture in the package *mixdist* (Macdonald, 2012). The maximum likelihood estimates were found, and the bootstrap with 1,000 replications was used to estimate standard errors of estimates. GLM model with the binomial link function was used to estimate logistic regression (1).

2 Data and Results

In the data, 58,030 respondents with the mean age 66.9 (standard deviation 9.8) were included. There is 55.1 % women with the mean age 66.8 (10.0), and 44.9 % of men with the mean age 67.1 (9.5). The mean value of the index is equal to 0.135 (0.189), for positive values 0.277 (0.183). For particular countries, these values are given in Table 1 for all observations (left) and positive values (right). Moreover, in the first column, the percentage of respondents without any problems (*depmat*=0) are given. The countries are sorted according to the mean of the

material deprivation index, from Denmark with the lowest level of deprivation to Estonia with the highest level.

In Table 1, the first column is given a percentage of respondents with zero value of the material deprivation index the complement percentage refers to the proportion of positive values (right part of the table).

Tab. 1: Characteristics for *Y* and *Y*>0 (*sd*=standard deviation)

	Material deprivation index *Y*				*Y* > 0	
country	*Y=0*	*n*	*mean*	*sd*	*mean*	*sd*
Denmark	76.94%	3,682	0.043	0.104	0.188	0.141
Netherlands	74.03%	3,516	0.053	0.122	0.206	0.160
Sweden	73.87%	4,042	0.049	0.110	0.187	0.142
Luxembourg	67.03%	1,468	0.068	0.136	0.206	0.165
Switzerland	65.83%	2,836	0.061	0.118	0.179	0.140
Austria	63.51%	4,127	0.084	0.146	0.231	0.158
Belgium	60.40%	5,093	0.090	0.153	0.227	0.168
Germany	56.99%	4,989	0.107	0.169	0.248	0.177
France	48.69%	4,227	0.131	0.178	0.255	0.173
Czech Republic	40.40%	4,352	0.159	0.186	0.266	0.171
Spain	39.86%	5,727	0.177	0.199	0.294	0.177
Israel	39.04%	1,665	0.191	0.210	0.313	0.185
Italy	35.45%	4,206	0.220	0.227	0.341	0.196
Slovenia	32.44%	2,731	0.188	0.190	0.278	0.168
Estonia	15.28%	5,369	0.309	0.219	0.365	0.190

Source: own computations

In Table 2, the results of the logistic model pro $\pi_0(\mathbf{x}) = P(Y=0|\mathbf{x})$ are given. All explanatory variables are treated as factors; for this reason, we may easily interpret odds ratios given in the last column. The odds ratios higher than 1 (significant for 5%) are stressed in red, odds ratios significantly lower than 1 in blue. For men, the chance is 1.134 times higher than for women, and the positive impact of living in a pair is well visible (odds ratio 1.678). Also, households with 3 and 4 members have a statistically significant odds ratio (1.335 and 1.420) concerning the respondents living alone. Odds ratios add information on differences between countries to the means in Table 1 and agree with these values.

Information about the continuous part of the distribution is included in the empirical distributions given in Figure 1; there is no correction for a percentage of zero values, given in the first column in Table 1. The shape of all distributions is similar; there is one component for small values (0 - 0.15) meaning a very low level of deprivation and another part for higher values possibly with one side mode. The distributions are positively skewed. In Switzerland,

