TERRITORIAL DIVISION AND INCOME AFFLUENCE – ANALYSIS USING TWO-LEVEL LOGIT MODELS

Anna Sączewska-Piotrowska

University of Economics in Katowice, Katowice, Poland
e-mail: anna.saczewska-piotrowska@ue.katowice.pl

ORCID: 0000-0001-9760-3649

Abstract: The aim of the paper was to assess differentiation of the occurrence of households' income affluence in Poland between subregions. An analysis was conducted using two-level logit models without explanatory variables (null model) and with explanatory variables at household level (random intercept model and random slope model). The variables were related to the characteristics of the household and its head. The conducted analysis allowed to state that the occurrence of affluence is differentiated between subregions in the null model as well as in the model with explanatory variables.

Keywords: two-level logit model, multilevel modelling, affluence, income, household.

1. Introduction

Many studies are focused on income poverty. Income affluence is, however, a rarely mentioned phenomenon in the literature. It should be noted that in each country the affluent have a big impact on socio-economic development and they are the reference group for the less affluent. The affluent accumulate a large part of wealth and income, and therefore they play a key role in the economy. The most affluent people in society have a big influence on trends in production and consumption. Research on affluence allow to indicate affluent individuals, and to obtain more information about this group which has, after all, the greatest economic, political and cultural impact on society. The affluent should be analysed and diagnosed continuously, and the conclusions should be drawn because this group of people is very important from the point of view of the whole of society.

The most popular indicators of financial well-being are income, wealth and luxury products and services. There is no single best indicator, and researchers use different choices to study financial well-being (e.g. [Brzeziński 2010; *Zasobność*

gospodarstw domowych… 2017; *Rynek dóbr luksusowych…* 2017]). In this study, monthly income was selected as the proxy for financial well-being. It should be noted that financial well-being is a narrower category than well-being itself because attention is paid only to the one dimension of it, not considering the other aspects of life.

There is a distinction between affluence and richness in the literature. Being rich is defined as the highest level of affluence [Radziukiewicz 2006], but in practice the affluence and richness lines proposed by different authors have different values and thus this same group of people may be considered by one author as affluent, but by another author – as rich.

The poor and the affluent (the rich) can be defined in different ways. The poverty line is defined as the cut-off income point below which a household is considered to be poor, and analogically the affluence (the richness) line as the cut-off income point above which a household is considered to be affluent (rich). Additionally, in some studies the middle class is defined as those living between the poverty and affluence (richness) thresholds. The lines can be defined in absolute and relative terms. We will focus only on the affluence (richness) lines in relative terms – calculated as a percentage of the median income. Among the authors considering the affluence (the richness) lines in absolute terms we can mention Di Maggio [2003], Hutton [2006], Bose et al. [2014].

One of the primary indicators used by Eurostat is the at-risk-of-poverty rate calculated using the 60% threshold (60% of the median income). However, it should be noted that the at-risk-of-poverty rate is also calculated using the 40% threshold, the 50% threshold and the 70% threshold. These three rates are part of one of the secondary indicators – dispersion around the at-risk-of-poverty threshold [European Commission, 2010]. Analogically, the affluence (the richness) line may be defined as some percent of the median income. This line is often defined as 200% of the median, 300% of the median or 400% of the median. Some authors [Franzini et al. 2016] defined three categories: affluent (income above 300% of the median), rich (above 500%) and super-rich (above 1000%). Relative thresholds of affluence (richness) were used by Brzeziński [2010], Peichl et al. [2010], Sączewska-Piotrowska [2015], Franzini et al. [2016], Törmälehto [2017]. In the analysis the author decided to use the threshold 200% of the median and the name "affluence" (not "richness"), because the threshold is not set very high. Setting a higher threshold may be risky – only a small part of society achieves an income higher than 300%, 400% etc. and therefore the scheduled estimation of the models with many variables could be impossible.

