# Quantifying the impact of relevant factors on equivalised disposable income in Slovak households based on EU SILC 2012 data

Erik Šoltés[1], Tatiana Šoltésová[2]

**Abstract**

In this article we analyse equivalised disposable household income defined for EU SILC (The European Union Statistics on Income and Living Conditions) survey. The main goal is to determine relevant factors affecting the equivalised disposable income in Slovak households and to quantify their influences. The article provides results that were gained from the sample statistical survey EU SILC realized by the Statistical Office of Slovak Republic in 2012. The main tools of the statistical analysis are multivariate analysis of variance, tests for the equality of marginal means, regression analysis, partial correlation and general linear models within GLM Procedure in SAS Enterprise Guide. The impact of relevant factors selected by means of the Backward elimination method is quantified through regression coefficients. Due to occurrence of heteroskedasticity we used the weighted least squares method to estimate appropriate regression model. Strength of relationship between the equivalised disposable income and individual independent variables is judged by semipartial correlation.

***Keywords:*** *EU SILC − European Union statistics on income and living conditions, equivalised disposable household income, marginal mean, general linear models*

***JEL Classification:*** C51; C52; R29

## 1. Introduction

The presented analysis is based on EU SILC 2012 data that were collected in 5 291 Slovak households. Period of time analysed (income reference period) is calendar year preceding the year of the survey, i.e. for EU SILC 2012 income reference period was calendar year 2011. EU SILC provides comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions. There are several scientific works that focus on analysis of income and poverty in the Slovak republic. Among the most important belong (Bartošová & Forbelská, 2010), (Labudová, Vojtková & Linda, 2010), (Labudová & Vojtková, 2010), (Sipková & Sipko, 2012) and (Tartaľová, 2012) that were inspiration to write this article.

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of statistics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, soltes.euba@gmail.com

[2] University of Economics in Bratislava, Faculty of Economic Informatics, Department of mathematics and actuary science, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, soltesova.euba@gmail.com

Our main goals were to determine factors that have significant impact on the equivalised disposable income of Slovak households and to quantify their influence on the income. We used the multivariate GLM through SAS Enterprise guide [see (Garson, 2012) and (Šoltés, 2008)] to reach these goals.

The main tools used in the analysis were general linear models where the dependent variable was *Equivalised disposable income of Slovak households (in € per month).* We will use abbreviation "EDI" in the following text. These explanatory variables were considered at the beginning of our analysis: *Activity status* (PX050), *Household type* (HT), *Education* (PE040), *Managerial position* (PL150), *Urbanisation* (DB100), *Region, Health* (PH010), *Marital status* (PB190), *Dwelling type* (HH010), *Tenure status* (HH020) and *Age* (age at the end of income reference period). Except variables *Household type, Urbanisation, Region,* all the variables are related only towards the household head.

## 2. Selection of relevant factors and database adaptation

To select factors that have a relevant influence on *Equivalised disposable income of Slovak households* the Backward elimination method in the procedure of linear regression in SAS Enterprise Guide was used. The selection of explanatory variables was as follows: *Activity status, Household type, Education, Managerial position, Urbanisation, Region, Health* and *Marital status.* In 2011, each of these variables had significant impact on equivalised disposable Slovak household's income at the usual significance level.

We quantified an impact of particular categorical variables on the EDI by marginal means (LS means). We do not use the "classic" arithmetic mean because the arithmetic mean for a group may not accurately reflect the "typical" response for that group, since it does not take other effects into account as opposed to the LS mean. Since the EDI is influenced by more than one effect as we showed in previous analysis, it is better to use LS means to compare the EDI for various categories of a categorical variable.

According to the test of equality of LS means of the EDI, we determined statistically significant differences of LS means between pairs of categories for each categorical variable. To explain our next procedure we will briefly address issues of regional differences in the EDI in the Slovak republic. We found out that there were only significant differences between Bratislava region and any other Slovak region and between Presov region and any other Slovak region. There were not relevant differences in the EDI between another pairs of regions (Fig. 1).

On the basis of above stated results, we merged all the regions except Bratislava region and Presov region. New categories of the variable *Region* were created this way. Point

estimates of LS means and *p*-values related to test of equality of LS means of Slovak households' EDI for new categories of the variable *Region* are in Table 1.

