

THE USE OF L-MOMENTS IN THE MODELLING OF DEPRIVATION FOR THE POPULATION AGED OVER 50 IN THE CZECH REPUBLIC

IVANA MALÁ

University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability, W. Churchill Sq. 1938/4, Prague, Czech Republic
email: malai@vse.cz

Abstract

In the Survey of Health, Ageing and Retirement in Europe (SHARE) the European population aged over 50 is of interest; SHARE provides a multidisciplinary approach that delivers the comprehensive picture of the ageing process in Europe. In the fifth wave of the survey (carried out in 2013) two composite indicators of deprivation are evaluated. The answers to the selected questions (11 items for the material deprivation index and 15 items for the social deprivation index) from the survey are weighted into two composite indices with values transformed to the interval from 0 to 1. In the contribution, the empirical distributions of these indices in the Czech Republic are analysed and dependence on gender and size of household is treated. A mixture of two normal distributions for social inclusion index and a mixture of one discrete and two normal distributions for the index of material deprivation are fitted into the data. The models identify two homogenous subpopulations for social deprivation index and for positive values of material deprivation. 40 percent of respondents without any material deprivation were found in our data in comparison with less than 1 percent for the social deprivation index. Maximum likelihood estimates are compared to estimates obtained by the moment method based on robust moments.

Key words: deprivation, SHARE data, finite mixture of distributions, L-moments

JEL Codes: C14, I31, J17

DOI: 10.15611/amse.2017.20.23

1. Introduction

The European Union 2020 targets [Europe 2020 Strategy, 2017] set the ambitious goals of reducing the risk of poverty and social exclusion in the EU. In order to treat this problem, it is useful to quantify such subjective feelings as the quality of life, life satisfaction or material or social deprivation [Bellani, D'Ambrosio, 2011]. There are more composite indicators that are able to reliably reflect the situation in the Czech households, these indicators try to include more characteristics than just household income. In this text, the population of inhabitants above 50 is of interest. For the targeted population, the social exclusion and isolation is an important problem together with material deprivation. For this reason, it is worth analysing not only material deprivation but also social deprivation, as social exclusion seems to be a serious problem especially for elderly people. Indices analysed in this text are adapted to the surveyed population and reflects the life of people above 50 (to the age of 104 in the data).

The index of the material deprivation is regularly published by the Czech Statistical Office for the Czech Republic [CZSO, 2013] or by Eurostat for the whole European Union [EUROSTAT, 2017]. In this approach, households with 3 or 4 items from selected 9 material indicators are detected as deprived. There were 15.1 % of deprived in the age group 50 - 64 and 16.6 % in the group above 65 for 3 items and 6.6 % and 5.3 % for 4 items (in 2013 in the

Czech Republic). In terms of composite indicators [Saisana, et al. 2005], this approach uses constant weights (equal to $1/9$) and the indicator is equal to the relative frequency of positive items with the threshold of deprivation set in $3/9 = 0.33$ or $4/9 = 0.44$.

In this text, two composite deprivation indicators (index of the material deprivation (referred as *depmat*) and index of the social deprivation (referred as *depsoc*)) given in the fifth wave of the Survey of Health, Ageing and Retirement in Europe ([SHARE, 2017], [Börsch-Supan, 2016]) are analysed. The survey focuses on the European population aged above 50 (and their younger partners). Our results refer to an ageing population (after 50) in the Czech Republic. Indices of both material and social deprivation are treated and people deprived in both (material and social) domains are considered to be seriously deprived [SHARE Release Guide, 2017]. These two indices represent a possible quantification of deprivation among people above 50. In ([Adena, et al. 2015], [Myck, et al. 2015]) the analysis of the defined indices and relationship with (for example) subjective life satisfaction is given from the point of view of sociology and theory of composite indicators and surveys; the statistical analysis of data is given for the whole dataset and all participating countries.