In many areas the data have a hierarchical or clustered structure. For example, respondents in a complex large survey are naturally clustered in geographical units and may be grouped into smaller units [Moineddin et al. 2007]. If individuals are clustered within classes – this is a two-level data structure. Not only individuals are clustered within classes, but the classes may be clustered within other bigger classes –

this is a three-level data structure. We have also a four-level data structure and so on, but such data structures are not considered in practice. Generally this kind of data have a multilevel structure and there are used special statistical models designed for this multilevel structure, known as multilevel, mixed, hierarchical or random effects models. Single-level ("naive") regression assumes that all observations are independent, multilevel analysis takes into account the dependency of the observations. Outcomes of interest in many fields do not only reflect continuous measures, other kinds of outcomes are also of interest. One of the biggest advantages of multilevel analysis is that it can be used for the analysis of different kinds of outcome variables [Twisk 2006]: continuous (linear multilevel analysis), 'count' (Poisson multilevel analysis), categorical (multinomial logistic multilevel analysis) and dichotomous (logistic multilevel analysis).

Multilevel models were used in poverty studies, for example by Brady et al. [2009], Kim et al. [2010], Amara and Jemmali [2018]. To the best of the author's knowledge there has been no use of multilevel models in affluence analysis so far.

The aim of this paper was to assess the differentiation of the occurrence of households' income affluence in Poland between subregions controlling some characteristics of the household and its head. In this study, households in Poland are clustered within subregions (NUTS 3 level) – the data have a two-level structure. The study allows to answer the question if random effects, random intercepts and random slopes are statistically significant. In other words this study should allow to answer the question of whether the relationship between the existence of affluence and the characteristics of a household and its head varies between the subregions, and whether the relationship between the affluence and the characteristics is different for different subregions. In the analysis, logistic multilevel models are used because the outcome variable is dichotomous – the household may be affluent or not.

2. Material and methods

The study was conducted based on data from the Social Diagnosis Project [Council for Social Monitoring 2016]. This project is a panel study conducted in Poland. Each subsequent wave involves all available households from the previous wave and households from a new representative sample. Eight waves were conducted from 2000 to 2015. Data from the 2015 wave were used in the study. The sample was 11 738 households from 66 subregions (in Poland there are now 73 subregions, but the Social Diagnosis database contains information about belonging to 66 subregions – a territorial division valid until 31.12.2014).

As the proxy for financial well-being, monthly income was chosen, the examined unit was household. Equivalised income was calculated in order to take account of the differences in a household's size and its composition. The modified OECD (Organisation for Economic Co-operation and Development) equivalence scale was used. This scale assigns 1 to the first adult of the household, 0.5 to each subsequent

adult aged 14 or above and 0.3 to children (each person under 14). The affluence threshold was set at 200% of the median income.

In the analysis the binary outcome variable was defined as affluent household (=1) or not (=0).

First of all, the relationship between affluence and some characteristics of the household head was tested by the "naive" logistic regression to be able to compare the results of multilevel modelling with the results of "naive" analysis. These types of "naive" models were used[1]:

- model A ("naive" model without explanatory variables):

$$logit(\pi_i) = \beta_0,$$

- model B ("naive" model with explanatory variable):

$$logit(\pi_i) = \beta_0 + \beta_1 X_{1i},$$

where β_0, β_1 – parameters, X_{1i} – explanatory variable, $\pi_i = \Pr(y_i)$, where y_i is the response for household i.

To take into account the dependency of the observations the multilevel models were estimated. In this study, households in Poland are clustered within subregions – this is a two-level structure (Figure 1).
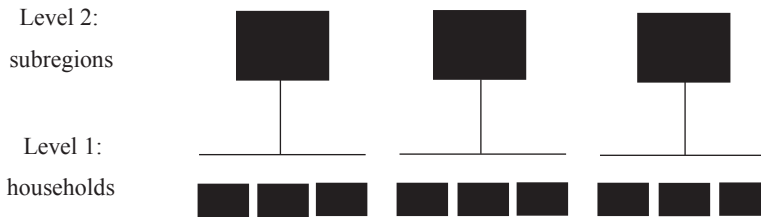
Fig. 1. Two-level structure – households clustered in subregions

Source: own work.

The parameters of three types of multilevel models were estimated: the null model (model 1), the random intercept model (model 2) and the random slope model (model 3) (Table 1). In model 1 the single variance parameter for the different intercepts for the different subregions is estimated, model 2 includes additionally household-level variable, and in model 3 one variance parameter for the different slopes of the regression lines for the different subregions is additionally estimated. This follows the "step-up" strategy described by Raudenbush and Bryk [2002].