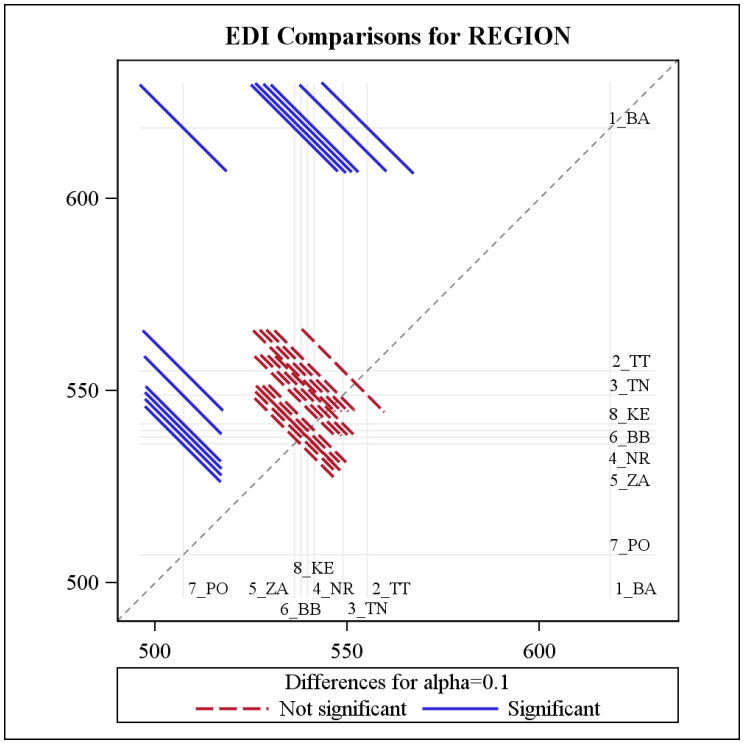


**Fig. 1.** Confidence intervals for LS means of Slovak households' EDI in dependency on Region.

*Source:* EU SILC 2012, authors' own elaboration within SAS Enterprise Guide

**Least Squares Means for effect REGION Dependent Variable: EDI**

**Pr > |t| for H0: LSMean(i)=LSMean(j)**

| **i/j** | | **Bratislava** | **Other** | **Presov** |
|---|---|---|---|---|
| | **point estimate** | **616.12** | **560.45** | **523.65** |
| **Bratislava** | **616.12** | | <.0001 | <.0001 |
| **Other** | **560.45** | <.0001 | | <.0001 |
| **Presov** | **523.65** | <.0001 | <.0001 | |

**Table 1** Matrix of *p*-values related to test of equality of LS means of Slovak households' EDI for new Region.

*Source:* EU SILC 2012, authors' own elaboration within SAS Enterprise Guide

We found out that the highest EDI in 2011 had households from Bratislava region. The lowest EDI had households from Presov region. Their monthly EDI was in average by around 92€ lower than in Bratislava region, assuming the same categories of other relevant factors that

are stated in Table 1. Table 1 contains all the relevant factors (except the variable Region) that were determined with previous analysis and their original categories as well as new categories that were created by the similar way as new categories of the variable Region. We must mention that highlighted categories are reference categories to which we will compare other categories of corresponding categorical variable. We tried to choose a typical or the most frequent category for each variable so that a number of observations was appropriate to get relevant results.

| Variable | Variable name | Original categories | New categories |
|---|---|---|---|
| PX050 | ACTIVITY STATUS (the most frequent status of economic activity) | At work | **AS_at_Work** |
| | | Unemployed | AS_Unemployed |
| | | Retired | AS_Retired |
| | | Other inactive persons | AS_Inactive_person |
| HT | HOUSEHOLD TYPE[3] | 5 | HT_1Adult |
| | | 6 | HT_2A_0Ch |
| | | 7 | HT_2A_1Retired |
| | | 8 | HT_Other_0Ch |
| | | 9 | HT_1A_at_least_1Ch |
| | | 10 | HT_2A_1Ch |
| | | 11 | **HT_2A_2Ch** |
| | | 12 | HT_2A_at_least_3Ch |
| | | 13 | HT_Other_with_Ch |
| PE040 | EDUCATION (highest ISCED level attained) | Pre-primary education | – |
| | | Primary education | EDU_Primary |
| | | Lower secondary education | |
| | | Upper secondary education | **EDU_Secondary** |
| | | Post-secondary education, | |
| | | 1th stage of tertiary education | EDU_Tertiary1 |
| | | 2 nd stage of tertiary education | EDU_Tertiary2 |
| PL150 | MANAGERIAL POSITION | Supervisory | SUPERV_Yes |
| | | Non-supervisory | **SUPERV_No** |

