Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(330) 2017

DOI: http://dx.doi.org/10.18778/0208-6018.330.08

Józef Dziechciarz
Wroclaw University of Economics, Faculty of Management, Computer Science and Finance, Department of Econometrics and Operations Research, jozef.dziechciarz@ue.wroc.pl

Marta Dziechciarz-Duda
Wroclaw University of Economics, Faculty of Economics, Management and Tourism, Department of Econometrics and Computer Science, marta.dziechciarz@ue.wroc.pl

Non-Metric Data in Household Durable Goods Analysis. Selected Aspects

Abstract: Measurement of household endowment with durables is crucial in many aspects of assessing the social and economic situation of a country and its citizens. The demand (sales) for durables is regarded as one of the key indicators of economic conditions. Similarly, analysis and evaluation of household durable goods are usually considered in the context of measuring the quality of life. The possession of durables is measured by means of the number and quality of goods in households. Measurement of household endowment is conducted usually by means of weak measurement scales, namely nominal and ordinal. Such data require the use of specialised tools for analysis and modelling. This study discusses the possibilities of statistical analysis of such data. Additionally, modelling and problems of inference on the basis of obtained results are discussed.

Keywords: non-metric data analysis, durable goods, correspondence analysis, logistic regression, classification trees, cart, households

JEL: C1, F2

1. Introduction

The situation of households can be described by income and other socio-economic factors, as well as the type of household or family life cycle phases. An important factor describing the economic and social family status is ownership of durable goods. Various research centres continuously monitor the state of saturation of households with particular products (e.g.: the Central Statistical Office of Poland (GUS) performs surveys of household budgets and the Council for Social Monitoring develops its *Social Diagnosis* (2015) project based on panel research of households). Material resources, including durable goods, describe the economic condition of a household, characterise the standard of living, and are evidence of household modernity. Therefore, its saturation is an important factor characterising households' well-being. Typically, at a certain level of income, some households are not satisfied with possession of just one unit of a certain good, and are willing to acquire more than one, usually each next purchase being of better quality (Dziechciarz, Dziechciarz-Duda, Przybysz, 2010; Dziechciarz-Duda, 2007: 395).

Measurement of endowment of households is usually conducted by weak measurement scales, nominal and ordinal. Such data require the use of dedicated statistical methods for analysis and modelling.

2. The problem

Endowment of households with durable goods is particularly important in the measurement of quality of life. Measurement of the number of units and the estimated value of durable goods in households' possession is difficult and faces many practical problems. Thus, the theoretical basis for the measurement of the number and value of durable goods in households' possession has to provide specialised techniques and estimation methods for determining the value of consumer durables in each household (Amendola, Vecchi, 2014).

The issue of measurement and analysis of the number and value of durable goods in households' possession is analysed here in the broader context of social well-being, poverty and inequality (*Framework*…, 2013; Deaton, Zaidi, 2002).

It is not possible to measure a household's social well-being directly, nor its poverty and inequality. Similarly, it is impossible to measure a household's income directly. Extensive discussion is devoted to the question whether consumption may be used as a proxy for income measurement. There is a general consensus on the value of using consumption as a summary measure of living standards. Among others, researchers at the World Bank developed consumption-based measures constructed from survey data to measure poverty, and thus to measure well-be-

ing and inequality. Consumption may be further used to analyse changes in living standards, and to assess the distributional impacts of anti-poverty policies.

There are several reasons why consumption can be a better measure of welfare than income (Slesnik, 2000). Consumption can provide a better picture of actual standards of living, while income fluctuates significantly. In particular, consumption takes into account whether households can access credit markets or household savings at times when current income is low due to, for instance, seasonal variations. Furthermore, consumption may be better measured than income in poor economies with large informal sectors, where households consume their own production or exchange it for some other goods.

Household surveys typically give information on household expenditure for several product categories. However, household expenditure is a satisfactory measure of living standards only if everyone faces the same prices and everyone lives in households with the same size and composition. In reality, in most cases, households are of different sizes, and are located across regions or neighbourhoods with different prices. Therefore, some adjustments are necessary to obtain a proper consumption-based welfare measure value (estimate). Measurement and assessment of households' durable goods endowment are even more complicated. Any practical solution needs to provide a harmonised and standardised framework for measuring the number and value of consumer durables which are at the disposal of a household. The list of methods and techniques includes approaches based on the purchase price and the valuation of the cost based on the rent or lease equivalent. Each of the approaches has major advantages but also disadvantages. Valuation of ownership is crucial in the context of social well-being, poverty and inequality analysis. Measuring social well-being, poverty and inequality requires the prior selection and determination of the appropriate welfare definition. It is not a simple task. For such assessment, a number of technical decisions have to be made. Often, the adopted definitions (norms) are controversial. In practice, the decision-making process is based on a more or less well-established practice. The theoretical arguments are rare and often difficult to apply. Reports prepared in the World Bank[1] list extensive examples in this respect (Lanjouw, 2009; Deaton, Zaidi, 2002; Ravallion, 1998). Two measures are the most commonly used to adjust for differences in prices: the money metric utility, proposed by P. Samuelson (1974); and the welfare ratio, based on C. Blackorby and D. Donaldson (1987). A. Deaton and S. Zaidi (2002) provide a discussion of the theoretical foundations and the practical implementation of these two measures.

It is widely acknowledged that durable goods possession has a strong, positive impact on the perceived social well-being, poverty and inequality (Offer, 2005).

[1] Additional reading references include, e.g.: Odekon, 2015; Deaton, 1997; Blackorby, Donaldson, 1988.

It is also widely appreciated that the level of spending on durables and the level of current expenses do not have the information value for disposable income estimation. An in-depth discussion of advantages and disadvantages of selected measurements and recommendations for methods of analysis may be found in the literature (Diewert, 2009). The conclusion is that spending on consumer durables does not directly reflect real situations, and thus should not be included in the aggregate welfare indicator. Instead, the estimate of benefit (value) coming from service-like use of durable goods should be included. Such a method offers good information that is easier to obtain in household budget surveys.