Tab. 2: Logistic regression model for $Y = 0$

variable	Estimate	Std. Error	z value	p-value	Logit
Intercept	0.150	0.047	3.216	0.0013	
genderMale	0.126	0.018	6.883	0.0000	1.134
age5 55-59	0.007	0.036	0.200	0.8418	1.007
age5 60-64	0.083	0.036	2.320	0.0203	1.087
age5 65-69	0.111	0.037	3.030	0.0024	1.118
age5 70-74	0.056	0.038	1.454	0.1458	1.057
age5 75-79	-0.040	0.041	-0.978	0.3282	0.961
age5 80-84	-0.094	0.045	-2.095	0.0362	0.910
age5 85+	-0.064	0.051	-1.264	0.2061	0.938
houhold size 2	0.518	0.024	21.979	0.0000	1.678
houhold size 3	0.289	0.034	8.619	0.0000	1.335
houhold size 4	0.351	0.045	7.828	0.0000	1.420
houhold size 4+	-0.052	0.062	-0.839	0.4014	0.950
country Belgium	-0.153	0.044	-3.516	0.0004	0.858
country Czech Republic	-0.987	0.045	-21.838	0.0000	0.373
country Denmark	0.607	0.051	11.854	0.0000	1.836
country Estonia	-2.298	0.050	-45.817	0.0000	0.100
country France	-0.628	0.045	-13.924	0.0000	0.534
country Germany	-0.332	0.044	-7.596	0.0000	0.718
country Israel	-1.046	0.060	-17.309	0.0000	0.351
country Italy	-1.189	0.046	-25.661	0.0000	0.305
country Luxembourg	0.121	0.065	1.860	0.0628	1.128
country Netherlands	0.444	0.051	8.754	0.0000	1.558
country Slovenia	-1.319	0.053	-25.041	0.0000	0.268
country Spain	-1.009	0.043	-23.545	0.0000	0.365
country Sweden	0.429	0.049	8.806	0.0000	1.536
country Switzerland	0.056	0.052	1.093	0.2744	1.058

Source: own computations

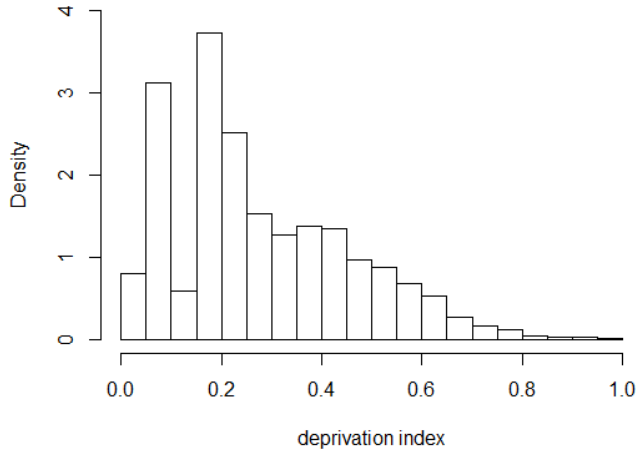
where a high percentage of respondents are not deprived at all, but if they are, usually the level of deprivation is higher than 0.175. For this reason, the component weight of the first component for this country should be relatively low with respect to all other countries. For Estonia, a country with the worst situation according to all analysed characteristics, the distribution is more uniform. In this contribution, the distribution is fitted only to the whole sample.

In Figure 2 the empirical distribution of the whole sample is shown. All previous interpretations concerning components are valid. In this text, the finite mixture

$$\sum_{j=1}^{K} \pi_j f_j(y; \mu, \sigma^2)$$

from (2) is fitted to all positive observations of the index. The mixtures of 2-4 (K = 2-4) components of normal and lognormal densities were fitted to data ($f_j(.)$, $j = 1, 2, ..., K$ are normal or lognormal densities parametrised with two parameters $\mu \in R$ and $\sigma^2 > 0$).