---

[1] For the purpose of presenting different types of estimated models, one explanatory variable was included. In practice 18 variables were included.

Table 1. Types of estimated two-level logit models*

Type of model	Household level (level 1)	Subregion level (level 2)
Model 1: null model	$logit(\pi_{ij}) = \beta_{0j}$	$\beta_{0j} = \gamma_{00} + u_{0j}$
Model 2: random intercept model	$logit(\pi_{ij}) = \beta_{0j} + \beta_{1j}X_{1ij}$	$\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10}$
Model 3: random slope model	$logit(\pi_{ij}) = \beta_{0j} + \beta_{1j}X_{1ij}$	$\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10} + u_{1j}$

* γ_{00}, γ_{10} – parameters, u_{0j}, u_{1j} – random effects, X_{1ij} – explanatory variable, $\pi_{ij} = \Pr(y_{ij} = 1)$, where y_{ij} is the response for household i in subregion j

Source: own work based on [Domański, Pokropek 2011].

It is possible to test the significance of random effects variance. In the likelihood ratio (LR) test the mixture of χ^2 distributions is used. In the simplest case (testing the addition of a single random effect) the distribution of the deviance under the null hypothesis is approximately a mixture of χ_0^2 and χ_1^2 distributions, which translates into simply dividing the *p*-value from a χ_1^2 by 2 [*Random effects testing*… 2009].

The estimated models included explanatory variables related to some characteristics of the household and its head. Categorized and dichotomized covariates were included in the estimated models. In each case one category is a reference category. The variables included in the models are shown in Table 2.

Table 2. List of variables in logit models

Variable	Description*	
1	2	
place_city_0	place of resident	1 – rural areas
edu2015_4_fin_1	education level of household head	lower secondary and less (ref.)
edu2015_4_fin_2		basic vocational
edu2015_4_fin_3		Secondary
edu2015_4_fin_4		Tertiary
seg_2015_1	socio-economic group	employees (ref.)
seg_2015_2		Farmers
seg_2015_3		self-employed
seg_2015_4		retirees and pensioners
seg_2015_5		living on unearned sources
age2015_4_fin_1	age group of household head	34 and less (ref.)
age2015_4_fin_2		35-44
age2015_4_fin_3		45-59
age2015_4_fin_4		60 and above

Table 2, cont.

1		2
sex2015_male_1	sex of household head	1 – male
hand2015_fin_1	presence of persons with disability	1 – household with persons with disability
unemp2015_fin_1	status of household in the labour market	1 – household with unemployed persons
number_5cat_2015_1	number of household's members	1 member (ref.)
number_5cat_2015_2		2 members
number_5cat_2015_3		3 members
number_5cat_2015_4		4 members
number_5cat_2015_5		5 members or more

* ref. – reference category

Source: own work.

To estimate the effect of a subregion itself on household outcomes, some measures of components of variance and of heterogeneity may be calculated.

The total variance in the outcome variable is the sum of the household-level and subregion-level variances. The intraclass correlation coefficient (ICC) is defined as [Twisk 2006] the variance between subregions divided by the total variance. Due to the household-level variance not available in a logistic model, the variance of a logistic distribution with scale parameter equal to one is used [Evans et al. 1993]. On this basis ICC is calculated as [Snijders, Bosker 1999; Austin, Merlo 2017]:

$$ICC = \frac{\tau^2}{\tau^2 + \frac{\pi^2}{3}}, \tag{1}$$

where τ^2 is the estimated variance of the random effect of subregion and $\frac{\pi^2}{3}$ denotes the variance of a standard logistic distribution. The ICC for logistic models is difficult to interpret and because of that the other measures can be used as an alternative of ICC.