3. Definitions

Durable goods are products that need special treatment while used for measurement of the quality of life. Many issues have to be discussed before creating theoretical approaches dealing with consumer durables and their role in the measurement of quality of life.

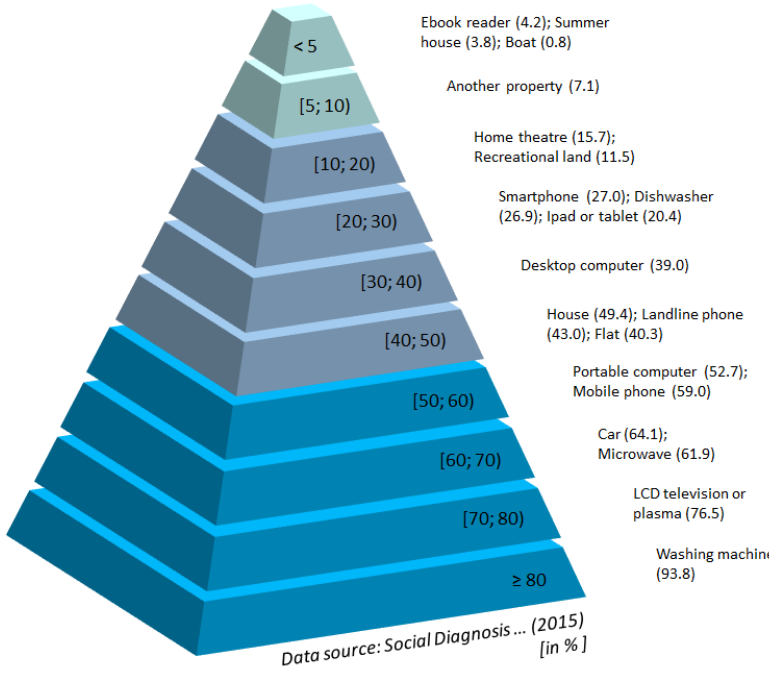

Figure 1. Classification of consumer durables

Source: own elaboration

According to J. Kramer's definition (1997), *durable goods are objects of consumption characterised by long lifetime and high frequency of use*. In the literature, definitions of durable goods stress their indivisibility, distinctiveness, long life cycle and significant value. W. Diewert (2009) says in his description that goods could provide useful services to consumers by repeated use for a long period of time.

When analysing household endowment with durable goods, one can distinguish groups of goods according to their occurrence among households:

1) standard goods, which are available for every household and are owned in more than fifty percent of households,
2) higher standard goods (owned by 10–50% of households),
3) luxury goods (owned by less than 10% of households).

This classification was introduced by J. Kramer (1993: 161–164), and further developed to the present state by M. Dziechciarz Duda (2007).

4. Measurement of the value of durables in households

Normally, households use durable equipment for a significant period of time. The time gap between successive purchases of the equipment often lasts for many years. Because of their substantial cost, durable goods are usually purchased more rationally than other products, and the purchases are usually planned. The main feature of the durables does not depend solely on their physical stability; physical stability is a common property of many other consumer goods, e.g.: capital goods. The ability to be consumed (used) for two or more periods is the key feature. The *System of National Accounts*[2] (2008) bases the distinction on whether the goods can be used, for the purpose of production or consumption, only once, repeatedly or continuously.

There exists inconsistency between the so-called *reference period* chosen for the welfare aggregate (normally a calendar year, a month or a week) and the period of time during which an individual durable product is used by a consumer or a household (Deaton, 1997: 151). Obviously, the purchasing price is clearly an inadequate valuing concept. A similar issue arises in the construction of a consumer price index[3] and a cost-of-living index[4]. The purchasing price shows the value

[2] An internationally agreed set of recommendations on how to compile measures of economic activity.

[3] An extensive list of topic references includes: Diewert, 2009 and 2004; Lippe, 2007; Dziechciarz, 2004a and 2004b; Goodhart, 2001; Fisher, 1911.

[4] Amendola, Vecchi (2014: 8) claim that the use of the purchase price instead of the consumption flow leads to overestimating the effect of economic cycle on the household welfare, to underesti-

of the product for its entire life, and not the value for a reference period. In the literature, there is a concept of *consumption flow* of durable goods (*General*…, 2014; Moulton, 2004; Katz, 1983).

Acquisition approach[5]. The solution where the entire value of a purchased good is attributed to the household expenditure is known as the acquisition approach. According to this approach, null spending is attributed if a household does not purchase any durable products during the survey period. Additionally, it does not take into account whether the acquired product is new or used. In reality, this means that the price paid illustrates (measures) the value of all services provided by the durable over the entire period of its presence in the household. The acquisition approach underestimates the welfare of households that own *used* durable goods with respect to households that happened to purchase durable goods in the current year. It mirrors the business cycle. Households tend to postpone the purchase of durables when the economy slows down, and to increase purchasing when it speeds up (Amendola, Vecchi, 2014: 10–11; Diewert, 2009 and 2004: 419).

Rental equivalence. To overcome the listed disadvantages of the acquisition approach, it is recommended to use rental or leasing prices for estimation of consumption flow value. For the rental equivalence method, it is assumed that a household can purchase the *services* of durable goods. The disadvantage of this technique is that homogeneity of durable goods is assumed, i.e. that all goods are of the same type and quality. Additional assumptions are that markets are competitive and the economy is in equilibrium. Only then the market rental price depicts the value that measures the consumption flow (Amendola, Vecchi, 2014: 11–12; Diewert 2009 and 2004; Gillingham, 1983). The obvious practical solution for determining rental price equivalents for consumer durables, i.e. asking households what they think their durables would rent for, is extremely arbitrary. Perception does not necessarily match reality.