The population of Czech inhabitants over 50 is not homogeneous with respect to analysed indices, we try to identify more homogeneous subpopulations. For this reason, we try to construct a model for a probability distribution of these indices, describe the relationship between both indices and compare the situation in subgroups given by gender and size of household. For the fit of distributions we use moment matching method applied to robust L-moments and dependence on observable explanatory variables is shown in figures. Our approach is only descriptive without any statistical inference. If both indices well quantify deprivation (and in this respect, we rely on authors' abilities and the analysis of included items given in ([Adena, et al. 2015], [Myck, et al. 2015])), the mixture models provide information about components as well about the overall distribution. For the comparison, we fit distribution also by maximum likelihood method. All computations are done in the program R [R Core Team, 2017].

2. Definition of indices and data description

The Survey of Health, Ageing and Retirement in Europe SHARE is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of European population aged over 50. Data from six waves (from 2004 to 2015) are available at present. A generated variable module Deprivation [SHARE Release Guide, 2017] is described in the paper of [Adena, et al. 2015]. The module is included only in the fifth wave of the survey from 2013; for this reason, it is not possible to use panel arrangement of the SHARE survey or validate definitions or results on next waves. Values of two composite indicators *depsoc* for social deprivation and *depmat* for material deprivation from this module are used to model the distribution of these indices in the Czech Republic. *Depmat* is an aggregate measure of material conditions of individuals aged over 50 in Europe using a set of 11 items that refer to two broad domains: the failure in the affordability of basic needs and financial difficulties (details are given in Table 1). *Deptsoc* is an index for measuring social deprivation and exclusion, for this purpose 15 items from the survey were used (Table 2).

Both composite indicators are transformed into a 0-1 scale, 0 means no deprivation and 1 indicates the most serious deprivation. All questions, included in indices, are of the alternative type with possible response yes/no. Values 1 corresponding to answers yes (Tables 1 and 2) are weighted to the composite indices [Adena, et al. 2015, SHARE Release Guide, 2017]. Weights used in the SHARE module are hedonic-based on probit regression [Adena, et al. 2015]. In Tables 1 and 2 items included in analyzed indices are given together with the

weights. To compare, constant weights assigning the same weight to all items are 0.09 for the material index and 0.067 for the social deprivation index.

Table 1: Definition of the index of material deprivation

Variable	Weight	Problem
Meat	0.120	Your household does not eat meat, fish or chicken more often than three times per week because you cannot afford it.
Fruit	0.128	Your household does not eat fruits or vegetables more often than three times per week because you cannot afford to eat it.
Groceries	0.119	Can your household afford to regularly buy necessary groceries and household supplies?
Holiday	0.194	Could your household afford to go for a week long holiday away from home at least once a year?
Expense	0.086	Could your household afford to pay an unexpected expense without borrowing any money?
<i>In the last twelve months, to help you keep your living costs down, have you</i>		
Clothing	0.054	continued wearing clothing that was worn out because you could not afford replacement?
Shoes	0.029	continued wearing shoes that were worn out because you could not afford replacement?
Heating	0.064	put up with feeling cold to save heating costs?
Glasses	0.064	gone without or not replaced glasses you needed because you could not afford new ones?
Dentist	0.042	postponed visits to the dentist?
Doctor	0.100	Was there a time in the past 12 months when you needed to see a doctor but could not because of cost?

Source: SHARE Release Guide 6.0.0 2017.

Table 2: Definition of the index of social deprivation

Variable	Weight	Problem
Room	0.044	Less than one room per person in the household.
Literacy	0.077	Poor reading or writing skills.
IT	0.041	Poor computer skills or never used a computer.
Feeling	0.104	Not feeling part of the local area.
Vandalism	0.035	Vandalism in the local area.
Clean area	0.053	Local area not clean.
Help in area	0.090	No helpful people in local area.
Bank access	0.005	Difficult access to bank.
Shop access	0.041	Difficult access to grocery shop.
Pharmacy access	0.017	Difficult access to pharmacy.
Doctor	0.088	Waiting too long to see a doctor.
Course	0.055	Not attending any course in the past 12 months.
Organization	0.037	Not taking part in any organization in the past 12 months.
Trust	0.076	People cannot be trusted.
Isolation	0.237	Feeling left out of things.

Source: SHARE Release Guide 6.0.0 2017.

