Bachelor's Studies

Field: Quantitative Methods in Economics and Information Systems

Kacper Kalinowski
Student No. 76975

Determinants of subjective financial situation of Polish households

Bachelor's thesis
written in Institute of Econometrics
under scientific supervision of
dr hab. Bartosz Witkowski, prof. SGH

Warsaw 2020

Table of contents

Introduction

In recent decades Polish economy has undergone fundamental transformation and its growth has been very dynamic. The same goes for Polish households. Their financial situation resolutely improved from year to year.

The objects of this study were Polish households and their members who took part in the social diagnosis research (Diagnoza społeczna, 2015). The data which was under analysis was collected from households in 2015, which is the newest edition of social diagnosis project. The subject of study was subjective financial situation of Polish households, which was evaluated on the basis of results obtained from respondents.

The goal of this thesis was to set down econometric model which may allow to evaluate particular socio-demographic factors influencing subjective financial situation and to make effective classifications of households by their perception.

In the first chapter appears the review of literature, which concerns subjective and objective indicators and formulation of research hypotheses. The second chapter contains description of applied methods which enable to define determinants of subjective financial situation and to make sufficient predictions of household condition. The next chapter involves empirical analysis which contains the description of the data, justification for applying particular variables, validation of logistic regression estimations, analysis of determinants of subjective financial situation, cut-off point selection, model evaluation and remarks which arose from the analysis. In the last chapter conclusions, discussion and also recommendations for future research can be found.

1. Subjective financial condition of household and its determinants as a research problem

Financial situation of household is an important factor influencing consumption decisions, but also is an object of social and policy interests. Traditional approach for assessing financial condition is exploration of objective measures. However objective indicators do not fully describe the family position. Subjective opinions of household individuals potentially impact purchasing process. What is interesting that in some situations evaluation of objective financial situation might be associated with completely different assessment of subjective condition (Białowolski, Weziak-Białowolska, 2013, p.366). That is why exploring only objective and forgetting about subjective measures is insufficient.

1.1. Literature review

In recent years much interest in literature was dedicated to mentioned indicators. Mainly researchers focused on determinants of subjective and objective financial condition, some of them studied changes over time.

Analysis of income which was conducted by authors of Social Diagnosis in 2015 showed that in recent years financial situation of Polish households improved. At the same time economic stratification was reduced what means that income inequalities between households decreased. In regards to subjective financial situation, in the last four years percentage of respondents who declared that their household make ends meet with great trouble or with difficulty significantly decreased (Diagnoza społeczna, 2015).

There are few studies which focus on determining factors influencing subjective financial situation of Polish households. In two papers researchers investigated financial condition by using households' responses from Social Diagnosis study concerning question "Do your household make end meets end with your actual income?". Influence of particular socio-demographic features on those responses was examined. Researchers claimed that socio-economic group of household, place of living and biological type had significant impact on subjective financial situation (Dudek, 2015; Genge, Trzęsiok, 2017).

Similar type of data was used in research where author explored responses from a question concerning evaluation of financial condition of household. There were five possible answers to mark: budget very tight, tight, modest, fair, good, very good. Data containing several socio-economic features was also collected. Author found that factors like disposable income, possesion of savings, number of individuals in family and satisfaction with work and level of

life significantly impacts respondents' perception of financial situation (Kasprzyk, 2016, p. 249).

Researchers also made use of Household Budget Surveys for Poland in 2008 conducted by the Central Statistical Office. The question from Polish Household Budget Survey "What level of income would be sufficient for satisfying household needs?" was used. Then the results were compared with observed level of income and new ordinal variable called income perception was constructed. Actual income was ranked into classes of subjective income perception. When current income was higher than sufficient income which would satisfy household needs then household was classified into group with good subjective financial situation (bad in opposite case). In this research socio-economic group, biological type and place of residence were also found to have significant influence on subjective financial situation. Moreover gender and education of head of the family were also claimed to be factors which impacted financial condition of houesholds (Liberda, Pęczkowski, Gucwa-Lesny, 2012).

However there are also papers which focus on objective measures evaluating financial situation of households (Kozera, Wysocki, 2014; Panek, Czapiński, 2015; Kozera, Stanisławska, 2014; Wereda, Prokopowicz, 2017; Bialowolski, Weziak-Bialowolska, 2013; Sączewska-Piotrowska, 2015). Due to the fact that correlation between subjective and objective measures appears (Gąsiorowska, 2010, p. 177), conclusions from studies concerning objective condition are similar to outcomes from those which focus on subjective situation. However this effect is not strong so results from papers which concerned objective situation do not have to be the same and it has to be taken into consideration.

In most cases researchers tried to anwer a question how socio-demographic features influence level of disposable income of households. Results are very comparable to those which were mentioned above. Socio-economic group, biological type, place of residence of household, eduaction and age of head of family and number of individuals in household were found to be drivers of objective financial situation.

In order to evaluate determinants of financial situation of households researchers most frequently applied logistic regression. In some of those papers mentioned algorithm was validated and it gave satysfing results of estimations.

Some of the researchers also claimed that subjective financial situation is a phenomenon which is not fully measurable (Kasprzyk, 2016, p. 238). There are many psychological, social and economic factors that impact this perception. One of this variables influencing subjective

financial condition is “reference group”. Individuals consciously or unconsciously compare their situation to people from immediate vicinity (van Praag 1968; Kapteyn 1977). This idea is also broadly discussed in other papers. It was shown that peoples’ happiness depends on what they observe around them (D’Ambrosio, Frick, 2007, p. 497–519).