[3] „5“ − One person household; „6“− 2 adults, no dependent children, both adults under 65 years; „7“− 2 adults, no dependent children, at least one adult 65 years or more; „8“− Other households without dependent children; „9“ − Single parent household, one or more dependent children; „10“ − 2 adults, one dependent child; „11“ – 2 adults, two dependent children; „12“ − 2 adults, three or more dependent children; „13“ − Other households with dependent children.

| | | | |
|---|---|---|---|
| DB100 | URBANISATION (degree of urbanisation) | Densely populated area | URBAN_Dense |
| | | Intermediate area | **URBAN_Intermediate** |
| | | Thinly populated area | URBAN_Sparse |
| PH010 | HEALTH (general health) | Very good | HEALTH_Very_good |
| | | Good | HEALTH_Good |
| | | Fair, Bad, Very bad | **HEALTH_Other** |
| PB190 | MARITAL STATUS | Never married, Married, Separated, Widowed | **MS_Other** |
| | | Divorced | MS_Divorced |

**Table 2** Original and new categories of explanatory variables.

*Source:* EU SILC 2012, authors' own elaboration

The next analysis showed that there are some influenced observations that were detected using Cook's statistics. On the basis of the analysis there were detected and subsequently deleted 58 observations that accounted about 1% of the whole data set.

## 3. The model estimated by the weighted least squares method

Since the error terms were not homoskedastic (as it is obvious from studentized residuals displayed in Fig. 2a) we used feasible GLS procedure [see (Hebák, Hustopecký & Malá, 2005) and (Wooldridge, 2009)] to calculate weights. It seems that calculated weights were suitable since the studentized residuals did not exhibit heteroskedasticity again (Fig. 2b). We found out that the error terms associated with the model estimated by the weighted least squares method fulfils all the assumptions to use them for our purposes.

a) b)

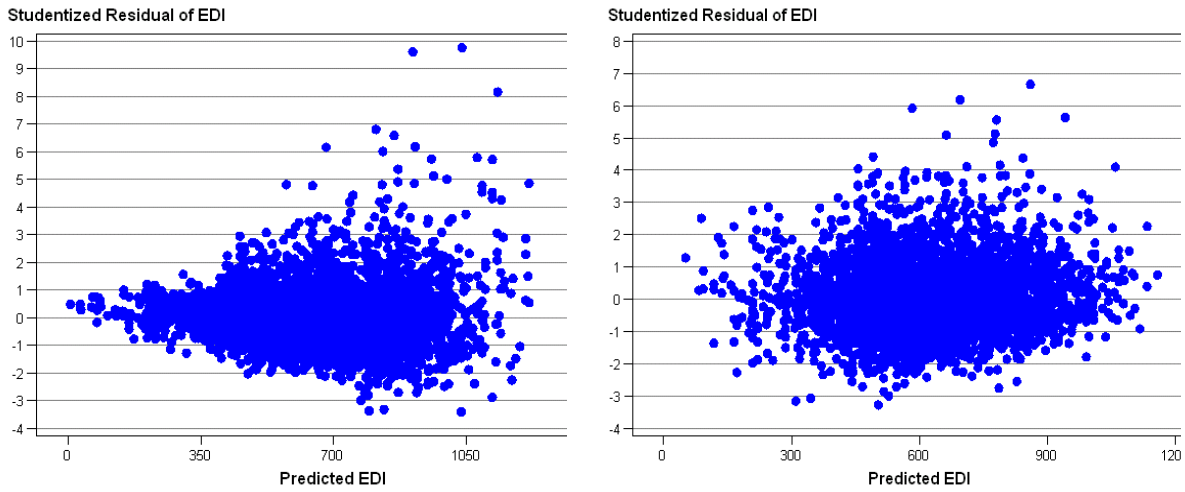


**Fig. 2.** Studentized residuals for the regression model estimated by a) OLS b) WLS.

*Source:* EU SILC 2012, authors'own elaboration within SAS Enterprise Guide.

Concerning determination of strength dependency among explanatory variables we computed the Condition index and the Variation inflation factor (*VIF* in Table 3). The highest value of the condition index (11.157) pointed out a weak degree of multicollinearity. We can make similar conclusion on basis of fairly low values of *VIF* (less than 5).

The EDI has the closest relationship with the household head's economic activity. The variable *Activity status* explains more than 26.5% variability of the EDI (see *Squared Semi-partial Corr Type I* in Table 3), followed by variables: household type, the highest level of education attained and the managerial position of the household head that explain 14.3%, 3.3% and 2.56% in the EDI variation respectively. Each of other variables (*Urbanisation*, *Region*, *Health* and *Marital status*) explains less than 1% in the EDI variation but each of them has significant impact on the EDI.