There are 4,091 respondents aged above 50 from the Czech Republic (2,366 females (58%) and 1,725 males (42%)) with observed values of both indices of interest in the SHARE data. For *depmat* index, there is 40.6% of zeroes (no deprivation) in the data, for the *depsoc* index only 0.6%. For this reason, the model with a discrete part (value 0 with positive probability) and a continuous part for positive values was selected for the modelling of the distribution of *depmat*.

In the SHARE project, the threshold for the material or social deprivation is proposed to be an upper quartile of all data (across European population). These values are 0.220 for material deprivation and 0.224 for social deprivation. Both values are lower than country-specific values for the Czech Republic (Table 3) 0.261 and 0.375. In the data for the Czech Republic, 26% (0.7%) of respondents overcome the threshold for material deprivation and more than 50% (0.7%) for social deprivation.

In Table 3, descriptive statistics for both indices are shown; for the *depmat* index, the characteristics are shown also for positive values of the index to show the impact of zeroes on the deprivation index.

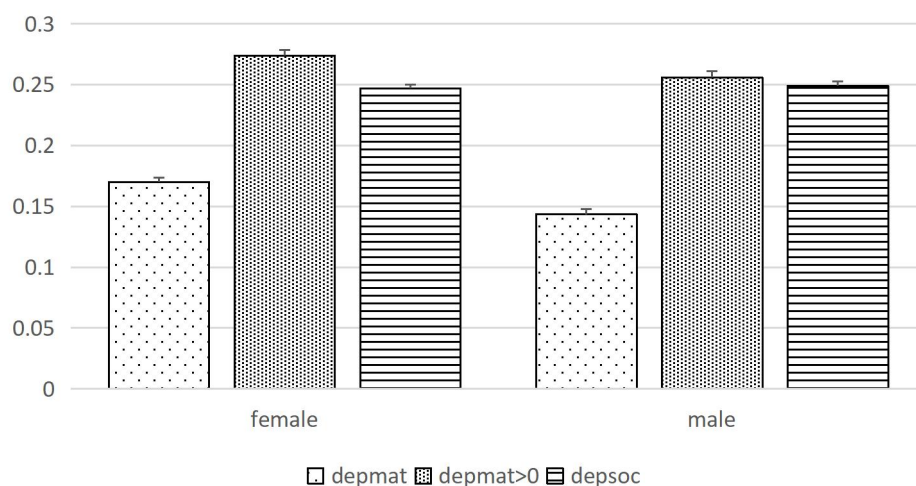
Table 3: Descriptive statistics of analysed indices (SD means standard deviation, lq is lower quartile, uq is upper quartile).

variable	mean	median	SD	lq	uq
<i>depmat</i>	0.159	0.114	0.186	0.000	0.261
<i>depmat</i> > 0	0.267	0.220	0.171	0.163	0.383
<i>depsoc</i>	0.248	0.243	0.144	0.126	0.375

Source: the author's work

From Table 3 the levels of both indicators seem to be different with a higher level of social deprivation. But from the origin of data as composite indicators, it is not fruitful to compare the levels. More interesting is to compare the whole distribution, variability or to evaluate a correlation. The Spearman correlation coefficient between both indices is relatively low, it is equal 0.222 (due to a large number of respondents this value is significantly positive).

Figure 1: Means and standard errors of indices in subpopulations defined by gender.