The aim of the median odds ratio (MOR) is to translate the subregion level variance into the widely used odds ratio scale, which has a consistent and intuitive interpretation. The MOR is defined as [Larsen et al. 2000; Larsen, Merlo 2005] the median value of the distribution of ORs obtained when randomly picking two households with the same covariate values from two different subregions, and comparing the one from the higher risk subregion to the one from the lower risk subregion. The MOR is calculated as:

$$MOR = \exp\left(\sqrt{2\tau^2} \cdot \Phi^{-1}(0.75)\right), \tag{2}$$

where τ^2 is the subregion-level variance, $\Phi^{-1}(\cdot)$ denotes the inverse cumulative standard normal distribution function. If the MOR is 1, there is no variation between subregions.

All the calculations were made in R [R Core Team 2017] with package lme4 [Bates et al. 2015]. Creating the charts also requires additional packages as ggplot2 [Wickham 2009], sjmisc [Lüdecke 2018a], sjPlot [Lüdecke 2018b] and maptools [Bivand, Lewin-Koh 2017]. The map of Poland with a division into subregions is available on the Eurostat website [www1].

3. Results

The basic information about the income affluence of households in Poland is that 9.77% households are affluent. The occurrence of affluence differs between subregions – 33.8% households from the capital city of Warsaw are affluent and only 2.5% and 2.3% households from the Sieradzki subregion and Pilski subregion are affluent respectively. The range of affluence in subregions in Poland is presented in Figure 2.

less than 6.95
6.95-9.50
9.50-12.05
12.05 and more
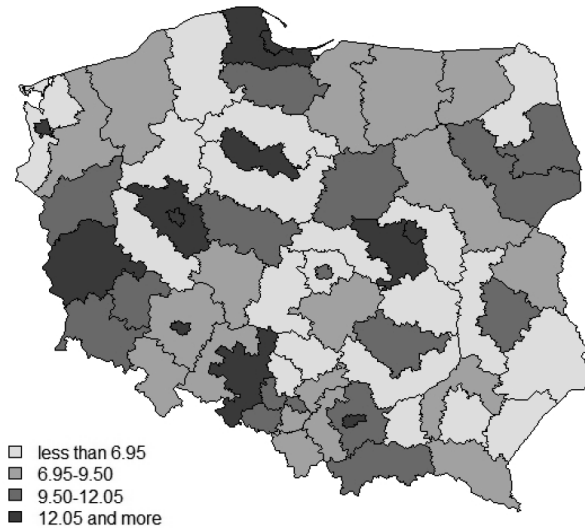
Fig. 2. Range of income affluence in subregions in Poland

Source: own work [Council for Social Monitoring 2016], © EuroGeographics for the administrative boundaries.

First of all, the parameters of the "naive" logit models A and B were estimated (Table 3). Based on the likelihood ratio test it can be stated that model B (the model with the intercept and coefficients) is better fitted to the data than model A (the model

with the intercept). These models are the "reference models" to multilevel logistic regression whose results are shown in Table 4.

The multilevel models were estimated using the Laplace approximation. Due to problems with convergence in model 3, the results of estimation are not presented in Table 4. Based on the comparison of model 1 with model A, and model 2 with model B, it can be stated that the random effect in model 1 is statistically significant as well in model 2. This means that occurrence of income affluence is varied between subregions both in model 1 and in model 2. Model 2 is better fitted than model 1, this situation is indicated by the lower Akaike information criterion (AIC) in model 2 and the results of the LR test ($\chi^2 = 748.98$ ($p < 0.001$)

Table 3. Estimated regression coefficients for the single-level logistic regression models

Variable	Model A			Model B		
	Coefficient	S.E.	*P*-value	Coefficient	S.E.	*P*-value
Intercept	−2.235	0.034	<0.001	−3.599	0.210	<0.001
place_city_0				−0.245	0.086	0.004
edu2015_4_fin_2				0.218	0.154	0.157
edu2015_4_fin_3				0.698	0.151	<0.001
edu2015_4_fin_4				1.960	0.152	<0.001
seg_2015_2				0.222	0.158	0.159
seg_2015_3				0.582	0.131	<0.001
seg_2015_4				−1.046	0.121	<0.001
seg_2015_5				−0.805	0.296	0.007
age2015_4_fin_2				−0,165	0.138	0.231
age2015_4_fin_3				0.287	0.126	0.022
age2015_4_fin_4				0.442	0.147	0.003
sex2015_male_1				0.381	0.086	<0.001
hand2015_fin_1				−0.200	0.092	0.031
unemp2015_fin_1				−0.081	0.125	0.518
number_5cat_2015_2				0.681	0.125	<0.001
number_5cat_2015_3				0.605	0.139	<0.001
number_5cat_2015_4				0.300	0.149	0.047
number_5cat_2015_5				0.437	0.161	0.007
LR test				$\chi^2 = 161.29$ ($p < 0.001$)		
AIC	6471.0			5648.5		

Source: own work based on [Council for Social Monitoring 2016].