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 25 | 20 081.00 | 803.22 | 182.39 | <.0001 |
| **Error** | 5 192 | 22 864.00 | 4.40 | | |
| **Corrected Total** | 5 217 | 42 945.00 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 2.0985 | **R-Square** | 0.4676 |
| **Dependent Mean** | 503.8999 | **Adj R-Sq** | 0.4650 |

| Variable | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Squared Semi-partial Corr Type I | VIF |
|---|---|---|---|---|---|---|
| **Intercept** | 560.36 | 10.26 | 54.62 | <.0001 | . | 0 |
| AS_Retired | -142.99 | 7.87 | -18.18 | <.0001 | 0.005 | 2.66 |
| AS_Inactive_person | -203.88 | 9.66 | -21.10 | <.0001 | 0.023 | 1.31 |
| AS_Unemployed | -349.54 | 8.76 | -39.91 | <.0001 | 0.218 | 1.29 |
| **AS_at_Work** | | | | | | |
| HT_Other_0Ch | 245.45 | 11.78 | 20.83 | <.0001 | 0.066 | 1.88 |
| HT_2A_0Ch | 162.84 | 12.81 | 12.71 | <.0001 | 0.020 | 1.64 |
| HT_Other_with_Ch | 111.65 | 10.48 | 10.65 | <.0001 | 0.021 | 2.12 |

| | | | | | | |
|---|---|---|---|---|---|---|
| HT_2A_1Retired | 84.94 | 12.24 | 6.94 | <.0001 | 0.013 | 2.83 |
| HT_2A_1Ch | 73.78 | 12.17 | 6.06 | <.0001 | 0.010 | 1.59 |
| HT_1Adult | 10.04 | 10.49 | 0.96 | 0.3388 | 0.004 | 4.21 |
| HT_1A_at_least_1Ch | -61.91 | 14.49 | -4.27 | <.0001 | 0.001 | 1.63 |
| HT_2A_at_least_3Ch | -115.63 | 12.47 | -9.28 | <.0001 | 0.009 | 1.53 |
| **HT_2A_2Ch** | | | | | | |
| EDU_Primary | -38.46 | 21.72 | -1.77 | 0.0767 | 0.003 | 1.03 |
| EDU_Tertiary1 | 127.72 | 9.17 | 13.93 | <.0001 | 0.027 | 1.09 |
| EDU_Tertiary2 | 259.17 | 66.19 | 3.92 | <.0001 | 0.003 | 1.01 |
| EDU_NotSpec | 21.16 | 31.62 | 0.67 | 0.5035 | $2 \cdot 10^{-5}$ | 1.01 |
| **EDU_Secondary** | | | | | | |
| SUPERV_Yes | 69.82 | 8.56 | 8.16 | <.0001 | 0.011 | 1.07 |
| SUPERV_NotSpec | -91.80 | 7.53 | -12.20 | <.0001 | 0.015 | 1.10 |
| **SUPERV_No** | | | | | | |
| URBAN_Dense | 39.49 | 7.49 | 5.27 | <.0001 | 0.006 | 1.35 |
| URBAN_Sparse | -12.42 | 5.54 | -2.24 | 0.0251 | 0.001 | 1.33 |
| **URBAN_Intermed.** | | | | | | |
| REGION_BA | 55.65 | 11.22 | 4.96 | <.0001 | 0.003 | 1.07 |
| REGION_Presov | -36.51 | 6.53 | -5.59 | <.0001 | 0.003 | 1.03 |
| **REGION_Other** | | | | | | |
| HEALTH_Very_good | 51.15 | 11.50 | 4.45 | <.0001 | 0.001 | 1.20 |
| HEALTH_Good | 35.17 | 6.00 | 5.86 | <.0001 | 0.004 | 1.40 |
| HEALTH_NotSpec | 273.24 | 457.20 | 0.60 | 0.5501 | $4 \cdot 10^{-5}$ | 1.00 |
| **HEALTH_Other** | | | | | | |
| MS_Divorced | -43.43 | 7.48 | -5.80 | <.0001 | 0.002 | 1.22 |
| **MS_Other** | | | | | | |

**Table 3** The regression model estimates by weighted least squares method.

Source: EU SILC 2012, authors' own elaboration within SAS Enterprise Guide

If we do not take into account categories *Not specified* there is only one non-significant regression coefficient (significance level 0.1) in the regression model and i.e. the regression coefficient for the category "household with one adult". That means there is no significant difference in the EDI between households with one adult and households with 2 adults and 2 dependent children. If we consider the significance level 0.05 there is one more non-

significant regression coefficient: *Education_primary*. According to this result, the difference in the EDI between households where the household head has primary vs. secondary education is not relevant (at stated significance level).