1.2. Research hypotheses

The essence of research process is verification of hypothesis. For this reason the following theses were stated: there are specific features which differ households with regard to subjective financial situation. The second hypothesis states that logistic regression is an effective algorithm which determines those features and also gives an opportunity to predict whether given household is in bad or good financial situation.

The data used to analyze subjective financial situation and to verify stated hypotheses was taken from Social Diagnosis (Diagnoza społeczna, 2015) research. The project contains indicators associated with living conditions and quality of dwellers’ life in their households. It is also frequently cited in research papers concerning subjective and objective financial indicators. Authors of Social Diagnosis examined Polish households and all their members aged 16 and above with the aid of two separate questionnaires.

Logistic regression was choosen to be a research method and to verify stated hypotheses. It is widely used algorithm applied to examine and describe relationship between binary response variable and set of predictors. Logistic regression has few main objectives: defining significance of determinants, their strength and direction of influence of predictors on dependent variable and predicting the outcome based on particular independent variables.

Assesment of subjective financial condition of household and its determinants gives an information about strength and direction of relationship between independent variable and particular socio-demographic characteristics. By applying logistic regression algorithm there is also possibility to classify households into groups with different perception of financial condition.

Knowledge of factors that impact perception of financial situation is important and interesting due to the fact that perception of family budget is one of the main drivers of consumption. Thanks to the awareness of determinants marketing specialists can better customize their strategies to better pick their audience target.

Moreover subjective financial situation is an object of policy interests. Thanks to the verification of stated hypotheses government can better accommodate their policies to the given conditions. Learning about results of this research can also lead to better evaluation of particular household situation.

Social interest aspect is also important. People want to know which factors impact their financial perception. Taking into account that correlation between subjective and objective measures occurs, knowledge of those determinants can contribute to better understanding the financial situation of households. People who want to improve their financial condition can benefit from learning about results of this study.

Having econometric model which differentiates households on the basis of subjective financial indicator can be very useful. Polish economy may face a downturn in the near future and the government will be probably forced to reduce social transfers. Undoubtedly information about the household which financially struggle, knowing basic details concerning families will be useful to determine receivers of social transfers.

Studying subjective financial condition is not frequently undertaken research problem especially in Poland. Moreover most of the papers which concern this topic were written couple of years ago and are not up-to-date. Situation of Polish households changes rapidly due to the economic growth. That is why researching subjective assement of those households and veryfing stated hypotheses is significant and worth conducting.

2. Research method

In order to conduct econometric analysis of subjective financial situation binary logitic regression was applied. This model is appropriate to explore connections between dichotomous independent variable and both continuous as well as categorical independent variables (Garson, 2014, p.13). Moreover researechers proved that applying this algorithm in exploring determinants of subjective financial situation is effective and correct. Logistic regression model is based on logistic function which can be given as:

Equation 1. Logistic function

$$P(Y) = \frac{e^{(\theta_0+\theta_1 x_1+\theta x_k)}}{1+e^{(\theta_0+\theta_1 x_1+\theta_k x_k)}}, \tag{2.1}$$

where:

Y - dichotomous independent variable,

$P(Y)$ - probability that Y is equal to 1 for values of dependent variables,

x_k - k dependent variable,

θ_k - structural parameter of the model for k predictor.

After logistic function transformation, the following equation is given:

Equation 2. Logistic regression equation.

$$Logit\ P = \ln\frac{P(Y=1)}{1-P(Y=1)} = \theta_0 + \theta_1 x_1 + \theta_k x_k \tag{2.2}$$

Most frequently interpreted parameter in logistic regression is odds ratio that is proportion of probability that given phenomenon occurs to probability that given phenomenon does not occur.

Equation 3. Odds ratio.

$$OR = \frac{P_i}{1-P_i} \tag{2.3}$$

In case of logistic regression, after estimation of structural parameters of the model, it is possible to write down odds ratio according to following formula:

Equation 4. Odds ratio in logistic regression.

$$OR = e^{(\theta_0+\theta_1 x_1+\theta_k x_k)} \tag{2.4}$$

Parameters of the model $\hat{\theta}_0$, $\hat{\theta}_1$, ... , $\hat{\theta}_k$ are estimated by maximum likelihood method which can be applied to find model parameters by maximazing log-likelihood function which can be given as follows:

Equation 5. Log-likelihood function

$$LL(\theta) = \sum_{i=1}^{n} y^{(i)} \log \sigma\left(\theta^T x^{(i)}\right) + \left(1 - y^{(i)}\right) \log[1 - \sigma\left(\theta^T x^{(i)}\right)] \tag{2.5}$$

where:

$\sigma(\theta^T x)$ - logistic function $\frac{1}{1+e^{-\theta^T x}}$,

$y^{(i)}$ – observed value for i observation.

The estimates are chosen to maximize this function. In other words, the parameters are found in a manner that the predicted probability $\hat{P}(Y)$ using logistic function corresponds as closely as possible to the observed value.

Maximum likelihood estimator has few properties. It is consistent, which means that if sample size increases then difference between estimator and observed value is infinitely close to zero. Moreover this estimator appears to be asymptotically unbiase so expected value of estimator distribution equals to estimated parameter. Maximum likelihood estimator is also effective. Hence, it has low variance. Lastly mentioned estimator is sufficient because it gives as much information about parameter as possible.