Table 4. Estimated regression coefficients and variance components for the two-level logistic regression models

Variable	Model A			Model B		
Fixed effects	Coefficient	S.E.	*P*-value	Coefficient	S.E.	*P*-value
Intercept	−2.234	0.067	<0.001	−3.780	0.218	<0.001
place_city_0				−0.116	0.091	0.205
edu2015_4_fin_2				0.212	0.154	0.171
edu2015_4_fin_3				0.687	0.152	<0.001
edu2015_4_fin_4				1.906	0.153	<0.001
seg_2015_2				0.251	0.160	0.117
seg_2015_3				0.573	0.133	<0.001
seg_2015_4				−1.031	0.122	<0.001
seg_2015_5				−0.821	0.297	0.006
age2015_4_fin_2				−0,151	0.140	0.278
age2015_4_fin_3				0.326	0.127	0.011
age2015_4_fin_4				0.469	0.149	0.002
sex2015_male_1				0.374	0.086	<0.001
hand2015_fin_1				−0.175	0.093	0.061
unemp2015_fin_1				−0.045	0.126	0.721
number_5cat_2015_2				0.744	0.127	<0.001
number_5cat_2015_3				0.673	0.141	<0.001
number_5cat_2015_4				0.371	0.152	0.014
number_5cat_2015_5				0.521	0.163	0.001
Random effects	Variance			Variance		
Intercept	0.200			0.094		
Test for random effects	$\chi^2 = 154.33$ ($p < 0.001$)			$\chi^2 = 40.634$ ($p < 0.001$)		
ICC	0.057			0.028		
MOR	1.530			1.338		
AIC	6318.7			5605.7		

Source: own work based on [Council for Social Monitoring 2016].

Using model 2, based on ICC it can be stated that 2.8% of the total variation in the existence of affluence is due to between-subregion differences (after adjusting for some characteristics of the household and its head). The remaining 97.2% is due to unmeasured differences between households.

The MOR for model 2 was equal to 1.338, which shows that if a household moved to another subregion with a higher probability of affluence, the median increase in their odds of affluence would be 1.338 times.

Interpreting the regression coefficients in model 2, special attention should be paid to education, the socio-economic group and the number of household's members. Through the prism of the odds ratio (Table 5), within each subregion, the estimated odds of affluence are 6.7 times (e1.[906]) as high for households with a high-educated head (tertiary education) in comparison to households with a low-educated head (lower secondary and less). The estimated odds of affluence for households of retirees and pensioners are approximately 36% of the odds for the households of the employed. We can also note that the odds of affluence in households with more than one member were significantly higher in comparison to single-person households.

**Table 5.** Estimated odds ratios for model 2 for statistically significant variables

Variable	Odds ratio	95% CI
edu2015_4_fin_3	1.988	1.476-2.677
edu2015_4_fin_4	6.725	4.979-9.084
seg_2015_3	1.773	1.365-2.302
seg_2015_4	0.357	0.281-0.453
seg_2015_5	0.440	0.246-0.787
age2015_4_fin_3	1.385	1.079-1,777
age2015_4_fin_4	1.598	1.193-2.141
sex2015_male_1	1.454	1.228-1.723
number_5cat_2015_2	2.104	1.640-2.698
number_5cat_2015_3	1.960	1.488-2.583
number_5cat_2015_4	1.449	1.077-1.951
number_5cat_2015_5	1.683	1.223-2.316

Source: own work based on [Council for Social Monitoring 2016].

The probability plots for each covariate can be plotted depending on the grouping level from the random intercept. Figure 3 shows the probability plots for two selected covariates – tertiary education of the household head and households of retirees and pensioners. The lines are not straight because this is a logit model with plotted probabilities, not the logarithm of probabilities. In Figure 3 one line is much higher than the other lines – this line is plotted for the capital city of Warsaw. It means that in the capital city the probability of affluence is much higher than in the other subregions. This situation takes place in the case of all covariates[2].