User cost. The concept was first introduced by M. Keynes (1936: 53), and then reformulated by D. Jorgenson (1963). The user cost approach uses the purchasing price and allocates the portion of it as a difference between the purchasing price and the possible (hypothetical) price which would be obtained while selling the durable at the end of using period. Additionally, the calculated value is corrected for the Net Present Value (NPV), meaning that the nominal interest rate is to be assumed (Diewert, 2004). N. Amendola and G. Vecchi (2014) compare this approach to the situation where the user has the choice of either owning the durable product and using it for the measurement (cost equivalence), or selling the durable at the

mating absolute poverty, and most likely to a bias in the poverty profile. They call upon a well-documented study on Russia where it is shown that the impact on inequality can be very large: the Gini index of expenditure increases from 32 percent to 44 percent when the full purchase value of durables is included instead of their use value.

[5] In the literature, this is also referred to as the net acquisition approach.

beginning of the period. This may be interpreted as potential income. In a more sophisticated approach, the cost of capital is introduced. Usually only economic depreciation of durables is taken into account, and not their physical depreciation (technical ageing). C. Hulten and F. Wykoff (1981) described the mechanism of economic depreciation. An extensive discussion of practical consequences of the use of described techniques may be found in the World Bank report (Amendola, Vecchi, 2014: 16–18).

5. Problems of statistical measurement

Prior to estimation of the value of services provided by a durable product for a given household, several problems of statistical measurement must be solved. Collected data are seldom measured on strong (metric) measurement scales. Primary sources of survey data often record attitudes, intentions, opinions, etc. This leads to measurement on weak, non-metric scales. The fundamental problem arising during the measurement of respondents' attitudes lies in the fact that it attempts to quantify immeasurable and hidden phenomena. Therefore, the key issue of accurate and precise measurement is that the variable which measures the attitude should be described on an adequate scale. In most studies, the non-metric measurement levels of analysed characteristic are coded as numbers. In such cases a precise coding criterion is vital (Mayntz, Holm, Hübner, 1985: 60). The better the measurement tool, the more accurate the measurement results for the strength, validity, intensity, and complexity of respondents' attitudes. Unfortunately, the quantitative scale commonly applied to measure attitudes fails to identify the deeper, more elusive aspects of consumers' behaviours. The measurement is a fundamental operation in the research process. Respondents' attitudes are known only to the degree to which the researcher can accurately measure them. The key issue when defining a variable to be examined in the course of research is the choice of measurement scale (Stevens, 1959). M. Walesiak (2000: 45) states that in the measurement theory there are four basic levels of measure that include the following scales (in the rising order of strength):

1) nominal scale (owning of goods, choice of products),
2) ordinal scale (ranking of quality of goods, ranking of products),
3) interval scale (scores of products),
4) ratio scale (e.g.: the choice on the scale of fixed sums, the probability of an individual choice).

The applied measurement scale of the value of endogenous variables affects the reliability of the measurement process. Measurement values belonging to the metric scale are obtained in procedures using graduated pairwise comparisons and the scale of fixed sums. Among non-metric scales, the most common ways to measure

the value of variables are the ranking and the pairwise comparisons. There is also an additional, popular method of measurement using the rating scale, representing the position scale, which could have a non-metric or metric construction (Walesiak, 1996: 189–191; Adams, Fagot, Robinson, 1965; Stevens, 1959: 25–27).

Metric scales are commonly used in preference analyses. However, due to their more intuitive nature, more possibilities are offered by applying non-metric scales, including verbal scales. In the case of verbal scales, the respondent uses expressions of a natural language. The main disadvantage of this measuring method is that (seemingly) it is far less precise than the metric scale. The problem of verbal scale application is the need to identify quantitative equivalents for verbal expressions used to mark various levels. It is not possible to arbitrarily and clearly define how to interpret the differences in ratings that have been expressed in linguistic terms. Therefore, it is necessary to apply tools which mitigate these undesirable characteristics of verbal measurement. One of the possible techniques is measurement using fuzzy numbers, usually triangular or trapezoidal. Linguistic variables may be used to describe users' preference towards analysed products. The fundamental work on sociological and utility measurement is presented in S. Stevens (1959). The specifics of linguistic variables are thoroughly discussed in the literature[6].

6. Analysis of statistical measurement results when non-metric data are involved

6.1. Introduction

Analysing data containing measures on weak measurement scales has become easier over the last few years. Although classical statistics offer some measurement techniques, new publications provide extensions and improved techniques for such analyses (Dziechciarz-Duda, Król, 2017).

Non-metric variables can be used for any type of relation. For dependence analysis techniques of non-metric data, non-metric variables can be used as dependent or independent variables. In cases when there is one dependent variable in a single relationship, the most popular methods are discriminant analysis and logistic regression. A less common situation is when several dependent variables in a single relationship are analysed. In such a case, canonical correlation analysis with dummy variables can be employed. Classification trees constitute an important group of methods that provide possibilities of analysis of non-metric var-

[6] The list of references includes: Borovicka, 2014; Abdullah, Khadiah, 2011; Chu, 2010; Liou, Wang, 1994; Zadeh, 1975.

iables (Gatnar, Walesiak, 2011). Another type of methods are those investigating independence. The most common method for non-metric data is correspondence analysis. There are specialised methods for analysing categorical predictors, e.g.: log-linear analysis of multi-way frequency tables.

The following section shows example techniques that could be used when non-metric data are employed. A case of a dishwasher was chosen for illustration. Data show that approximately 26% of households had the selected durable product. For the test study, a sample of Polish households was analysed[7].