2.1. Strengths and weaknesses of logistic regression

Logistic regression algorithm belongs to the group called white box classifiers, which permit an interpratation for model parameters. It is very important because logistic regression not only provides ability to predict but also enable to draw conclusions from the estimated parameters, analyse direction and strength of distinct factors on dependent variable (Dreiseitl, Ohno-Machado, 2002, p. 357). It gives classification but also probabilities which are very useful in evaluating differences in distinct observations. Moreover logistic regression does not require too many computional resources.

However, there are also disadvantages of using logistic regression algorithm. Discriminant power for mentioned algorithm is often significantly worse than that of black box

models, such as supported vector machines and artificial neutral networks (Dreiseitl, Ohno-Machado, 2002, p. 357). Besides the balanced structure of dependent variable is considered a preferable arrangement in binary logistic regression (Meyers, Gamst, Guariano, 2005, p. 222). It means that for datasets with imbalanced distribution for target variable algorithm is not effective. Moreover in case of multicollinearity of independent variables conclusions from estimates are biased (Greene, 2002, p. 57). Thus the assumption of no multicollinearity has to be checked if interpretation of structural parameters is a goal of the research.

2.2. Model validation methods

In order to check logistic regression fit to the data, the overall test of model coefficients, Hosmer Lemeshow test and pseudo R-squared measures can be applied.

Overall test of model coefficients is based on chi-square statistic and it is a likelihood-ratio test based on the maximum likelihood method. Null hyphothesis for the test states that proposed model is no better than baseline model at predicting the positive class.

More information how model fits the data is provided by pseudo R^2 measures. These statistics were developed to be analogs of R^2 as used in ordinary least squares. Both scores make use of likelihood functions for intercept and full model. However Nagelkerke R^2 was developed not to exceed value of 1, which is not always a matter for Cox & Snell R^2 (Smith, McKenna, 2013, p. 17).

The goodness of fit of the logistic regression model can be also measured by the Hosmer Lemeshow test which is conducted by comparing expected probabilities with actual membership of observed values after being divided into 10 groups, ordered by the probability estimated from the regression equation (Agresti, 2002, p. 175-176). The procedure tests the hyphothesis about the difference between observed and expected events is simultaneously zero for all the groups (Shah, Barnwell, 2003, p. 3778). It means that non-significant outcome indicates a well-fitting model.

Another important matter to check is multicollinearity, which can cause unstable estimates and inaccurate variances that affect confidence intervals and hypothesis tests (Greene, 2002, p. 57). The multicollinearity is not a big problem while predicting outcome from a set of independent variables. However, if the goal is to understand how those predictors impact dependent variable then conclusions can be misleading (Paul, 2014, p. 3). In multiple regressions the variance inflantion factor (VIF) is used as indicator of multicollinearity. The increase in VIF value can be associated with the fact that variable is not orthogonal to the other

variables in the model what means that they are highly correlated. Values higher than 10 indicate the problem (Greene, 2002, p. 57-58).

3. Empirical results

In order to investigate relationship between subjective financial situation and socio-demographic factors, mentioned methods and described below materials were applied.

3.1. Source materials

As mentioned before, the data which was applied to analyze subjective financial indicator was taken from Social Diagnosis research (Diagnoza Społeczna, 2015). Results from this study are broadly mentioned and analyzed in research papers. In most cases quality and conditions of life and financial behavior of households are under analysis.

The final dataset, which was used for modelling, consists of two parts. The first part is a dataset concerning individuals, which are heads of the family and took part in wave 2015. The second part is a dataset concerning households which took part in wave 2015. Those two parts are merged by number of household, which is a variable with unique values, occurring in both datasets.

Observations with unavailable values were omitted. Moreover dataset was narrowed down – households in which head of the family was unemployed were excluded from the research. Furthermore households which were classified as “non-family multi-person”, “multi-family household” and “living on unearned sources” were also ruled out because they are not in the scope of this analysis.

3.2. Variables included in models

In order to evaluate financial situation of households, one question from the questionnaire was used. Member of the household was asked: “Do your household make end meets end with your actual income?”. There were six possible answers to mark:

- with great difficulty,
- with difficulty,
- with some difficulty,
- rather easily,
- easily.

This kind of question is applied in most socio-economic survyes, among others in European Community Household Panel (HICP) and The European Union Statistics on Income and Living Conditions (EU-SILC). It can be assigned to studying subjective financial situation (Dudek, 2015, p. 2)

Mentioned responses were transformed to binary outcome – households members who answered that they make ends meet with great difficulty, with difficulty or with some difficulty were appointed as those who have poor subjective financial condition. All the others were recognized as households which have good financial situation. The structure of the created binary variable is presented below.

Table 1. Frequency distribution of dichotomized subjective financial situation.

	Frequency	%
Good subjective financial situation (0)	3171	33%
Poor subjective financial situation (1)	6426	67%

Source: own elaboration.

There are several factors from social diagnosis dataset, which could be used to predict and evaluate subjective financial condition of households. There is theoretical background to include below mentioned variables in the modelling:

- Socio-economic group – 6 categories, (nominal scale):
    - farmers,
    - self-employed,
    - employed,
    - pensioners,
    - disability pensioners.

Membership in socio-economic group seems to have influence on financial situation. Research proves that despite the improvement of financial conditions of farmers, there is still relative income inequality between them and white-collar workers (Kozera, Wysocki, 2014, p. 97). Moreover houseolds with the best financial position are those which head is a manager or is self-employed. In addition households with blue-collar worker as head of the family, have only slightly better situation than those with head being not employed (Bialowolski, Weziak-Bialowolska, 2013, p. 379). Researchers also prove that pensioners and disability pensioners have lower probability of making ends meet than other socio-economic groups (Dudek, 2013, p. 9), (Panek, Czapiński, 2015, p. 42).

