THE INFLUENCE OF SELECTED FACTORS ON THE AT-RISK-OF-POVERTY RATE OF SLOVAK HOUSEHOLDS

Helena Glaser-Opitzová[1]
Mária Vojtková[2]

DOI: https://doi.org/10.31410/ITEMA.S.P.2020.107

Abstract: *Since the goal of any advanced society is to reduce poverty and improve the social status of the population, it is important to know the causes of its emergence. In connection with Slovakia's membership in the European Union, we have taken over European legislation in this area. The Europe 2020 strategy is currently in force in the countries of the European Union, while one of its five main objectives is "Fight against poverty and social exclusion". Poverty research is undoubtedly a topical, multidimensional problem. One of the issues it focuses on is the so-called income poverty. The poverty line is considered to be 60% of the median national equivalent disposable household income. In order for assistance to those at risk to be truly targeted at those who need it most, it is necessary to map the situation in detail and identify the factors that have the greatest impact on the incidence of poverty. In our paper, the subject of analysis will be the quantification of the influence of selected factors from The European Union Statistics on Income and Living Conditions (EU-SILC) database on the at-risk-of-poverty rate in Slovak households. The at-risk-of-poverty rate represents the proportion of people (in percent) in the whole population, whose equivalent disposable income is below the at-risk-of-poverty line. We will verify the impact of selected factors on the at-risk-of-poverty rate using a logistic regression model in the SAS Enterprise Guide statistical tool.*

Keywords*: Income poverty, Risk of poverty rate, EU-SILC, Logistic regression.*

INTRODUCTION

Poverty reduction is one of the key challenges of the Europe 2020 strategy. By setting a poverty reduction target, the EU has put this social problem at the same level with economic goals. Achieving the goal of reducing the number of people at-risk-of-poverty and social exclusion will depend on the successful implementation of other priorities, such as providing better employment and education opportunities.

The main indicator “people at-risk-of-poverty or social exclusion” shows the number of people affected by at least one of the three forms of poverty: income poverty, material deprivation, and low work intensity. The most widespread form of poverty is income poverty which seems to be one of the main challenges for achieving the objectives of the Europe 2020 strategy. The proportion of people at-risk-of-income poverty is closely linked to income inequality. As stated in the synthesis report Peer Review in Social Protection and Social Inclusion 2011 (Kenneth, 2011, p. 30), reducing income inequality cannot be achieved simply by rising the average wage.

[1] University of Economics in Bratislava, Faculty of Economic and Informatics, Dolnozemská cesta 1, Slovakia
[2] University of Economics in Bratislava, Faculty of Economic and Informatics, Dolnozemská cesta 1, Slovakia

Social protection measures shall be taken as well as the efficiency and effectiveness of employment and income support shall be improved.

EU-SILC

In this article, there is income poverty as the subject of analysis. One of the sources for calculating indicators and measuring income poverty is data from EU-SILC. The harmonized statistical survey on household income and living conditions EU-SILC (*The European Union Statistic on Income and Living Conditions*) is carried out in all countries of the European Union which are currently in the number of 27. It has been implemented regularly in Slovakia since 2005. It is carried out by the Statistical Office of the Slovak Republic at annual intervals pursuant to a comparable international methodology within the project of European statistical surveys. The data are compared based on a uniform list of mandatory indicators together with their definitions, uniform procedures for the application of statistical methods, guidelines, rules, and calculations of poverty indicators. Such a comparison allows us to determine the social situation of the household in Slovakia, as well as an international comparison of Slovakia with the rest of the European Union and the countries participating in the survey.

The statistical unit of the survey is a private household and persons (i.e. current and former household members), the interconnection of data enables multidimensional analysis at the level of households and persons (The European Commission, 2008). The data obtained are recorded in four types of questionnaires - two questionnaires concern the household and the other two pertain to its members. The main areas of the survey at the level of households and individuals are listed in (Ivančíková, 2004).

One of the main indicators of income poverty is the **At-Risk-of-Poverty Rate** (ARPT60i), which represents the share of people (in percent) from the whole population whose equivalised disposable income is below the at-risk-of-poverty threshold (60% of the median of the yearly national equivalised disposable income) (Gerbery, 2011). *The equivalised disposable income of households* in the total disposable household income divided by the equivalent household size, which takes into account the size and composition of the household. Household disposable income is the sum of all monetary incomes received from any sources by each member of the household, including income from work, investments, and social benefits after deduction of taxes and social contributions paid. To calculate the equivalent size, there is used the so-called modified OECD scale.