6.2. Correspondence analysis

The first step was to conduct multiple correspondence analysis (10,876 Polish households were analysed). As many as 24,545% of the relationships between the categories of analysed variables in two-dimensional space are explained. Category links resulting from correspondence analysis are shown in Figure 2. The data contain coexistence categories of variables from the nominal and ordinal scale. The analysis took into consideration: possession of a dishwasher (possesses, does not possess), the socio-economic group (workers, farmers, self-employed, retirees), the household type, the existence of a loan in the household (1 – a household repaying a mortgage credit, 0 – not repaying), the age of the head of household (6 categories), the level of education of the household head (4 categories), and assessment of income (enough for current expenses or not sufficient for current expenses). The analysis enabled to outline general trends that determine variations in possession of the analysed good. It is clear that dishwasher possession is more often attributed to the younger and better educated, along with entrepreneurs and married couples with children. A group that does not possess a dishwasher includes typically smaller households, the elderly and those assessing their own income as insufficient. There exists a group with a profile significantly different from the average profile (the group of farmers, characterised by a lower than secondary education level and households with more than two children).

Correspondence analysis is a recommendable technique to identify coexistence of two or more variables measured on the weak (non-metric) scales. The advantage of this method is a tool for visualisation involving presentation of co-occurrence of categories of two (or more) variables. This provides the ability to quickly interpret the results and allows general indication of the existing relationship between the categories. Due to significant simplification of existing dependencies, there is often a need to supplement the analysis by other methods of multivariate statistical analysis (e.g.: the classification methods).

[7] Data source: *Social Diagnosis*, 2015, integrated database: www.diagnoza.com.

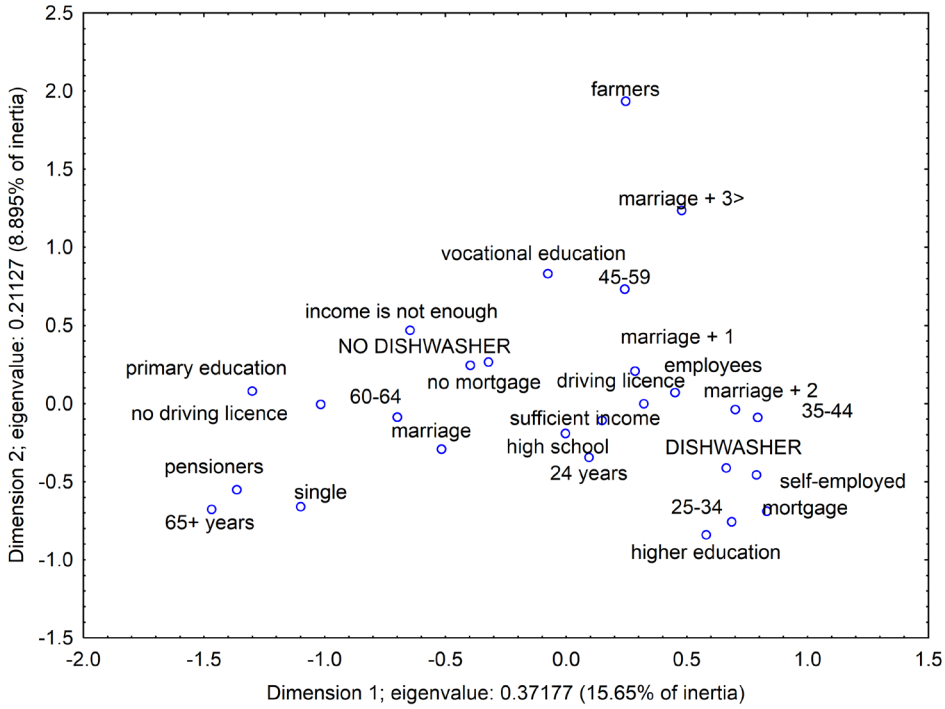

Figure 2. Links between categories. Multiple correspondence analysis results

Source: own elaboration (Statistica)

To gain fuller insight and supplement graphical presentation of correspondence analysis results, classification according to hierarchical agglomeration (Ward; Euclidean distance) was applied.

As a result, a tree linking analysed categories of characteristics was obtained. An analysis of the dendrogram allowed clear identification of categories co-existing with the fact that a household possesses or does not possess a dishwasher (Fig. 3). Along with three groups identified in correspondence analysis, a separate group (the group of farmers characterised by a lower than secondary education level and households with more than two children), difficult to unequivocally assign to one of the main groups, may be seen. Based on the dendrogram, one can claim that these households are more similar to the group with dishwasher possession.

The drawback of the correspondence analysis method is a large number of variables that could hamper the interpretation of results. In addition, in the case of a very complex structure, it is difficult to determine all dependencies. This is due to the fact that with a large number of variables, the degree of explanation in two-dimensional space is reduced. Another drawback of correspondence analysis is the requirement that variables are only non-metric (so called qualitative var-

iables). One may analyse metric variables, but they must be pre-encoded, which always means loss of information.