- Categorized age of head of family – 6 categories, (ordinal scale):
    - under 24
    - between 25 and 34,

- between 35 and 44,
- between 45 and 59,
- between 60 and 64,
- above 65+.

The older the head of the family the lower probability of financial problems of the household. Moreover households led by people in group age 25-34 have the biggest chance of not repaying current liabilities (Anioła, Gołaś, 2012, p. 51), (Bialowolski, Weziak-Bialowolska, 2013, p. 379).

- Biological type of households - 8 categories (nominal scale):
  - marriage without children,
  - marriage with 1 child,
  - marriage with 2+ children,
  - non-family one-person ,
  - one-parent families,
  - multi-family households.

Research leads to conclusion that more problems with budget were observed among single person households, households of four or more members and one-parent families. On the other hand higher disposable income was achieved in childless families (Panek, Czapiński, 2015, p. 42-43), (Bialowolski, Weziak-Bialowolska, 2013, p. 379), (Wereda, Prokopowicz, 2017, p. 247).

- Level of education of the head of the family - 4 categories (ordinal scale):
  - primary and lower education,
  - vocational education,
  - secondary education,
  - higher and post-secondary education.

Researchers point out the significance of level of education in terms of financial condition. The higher level of education of household head, the higher probability of being in sphere of wealth (Sączewska-Piotrowska 2015, p. 253).

- Class of place of residence – 2 categories (nominal scale):
  - rural area
  - town

Authors of papers suggest that place of residence influence financial situation. Households living in rural areas and smaller towns appears to have problems with fullfilling current needs (Panek, Czapiński, 2015, p. 44).

- Number of individuals in household, numeric variable.

Relationship between number of household members and financial condition have also been found. In every socio-economic group with the increasing number of household members the income decreases (Kozera, Stanisławska, 2014, p. 140), (Bialowolski, Weziak-Bialowolska, p. 2013, 379).

Number of cases (households and individuals) with complete information (without lack of data) amounts to 9597.

All of the variables which were measured on nominal or ordinal scale, were transformed into n-1 dummies (n is number of categories in every category).

In order to better understand the relationship between dependent and independent variables, presented below table was made.

Table 2. Structure of households reporting financial problems with a breakdown per socio-demographic attributes.

Charasteristics		**Does household make ends meet?**	
		Yes	**No**
		In %	
Age of head of the family	Below 34	9%	6%
	Between 35 and 44	19%	13%
	Between 45 and 59	30%	30%
	Between 60 and 64	11%	14%
	Above 65+	31%	37%
Biological type of households	Marriage without children	34%	23%
	Marriage with 1 child	20%	16%
	Marriage with 2+ children	22%	22%
	Non-family one-person	18%	25%
	One-parent families	6%	14%

Education level of head of the family	Primary and lower education	8%	23%
	Vocational education	24%	35%
	Secondary education	30%	27%
	Higher and post-secondary education	38%	15%
Class of place of residence	Rural area	31%	45%
	Town	69%	55%
Socio-economic group	Employed	51%	41%
	Pensioners	35%	40%
	Self-Employed	8%	3%
	Farmers	4%	7%
	Disability pensioners	2%	9%
Number of individuals in household	1	18%	25%
	2	36%	30%
	3	21%	19%
	4	18%	16%
	5	5%	7%
	6+	2%	3%

Source: own elaboration based on Social diagnosis [Diagnoza społeczna 2015].

The structure of household reporting the subjective financial situation, divided by the variables available in the Social diagnosis questionnaire was presented in the table 1. The lowest percentage point differences between dependent variable outcomes in particular groups were found age characteristic and the biggest in education level of head of the family. It is probable, that age variable will be found insignificant in logistic regression. All the other characteristics had vivid discrepancies within the groups.

In order to provide unbiased evaluation metrics of model's performance, it is very important to calculate them from a data set not used in the model building process (Dreiseitl, Ohno-Machado, 2002, p. 355, 357). Accordingly, the split into training and test sets was applied - 80% of observations were randomly assigned to training set, which was applied for creating

model and remaining 20% were allocated to test set, which was used for making unbiased evaluation. It is the most common practice of model assessment in machine learning because it is the easiest and the least computationally expensive way to provide reliable information about fitted models.

Since the validity of using proposed controlling variables were checked with the literature and dataset preparation was completed, the modelling process was ready to start.

3.3. Logistic regression model

When choosing the independent variables, significance criteria and economic corectness (valid sign of estimated parameters) were applied. Final logistic regression model consisted of 17 independent variables (most of them were dummy variables). Some of the variables, like county in which household lives and gender of the head of family were found insignificant and they were not included in the modelling process. Those variables did not fulfill mentioned requirements, mainly concerning statistical validity.

First of all, the goodness of fit of logistic regression model was checked. Test of significance for the overall equation, Pseudo R-square measures, Hosmer-Lemeshow test and Variance inflation factor were used.

Table 3. Test of significance for the overall equation

	Chi-square	**df**	**p-value**
Overall test of model coefficients	1125.57	17	p<0.01

Source: own elaboration.

The proposed model was substantially better at predicting households being in bad financial situation than the baseline model containing only constant term. Decision was made for 0.01 significance level.

Table 4. Pseudo R-square coefficients.