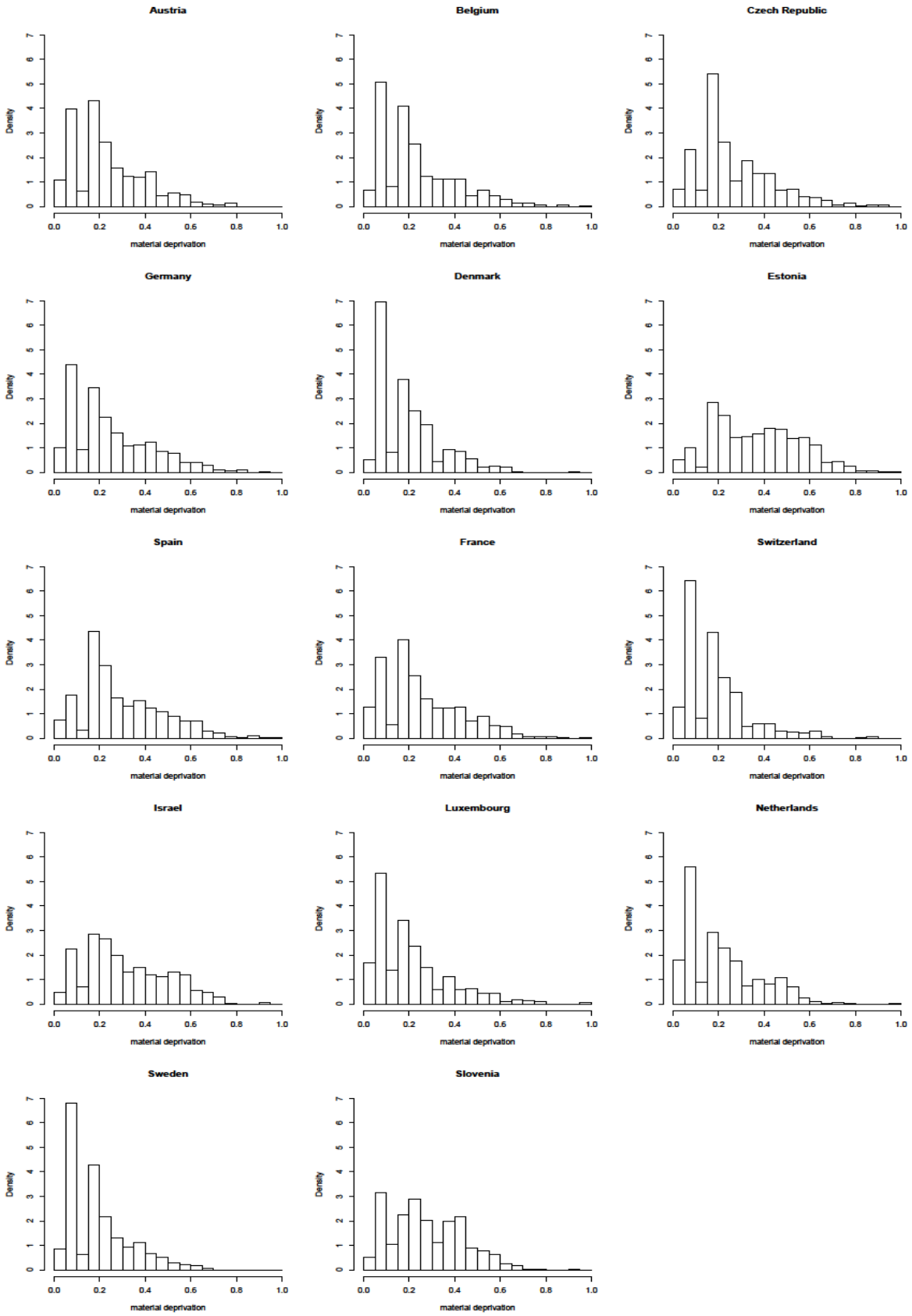
Fig. 2: Histogram of the material deprivation index for the whole sample.



Source: own computations

In Table 3, the estimated values of parameters are given for the whole data without taking into account the country of living. Using the AIC criterion, the mixture of three lognormal components was selected (Figure 3 left). The first and second component have similar variance parameter; the last component has approximately 10 times the higher standard deviation. We can interpret these components as more homogenous subgroups of respondents with low, medium and high (severe) deprivation. The weight for the third component is almost 0.5 from Table 3. In the model, all component weights must be multiplied by the estimated probability $\hat{\pi}_0$ from the logistic model; the fitted distribution in Table 3 is presented for positive values.

Fig. 1: Distributions of the index for participating countries



Source: own computations

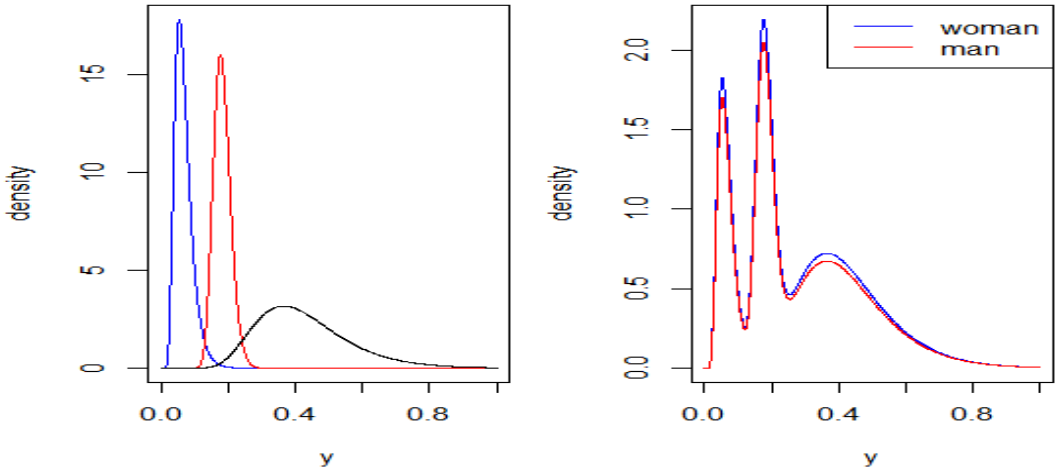
In Figure 3 (right), the estimated density is plotted for a man and a woman from Austria (baseline distribution). Estimated probabilities of *depmat*=0 are 0.569 for men and 0.538 for women.

