QUANTITATIVE METHODS IN ECONOMICS
Vol. XIV, No. 1, 2013, pp. 212 – 221

APPLICATION OF LOGISTIC REGRESSION IN THE SUBJECTIVE ASSESSMENT OF THE QUALITY OF LIFE OF POLES

Beata Kasprzyk, Jolanta Wojnar
Department of Quantitative Methods, University of Rzeszow
e-mail: bkasprzy@univ.rzeszow.pl, jwojnar@univ.rzeszow.pl

Abstract: The paper discusses the application of tools of multivariate analytical methods i.e., logistic regression, to model the subjective assessment of satisfaction/dissatisfaction with quality of life achieved. Based on empirical data from a random sampling of households in Podkarpacie province, the statistical significance of variables was estimated, which enabled the estimation of the value of life quality based entirely on qualitative variables, that are characteristic for a given household, with incomes being excluded. The essence of the regression is to seek such household related parameters, that could have varying degrees of impact on the occurrence of either negative or positive values. The models have been designed based on varied classification categories for households, thus allowing for both the quantitative analysis and comparison of the impacts of key cardinal determinants of values of life.

Keywords: logistic regression analysis, households, value of quality of life, qualitative variable, probability

INTRODUCTION

"Quality is a fulfillment of requirements and expectations, the way that leads to satisfaction whilst satisfaction serves as very good and reliable measure of quality".

(E. Skrzypek)

A key phenomenon of contemporary times is the noticeable increase in significance of the quality factor [Skrzypek 2002]. The concept of the quality of life is very complex, both within economic studies and beyond. One of the first definitions of life quality states that "it is the resultant integration of personal traits

as well as the subjective and objective indicators of well-being[1]". Quality of life is generally defined as the extent to which material and non-material needs of individuals, families and communities are met. Quality, according to Skrzypek, should be understood as something that can be improved upon. Thus, it can refer to life and be considered a tool for improving all aspects of human endeavors [Skrzypek 2002]. Rampley claims that in speaking about quality of life, it is best to assume that it is a multidimensional concept [Rampley 2003].

The concept of quality of life, in contemporary economic studies, is very complex. It covers conditions construed as objective factors, that include, among other things economic circumstances, free time, social security, living conditions, natural environment, health, and social environment etc. Due to its varied aspects, studies on the quality of life are conducted according to applicable methodologies in given fields, which in economics are categorized by indicators that summarize the common aspects of the measured objective. The best known indicator that reflects standard of living and life satisfaction is the *Quality of Life Index*, published since 2005 by *„The Economist*" magazine for needs of various countries, which combines the results of subjective life satisfaction surveys with objective factors of quality of life in various countries[2]. This, in economic studies, means that the private perception of all individuals impacts on the quality of life. The subjective approach is manifested in self-esteem, which is accounted for in psychological categories, including the feeling of satisfaction, happiness, anxiety, fears, hopes, etc. [Domański et.al. 2000]. Subjective measurements of quality of life have no clearly defined methodology, although attempts are being made to quantify them. R. Kolman postulates, that both theoretical and practical knowledge on quality ought be accurate. He claims that the reason for the poor performance of quality control systems is ignorance in areas of quantitative evaluation methods of levels of quality [Kolman 1992].

The aim of the study is to show the possibilities of applying logistic regression in issues concerning the assessment of subjective quality of life based on households in Podkarpacie province. The essence of modeling is to decide which statistically significant factors determine the divide into two emerging groups of households, i.e., those satisfied or not satisfied with achieved levels of quality of life. The estimated value of certain parameters of the logistic model will be determined on the basis of empirical analysis which should allow the classification

[1] Nowa encyklopedia (1996), PWN, Vol. III, Warszawa, pp. 121.

[2] Economist Intelligence Unit, An annual index that comprises of partial indicators for 9 categories, i.e., cost of living, culture and leisure, economy, environment, freedom, health, safety and risks and climate. http://internationalliving.com/2010/02/quality-of-life-2010/. The Quality of Life Index rating for 2010 places France as first (82 points), with Australia, Switzerland and Germany coming next with 81 points. Poland is placed further in the rating with 70 points.

of households into two separable groups and also to predict, for each group, the chances of positive or negative assessment.