	-2 Log likelihood	**Cox & Snell R Square**	**Nagelkerke R Square**
Model	8679.16	0.14	0.19

Source: own elaboration.

Nagelkerke R^2 coefficient suggests that model was characterized by approximately 19% error reduction in comparison to the baseline model.

Table 5. Results of Hosmer-Lemeshow test.

	Chi-square	df	p-value
Hosmer-Lemeshow test	5.13	8	0.74

Source: own elaboration.

There were no ground to reject null hypthothesis, he p-value estimated for the test was higher than 0.05. It means that model's estimates fit the data at an acceptable level.

Table 6. Variance inflantion factor for independent variables.

Variance inflation factor	
Farmers	1.14
Self employed	1.06
Pensioners	3.75
Disabled pensioners	1.16
Below 34	1.4
Between 45 and 59	2.04
Between 60 and 64	2.09
Above 65	5.3
Rural	1.21
Marriage one child	2.18
Marriage 2+ child	5.36
One parent family	1.27
One person household	2.03
Primary_and_lower	1.49
Vocational	1.69
Secondary	1.55
Number of individuals	6.37

Source: own elaboration.

Multicollinearity is not a problem in the model. Thus estimates are stable and variances are accurate.

Since the model offered adequate fit with empirical data, it is appropriate to start interpretention of the obtained parameters.

Estimation results in logistic regression model for probability of poor subjective financial situation of household are presented below.

Table 7. Estimation of structural parameters of the model.

Characteristic	Coefficient	Estimate (β)	Std. error	z value	p value	Odds ratio
	Intercept	-1,25	0,14	-8,79	0,00	0,29
Socio-economic group	Reference group: employed					
	Farmers	0,29	0,13	2,23	0,03	1,34
	Self employed	-0,56	0,12	-4,86	0,00	0,57
	Pensioners	0,34	0,10	3,21	0,00	1,40
	Disabled pensioners	1,14	0,17	6,79	0,00	3,13
Biological type	Reference group: marriage without children					
	Marriage one child	0,21	0,10	2,12	0,03	1,23
	Marriage 2+ children	0,38	0,14	2,67	0,01	1,47
	One parent family	1,22	0,11	10,99	0,00	3,39
	One person household	0,80	0,09	8,71	0,00	2,23
Education of head of the family	Reference group: higher and post-secondary					
	Primary and lower	1,80	0,10	17,87	0,00	6,05
	Vocational	1,24	0,07	16,62	0,00	3,44
	Secondary	0,81	0,07	11,73	0,00	2,26
Place of living	Reference group: town					
	Rural	0,21	0,06	3,51	0,00	1,23
Age of head of the family	Reference group: between 35 and 44					
	Below 34	0,25	0,11	2,20	0,03	1,28
	Between 45 and 59	0,19	0,08	2,37	0,02	1,21
	Between 60 and 64	0,28	0,12	2,39	0,02	1,32
	Above 65	0,06	0,13	0,50	0,62	1,07
Number of individuals		0,11	0,05	2,22	0,03	1,12

Source: own elaboration based on Social diagnosis [Diagnoza społeczna 2015].

All variables were found statistically significant (p value for z statistic were lower than 0.05), except for “Above 65” dummy variable. In this case p value were higher than 0.05, so there were no grounds to reject null hypthothesis which meant that estimation of the parameter was statistically equal to 0.

In order to more precisely evaluate effect sizes for coefficient estimations it is convenient to use odds ratios, which are exponentiated values of β estimates. Odds ratios are interptreted as the amount by which the probability of success (i.e. ending up in households being in poor financial situation) over the probability of failure (i.e., belonging to households, which report good subjective financial condition) grows or falls with one unit increase in an independent variable, controlling for effects of all other explanatory variables (James, Witten, Hastie, Tibshirani, 2013, p. 132-133).

The presented estimation results point out that only households with self-employed members had lower chances for being in bad financial situation, as compared to employed group. Odds ratio for this group equaled to 0.57. It means self-employed households had fewer chances for being in bad financial situation by a multiple of 43%. All the others group were found to had higher chance of being in poor financial condition. Farmers, pensioners and disabled pensioners had respectively 1.34, 1.40 and 3.13 times higher chance of having poor financial situation than households, which members were assigned to employed group. Accordingly, the multi-variate analysis shows that being in employed or self-employed group had an important impact on reducing the probability that a household is poor. On the other hand disabled pensioners households had the highest chance to have bad subjective financial situation.

Biological type was found to significantly explain subjective financial situation. The lowest risk of having bad financial situation was found in marriages without children. Marriages with one and 2 or more children had respectively 23% and 47% higher chance of being in poor financial condition than marriages without children. In addition, one parent families and one person households were respectively characterized by increased odds of being in bad financial condition by 239% and 123% than marriages without children. It means that marriages had lower probability of having poor subjective financia situation in comparison to one parent families and one person households.

The most important factor influencing subjective financial situation is education of head of the family. Households led by better educated people considerably less often reported bad

financial situation. This conclusion was verified with logistic regression estimates. As the level of education rose, odds of being in poor financial situation decreased. Households which head of the family gained primary and lower education had a 6,05 times higher chance of having poor financial situation than those, which head of the family had higher and post-secondary education. Moreover, households with head of the family having vocational and secondary education had respectivey 244% and 126% higher chance of being in poor financial condition than households from the reference group.

Results of estimated logistic regression parameters also confirm significant impact of place of living on subjective financial situation. Families which lived in rural area had 23% higher chance of being in bad financial situation, as compared to those who lived in town.