Intercept (560.36€) is an average EDI of households that belong to the reference category for each categorical variable, i.e. households with 2 adults and 2 dependent children that did not live in neither Bratislava region nor Presov region and lived in intermediate populated area, while their secondary educated household head worked in non-supervisory position, his/her general health was fair or worse and he/she was not divorced.

Estimates of regression coefficients brought mainly following findings:
Regarding the activity status of the household head, the highest EDI had households where the household head was employed and the lowest EDI with unemployed household head. The difference in the EDI between such households was about 350€. Differences between households with employed head vs. households where the head has any other activity status are quantified through corresponding regression coefficient in Table 5.

If we focus on household type, there were only two types of households that had worse income situation than households with 2 adults and 2 dependent children. Such households were single parent households with one or more dependent children and households with 2 adults and 3 or more dependent children. The highest EDI had households without children. Generally we can say that EDI is positively dependent on adults vs. children/retired ratio.

It is natural that higher attained education of the household head has positive impact on income. If we take into account the significance level 0.1 there is statistically significant difference between EDI of households with primary and households with secondary educated household heads. The former had the EDI lower than the latter by around 38€. The highest EDI had households where the head attained tertiary education that directly leads to an advanced research qualification (tertiary 2). Their EDI was in average by nearly 260€ higher than the EDI of households where the head was secondary educated.

Concerning the place where the household lived we considered two factors: degree of urbanisation and region (NUTS2) at the beginning of our analysis and both showed as significant. The highest EDI had households which lived in densely populated areas in Bratislava region whereas the lowest EDI had households which lived in thinly populated areas in Presov region. Differences between these households vs. reference categories i.e. intermediate area for factor Degree of urbanisation and other region for factor Region, are quantified by means of regression coefficients in Table 3. We would like to notify that we

gained the same differences between regions using the regression analysis and LS means stated in Table 1.

Similarly we can interpret other estimates of regression coefficients. Let us remind that above mentioned outcomes are valid under the condition of constancy of values of other explanatory variables.

## Conclusion

On the basis of EU SILC 2012 data using the Backward elimination method we discovered that the equivalised disposable income of Slovak households in 2011 was significantly dependent on household type, region and degree of urbanisation of the place where the household lives and factors related to the household head: economic activity status, educational attainment, managerial position, health and marital status. We constructed adequacy regression model by means of which we could estimate partial contribution of individual factors towards the explanation of variability in the EDI. We quantified the impact of individual relevant factors through marginal means and regression coefficients. We found out that the highest EDI had households without children living in Bratislava region in densely populated area with the household head who works in supervisory position, has tertiary education, is not divorced and furthermore he/she classifies his/her health as very good. On the other hand, the lowest EDI had households with 2 adults and 3 or more dependent children living in Presov region in thinly populated area with unemployed household head who has primary education, simultaneously he/she is divorced and classifies his/her health as fair or worse.

## References

Bartošová, J., & Forbelská, M. (2010, October). *Comparison of regional monetary poverty in the Czech and Slovak republic*. Conference on Social Capital, Human Capital and Poverty in the Regions of Slovakia, Herľany, Slovakia. Retrieved from books.google.sk/books ?isbn=8055305730.

Garson, G. D. (2012). *General linear models: Multivariate GLM, MANOVA and MANCOVA*. Asheboro, N. C.: Statistical Associates Publishers.

Hebák, P., Hustopecký, J., & Malá, I. (2005). *Vícerozměrné statistické metody (2)*. Praha: Informatorium.

Labudová, V., Vojtková, M., & Linda, B. (2010). Application of multidimensional methods to measure poverty. *E & M Ekonomie a Management* , *13*(1), 6-22.

Labudová, V., & Vojtková, M. (2010). Regional analysis of expenditure and income of households in the slovak republic. *Journal of Economics, 58*(8), 802-820.

Sipková, Ļ, & Sipko, J. (2012, September). *Analysis of income inequality of employees in the Slovak republic*. 6th International Days of Statistics and Economics, Prague, Czech Republic. Retrieved from http://msed.vse.cz/files/2012/Sipkova_2012.pdf

Šoltés, E. (2008). *Regresná a korelačná analýza s aplikáciami*. Bratislava: Iura Edition.

Tartaľová, A. (2012, September). *Modelling Income Distribution in Slovakia*. 6th International Days of Statistics and Economics, Prague, Czech Republic. Retrieved from http://msed.vse.cz/files/2012/Tartalova_2012.pdf.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Mason: South-Western Cengage Learning.