MATERIALS AND RESEARCH METHODS

A survey, using the direct interview method (*mall intercepts method*), was conducted to measure the quality of life. The research period covered 2012. The total sample size was set at *n=1050,* but having taken account of refusals and wrong answers, the final random sample obtained was *n = 835* responses.

The questionnaire contained a fundamental question of the type: „Are you satisfied with the current level of quality of life of your household?" (of course taking into account material situation, work, health, home, life style etc.)?. Possible variation of yes or no answers were anticipated. Beside the responses to this question, additional information on specifically chosen socio-economic characteristics of the sampled households was collected. The spatial extent of the study covered south-eastern Poland, mainly Podkarpacie province.

The primary issue, for a strictly defined methodology of econometric research, i.e., logistic regression model, was to define the dependent variable as well as the input set of potential independent variables. The dependent variables being nominal variables served as responses to the subjective rating of quality of life, with possible variation of yes or no. The independent variables served as variables that characterized prevailing living conditions of a given household. They were unitary data, which, due to the scope of the study were essentially socio-demographic for purposes of quantitative measurements (income levels, total number of persons per household, number of income earning persons, number of dependent children). The qualitative variables were, on the other hand, types of home management in a household[3], satisfaction with achieved wages levels, the fact of having or not having savings[4] including sex, age, education of head of household, place of residence, membership of any socio-economic group and the biological type of household. Measurements of the qualitative variable was, depending on the question, relative to the nominal or ordinal scale.

The subjective value of the quality of life is strongly linked with material situation, which is a derivative of earned incomes. It was assumed, a priori, that the key, strict economic factor that influences satisfaction or its lack with quality of life is the attainable household income, with improvements in the level of satisfaction with varied aspects of life increasing with rising levels of incomes [Bywalec et al.

[3] As regards the question *how finance are handled in the household*, the following categories of responses were possible: very modest-not enough for basic needs; modest-very economical in spending; average-enough, but have to save up for major purchases; very good- we can indulge in some luxuries.

[4] The question was: *Does the household have savings?* (with yes/no category of responses) and *Is your remuneration satisfactory?* (with yes/no category of responses).

2002]. The strongest predictor of well-being is the evaluation of material conditions. The more frequent positive responses in the evaluation of quality of life, the higher the state of well-being. The value of the material situation is thus relative to the level of life of the immediate environment as positivism of the comparison increases, the higher the satisfaction. A negative rating for levels of wealth, i.e., foods, incomes, consumption and social contacts, means higher satisfaction [Długosz 2008]. The quantitative variable „income" that is highly correlated with the explanatory (independent) variable was hence deliberately omitted in the model, thus exacting more efforts in seeking other determinants of the quality of life assessments during the analysis.

The binomial logistic regression model (logit model) is applied to explain the dichotomous qualitative variable Y depending on the level of exogenous variables X_1, X_2, ..., X_k (be they qualitative or quantitative). The explained variable is, in such models, quantified using dummy (0-1) variables. The variable Y_i assumes the value of 1 with a probability p_i, and value 0 with probability $1\text{-}p_i$, that can be illustrated thus:

$$P(y_i = 1) = p_i,\ P(y_i = 0) = 1 - p_i$$

Probability is a function of the vector of dependent variables $\boldsymbol{x_i}$ as well as vector parameters $\boldsymbol{\beta}$, hence:

$$p_i = P(y_i = 1) = F\left(x_i^T \beta\right) \qquad \textit{for } i=1,2,...,n \tag{1}$$

In this model, the transform function is the cumulative logistic distribution [Cramer 2004]:

$$p_i = F(x_i^T \beta) = \frac{1}{1+\exp(-x_i^T \beta)} = \frac{\exp(x_i^T \beta)}{1+\exp(x_i^T \beta)}, \tag{2}$$