Finally, the normality assumption of random effects was tested (Figure 4). To test this assumption QQ-plots were made (this kind of plot shows random against standard quantiles). In the QQ-plots for model 1 and for model 2 not all the vertical segments cross the straight line which suggests rejecting the hypothesis: the random

---

[2] Due to limited space only two plots were presented in the paper.

a) tertiary education of household head   b) households of retirees and pensioners
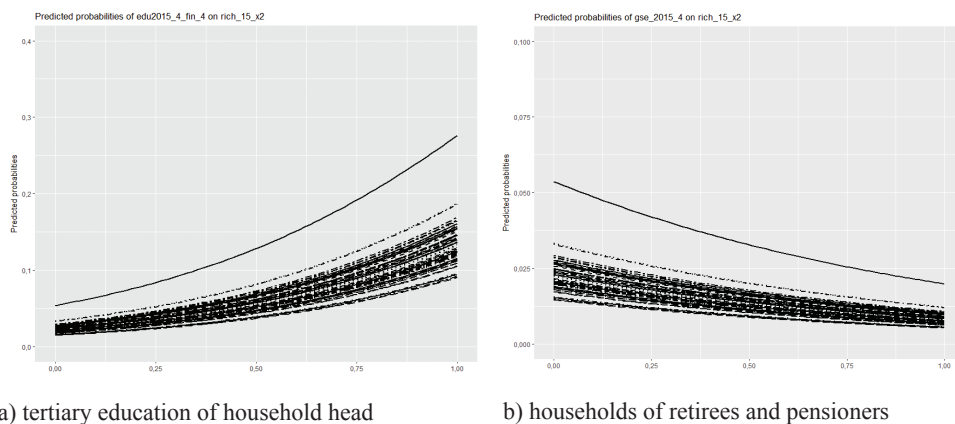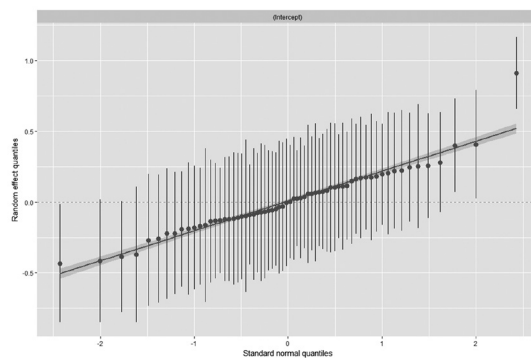
Fig. 3. Probabilities of fixed effects in model 2 grouped by random intercepts – selected covariates

Source: own work based on [Council for Social Monitoring 2016].
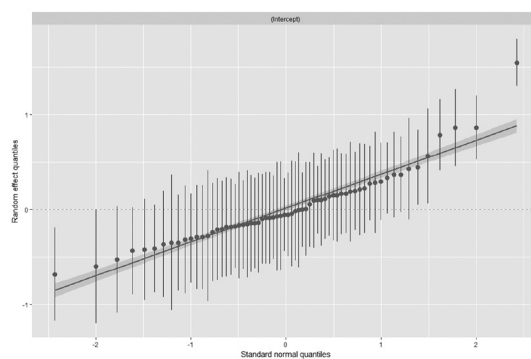
a) model 1

b) model 2

Fig. 4. QQ-plot for random effects in model 1 and in model 2

Source: own work based on [Council for Social Monitoring 2016].

effect is normal. We can see that including the covariates in the model improves the situation, but the last segment still does not cross the straight line and thus the normality assumption is not met.