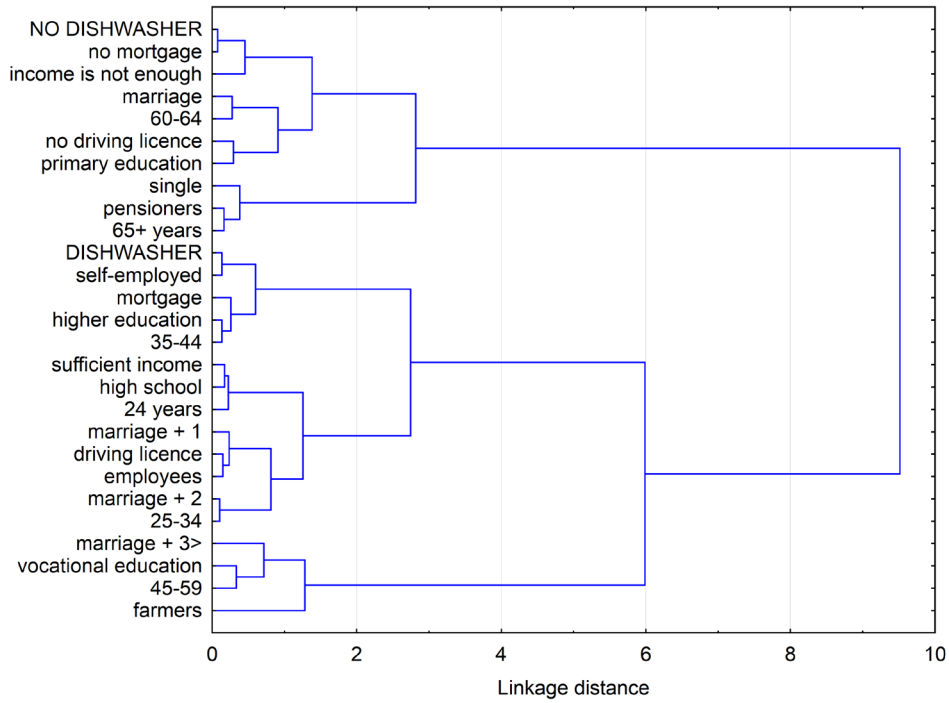

Figure 3. Dendrogram. Hierarchical agglomeration; Ward; Euclidean distance

Source: own elaboration (Statistica)

6.3. Logistic regression

In the analysed logistic regression model specification, the selected dependent variable[8] described dishwasher possession (0: does not own a dishwasher, 7,814 households (74.092%); 1: owns a dishwasher, 2,736 (25.907%)). Twenty six independent variables were included in the model. These were: characteristics of the household income equivalent, savings, mortgage payments, the size of the place of residence, Internet access, the socio-economic group, the head of household education level, possession of a driving license by the head of household, and possession of other durable goods. The model specification with all statistically significant parame-

[8] Based on the same data source (*Social Diagnosis*, 2015).

ters is characterised by McFadden R^2 value, which is equal to 27.1%. The classification matrix showed that the model predicted rate was 80.2%. Based on the estimated logistic regression model, it can be seen that in households where the head of household is an entrepreneur, the chance of owning a dishwasher is more than twice higher than in other households. In households where the head of household has a college education, the chance of owning a dishwasher is, again, twice higher than in other households. Other important factors that increase the chance of owning a dishwasher are the municipality class and the type of household (marital status). Single-person households rarely own them (chance lower by 25%). Dishwashers are also more than twice as common in households with a mortgage. By *ceteris paribus*, each additional one thousand of income increases the chance of having a dishwasher by about 30%.

Table 1. Logistic regression model[9] (dependent variable: *owning a dishwasher*)

Variables	Coefficients	Standard error	*z*	*p*	Odds ratio
CONSTANT	–5.187	0.142	–36.507	0.000	0.006
EDUCATION_HIGHER	0.675	0.076	8.851	0.000	1.964
EDUCATION_SECONDARY	0.327	0.065	4.995	0.000	1.386
EQUIVALENT INCOME [PLN 1000]	0.256	0.034	7.554	0.000	1.292
SAVINGS	0.182	0.056	3.257	0.001	1.200
MORTGAGE CREDIT	0.821	0.093	8.845	0.000	2.272
SOCIO-EC_GR_SELF-EMPLOYED	0.749	0.115	6.541	0.000	2.116
TYPE OF HH_MARRIAGE	0.194	0.077	2.527	0.011	1.214
TYPE OF HH_MARRIAGE+2	0.230	0.074	3.116	0.002	1.259
TYPE OF HH_MARRIAGE+3	0.435	0.103	4.234	0.000	1.545
TYPE OF HH_SINGLE	–0.293	0.108	–2.706	0.007	0.746
PLACE OF RESIDENCE_>500	0.433	0.119	3.643	0.000	1.542
PLACE OF RESIDENCE_200–500	0.279	0.113	2.470	0.014	1.322
PLACE OF RESIDENCE_100–200	0.366	0.119	3.084	0.002	1.442
PLACE OF RESIDENCE_20–100	0.260	0.083	3.134	0.002	1.297
PLACE OF RESIDENCE_<20	0.240	0.090	2.676	0.007	1.271
DRIVING LICENCE	0.342	0.078	4.376	0.000	1.408
MICROWAVE OVEN	0.316	0.106	2.993	0.003	1.372
IPAD OR OTHER TABLET	0.224	0.059	3.780	0.000	1.251
LCD OR PLASMA TV SET	0.860	0.067	12.759	0.000	2.362
CAR	0.584	0.062	9.421	0.000	1.793
SATELLITE PAY TELEVISION	0.584	0.086	6.773	0.000	1.793
OWN HOUSE/FLAT	0.489	0.086	5.673	0.000	1.630

[9] Likelihood ratio test: $c^2(26) = 3277.79$ [0.0000], $N = 10546$.

Variables	Coefficients	Standard error	*z*	*p*	Odds ratio
HOME CINEMA	0.484	0.068	7.141	0.000	1.622
INTERNET	0.396	0.066	6.001	0.000	1.486
COMPUTER	0.178	0.061	2.924	0.003	1.195
PORTABLE COMPUTER	0.479	0.077	6.238	0.000	1.614
Average value of dependent variable	0.260	Standard deviation of dependent variable			0.438
McFadden R^2	0.271	Adjusted R^2			0.267
Log-Likelihood	–4399.371	Akaike criterion			8852.742
Schwarz criterion	9048.857	Hannana-Quinn			8918.951

Source: own elaboration (Gretl), data source: *Social Diagnosis*, 2015

Logistic regression models can be used to study factors that determine the existing differences in the level of endowment with durable goods. This is due to the fact that the logistic regression model allows the description of the effect of several variables influencing the value of dichotomous dependent variable. The resulting evaluation of the model parameters allows assessment of the individual impact of each explanatory variable on the probability of the studied phenomenon (e.g.: owning the product or not). The parameter sign informs about the impact of an individual variable on the analysed phenomenon value (see Rószkiewicz, 2002: 90). If the parameter estimate resulting from the sample has a positive value, it can be concluded that when the independent variable increases its value, the probability of occurrence of analysed phenomenon grows. An analogous mechanism is in place for negative parameter values. The quality of the model is assessed with the use of odd ratio. For details, refer to the topic literature, e.g.: M. Rószkiewicz (2002: 92). In the analysed model, the dependent variable is a non-metric variable, while the independent variables are characteristics of the household and characteristics of the head of household, which may come from different (weak or strong) measurement scales. The metric characteristics are used after encoding to a non-metric character. One may use a special case of logistic regression, the so-called *log-linear analysis*, which allows the search for relationships and interactions between qualitative variables. It is meant for a situation where both the dependent and all the independent variables are derived from a non-metric scale.