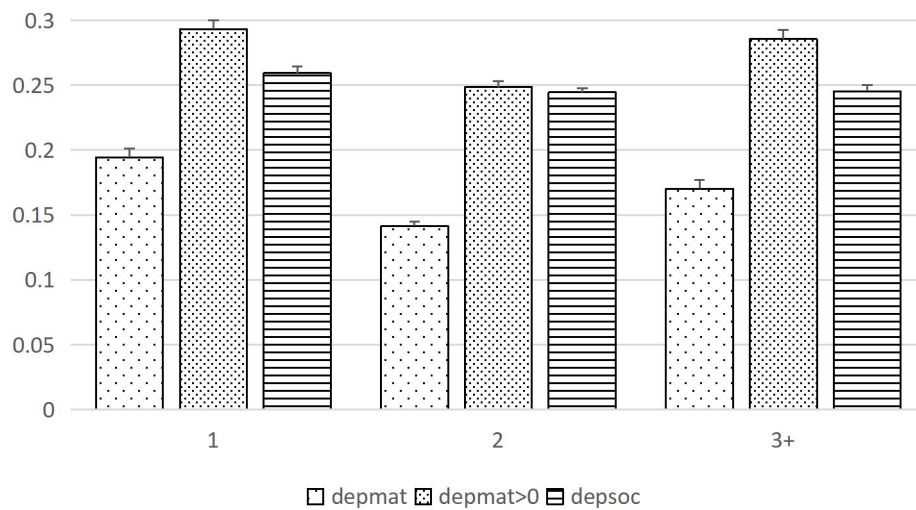


Source: Own computations.

In Figure 1 the means of both indices for subpopulations defined by gender are given, in Figure 2 households are divided into three groups (single member, two members and more

than 2 members) to show the impact of sharing a household on material and social deprivation. Due to the large sample, the standard errors of the means are small in limits 0.002 - 0.008 (shown in both figures). The relationship between men and women in material deprivation is difficult to be assessed, as this index depends only on the household situation (Table 1) and its value is the same for couples living in one household. This is not true for the index of social deprivation that is more personal (Table 2). Both means for material deprivation are higher for women, the mean social deprivation index is equal to man and women. In Figure 2, the positive impact of living in larger than single member households is shown, the mean index of material deprivation for couples (or households with two members) is equal to 2/3 of the mean for single member households. The difference in positive value index is smaller but well visible. For households with more than two members this effect is not so strong, but still apparent.

Figure 2: Means and standard errors of indices in subgroups defined by the size of household.



Source: Own computations.

3. Model

For the analysis of distributions of variables *depsoc* and *depmat* we use mixture models [Titterington, et al. 1985] in the form (*soc* refers to *depsoc* and *mat* to *depmat*)

$$f_{soc}(x) = \pi_{soc} f_{soc,1}(x; \mu_{soc,1}, \sigma_{soc,1}^2) + (1 - \pi_{soc}) f_2(x; \mu_{soc,2}, \sigma_{soc,2}^2), \quad (1)$$

$$f_{mat}(x) = \pi_{mat,1} f_1(x; \mu_{mat,1}, \sigma_{mat,1}^2) + \pi_{mat,2} f_2(x; \mu_{mat,2}, \sigma_{mat,2}^2) + \pi_{mat,3} \delta_0, \quad (2)$$

where $f_{soc}(x)$, $f_{mat}(x)$ are probability densities of analysed variables, $f_{mat,1}(x; \mu_{mat,1}, \sigma_{mat,1}^2)$, $f_{mat,2}(x; \mu_{mat,2}, \sigma_{mat,2}^2)$, $f_{soc,1}(x; \mu_{soc,1}, \sigma_{soc,1}^2)$, $f_2(x; \mu_{soc,2}, \sigma_{soc,2}^2)$ are component normal densities, δ_0 is the point mass distribution located at zero, $\pi_{soc}, \pi_{mat,1}, \pi_{mat,2}, \pi_{mat,3}$ are component weights fulfilling constraints

$$0 \leq \pi_{soc} \leq 1, 0 \leq \pi_{mat,1}, \pi_{mat,2}, \pi_{mat,3} \leq 1, \pi_{mat,1} + \pi_{mat,2} + \pi_{mat,3} = 1. \quad (3)$$