4. Conclusions

The occurrence of affluence is varied between subregions (null model) and this differentiation persists in controlling many variables associated with a household and its head (random intercept model). It can be stated that there is no possibility to estimate the parameters of the random slope model, because the model could not achieve convergence. Thus, the answer to the one of the research questions – if the relationship between affluence and some characteristics of a household and its head is different for different subregions – was impossible. The reasons of non-convergence may be different. Too many random effects may lead to non-convergence [Vorbeke, Molenberghs 2000]. This finding was confirmed by other authors who stated that the likelihood of convergence failure tends to increase with the complexity of the model, especially the random effects structure [Barr et al. 2013]. Other authors proved that convergence problems arise when prevalence is low [Moineddin et al. 2007]. In model 3 (with random intercepts and random slopes) the causes of non-convergence are both: too many random effects and low prevalence. Some authors directly stated that the Laplace approximation can handle up to three random effects [*A practical guide*… 2017].

Based on the conducted study it can be said that subregion-level variance is significant, but is definitely lower than household-level variance. Households from Warsaw are characterized by definitely the highest probability of affluence. Within the same subregion, the odds of affluence are definitely higher for households with a high-educated household head (comparing to households with a low-educated head) and definitely lower for households of retirees and pensioners (compared to households of the employed). The odds of affluence were higher in households with two members and more in comparison to single-person households. The relationships between existence of affluence and the other covariates were not so clear, but some of them were statistically significant.

Further research will focus on attempts to fit models with more random effects using other estimation methods. We will also estimate the parameters of two-level models with a binary outcome variable concerning the occurrence and absence of richness to answer the question about the differentiation of richness between subregions.

Bibliography

Amara M., Jemmali H., 2018, *Household and contextual indicators of poverty in Tunisia: A multilevel analysis*, Social Indicators Research, vol. 137, is. 1, pp. 113-138.

A practical guide to mixed models in R, 2017, https://ase.tufts.edu/gsc/gradresources/guidetomixed-modelsinr/mixed%20model%20guide.html (16.06.2018).

Austin P.C., Merlo J., 2017, *Intermediate and advanced topics in multilevel logistic regression analysis*, Statistics in Medicine, vol. 36, no. 20, pp. 3257-3277.

Barr D.J., Levy R., Scheepers C., Tily H.J., 2013, *Random effects structure for confirmatory hypothesis testing: keep it maximal,* Journal of Memory and Language, vol. 68, no. 3, pp. 1-43.

Bates D., Maechler M., Bolker B., Walker S., 2015, *Fitting linear mixed-effects models using lme4*, Journal of Statistical Software, vol. 67, is. 1, pp. 1-48.

Bivand R., Lewin-Koh N., 2017, *Maptools: Tools for reading and handling spatial objects,* R package version 0.9-2, https://CRAN.R-project.org/package=maptools.

Bose A., Chakravarty S.R., D'Ambrosio C., 2014, *Richness orderings*, The Journal of Economic Inequality, vol. 12, is. 1, pp. 5-22.

Brady D., Fullerton A.S., Cross J.M., 2009, *Putting poverty in political context: a multi-level analysis of adult poverty across 18 affluent democracies*, Social Forces, vol. 88, no. 1, pp. 271-299.

Brzeziński M., 2010, *Income affluence in Poland*, Social Indicators Research, vol. 99, no. 2, pp. 285-299.

Council for Social Monitoring, 2016, *Integrated database*, http://www.diagnoza.com (11.09.2016).

Di Maggio F., Romanowski P., Walter C., 2003, *Eastern European banking matures*, The McKinsey Quarterly, no. 2.

Domański H., Pokropek A., 2011, *Podziały terytorialne, globalizacja a nierówności społeczne. Wprowadzenie do modeli wielopoziomowych*, Wydawnictwo IFiS PAN, Warszawa.

European Commission, 2010, *Algorithms to compute Social Inclusion Indicators based on EU-SILC and adopted under the Open Method of Coordination (OMC)*, Doc. LC-ILC/39/09/EN-rev.1, http://www.dst.dk/ext/747139308/0 (15.06.2018).

Evans M., Hastings N., Peacock B., 1993, *Statistical Distributions*, John Wiley and Sons, New York.

Franzini M., Granaglia E., Raitano M., 2016, *Extreme Inequalities in Contemporary Capitalism. Should We Be Concerned about the Rich?*, Springer, Switzerland.

Hutton R., 2006*, Rich spurn ultra-luxury cars*, The Sunday Times, November 5, 2006.

Kim K., Lee Y., Lee Y., 2010, *A multilevel analysis of factors related to poverty in welfare states*, Social Indicators Research, vol. 99, iss. 3, pp. 391-404.