6.4. Classification tree

An important group of methods for research problems where both metric and non-metric data are involved is a group of analytical techniques called classification trees. In the present study, a classification tree was used with CART as a tree construction algorithm.

The CART method was created by a team of statisticians in the early eighties (Breiman et al., 1984). In the CART algorithm, the Gini index is used to split the tree. Alternatively, a measure of entropy is used as the rule of division into two parts (called *twoing rule*). The advantages of this solution include its low sensitivity to the presence of outliers, its resistance to missing data in independent variables, and the ability to use the same variables in different parts of the tree. This is important since it allows identification of the context dependencies and interactions between variables.

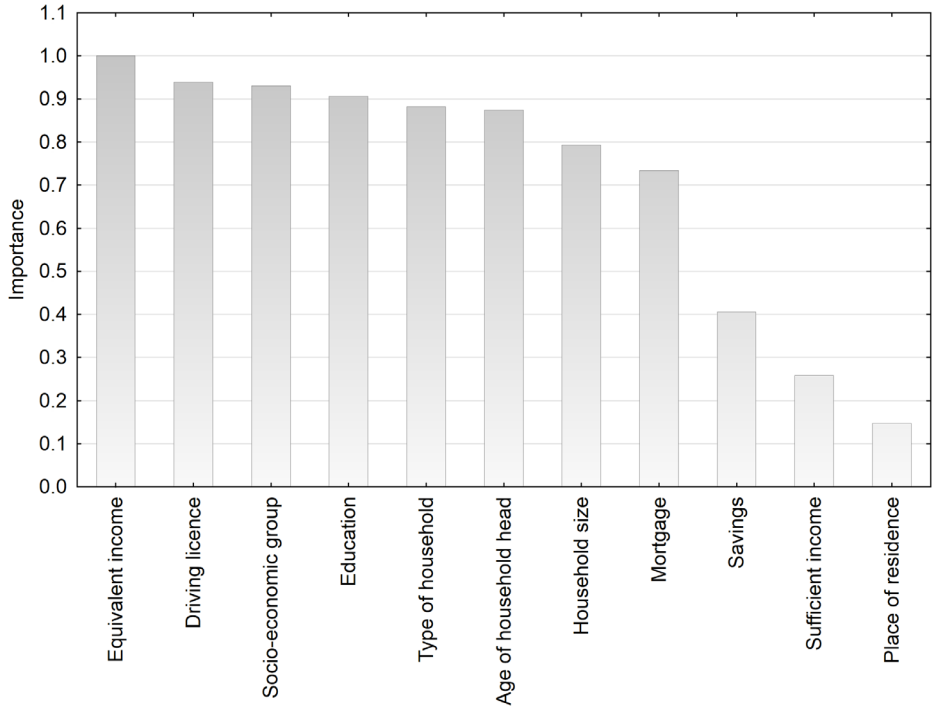

Figure 4. Importance of predictors

Source: own elaboration (Statistica)

As a continuation of the issue of dishwasher possession by households, a CART was constructed[10]. The non-metric variable where 1 means there is a dishwasher and 0 means there is no dishwasher was used as the dependent variable. The analysis was conducted using the classification tree where non-metric variables prevail among the independent variables: the socio-economic group, the type of house-

[10] Data source: the database *Social Diagnosis*, 2015. Analysed data set: 10,773 observations (heads of households).

hold, repayment of the mortgage loan by a farm the age of the head of household, the level of education of the head of household, savings, and possession of a driving licence. The only metric independent variable was the level of monthly income equivalent.

The CART tree algorithm allows the determination of the validity (importance) of individual predictor variables (independent variables). A given variable is considered to be important in the classification process, or for carrying information about the class, if the variable is involved in the classifying objects (in the training set, see Fig. 4).

In the analysed problem, the predictor importance ranking (the highest values represent the greatest impact) indicates that the most influential variables are: the equivalent income, possession of a driving licence, the socio-economic group, the level of education, and the type of household.

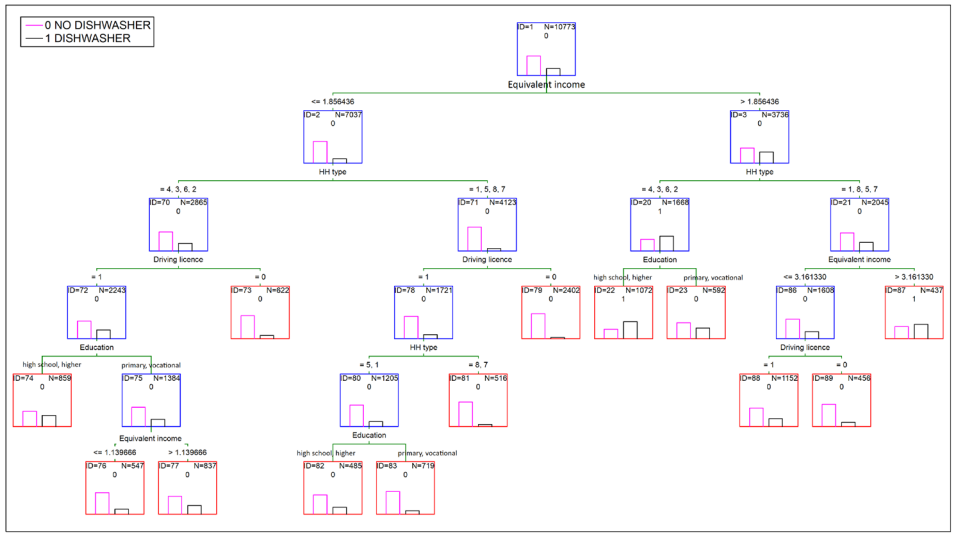

Figure 5. Classification tree – interactive CART

Source: own calculations (Statistica)