Tab. 3: Estimated parameters of the mixture – 3 lognormal components

component	π	μ	σ	*Expected value*	*Standard deviation*
1	0.221 (0.003)	-2.780	0.390	0.067 (0.00044)	0.027 (0.00050)
2	0.286 (0.004)	-1.716	0.140	0.182 (0.00041)	0.026 (0.00037)
3	0.493 (0.004)	-0.902	0.327	0.428 (0.00153)	0.144 (0.00134)

Source: own computations

Fig. 3: Distributions of the index for participating countries. Two-member households, Austria



Source: own computations

Conclusion

In the text, the distribution of the material index from the SHARE data is modelled. According to the histograms in Figure 1 for the distribution of the index for participating countries in Table 1, the empirical distributions are of a similar shape except for Switzerland.

In the European Union, the material security of citizens is an important task of the state administration. The relatively low values of the material deprivation index testify to the success of the welfare state and its concern for ageing citizens. The success of this care can be expressed as a percentage of people with the zero value of the index; this value represents the respondents

reporting no problem in any of the areas included in the index. However, in the analysed data, among positive index values, there are still approximately half of the respondents in the severe deprivation group (component 3 in Table 3).

From Figure 1, it is obvious, that the distribution of the analysed index is country specific and we can obtain more detailed information about participating countries fitting distributions for particular countries separately. The model presented in Table 3 and Figure 3 refers to the whole population of participating countries (13 countries of the European Union, Switzerland and Israel). Estimation of distributions with country-specific components will be subject to further investigation.

Acknowledgment

This paper is supported by the grant F4/80/2018 which has been provided by Internal grant agency of University of Economics, Prague.

References

Adena, M., Myck, M., Oczkowska, M. (2015). Innovation for better understanding deprivation index. In: Ageing in Europe - Supporting Policies for an Inclusive Society, De Gruyter.

Bellani, L., D´Ambrosio, C. (2011). Deprivation, social exclusion and subjective well-being. *Social Indicators Research*, 67-86.

Bellani, L.(2013). Multidimensional indices of deprivation: the introduction of reference groups weights, *Econ Inequal*, 495–515.

Benaglia, T., Chauveau, D., Hunter, D., Young, D. (2009) Mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32, http://www.jstatsoft.org/v32/i06/.

Börsch-Supan, A. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 5.0.0. SHARE-ERIC. Data set. 2016.

Dalrymple, M. L., Hudson, I.L., Ford, R.P.K. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Journal Computational Statistics & Data Analysis*, 41 (3-4), 491-504.

Hyde, M. (2003). A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). *Ageing and mental health*, 7 (3), 186-194.

Malá, I. (2017). The Use of L-Moments in Modelling of Deprivation of the Population Aged over 50 in the Czech Republic. In: *Applications of Mathematics and Statistics in Economics (AMSE 2017)* [online]. Szklarska Poreba, 30.08.2017 – 03.09.2017. Wroclaw : Publishing House of Wrocław University of Economics, 285–296. http://amse.ue.wroc.pl/papers/2017/Mala.pdf.

mixdist: Finite Mixture Distribution Models. [online]. (2012) Macdonald, P. http://ms.mcmaster.ca/peter/mix/mix.html

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. Retrieved from https://www.R-project.org/.

Saisana, M., Saltelli, A., S. (2005). Tarantola. Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 168, 307-323.

SHARE Release Guide 7.0.0. [online]. [25.4.2019] http://www.share-project.org/data-documentation/waves-overview/wave-7. html.

Contact

Ivana Malá
University of Economics, Prague
W. Churchill Sq. 1938/4
130 67 Prague 3 – Žižkov
Czech Republic
malai@vse.cz