Larsen K., Merlo J., 2005, *Appropriate assessment of neighbourhood effects on individual health: integrating random and fixed effects in multilevel logistic regression*, American Journal of Epidemiology, vol. 161, no. 1, pp. 81-88.

Larsen K., Petersen J.H., Budtz-Jorgensen E., Endahl L., 2000, *Interpreting parameters in the logistic regression model with random effects*, Biometrics, vol. 56, no. 3, pp. 909-914.

Lüdecke D., 2018a, *Sjmisc: Miscellaneous data management tools*, R package version 2.7.0, https://CRAN.R-project.org/package=sjmisc.

Lüdecke D., 2018b, *SjPlot: Data visualization for statistics in social science*, R package version 2.4.1, https://CRAN.R-project.org/package=sjPlot.

Moineddin R., Matheson F.I., Glazier R. II, 2007, *A simulation study of sample size for multilevel logistic regression models*, BMC Medical Research Methodology, vol. 7, no. 34, pp. 1-10.

Peichl A., Schaefer T., Scheicher C., 2010, *Measuring richness and poverty: a micro data application to Europe and Germany*, Review of Income and Wealth, vol. 56, is. 3, pp. 597-619.

Radziukiewicz M., 2006, *Zasięg ubóstwa w Polsce*, PWN, Warszawa.

Random effects testing, 2009, http://glmm.wikidot.com/random-effects-testing (15.06.2018).

Raudenbush S.W., Bryk A.S., 2002, *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, Thousand Oaks.

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

Rynek dóbr luksusowych w Polsce. Edycja 2017, 2017, KPMG, https://assets.kpmg.com/content/dam/kpmg/pl/pdf/2017/12/pl-Raport-KPMG-Rynek-dobr-luksusowych-w-Polsce-Edycja-2017.pdf (15.05.2018).

Sączewska-Piotrowska A., 2015, *Identyfikacja determinant bogactwa dochodowego z zastosowaniem modelu logitowego*, Zarządzanie i Finanse, vol. 13, no. 4/2, pp. 241-259.

Snijders T.A.B., Bosker R.J., 1999, *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modellin*g, Sage Publications, London.

Törmälehto V.-M., 2017, *High income and affluence: evidence from the European Union statistics on income and living conditions (EU-SILC)*, 2017 Edition, Statistical Working Papers, European Union.

Twisk J.W.R., 2006, *Applied Multilevel Analysis. A Practical Guide*, Cambridge University Press, New York.

Vorbeke G., Molenberghs G., 2000, *Linear Mixed Models for Longitudinal Data*, Springer, New York.

Wickham H., 2009, *Ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.

Zasobność gospodarstw domowych w Polsce. Raport z badania 2016 r., 2017, NBP, Departament Analiz Ekonomicznych i Departament Stabilności Finansowej, Warszawa, https://www.nbp.pl/aktualnosci/wiadomosci_2018/Raport_BZGD_2016.pdf (24.06.2018).

[www1] http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units (21.06.2017).

PODZIAŁ TERYTORIALNY A ZAMOŻNOŚĆ DOCHODOWA – ANALIZA Z WYKORZYSTANIEM DWUPOZIOMOWYCH MODELI LOGITOWYCH

Streszczenie: Celem pracy była ocena zróżnicowania występowania zamożności gospodarstw domowych w Polsce pomiędzy podregionami. Przeprowadzono analizę przy użyciu dwupoziomowych modeli logitowych bez zmiennych objaśniających (model pusty) i ze zmiennymi objaśniającymi na poziomie gospodarstwa domowego (model z losowym wyrazem wolnym i model z losowym nachyleniem). Zmienne odnosiły się do cech głowy gospodarstwa domowego, jak również do samego gospodarstwa. Przeprowadzona analiza pozwoliła stwierdzić, że występowanie zamożności jest zróżnicowane pomiędzy podregionami, zarówno w modelu pustym, jak i w modelu ze zmiennymi objaśniającymi.

Słowa kluczowe: dwupoziomowy model logitowy, modelowanie wielopoziomowe, zamożność, dochód, gospodarstwo domowe.