Outcome of the research also indicates on influence of head of the family age on financial condition. However, it is the least important factor influencing subjective financial situation. The lowest probability of being in bad financial condition appeared in families with head being older than 35 and younger than 44. Households which head of the family was aged, between 60 and 64 or below 34 had approximately 30% higher chance of having bad financial situation than households with family head being older than 35 and younger than 44. On the other hand, families which head of the family was aged between 45 and 59 were characterized by increased odds of being in poor financial situation by 21%, as compared to the reference group. However, the relationship between age group "above 65" and "between 35 and 44" was found insignificant. It means that there is no statistical difference between those classes in terms of impact on subjective financial situation.

Lastly, research results point out that increase in number of individuals in households by one person was linked to increased odds of being in poor financial situation by 12%.

Conducted analysis confirms existence of connection between distinct socio-demographic features and subjective financial situation. The biggest risk of being in bad financial condition have housholds led by disabled pensioners, living in one-parent families with head being in age group "between 60 and 64", having primary and lower education, dwelling in rural area. Moreover, problem of poor subjective financial situation escalates in multi-person households.

3.4. Cut-off point selection and model evaluation

The criterion applied to access the quality of logistic regression is distrimination, which is a measure of how well the two classes in dataset are separeted (Dreiseitl, Ohno-Machado,

2002, 355). In order to provide unbiased estimation of model's distrimination the test dataset was applied.

In order to measure predictive capacity the optimal cut point needs to be established because for every possible cut-off there are different values of evaluation metrics.

Since the dataset is unbalanced the method to determine optimal cut-off based on classification accuracy is not adequate because maximum accuracy is achieved by assigning every observation to majority class (Fawcett, Provost, 1996, p. 305-306). Hence there was a need of finding cut-point based on specificity and sensitivity values.

Commonly used method for cut-off selection is the Younden index (*J*) (Youden, 1950, p.32-55) which defines optimal cut-point as the point maximizing Younden function all over possible cut-off values (c). *J* is a function of sensitivity (*Se)* and specificity *(Sp),* such that

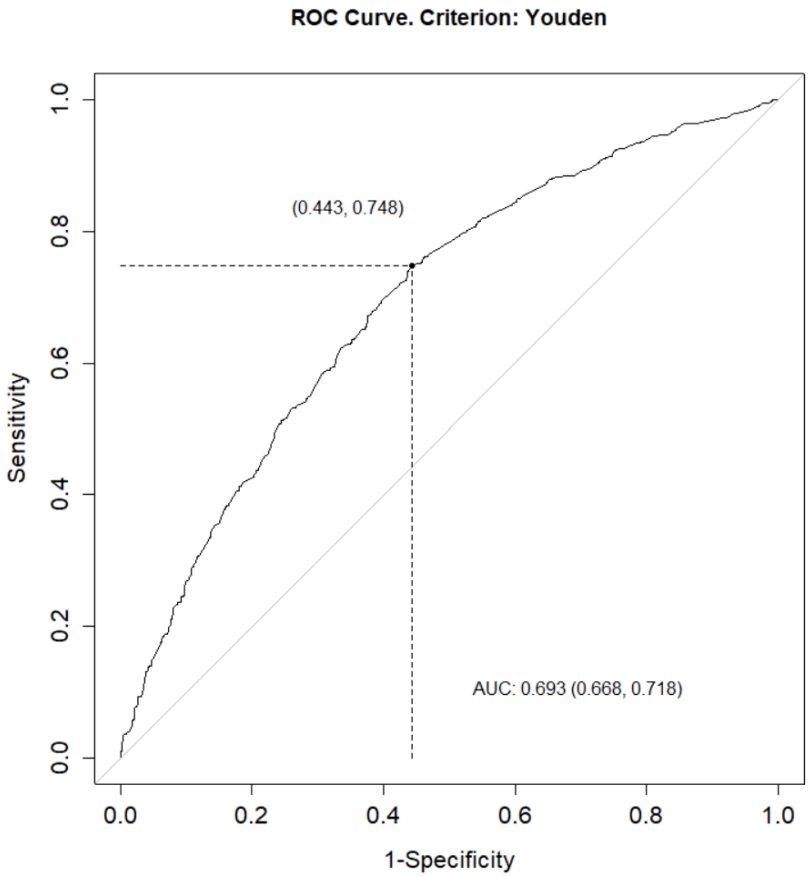
Equation 6. Younden index.

$$J(c) = Se(c) + Sp(c) - 1. \quad (3.1)$$

In order to better understand trade-off between true positive and true negative rate and visualize outcome of Younden index the receiver operator curve was plotted.

The ROC curve is a plot of sensivitiy versus 1-specificity across different cut-off points. The greater discriminant capacity of model the closer the line to the upper corner in "ROC space" (Tilaki, 2013, p. 630). Receiver operator curve not only validate the model by presenting AUC, specificity and sensitivity values but also help to determine suitable cutoff.

Figure 1. ROC curve.



Source: own elaboration.

Table 8. TPR and TNR for choosen cut-off point established by Younden criterion.

Cut-off	0.63
True positive rate	0.75
True negative rate	0.56
Accuracy	0.68

Source: own elaboration.

Table 9. Confusion matrix for cut-off point 0.63.

Classification table	Predicted
	Financial situation

			Good (0)	Bad (1)
Real	Financial situation	Good (0)	354	267
		Bad (1)	325	919

Source: own elaboration.