The article aims to assess and quantify the statistical significance of the influence of the considered factors on the probability that the Slovak household will be at-risk-of-poverty through a logistic regression model based on data from the database of EU-SILC 2018. Factors whose impact we decided to verify are sex (RB090), marital status (PB190), household type (HT), the highest ISCED level attained (EDUCATION), the most frequent status of economic activity in the income reference period (PX050), general health (PH010), degree of urbanisation (DB100) and region according to NUTS 3(REGION). A list of them with variations of individual factors is given in the annex.

METHODOLOGY

We decided to assess the statistical significance of the considered factor influence on the probability that the household will be suffering by income poverty via logistic regression model

with the logit link function (Šoltés et al**.,** 2020; Hurbánková, 2018; Hilbe, 2016; Hosmer and Lemeshow, 2013; Bagley et al., 2001)**:**

$$logit(p_i) = ln\frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \quad (1)$$

where
p_i is the probability, that the household will suffer by the income poverty, i.e. $y_i = 1$, $\beta_0, \beta_1, \dots, \beta_k$ are parameters of the logit model and $x_{i1}, x_{i2}, \dots, x_{ik}$, where $i = 1, 2, \dots, n$, are the values of explanatory variables $X_1, X_2, \dots, X_k$ observed for the i-th statistical unit (in our case, the household). To estimate the parameters of the logistic regression model, we used the standard applied method of maximum likelihood, which maximizes the likelihood function.

We used three Chi-square tests to verify the significance of the model as a whole (Likelihood ratio, Score statistics, Wald statistics). These tests verify the validity of the null hypothesis $\boldsymbol{\beta}^T = (\beta_1 \ \beta_2 \dots \ \beta_k) = \mathbf{0}^T$ against alternative hypothesis, which states, that at least one regression coefficient is non-null one. For large samples, there is no reason to prefer any of these tests and they generally provide comparable results. (Allison, 2012).

The Wald test was used to verify the null hypothesis that explanatory variable does not affect the probability of the occurrence of the observed phenomenon. We verified the significance of the influence of the explanatory variables on the probability p in the SAS Enterprise Guide using test statistic:

$$Wald = \widehat{\boldsymbol{\beta}}^T . \boldsymbol{S_b^{-1}} . \widehat{\boldsymbol{\beta}} \quad (2)$$

where $\widehat{\boldsymbol{\beta}}$ is the vector of estimates of the regression coefficients standing for the artificial variables for the relevant factor – categorical explanatory variable and $\boldsymbol{S_b}$ is a covariance matrix of a vector $\widehat{\boldsymbol{\beta}}$. The Wald test statistic has asymptotically χ^2 distribution with degrees of freedom that are equal to the number of the estimated vector´s $\boldsymbol{\beta}$ parameters.

In logistic regression, the influence of the explanatory variable X_j on explained variable Y is quantified through odds ratio (OR), whose estimate is given by the relation $OR_j = e^{\widehat{\beta}_j}$ where $\hat{\beta}_j$ is the estimate of the relevant regression coefficient. The odds ratio in binary logistic regression expresses how is changed the odds that $Y = 1$ (in our case, the household will be at-risk-of-poverty) against the odds that $Y = 0$ (in our case that the household is not at-risk-of-poverty), for the unit increase of the explanatory variable under the condition of ceteris paribus. If the explanatory variable is the artificial variable, the odds ratio compares the odds of the event occurrence at two different levels of the predictor.

INFLUENCE OF SELECTED FACTORS ON INCOME POVERTY IN SLOVAKIA

We decided, at the beginning, to model the dependence between the probability, that the household is at-risk-of-poverty (the profile of the modelled variable is given in Table 1) and explanatory variables, using a complete logistic regression model, i.e. all explanatory variables – factors according to the Annex – entered the model.

As the number of households within some of the originally defined variations (categories) of individual factors was low, we adjusted the set compared to the original EU-SILC 2018 and

aggregated some categories. All explanatory variables except the sex variable are multinomial categorical variables that needed to be transformed into *s*-1 artificial variables, while the category for which no artificial variable was created is the so-called reference category (REF).