The tree was constructed using *V-fold cross-validation* (tenfold), and on this basis the depth of the tree was determined automatically. The surveyed population is dominated by households that do not possess the product. Since classification trees can be used to analyse sets of observations with a complex structure, they are a valuable tool for analysis of households with durable goods. Additionally, interpreting the results intuitively, it is possible to identify the most important characteristics of the surveyed households. To illustrate the interpretation mode, the node 22 ($N = 1022$) is used. In this node, households with a dishwasher dominate. These are multi-family households

whose head has a secondary or higher education, is often married with 1, 2 or 3 children, and has the monthly income equivalent higher than PLN 1,865 (see Fig. 5).

7. Final remarks

Measurement of endowment of households with durable goods cannot be limited to registering the number of items. Their quality, age, value, etc. have to be taken into consideration as well. Subjective assessment of the situation is often involved when measurement (data collection) is conducted.

An important issue which has to be considered is the reason why a household purchases a certain durable product. The assessments of necessity, striving for status manifestation or seeking luxury feeling are the most common motivation types.

The nature of durables requires diversity of measuring scales, including the fuzzy approach. As a result, a mixture of metric and non-metric data is to be analysed. Such data require specialised analytical tools, designed for non-metric data, a mixture of metric and non-metric data, or fuzzy data analysis techniques.

References

Abdullah L., Khadiah S. (2011), *Fuzzy Linguistic for Measuring Customer Satisfaction*, "International Journal of Latest Trends in Computing", no. 2(2), pp. 220–224.

Adams E., Fagot R., Robinson R. (1965), *A theory of appropriate statistics*, "Psychometrika", vol. 30, pp. 99–127.

Amendola N., Vecchi G. (2014), *Durable Goods and Poverty Measurement*, World Bank Policy Research Working Paper No. 7105, Washington.

Blackorby C., Donaldson D. (1987), *Welfare Ratios and Distributionally Sensitive Cost-Benefit Analysis*, "Journal of Public Economics", no. 34, pp. 265–290.

Blackorby C., Donaldson D. (1988), *Money Metric Utility. A Harmless Normalization?*, "Journal of Economic Theory", no. 46, pp. 120–129.

Borovicka A. (2014), *Fuzzy Weights Estimation Method Based on the Linguistic Expression of Criteria Relevance*, "Economic Review. Central European Review of Economics", no. 17, pp. 13–23.

Breiman L., Friedman J., Olshen R., Stone C. (1984), *Classification and regression trees*, Chapman & Hall, London.

Chu K. (2010), *The Application of Fuzzy Linguistic Scale on Internet Questionnaire Survey*, "International Journal of Organisation and Innovation", no. 2(4), pp. 1–10.

Deaton A. (1997), *The Analysis of Household Surveys. Microeconometric Analysis for Development Policy*, Johns Hopkins University Press, Baltimore.

Deaton A., Zaidi S. (2002), *Guidelines for Constructing Consumption Aggregates for Welfare Analysis*, Living Standards Measurement Study World Bank Working Paper no. 135, Washington, https://openknowledge.worldbank.org/handle/10986/14101 [accessed: 12.10.2016].

Diewert W. (2004), *Durables and User Costs in: Consumer Price Index Manual. Theory and Practice*, chapter 23, ILO, IMF, OECD, UNECE, Eurostat, World Bank.

Diewert W. (2009), *Durables and Owner-Occupied Housing in a Consumer Price Index*, [in:] W. Diewert, J. Greenlees, C. Hulten (eds.), *Price Index Concepts and Measurements*, University of Chicago Press, Chicago.

Dziechciarz-Duda M. (2007), *Klasyfikacja konsumentów na rynku dóbr trwałego użytkowania w Polsce*, "Taksonomia", no. 13, pp. 389–396.

Dziechciarz-Duda M., Król A. (2017), *An application of multivariate statistical analysis for the valuation of durable goods brands*, "Statistics in Transition New Series", vol. 18, no. 1, pp. 75–90.

Dziechciarz J. (2004a), *O możliwości doskonalenia polityki cenowej przy pomocy metody conjoint measurement oraz regresji hedonicznej*, [in:] M. Rószkiewicz (ed.), *Identyfikacja struktur rynkowych. Pomiar, modelowanie, symulacja*, SGH, Warszawa.

Dziechciarz J. (2004b), *Regresja hedoniczna: próba wskazania obszarów stosowalności*, [in:] A. Zeliaś (ed.), *Przestrzenno-czasowe modelowanie zjawisk gospodarczych*, Akademia Ekonomiczna, Kraków.

Dziechciarz J., Dziechciarz-Duda M., Przybysz K. (2010), *Household Possession of Consumer Durables on Background of Some Poverty Lines*, [in:] H. Locarek-Junge, C. Weihs (eds.), *Classification as a Tool for Research*, Springer, Berlin, pp. 735–742.

Fisher I. (1911), *The Purchasing Power of Money, its Determination and Relation to Credit, Interest, and Crises*, McMillan, New York.

Framework for Statistics on the Distribution of Household Income, Consumption and Wealth (2013), OECD, Paris.

Gatnar E., Walesiak M. (2011), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, Beck, Warszawa.