Accuracy for given cut-off point is aprroximately 0.7, so there was no particular improvement in accuracy comparing to the baseline model. However there was substantial difference in ability of correctly predicting negative class (specificity value). The increase in true negative rate amounted to more than 50 percentage points.

3.5. Concluding remarks

Proposed model was found to be statistically better than baseline model. Moreover it also appeared that model fits the data. It was validated by overall test of model coefficients, Hosmer-Lemeshow test and Nagelkerke R^2.

The constructed logistic regression model brought results, which provide grounds for applying socio-demographic factors for predicting subjective financial situation of Polish households. These include: socio-economic group, biological type of family, education level and age of family head, place of living and number of individuals in household.

In addition, estimated predictive capacity gave satysfing results for the cut-off point selected by Younden index. Suggested model is good in distriminating households with bad subjective financial situation from the rest.

4. Conclusions and discussion

The main objective of this research was to construct econometric model which allows to determine factors impacting subjective financial situation of Polish households and also to make sufficient predictions. The goal was undoubtedly achieved. Thanks to the estimation of logistic regression particular socio-demographic factors were set to be drivers of perception of financial condition. Moreover the alghoritm was evaluated on test dataset in terms of classification and it gave satisfying results.

4.1. Key findings

Comparing to other studies obtained results concerning determinants of financial situation of Polish households are consistent with the results of researchers, reffering to biological type – (Panek, Czapiński, 2015), (Bialowolski, Weziak-Bialowolska, 2013), (Wereda, Prokopowicz, 2017), socio-economic group and number of individuals - (Panek, Czapiński, 2015), (Kozera, Wysocki, 2014), (Bialowolski, Weziak-Bialowolska, 2013), (Dudek, 2013), education of head of the family – (Sączewska-Piotrowska, 2015) and place of residence – (Panek, Czapiński, 2015). However, they do not confirm the results concerning age of head of the household - (Anioła, Gołaś, 2012), (Bialowolski, Weziak-Bialowolska, 2013).

Those papers refers to objective financial situation measures. However, correlation between objective and subjective measures was found (Gąsiorowska, 2010, p. 178). It means that comparing those measures is sensible.

Regards to the age factor, researchers claim that the older head of family the lower probability of not repaying current liabilities, which is highly connected with financial condition of household. Results presented above indicate that age group 35-44 reports the best subjective financial situation. What is important the positive trend of having better financial situtation connected with older head of the family is not visible. There is the biggest increase in odds in 60-64 category in reference to 35-44 category which means that households with older head of the family performed worse than those with younger heads in case of subjective financial condition. It means that conclusions for this factor are not analogous. In this case, the fact that subjective financial situation was taken into consideration seems to be crucial. First of all the correlation between subjective and objective measures of financial situation is not strong (Gąsiorowska, 2010, p. 178). It implies that some factors differently influence on subjective and objective financial situation. Secondly, psychological factor has to be taken into consideration. It can influence perception of subjective evaluation (Gąsiorowska, 2010, p. 178).

Theory of relative standards says that people assess their personal success reffering to the past (Gąsiorowska, 2010, p. 178). Older people can compare their present financial situation to the household condition from the past, which could be worse in some cases. It may describe difference between impact on subjective and objective financial situation in case of age of head of the family.

Like in the other studies concerning financial situation in households, logistic regression algorithm proved its ability to be used in this kind of analysis. Model offered adequate fit to the data and satisfying results concerning predictive capacity. Moreover, there was found that there is no multicollinearity of independent variables. It means that proposed model can help to determine whether particular socio-demographic factor influences financial condition and how strong the effect is. Additionaly, featured logistic regression supports in classifying whether given household has good or bad subjective financial situation. Thus the model can be used in practical applications.

On the grounds of estimated model it can be concluded that odds of subjective financial situation depend on particular socio-demographic factors like membership to socio-economic group, biological type of family, age and education of head of family, class of place of residence and number of individuals in household. The most important factor influencing subjective financial situation was education. Households with head having higher and post-secondary eduacation had the lowest odds of being in bad financial situation and the odds increased together with lower education level. Households with head being in the age group "between 35 and 44" had also the lowest chance of being in poor financial condition compared to the households belonging to other age groups. As far as biological type, one parent and one person households seemed to have the highest odds of being in bad financial situation compared to the marriages without children. On the other hand, self-employed and employed group of households had the lowest chance of having poor financial condition in comparison to other socio-economic groups. In regard to place of living households which lived in the town had lower odds of being in bad financial situation than those which lived in rural area. Lastly, the more individuals in the household the higher odds of having bad subjective financial situation.

4.2. Recommendations for future research

The extension of the research can concern comparing subjective and objective measures. Several hypothesis can be made, for instance how level of education influence requirements of households concerning income and how this effect impact subjective financial situation. It is

probable that households with highly educated members have higher income requirements, which can contribute to the lower satisfaction with the same amount of money in comparison to less educated ones. Furthermore the influence of the place of living of household in case of county and comparison between subjective and objective measures can also be examined. The fact that county variable in logistic regression model was found insignificant does not mean that there is no relationship. Perhaps households in various counties have different expectations concerning income and it influences perception of their situation.

The correlation between objective and subjective measures can also be validated. Further research may help to evaluate whether there is relationship between them and how strong it is.

If social diagnosis is updated, new model should be estimated and results compared to older ones. The effect of new social beneftis may be vivid and it would be worth testing whether it influences subjective financial situation of households.

References

Agresti A., (2002), Categorical data analysis: Second edition, John Wiley & Sons, Inc., Hoboken, New Jersey.