The standard method of maximum likelihood was used to estimate the model parameters. Based on the p-value (Table 2) which is lower than the commonly used significance level, the significance of the model as a whole was confirmed by three different tests: Likelihood ratio, Score and the Wald test. To identify statistically significant explanatory variables, we used the *Stepwise selection* method, which resulted in a reduced logistic regression model. As the most important variables, which entered the model for modelling the probability that the household is at-risk-of-income poverty, were: the economic activity of the head of household in the income period, the type of household, and the level of education attained by the head of household. Potential 8 factors were reduced to 6 by stepwise selection (Table 3). Their significance for the model can be assessed based on Wald or score chi-square statistics (Table 3). Sex and general health were identified as statistically insignificant factors. They did not enter the model.

Table 1. Profile of the modelled variable

Response Profile		
Ordered Value	**ARPT60i**	**Total Frequency**
1	0	4964
2	1	654

Probability modelled is ARPT60i='1'.

Table 2. Model significance tests

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	946.6306	27	<.0001
Score	1167.4497	27	<.0001
Wald	648.1629	27	<.0001

Table 3. Outcome of the Stepwise selection for the model ARPT60i

Summary of Stepwise Selection							
Step	**Effect**		**DF**	**Number**	**Score**	**Wald**	**Pr > ChiSq**
	Entered	**Removed**		**In**	**Chi-Square**	**Chi-Square**	
1	**PX050**		3	1	707.1888	251.6755	<.0001
2	**HT**		8	2	278.0643	221.0341	<.0001
3	**EDUCATION**		4	3	104.8842	99.5970	<.0001
4	**PB190**		3	4	76.3092	82.0489	<.0001
5	**REGION**		7	5	35.8572	22.0810	<.0001
6	**DB100**		2	6	8.5900	8.4300	0.0121

Source: EU-SILC 2018, own calculation in SAS EG

Interpretation of the outcomes of the logistic regression will be made under the estimation of the model parameters as well as the odds ratios (Table 4), while all interpretations are under the conditions of ceteris paribus (which we will not be mentioned further). Resulting from the estimated odds ratios, it can be concluded, that the influence of the factors - economic activity

and the highest level of education attained – is dominating. The odds, that the household will be below the at-risk-of-poverty threshold, are 16.187 times higher if the head of household is an unemployed person in comparison with the household where the head of household is an employed person. In case the household is headed by a person with less than secondary education, the odds, that the household will be below the at-risk-of-poverty threshold, are 5.986 times higher than in the case that the household is headed by a person with tertiary education of the 2nd or the 3rd stage.

Table 4. Estimates of parameters of binomial logistics models and estimates of odds ratios

Analysis of Maximum Likelihood Estimates				
PARAMETER		**Model ARPT60i**		
		BETA	**Pr > ChiSq**	**Odds ratio**
Intercept		-3.3443	<.0001	
PB190	**1**	0.4974	0.0023	1.644
PB190	**3**	-0.7337	<.0001	0.480
PB190	**4**	0.7470	<.0001	2.111
EDUCATION	**1**	1.94	<.0001	5.986
EDUCATION	**2**	0.6551	<.0001	1.925
EDUCATION	**3**	0.3205	0.3883	1.378
EDUCATION	**4**	0.3054	0.4545	1.357
PX050	**2**	2.42	<.0001	16.187
PX050	**3**	0.5403	0.0001	1.716
PX050	**4**	1.97	<.0001	5.583
HT	**1**	-0.4479	0.0321	0.639
HT	**2**	-1.2760	<.0001	0.279
HT	**3**	-2.2351	<.0001	0.107
HT	**4**	-2.0379	<.0001	0.130
HT	**5**	0.6167	0.0221	1.853
HT	**6**	-0.5086	0.0183	0.601
HT	**8**	0.8589	0.0010	2.360
HT	**9**	-0.9668	<.0001	0.380
DB100	**2**	0.2906	0.0610	1.337
DB100	**3**	0.4537	0.0035	1.574
REGION	**2**	0.2095	0.3842	1.233
REGION	**3**	0.4320	0.0615	1.540
REGION	**4**	0.4895	0.0284	1.632
REGION	**5**	0.5786	0.0089	1.784
REGION	**6**	0.6638	0.0025	1.942
REGION	**7**	0.8691	<.0001	2.385
REGION	**8**	0.5648	0.0086	1.759