The values of the inverse function of F are known as logits (hence, the adoption of the expression „logit models" in books). Logit is the logarithm of the odds ratio and the occurrence of the incident, which is determined by the dependency [Guzik 2004]:

$$F^{-1}(p_i) = \ln \frac{p_i}{1-p_i}, \tag{3}$$

Based on this model, one can specify the marginal increase in probability, the so-called marginal effect, which describes how increases in the values of each explanatory variable influences the probability of its belonging to a given group.

$$\frac{\partial p_i}{\partial x_{ji}} = \beta_j p_i (1-p_i) = \beta_j \frac{\exp(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})}{\left[1+\exp(\beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki})\right]^2}, \tag{4}$$

Since $p_i(1\text{-}p_i) > 0$, hence the sign of the parameter by the variable X_j determines the direction of impact of variable X_j on variable Y. If the explanatory

variable X_j increases by one unit (*ceteris paribus*) then the odds ratio varies by $exp(\beta_j)$. In situations where $exp(\beta_j)>0$, growth is observed, when $exp(\beta_j)<0$ the odds ratio declines. If X_j is a binary zero-one variable, then $exp(\beta_j)$ provides by how many times the odds ratio $Y_i=1$ for category „1" increases in comparison with the same odds ratio for category „0".

Binomial model parameters are estimated using the maximum likelihood method. If the observations are independent, then the likelihood function of n-element sample should be seen as follows [Chow 1995, Kalouptsidis et al. 2009]:

$$L = \prod_{y=1} p_i \prod_{y=0} (1-p_i) = \prod_{i=1}^{n} [F(x_i^T \beta)]^{y_i} [1-F(x_i^T \beta)]^{1-y_i}, \tag{5}$$

The correctness of the estimated model can be investigated using the likelihood ratio test, aimed at verifying null hypothesis for models with the intercept only (all other parameters of the model except the intercept are equal to zero). The statistical test is expressed with the formula [Gruszczyński 2001]:

$$\chi^2 = 2(\ln L_{UR} - \ln L_R), \tag{6}$$

where:

L_{UR}– value of the likelihood function for a full model,

L_R– value of the likelihood function for models reduced to the intercept.

The statistic has a distribution χ^2 with degrees of freedom equaling the number of explanatory (independent) variables.

Another measure for fitting binomial models is the odds ratio, calculated as the product of properly classified cases to the product of improperly classified cases.

$$\boldsymbol{IS = \frac{n_{11} \cdot n_{00}}{n_{01} \cdot n_{00}}}, \tag{8}$$

The appropriate size n_{ij} is illustrated in a table of accuracy of classification of cases:

	$\hat{y}_i = 1$	$\hat{y}_i = 0$
$y_i = 1$	n_{11}	n_{10}
$y_i = 0$	n_{01}	n_{00}

In order to construct the table of accuracy, it is necessary to set theoretical values for the variable y_i. The transformation of probability into a dichotomous variable is achieved keeping with the standard rules of predictions: $y_i = 1$ if $p_i > 0{,}5$ and $y_i = 0$ if $p_i \leq 0{,}5$. The model is suitable for prediction since if $IS > 1$, then any classification based on the model is better than random classifications.

EMPIRICAL RESEARCH RESULTS

Based on the empirical material collected, models for several variants of combinations of explanatory variables were constructed. A description of the significant independent variables for the models is presented in table 1.

Table 1. Selected explanatory variables for the models

Variable	Description of variable	Size
Savings	has	425
	does not have	410
Ways of managing finances in household	very modest (not enough for basic needs)	46
	modest – very economical in spending	216
	average - enough, but has to save up for major purchases	337
	good – enough for most needs	203
	very good – can indulge in certain luxuries	33
Satisfaction with remuneration	yes	280
	no	555
Education of head of household (the person that brings in the highest incomes for the household)	elementary (or junior secondary)	26
	vocational	174
	secondary	367
	higher	268
Type of household	marriage	695
	single parent	62
	one-member household	78
Age of head of household	up to 30	148
	30-40	198
	40-50	281
	50-60	169

Source: own studies based on results of a survey

Table 2. Estimation results of the logit model (1)