For the mixture with two normal components, we have in general 5 unknown free parameters (4 component parameters and one weight). In this contribution, the model for the *depmat* index with three components was treated in a simple two-step procedure, we estimated the proportion $\pi_{mat,3}$ of the population without any deprivation (material deprivation equal to zero, $depmat = 0$) by the relative frequency of zeroes $\hat{\pi}_{mat,3}$ in *depmat* and then moment matching method based on L-moments was used for two normal components under the constraint $\pi_{mat,1} + \pi_{mat,2} = 1 - \hat{\pi}_{mat,3}$. In order to estimate 5 unknown parameters of the distribution of the social deprivation index $(\pi_{soc}, \mu_{soc,1}, \sigma_{soc,1}^2, \mu_{soc,2}, \sigma_{soc,2}^2)$ and of the distribution of the material deprivation index $(\pi_{mat,1}, \mu_{mat,1}, \sigma_{mat,1}^2, \mu_{mat,2}, \sigma_{mat,2}^2)$, first five L-moments are used in the moment method estimation. For the application of formula (4), the quantile function Q of the mixture is needed. Unfortunately, there is not a close formula for the quantile function of a mixture. In the text, quantiles $x_p, 0 < P < 1$, are evaluated on the grid $\left(P = \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}\right)$ for any (large enough) integer m from the definition of quantiles ($F(x_p) = P$, where F is a cumulative distribution function of the mixture and $Q(P) = x_p, 0 < P < 1$). A weighted average of component quantiles $x_{p,j}, j = 1, 2$ is a good initial guess for a numeric procedure. In this contribution, the function *uniroot* from basic *Stats* package in the program R [R Core Team, 2017] was used to evaluate theoretical quantile function of all mixtures. Another possible method is described in [Bernard, Vanduffel, 2015].

The formula for evaluation of L-moments based on a quantile function was given in [Hosking 1990] (see also [Karvanen, 2006], [Bílková, 2016]) as ($k = 1, 2, \dots$)

$$\lambda_k = \int Q(u) P_{k-1}^*(u) du, \tag{4}$$

where P_r^* is the r -th shifted Legendre polynomial given as

$$P_r^*(u) = \sum_{l=0}^r p_{r,l}^* u^l,$$

where coefficients $p_{r,l}^*$ in the polynomial are

$$p_{r,l}^* = (-1)^{r-l} \binom{r}{l} \binom{r+l}{l}.$$

The formula (4) was applied for $k = 1, 2, \dots, 5$ and the theoretical L-moments $(\lambda_1, \dots, \lambda_5)$ were evaluated together with the first five sample L-moments (l_1, \dots, l_5) (package *Lmoments* [Karvanen, 2006] in the program R was applied).

In the moment-matching method of estimation, five nonlinear equations $\lambda_j = l_j, j = 1, 2, \dots, 5$ should be solved with respect to the unknown parameters (and under constraints given in (3) and $\sigma_{soc,1}, \sigma_{soc,2}, \sigma_{mat,1}, \sigma_{mat,2} > 0$). Instead of that, the function