Source: EU-SILC 2018, own calculation in SAS EG

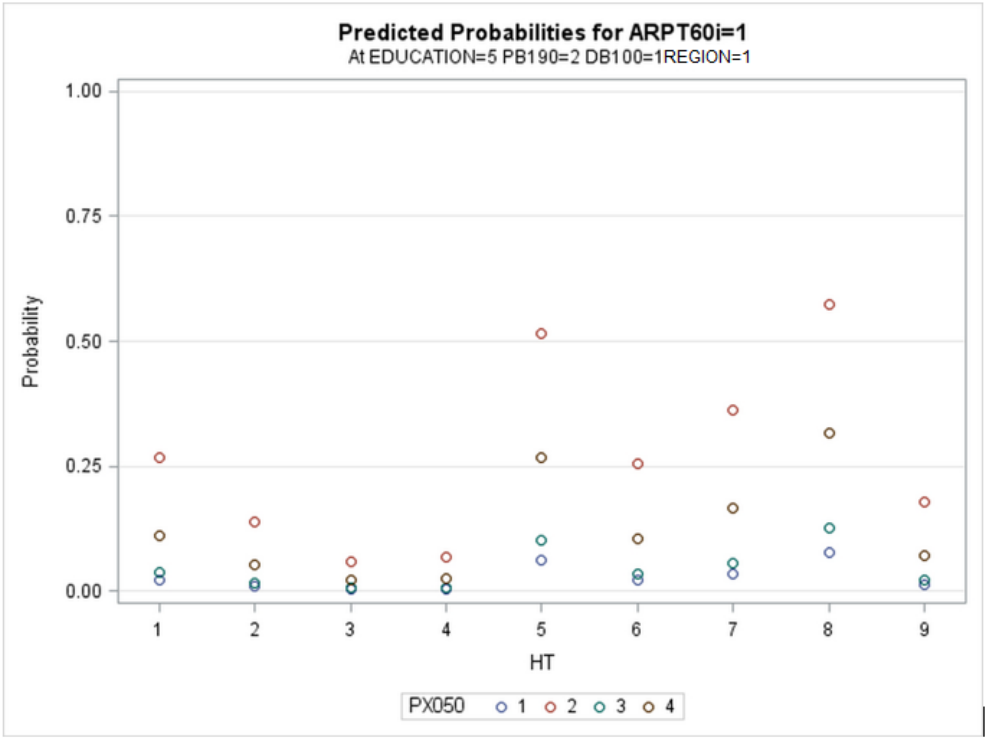
In terms of household composition, the worst situation is in families with two adults and more than two children where the odds of the risk-of-poverty are 2.36 times higher than in the case of a household consisting of 2 adults and 2 children. However, in a household of 2 adults without dependent children where one adult is over 65 years old, the odds that such a household will be poor are 9.345 times lower than the household of 2 adults with 2 children. According

to marital status, the household which is the most at-risk-of-poverty is that one where the head of household is divorced (odds are 2.111 times higher) compared to the household with the married head of household. On the contrary, they are lower if the household is headed by a widow/widower. From the regional point of view, the worst situation is in Prešov region (the odds that the household will be at-risk-of-poverty are 2.385 times higher than in the Bratislava region), which includes several districts with fewer job opportunities and subsequently low financial recognition in the labour market. In general, however, in all regions outside the Bratislava region (where the capital city is located), households face a higher risk-of-poverty. The odds that households located in thinly-populated areas are 1.574 times higher to be at-risk-of-poverty than households in densely populated areas.

CONCLUSION

Resulting from the estimated logistic model, the Slovak household which is the most at-risk-of-poverty is found in Prešov region, in a thinly populated area, headed by an unemployed, divorced person with less than secondary education, whose composition consists of 2 adults with three or more dependent children. An interesting fact is that the influence of the sex of the head of household, as well as his/her health condition, proved to be insignificant in quantifying the risk-of-poverty.

Figure 1. Estimates of the probability of the risk of poverty depending on the composition of the household and the status of economic activity of the head of the household for the reference categories of other explanatory variables



Source: EU-SILC 2018, own calculation in SAS EG

For two statistically most significant factors - the most frequent status of economic activity and household composition (Figure 1), the probability that the Slovak household will face the risk-

of-poverty was predicted based on the estimated model. Other factors remained at the level of the reference categories, i.e. we considered a household living in the region of Bratislava, in a densely populated area, headed by a married person with tertiary education of the second or third stage. If the person is unemployed, the household is significantly more likely to face the risk-of-poverty compared to households headed by the employed head of household. The most vulnerable group of households with the probability of the risk of poverty of almost 60% is that one which is composed of two adults and three or more dependent children where the head of household is the person, who although attained tertiary education of the 2nd or 3rd stage, is unemployed. By contrast, the least at-risk groups are the households of two adults without dependent children, where at least one person is over the age of 65. For these households, it can be assumed that the risk of income poverty is significantly reduced by receiving old-age pensions.