Anioła P., Gołaś Z., (2012), Zastosowanie wielowymiarowych metod statystycznych w typologii strategii oszczędnościowych gospodarstw domowych w Polsce. Narodowy Bank Polski, ISSN 2084–6258.

Białowolski P., Weziak-Białowolska D., (2013), The Index of Household Financial Condition, Combining Subjective and Objective Indicators: An Appraisal of Italian Households. Soc Indic Res (2014) 118:365–385 DOI 10.1007/s11205-013-0401-0.

Czapiński J., Panek T. (red.), Diagnoza Społeczna 2015 – warunki i jakość życia Polaków, raport z badania, Warszawa: Rada Monitoringu Społecznego, 2015.

D'Ambrosio C., Frick J., (2007), Income Satisfaction and Relative Deprivation: An Empirical Link, „Social Indicators Research", nr 81(3).

Dreiseitl, S. and Ohno-Machado L. (2002) Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. Journal of Biomedical Informatics, 35, 352-359. http://dx.doi.org/10.1016/S1532-0464(03)00034-0

Dudek H., (2015), Subiektywne postrzeganie sytuacji dochodowej – mikroekonometryczna analiza danych panelowych. Wydawnictwo SGGW, Warszawa, 2015.

Fawcett T., Provost F., (1997), Adaptive Fraud Detection. Data Mining and Knowledge Discovery 1, 291–316.

Garson G. D., (2014). Logistic regression: binary and multinomial. Statistical Associates Publishing.

Gąsiorowska A., (2010), Biedni czy bogaci? Wpływ dochodu i postaw na ocenę własnej sytuacji finansowej. Sopot: Gdańskie Wydawnictwo Psychologiczne.

Genge E., Trzęsiok J., (2017), Czy łatwiej związać koniec z końcem? Badanie sytuacji materialnej gospodarstw domowych w Polsce z wykorzystaniem modeli panelowych. Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 468.

Greene W. H., (2002). Econometric analysis 5th edition. Upper Saddle River, New Jersey 07458.

James G., Witten D., Hastie T., Tibshirani R., (2013), An Introduction to Statistical Learning: with Applications in R. New York :Springer.

Kapteyn A., (1977), A Theory of Preference Formation, Leyden Univesity, Leyden.

Kasprzyk B., (2016), Subiektywne oceny dobrobytu ekonomicznego w gospodarstwach domowych w świetle modelowania dyskryminacyjnego. Wyd. Uniwersytet Rzeszowski, Rzeszów.

Kozera A., Stanisławska J., (2014), Taksonomiczna ocena sytuacji finansowej gospodarstw domowych w Polsce w 2010 roku. Nierówności Społeczne a Wzrost Gospodarczy, nr 38 ISSN 1898-5084.

Kozera A., Wysocki F., (2014), Sytuacja finansowa gospodarstw domowych rolników na tle pozostałych grup społeczno-ekonomicznych. Journal of Agribusiness and Rural Development, 3(33), 85-99.

Liberda B., Pęczkowski M., Gucwa-Leśny E., (2011), How do we value our income from which we save?, University of Warsaw, „Faculty of Economic Sciences Working Papers", vol. 3(43), 1-19.

Meyers L. S., Gamst G. and Guarino A.J., (2005), Applied Multivariate Research: Design and Interpretation. SAGE Publications.

Paul R. K., (2014), Multicollinearity: Causes, Effects and Remedies. M. Sc. (Agricultural Statistics), Roll No. 4405 I.A.S.R.I, Library Avenue, New Delhi-110012.

Rada Monitoringu Społecznego (2015), Diagnoza społeczna 2000–2015: zintegrowana baza danych, http://www.diagnoza.com.

Sączewska-Piotrowska A., (2015), Identyfikacja determinant bogactwa dochodowego z zastosowaniem modelu logitowego. Zarządzanie i Finanse Journal of Management and Finance Vol. 13, No. 4/2/2015.

Shah B.V. and Barnwell B. G., (2003), Hosmer-Lemeshow goodness of fit test for Survey data. Babubhai Shah, Safal Institute, 22 Autumn Woods Drive, Durham, NC.

Smith J. T. and McKenna C. K., (2013) A Comparison of Logistic Regression Pseudo R^2 Indices. Multiple Linear Regression Viewpoints, Vol. 39(2).

Tilaki K., (2013), Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med. Spring;4(2):627-35.

Van Praag B.M.S., (1968), Individual Welfare Function and Consumer Behavior: A Theory of Rational Irrationality, North Holland, Amsterdam.

Wereda W., Prokopowicz D., (2017), Economic and financial situation of households in Poland – analysis and diagnosis. Polityki Europejskie, Finanse i Marketing, 18 (67).

Youden W.J., (1950). Index for rating diagnostic tests; pp. 32–35. Cancer. Vol. 3.

List of equations, tables and figures

Equations

Tables

Figures

Summary

The following thesis concerned determinants subjective financial situation of Polish households, which took part in Social Diagnosis research. The main purpose of the analysis was to construct model which may allow to evaluate particular socio-demographic factors influencing subjective financial situation and to make effective classifications of households by their perception. In order to achieve mentioned goal, logistic regression algorithm was applied. The obtained results were consistent with outcomes from other researches. Belonging to socio-economic group, biological type, age and education level of head of family, class of place of residence and number of individuals in households were found to be significant factors influencing subjective financial situation. Moreover it appeared that logistic regression was sufficient classification model. The algorithm was validated on test data set and it gave satisfying results, which means that model can be used in practical applications.