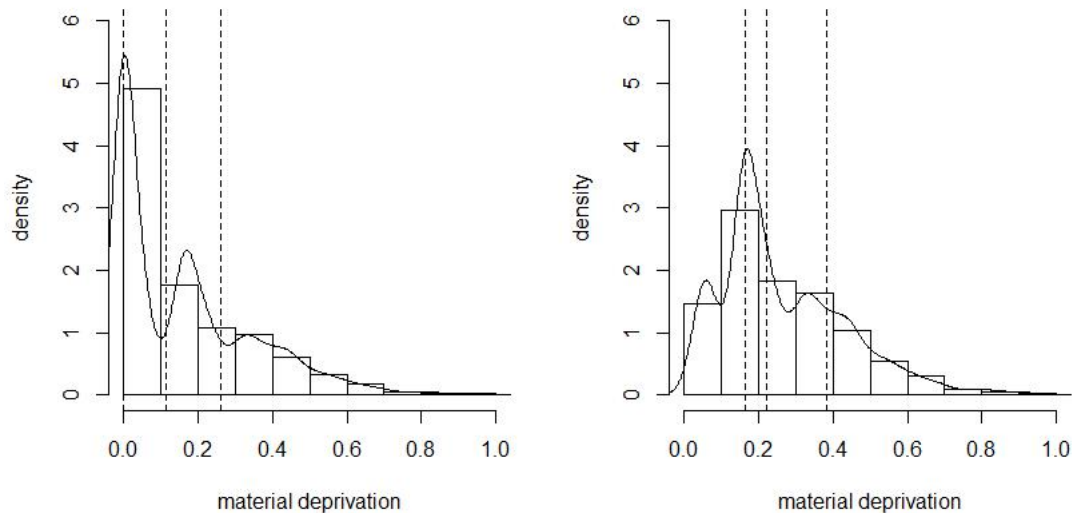
$$\sum_{j=1}^5 (\lambda_j - l_j)^2,$$

was minimised with respect to unknown parameters. Function *nls* (Nonlinear Least Squares) in the basic *Stats* package in the program R was used for the minimization. In Table 4 estimated parameters are given together with classical maximum likelihood estimates evaluated with the use of EM algorithm in the package *mixtools* [Benaglia, et al. 2009].

4. Results

In Figure 3, two histograms of *depmat* are shown and kernel estimates (the Epanechnikov kernel was applied) of the densities are added to characterise empirical distributions. On the left, all data are used; on the right, we use only positive values of *depmat*. The positive probability of zero deprivation index causes the positive probability of negative values (left figure), although the modelled variable is positive. On the right, this problem is almost eliminated. The difference in distributions is well visible, the latter distribution is skewed, unimodal with the mode close to the lower quantile.

Figure 3: An empirical distribution of *depmat* with quartiles (dashed lines) and kernel estimate of the density.



Source: Own computations.

In Figure 4, the histogram of *depsoc* shows two modes well coinciding with quartiles of data. There is an anti-mode close to the median value.

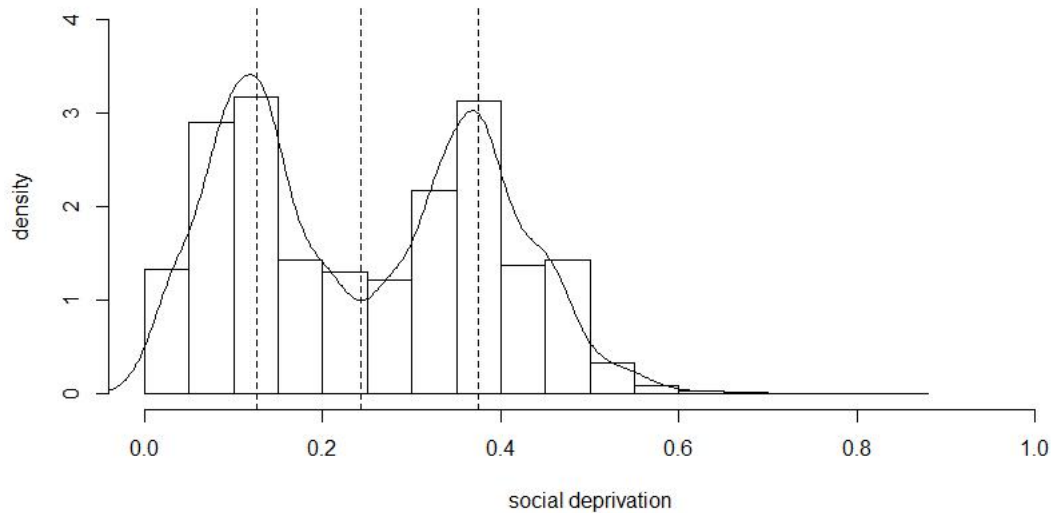
Distributions of analysed indices are very different, but for both the normal mixture model fits the distribution well; for the material index, it fits the skewed distribution with two highly overlapping components, for the social index it identifies two almost separated components.

In Table 4 estimated parameters of mixtures are given. The expected values of both components are higher for material deprivation index than for social deprivation index (but we compare only positive values of *depmat* (approximately 60% of values)).