Items	Independent variable				
	Constant	Savings	Way of managing finances in household	Number of income earning persons	Satisfaction with remuneration
Estimate	-5.915	1.392	1.508	0.300	2.204
Standard error	0.551	0.175	0.210	0.117	0.294
Statistics *t*	-10.727	7.976	7.190	2.561	7.496
Level *p*	0.000	0.000	0.000	0.011	0.000
-95%	-6.997	1.050	1.096	0.070	1.627
+95%	-4.832	1.735	1.920	0.530	2.781
Wald Chi-square	115.074	63.611	51.691	6.558	56.194
Level *p*	0.000	0.000	0.000	0.010	0.000
Odds ratio	0.003	4.024	4.518	1.350	9.059
–95%	0.001	2.857	2.993	1.073	5.087
+95%	0.008	5.669	6.819	1.698	16.131
Number of observations $n = 835$, The likelihood function 297.859, $\chi^2 = 560.83$; p=0.0000					

Source: own studies based on results of a survey

Based on the analyzed parameters of the model, it can be stated that savings, financial management in households, number of income earners as well as satisfaction with remuneration have, from the subjective perspective, statistically significant impacts on the value of quality of life.

Estimates based on random sampling of parameter values has allowed the final form of the logistic model to be written as:

$$p(Y) = \frac{e^{-5.92+1.40\ \text{savings}+1.51\ \text{financial managt}+0.30\ \text{income earners}+2.20\ \text{satisfaction with remuneration}}}{1+e^{-5.92+1.40\text{saving}+1.51\text{financial managt.}+0.301.\text{income earners}+2.20\ \text{satisfaction with remuneration}}}$$

Since the value sign of the parameter next to the independent variables of the estimated model is positive, it means that any increase in the values of these variables increases the chances of being satisfied with current quality of life. In interpreting the odds ratio for each variable, it is worthy of note that *savings* result in fourfold increased (odds) satisfaction with current level of quality of life (odds ratio value for this variable is 4.024). A similar dependency can be observed in respect of the *way of managing finances* variable. The highest value of the odds ratio (9.059) occurs in respect of the *satisfaction with remuneration* variable, which means that people who are satisfied with their remunerations are over nine times more likely to be *satisfied with current level of quality of life.*

Based on the estimated model, it is possible to identify the likelihood for a person to be classified as being *satisfied with its current level of quality of life*. If it is assumed that such a person has savings, that financial management in the household is average and that there are two income earners in the given household,

but there is lack of satisfaction with remuneration earned, then the estimated likelihood being classified as *satisfied with current level of quality of life* is 0.288. If for the same person, the fact of being satisfied with remuneration earned improves, then his estimated likelihood will rise to 0.783.

Table 3 illustrates the likelihood of classifying all respondents into groups of „unhappy" or satisfied" with current quality of life. The classification of individual cases indicated the number and fraction of those correctly or incorrectly classified. The number of correct prediction constituted 84.577% of the tested sample. Over 87% of the overall cases accounted for accuracy prediction value of *Y=1* (satisfied with *current quality of life*), while 81.886% was for the accuracy of prediction value of *Y=0* (unhappy with *current quality of life*). The high value 30.986 means that the classification is much better than that obtainable through a completely random classification.

Table 3. Propriety of classification of subjective evaluation of quality of using the logit model (1)

Actual	Predicted		Total	%	
	Y = 0	Y = 1		Odds ratio	30.986
Y=0	330	73	403	Y=0	81.886
Y=1	55	377	432	Y=1	87.269
Sum	385	450	835	Total	84.577

Source: own studies based on results of a survey

Using a different combination of explanatory variables, another model that correctly classifies 78.806% of cases can be evaluated. Estimated results of such a model are presented in table 4.

Based on the parameters assessed, it can be deduced, that key determinants in the subjective assessment of quality of life remains savings, education and type of household.