General household survey (2014), Statistics South Africa, Pretoria, https://www.statssa.gov.za/publications/p0318/p03182013.pdf [accessed: 28.09.2016].

Gillingham R. (1983), *Measuring the Cos of Shelter for Homeowners. Theoretical and Empirical Considerations*, "The Review of Economics and Statistics", no. 2(65), pp. 254–265.

Goodhart C. (2001), *What Weight should be given to Asset Price in Measurement of Inflation?*, "The Economic Journal", vol. 111, pp. 335–356.

Hulten C., Wykoff F. (1981), *The Measurement of Economic Depreciation*, [in:] C. Hulten (ed.), *Depreciation, Inflation, and the Taxation of Income from Capital*, Urban Institute Press, Washington.

Jorgenson D. (1963), *Capital Theory an Investment Behaviour*, "American Economic Review", no. 53, pp. 247–259.

Katz A. (1983), *Valuing the Services of consumer Durables*, "Review of Income and Wealth", no. 29(4), pp. 405–427.

Keynes J. (1936), *The General Theory of Employment, Interest, and Money*, Harcourt, New York.

Kramer J. (1993), *Konsumpcja. Prawidłowości, struktura, przyszłość*, PWE, Warszawa.

Kramer J. (1997), *Konsumpcja w gospodarce rynkowej*, PWE, Warszawa.

Lanjouw P. (2009), *Constructing a Consumption Aggregate for the Purpose of Welfare Analysis. Principles, Issues and Recommendations Arising from the Case of Brazil)*, Paper prepared in the World Bank for the OECD/University of Maryland Conference entitled: *Measuring Poverty, Income Inequality and Social Exclusion: Lessons from Europe*, March 16/17, Paris.

Liou T., Wang M. (1994), *Subjective Assessment of Mental Workload. A Fuzzy Linguistic Multi criteria Approach*, "Fuzzy Sets and Systems", no. 62(2), pp. 155–165.

Lippe P. von der (2007), *Index Theory and Price Statistics*, Lang Publishing, Oxford.

Mayntz R., Holm K., Hübner P. (1985), *Wprowadzenie do metod socjologii empirycznej*, PWN, Warszawa.

Moulton B. (2004), *The System of National Accounts for the New Economy. What Should Change?*, "Review of Income and Wealth", no. 50(2), pp. 261–278.

Odekon M. (ed.) (2015), *Encyclopaedia of World Poverty*, Sage, Thousand Oaks.

Offer A. (2005), *The Challenge of Affluence Self-Control and Well-Being in the United States and Britain since 1950*, Oxford University Press, Oxford.
Ravallion M. (1998), *Poverty Lines in Theory and Practice. Living Standards Measurement Study*, LSMS Working Paper no. 133, World Bank, Washington, http://documents.worldbank.org/curated/en/916871468766156239/pdf/multi-page.pdf [accessed: 28.09.2016].
Rószkiewicz M. (2002), *Narzędzia statystyczne w analizach marketingowych*, C.H. Beck, Warszawa.
Samuelson P. (1974), *Complementarity. An Essay on the 40th Anniversary of the Hicks – Allen Revolution in Demand Theory*, "Journal of Economic Literature", no. 15, pp. 24–55.
Slesnik D. (2000), *Consumption and Social Welfare. Living Standards and Their Distribution in the United States*, Cambridge University Press, Cambridge.
Social Diagnosis (2015), Council for Social Monitoring, Social Diagnosis integrated database, www.diagnoza.com [accessed: 11.05.2016].
Stevens S. (1959), *Measurement, Psychophysics and Utility*, [in:] C. Churchman, P. Ratoosh (eds.), *Measurement; Definitions and Theories*, Wiley, New York.
System of National Accounts (2008), Commission of the European Communities, IMF, UN, World Bank, Brussels–Luxembourg–New York–Paris–Washington.
Walesiak M. (1996), *Metody analizy danych marketingowych*, PWN, Warszawa.
Walesiak M. (2000), *Skale pomiaru zmiennych*, [in:] M. Walesiak, A. Bąk, *Conjoint analysis w badaniach marketingowych*, Akademia Ekonomiczna, Wrocław.
Zadeh L. (1975), *The Concept of a Linguistic Variable and its Application to Approximate Reasoning*, "Information Sciences", part I, no. 8, pp. 199–249, part II, no. 8, pp. 301–357.

Analiza danych niemetrycznych w badaniu wyposażenia gospodarstw domowych w dobra trwałe. Wybrane aspekty

Streszczenie: Pomiar wyposażenia gospodarstw domowych jest kluczowy w wielu aspektach oceny stanu społeczno-ekonomicznego kraju i jego obywateli. Zapotrzebowanie (sprzedaż) jest traktowane jako jeden z kluczowych wskaźników stanu koniunktury w gospodarce. Podobnie analiza i ocena poziomu wyposażenia gospodarstw domowych w dobra trwałe są rozpatrywane w kontekście pomiaru jakości życia. W badaniu wyposażenia gospodarstw domowych mierzy się liczbę i jakość dóbr, w jakie wyposażone są gospodarstwa domowe. Pomiar wyposażenia gospodarstw domowych prowadzony jest najczęściej za pomocą słabych skali pomiarowych, nominalnej i porządkowej. Takie dane wymagają stosowania wyspecjalizowanych narzędzi analizy i modelowania. W opracowaniu zostanie podjęta dyskusja o możliwościach statystycznej analizy takich danych i ich modelowania oraz o problemach wnioskowania na podstawie uzyskanych wyników.

Słowa kluczowe: analiza danych niemetrycznych, dobra trwałe, analiza korespondencji, regresja logistyczna, drzewa klasyfikacyjne, CART, gospodarstwa domowe

JEL: C1, F2

Received: 2017-01-14; verified: 2017-05-24. Accepted: 2017-10-11