All estimated densities are shown in Figure 5. Comparing Figures 3 and 4 with Figure 5 we realise that the model is better for the social deprivation than for the material deprivation. Both fits (moment-matching and maximum likelihood) are similar with comparable estimated characteristics of the level and variance, as well as of quartiles. The chi-squared goodness of fit tests are highly significant for all four fits due to both quality of fits and a high number of

observations. For our data, the method using L-moments gives comparable results (even in variances, done by bootstrapping and not included in this text). The moment method uses only basic methods from numeric mathematics and it enables us to avoid using the EM algorithm. L-moments are robust to outliers, skewness and heavy tails ([Hosking, 1990], [Karvanen, 2006]) and for its application, it needs only the first absolute component moments to be finite. In case of our data (with bounded support), we do not use this property, although we wanted to benefit from the robustness of the method and use the normal mixture model in presence of outliers. Moreover we showed, that in the method in our analysis gives comparative results with maximum likelihood even for normal mixtures.

Figure 4: Empirical distribution of depsoc with quartiles (dashed lines) and kernel estimate of the density.



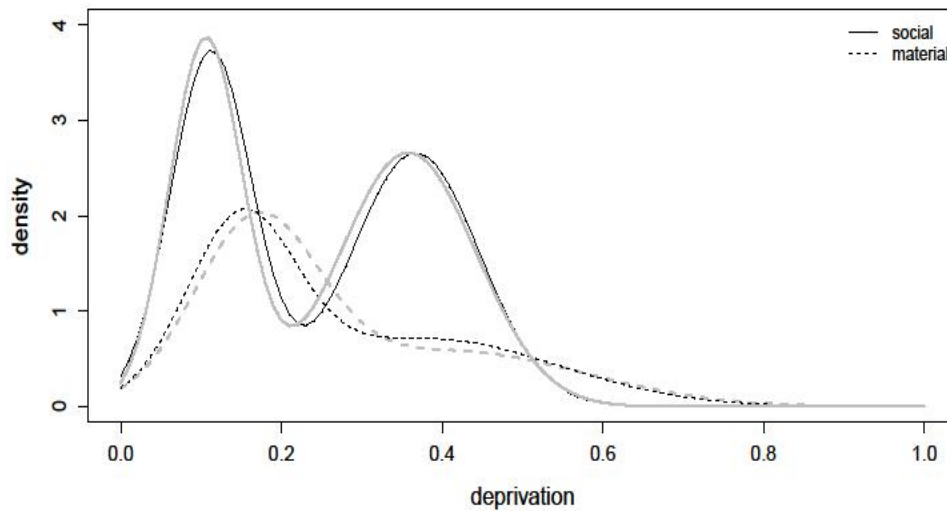
Source: Own computations.

Table 4: Estimated parameters of models given in (1) and (2). Moment method and MLE.

parameter	Moment method L-moments		Maximum likelihood	
	social deprivation	material deprivation	social deprivation	material deprivation
π_1	0.437	0.353	0.472	0.304
π_2	0.563	0.241	0.528	0.290
π_3	-	0.406	-	0.406
μ_1	0.1059	0.1683	0.1128	0.1502
σ_1	0.0453	0.0767	0.0507	0.0667
μ_2	0.3580	0.4112	0.3672	0.3815
σ_2	0.0843	0.1659	0.0792	0.1652

Source: Own computations.

Figure 5: Estimated distribution of indices *depmat* (unimodal, skewed distribution) and *depsoc* (bimodal distribution), MLE estimates in black, moment method in grey.



Source: Own computations.

4. Conclusion

In the text, the model of the finite normal mixture is applied in order to describe the distribution of two composite indicators (material deprivation and social deprivation). Two normal components are used; in case of material deprivation, one additional discrete component was necessary to include respondents without any deprivation. Analyzed indices are adapted to the target ageing population, especially in social deprivation. For example using computers is a relevant question for older people and useless for the young generation. An effort to create indices oriented on elderly people is important and useful for future of the ageing population.