Table 4. Estimated results using the logit model (2)

Items	Independent variable			
	Constant	Savings	Education	Type of household
Estimate	-2.635	2.531	0.628	-0.377
Standard error	0.393	0.174	0.115	0.143
Statistics *t*	-6.712	14.559	5.461	-2.627
Level *p*	0.000	0.000	0.000	0.009
-95%	-3.405	2.189	0.402	-0.658
+95%	-1.864	2.872	0.853	-0.095
Wald Chi-squares	45.051	211.961	29.821	6.899
Level *p*	0.000	0.000	0.000	0.009
Odds ratio	0.072	12.561	1.873	0.686
-95%	0.033	8.930	1.495	0.518
+95%	0.155	17.668	2.348	0.909
Number of observation *n = 835,* The likelihood function 412.791 $\chi^2 = 330.87$; $p = 0.0000$				

Source: own studies based on results of a survey

A negative parameter value by the variable, *„type of household",* means that modifying the type of household (e.g., a marriage for a single parent) leads to declining odds (probability) of being *satisfied with current quality of life*, which in other words means that satisfaction with actual level of quality of life in marriage is higher than in single parent families. Increasing education level of the highest income earner in a household by „one level", for example from secondary to higher, increases the odds of being satisfied with quality of life by almost two-fold (the odds ratio being 1.873). The relevance of case classification, based on this model is slightly lower i.e., 658 cases of accurate classification out of 835.

Table 5. Propriety of classification of subjective evaluation of quality of life using logit model (2)

Actual	Predicted		Total	%	
	Y = 0	Y = 1			
Y=0	318	85	403	Y=0	78.908
Y=1	92	340	432	Y=1	78.704
Sum	410	425	835	Total	78.806

Source: own studies based on results of a survey

SUMMARY

Logit models serve as very good tools in explaining developments in subjective assessment of the quality of life depending on prevailing socio-economic circumstances. Modeling results indicate that the dependence of the value of this variable on such factors as *savings, financial management, number of income earning persons as well as satisfaction with remuneration* is statistically reliable. This is understood to mean, that material factors are basic determinants in the rating of quality of life as either satisfactory or non-satisfactory.

It should be remembered, however, that the fore-going analysis was an attempt to proffer reasons for satisfaction with the quality of life, whilst ratings for the subjective quality of life are driven not only by economic situations but also by non-economic thus making it a complex category that involves both psychological and sociological aspects. The issue of measurement and assessment therefore, remains open and certainly not exhaustive in terms of the issues covered by the studies [Podolec 2008].

REFERENCES

Bywalec Cz., Rudnicki L. (2002) Konsumpcja, PWE, Warszawa, str. 43.

Chow G.C. (1995) Ekonometria, Wydawnictwo Naukowe PWN, Warszawa 1995.

Cramer J.S. (2004) The early origins of the logit model, „Studies in History and Philosophy of Biological and Biomedical Sciences", nr 35.

Długosz P. (2008) Czy w Polsce ‘B” można być szczęśliwym?, [w:] Jakość życia. Od wykluczonych do elity, (red.) R. Derbis, Wyd. Akademia im. J. Długosza, Częstochowa, str. 243-244.

Domański R., Ostrowska A., Rychard A. (2000) Jak żyją Polacy, Wyd. IfiS, Warszawa, str. 84.

Gruszczyński M. (2001) Modele i prognozy zmiennych jakościowych w finansach ibankowości, „Monografie i Opracowania”, Oficyna Wydawnicza SGH w Warszawie, nr 490, Warszawa.

Guzik B., Appenzeller D., Jurek D. (2004) Prognozowanie i symulacje, Wydawnictwo AE, Poznań, str. 176.

Kalouptsidis N., Psaraki V. (2009) Approximations of choice probabilities in mixed logit models, „European Journal of Operational Research”, XXX.

Kolman R. (1992) Inżynieria jakości, PWE, Warszawa, str. 31.

Podolec B. (2008) Społeczno-ekonomiczne uwarunkowania sytuacji materialnej gospodarstw domowych, [w:] Statystyka społeczna – dokonania, szanse, perspektywy. Biblioteka Wiadomości Statystycznych, Tom 57, Główny Urząd Statystyczny, Warszawa, str. 112.

Skrzypek E. (2002) Jakość i efektywność, Wyd. Uniw. M.C. Skłodowskiej, Lublin, str.10.

Rapley M. (2003) Quality of Life Research – a critical introduction, London, Sage Publications Incorporation, s.50.