REFERENCES

Allison, P. D. (2012). *Logistic Regression using SAS. Theory and Application* (Second ed.). North Carolina, USA: SAS Institute.

Bagley, SC., White, H., & Golomb, BA. (2001). Logistic Regression in the Medical Literature: Standards for Use and Reporting, with Particular Attention to One Medical Domain. *Journal of Clinical Epidemiology*, Vol 54, pp. 979-985.

Hilbe, J.M. (2016). *Practical Guide to Logistic Regression*. Boca Raton: Chapman & Hall/CRC.

Hosmer, D.W., & Lemeshow, S. (2013). *Applied Logistic Regression*. New York, USA: John Wiley & Sons.

Hurbánková, Ľ. (2018). Analýza vybraných ukazovateľov charakterizujúcich vylúčenie z trhu práce. In *Building of society based on knowledge*. Paper presented at 18th International Scientific Conference AIESA 2018. University of Economics in Bratislava, October 25-26, 2018 (s. 253-266). Bratislava: Letra Edu.

Gerbery, D. (2011). *Pretrvávajúca chudoba-analýza longitudinálnej databázy EU SILC*. Bratislava, Slovakia: Inštitút pre výskum práce a rodiny. [cit. 30th September 2020]. Available at: https://www.ceit.sk/IVPR/images/IVPR/vyskum/2011/Gerbery/gerbery_2251.pdf

Ivančíková, Ľ. (2004). Zisťovanie o príjmoch a životných podmienkach (EU SILC). In *Zborník. Otázky merania chudoby*. Bratislava: Friedrich Ebert Stiftung, e. V., representation in the Slovak republic, 2004, s 44-51. [cit. 30th September 2020]. Available at: http://library.fes.de/pdf-files/bueros/slowakei/04223.pdf

Kenneth, N. (2011). *Improving the efficiency of social protection*. Lisbon: Swedish Institute for social research.

Šoltés, E. Táborecká-Petrovičová, J., & Šipoldová, R. (2020) Targeting of Online Advertising Using Logistic Regression. *E + M. Ekonomie a management: vědecký ekonomický časopis*. - Liberec: The Faculty of Economics of the Technical University of Liberec, roč. 23, č. 4. (Vol 23, No 4).

The European Commission (2008). *Child poverty and well-being in the EU: Current status and way forward*. The Social Protection Committee. Luxembourg: Office for Official Publications of the European Communities.

Annex: Input explanatory variables

Variable	Target variable name	Values		Note
RB090	Sex	1		Man
		2		Woman
PB190	Marital status	1		Single
		2	REF	Married
		3		Widower / widow
		4		Divorced
EDUCATION	Highest level of education attained (according to ISCED)	1		Less_than_Secondary
		2		Upper_Secondary
		3		Post_Secondary
		4		Tertiary_1
		5	REF	Tertiary_2_3
PH010_a	General health	1	REF	Good
		2		Fair
		3		Bad
PX050	The most frequent economic activity status in the income reference period	1	REF	Employed person
		2		Unemployed person
		3		Old-age retiree, early retiree
		4		Other inactive person
HT	Household type	1		One person household
		2		2 adults no dependent children,
				- both adults under 65 years
		3		2 adults, no dependent children
				- at least one adult 65 years or more
		4		Other households without dependent children
		5		Single-parent household, one or more dependent children
		6		2 adults, one dependent child
		7	REF	2 adults, two dependent children
		8		2 adults, three or more dependent children
		9		Other households with dependent children
DB100	Degree of urbanization	1	REF	Densely-populated area
		2		Intermediate urbanised area
		3		Thinly-populated area
REGION	Region according to NUTS 3	1	REF	Bratislava
		2		Trnava
		3		Trenčín

		4		Nitra
		5		Žilina
		6		Banská Bystrica
		7		Prešov
		8		Košice
ARPT60i	Below the at-risk-of poverty threshold	1		Yes
		0		No

Source: EU-SILC 2018, own calculation in SAS EG