Our results suggest, that the social deprivation is a more serious problem for the ageing population in the Czech Republic (for the population above 50) than material deprivation. Attention should be paid to social relationships and social inclusion, not just to material needs of elderly people. There is almost 40 percent of respondents who feel no material deprivation in comparison with less than 1% in the case of social deprivation. If we take the limit of existence of deprivation based on the European population on upper quantiles, in the Czech Republic we obtain 26% of materially deprived, but more than 50% socially deprived. It means, that for material deprivation we well copy European mean, for social index the situation is worse and (based on analyzed indices) social exclusion is a serious problem of our elderly population. Moreover, the distribution of both indexes is different in shape even for positive values of both indices. The distribution of the social deprivation index is clearly bimodal with two components including 44% in the component of respondents with low deprivation and 56% of respondents in the higher level component (also with higher variability). The mixture models might be used for clustering data. The analysis of positive items for the second subpopulation can identify the most serious problems. Even if we take only positive values of the material deprivation index, the distribution is not symmetric after logarithmic transformation (one lognormal distribution is not an acceptable model). In this case, the mixture also enables fitting of the data, components are not separated.

Unfortunately, the deprivation module is included only in the fifth wave of the SHARE; in other waves all questions, included in the indices, are not asked. For this reason, we do not

have data to validate these indices for more time periods (and utilise longitudinal character of data) or test our model on new data of other respondents in different time period. From this point of view, our analysis has limited meaning. However, the proposed indices provide an interesting attempt to quantify (or even measure) deprivation specifically for the elderly. The indices seem to measure two different dimensions of deprivation as the relationship between them is surprisingly weak. The concept of serious deprivation based on both indices can describe more complex characteristics of the situation of elderly people.

Acknowledgements

The support of the funds of institutional support of a long-term conceptual advancement of science and research number IP 400 040 at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic is gladly acknowledged.

References

- [1] 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.
- [2] Bellani, L., D'Ambrosio, C. (2011). Deprivation, social exclusion and subjective well-being. *Social Indicators Research*, pp. 67-86.
- [3] Benaglia, T., Chauveau, D., Hunter, D., Young, D. (2009) Mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, vol. 32, pp. 1-29, <http://www.jstatsoft.org/v32/i06/>.
- [4] Bernard, C., Vanduffel, S. (2015). Quantile of a Mixture with Application to Model Risk Assessment. *Dependence Modeling*, 3(1), pp. 172-181.
- [5] Bílková, D. (2016). Modification of the Classical L-Moments: TL-Moments. In: The 10th International Days of Statistics and Economics (MSED 2016) [online]. Praha, 08.09.2016 – 10.09.2016. Slaný : Melandrium, 2016, pp. 173–194. ISBN 978-80-87990-10-0. Retrieved from https://msed.vse.cz/msed_2016/article/27-Bilkova-Diana-paper.pdf.
- [6] Börsch-Supan, A. (2016). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 5.0.0. SHARE-ERIC. Data set.
- [7] CZSO. Household Income and Living Conditions – 2013. (2013). [20.04.2017] <https://www.czso.cz/csu/czso/household-income-and-living-conditions-2013-ia0fwqxyxa>
- [8] Europe 2020 Strategy. (2017). [12.04.2017] https://ec.europa.eu/info/strategy/european-semester/framework/europe-2020-strategy_en
- [9] EUROSTAT. Depth of material deprivation - EU-SILC survey. (2017). [12.04.2017] <http://ec.europa.eu/eurostat/web/products-datasets/-/tessi150http://ec.europa.eu/eurostat/tgm/refreshTableAction.do?tab=table&plugin=1&pcode=tessi082&language=en>
- [10] Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of OrderStatistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52, pp. 105-124.
- [11] Myck, M., Najsztab, M., Oczkowska, M. (2015). Measuring social deprivation and social exclusion. In: Ageing in Europe - Supporting Policies for an Inclusive Society, De Gruyter.
- [12] Karvanen, J. (2006). Estimation of quantile mixtures via L-moments and trimmed L-moments. *Computational Statistics & Data Analysis*, vol. 51, pp. 947-959.

- [13] R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. Retrieved from <https://www.R-project.org/>.
- [14] Saisana, M., Saltelli, A., S. Tarantola. (2005). Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 168, pp. 307-323
- [15] SHARE Release Guide 6.0.0. (2017). [18.4.2017]
http://www.share-project.org/fileadmin/pdf_documentation/SHARE_release_guide_6-0-0.pdf
- [16] Titterington, D.M., Smith A.F., & Makov U.E. (1985). *Statistical analysis of finite mixture distributions*. Wiley & Sons.

