OVERREPRESENTATION MAPS AS A TOOL TO ANALYSIS OF EXPENDITURE STRUCTURE

Michał Gostkowski✉, Dominik Sieradzki

Warsaw University of Life Sciences – SGGW

ABSTRACT

In this article we made an approach to use tools of the grade data analysis to scrutinize structure of the consumer expenditures. Research material was taken from the Central Statistical Office of Poland database concerning expenditure of Polish households in the period of 1999–2015. Grade data analysis is treated as one of the multivariate data analysis methods and can be perceived as a method that is focused on fine visualization of the data in order to better understanding the interdependence that is contains. To do so, we used so-called overrepresentation maps. The analysis was conducted with taking into account time, class of the locality and socio-economic group. Conducted research showed the direction of changes in the average structure of the consumer expenditure.

Key words: grade correspondence analysis, GCA, structure, consumer goods and services expenditures

INTRODUCTION

Economic situation of Poland is constantly undergoing transformation. It is strictly related to overlapping processes, i.e. economic transformation and ongoing globalization. These processes had a significant impact on economic situation of all economic sectors, including household sector [Olejniczuk-Merta 2016].

Since the beginning of 1990s adaptive processes in the economic and social area have been observed, which led to adopting consumption patterns from American and European Union countries by Polish society. Situation of the Polish consumer changed significantly. Initially the Polish consumer functioned in the economy of shortage, whereas now he needs to make a choice from variety of goods and services, which in some situations may lead to inability to make one. This issue had been previously widely discussed by Schwartz [2013].

Additionally over the last years consumption of the Polish households has been undergoing perpetual changes and fluctuations, therefore constant analysis and observation is needed. Market subjects, including households, are functioning in the environment of various trends and macroeconomic factors, among which we can indicate: economic development, financial politics, inflation and level of redundancy. Moreover, the ongoing social changes, i.e. an influence of other cultures, constant raise the knowledge level, rational nutrition or simplified consumption, play also quite an important role. All of the above factors undoubtedly have enormous impact on building final household expenditure structure.

The purpose of this article is to show the most important changes and trends in the consumer expenditures structure in Polish households in the period of 1999–2015. The analysis was conducted for all Polish

✉ michal_gostkowski@sggw.pl

households and in distinction by both individual socio-economic groups and class of locality.

EMPIRICAL DATA

The empirical material used in this article is data taken from the Central Statistical Office of Poland concerning budget of the Polish households in the period of 1999–2015. It represents basic information source regarding revenue, consumption and expenditures within Polish households.

Households' expenditures were divided into groups according to the classification of expenditures based on COICOP/HBS [GUS 2011]. Finally there were 12 groups of expenditures received: food and non-alcoholic beverages (FOOD), alcoholic beverages, tobacco and narcotics (ALCO), clothing and footwear (CLOTH), housing, water, electricity, gas and other fuels (HOME), furnishings, household equipment and routine maintenance of the house (EQUIP), health (HEALTH), transport (TRAN), communication (COMMU), culture and recreation (CULT), education (EDUC), restaurants and hotels (HOTE), miscellaneous goods and services (OTHER).

METHODS

In order to better illustrate changes in the average structure of the consumer expenditures the grade correspondence analysis (GCA) was used [Kowalczyk et al. 2004, Gostkowski et al. 2015]. Grade correspondence analysis is a part of multidimensional data exploration [Borkowski and Szczesny 2005]. It represents a very important stage in order to reach statistics integration, data explorations, taxonomy, measurement theory and equal treatment of both continuous and discrete data. It can be seen as a method that is focused on very good visualisation of the data in order to better understanding the comprehensive data structure and its interdependencies [Borkowski et al. 2008]. All calculations were performed in GradeStat software developed and supported at the Institute of Computer Science Polish Academy of Sciences. Detailed description can be found at website http://gradestat.ipipan.waw.pl.

In this article the structure of given unit is understood as a vector of non-negative values in n dimensional Euclidean space, which sum of coordinates equals unity, where number of n dimensions is a number of groups concerning consumer goods and services. To evaluate any structural similarities between two units different similarity measures can be used [Kukuła 2000, Zeliaś 2000, Borkowski and Szczesny 2002, Malina 2004]. Among them we can distinguish concentration curve and associated concentration index [Koszela 2016, Koszela and Szczesny 2017]. Concentration curve of the distribution $q = (q_1, \ldots , q_k)$ in relation to distribution $p = (p_1, \ldots , p_2)$ is polygonal chain in the unit square joining points $(0; 0)$, (p_1, q_1), $(p_1 + p_2, q_1 + q_2)$, $\ldots$, $(p_1 + \ldots + p_k, q_1 + \ldots + q_k)$, $(1; 1)$. Slope of the further sections of the polygonal chain to OX axis shows quotient q_i / p_i. Quotient q_i / p_i can run quite freely. Concentration curve received after shifting co-ordinates of vectors q and p in such a way, that next quotients are non-decreasing is called the maximum concentration curve, while corresponding concentration index is called the maximum concentration index ar_{max}. Formally, it could be put in a form of pattern:

$$ar = 1 - 2 \cdot \int_0^1 C(t)dt \qquad (1)$$

where: C – concentration curve of the distribution q in relations to p.

The maximum concentration index and the distribution differentiation curves are basic tools of grade data analysis – GDA [Ząbkowski and Szczesny 2012]. Grade correspondence analysis algorithm strives to set rows and columns in the standardized P table ($\mathbf{P} = \{p_{ij} : i = 1, k, \ldots, m, j = 1, k, \ldots, k\}$, $\sum p_{ij} : i = 1, p_{ij} = 1$, $p_{ij} \geq 0$) in such a way, that all the ar index equal ar_{max}. It also should be noticed, that maximizing ar for one pair of variables causes decrease the ar value for other pairs of variables. Hence GCA algorithm attempts to make reached ar indexes are as close as possible to ar_{max}. To achieve that, GCA changes arrangement of rows and columns in the table of m rows and k columns at every step, trying to maximize the Spearman's rank correlation coefficient $\rho*$:

$$\rho* = 3 \sum_{i=1}^{m} \sum_{s=1}^{k} \left\{ \left(p_{is} \left[2S_{row}(i) - 1 \right] \left[2S_{col}(s) - 1 \right] \right) \right\} \qquad (2)$$

where:

$$S_{row}(i) = \left(\sum_{j=1}^{i-1} p_{j+} \right) + \frac{1}{2} p_{i+}$$

$$S_{col}(s) = \left(\sum_{y=1}^{s-1} p_{+t} \right) + \frac{1}{2} p_{+s}$$

$$p_{j+} = \sum_{s=1}^{k} p_{js}$$

$$p_{+t} = \sum_{n=1}^{m} p_{nt}$$

The number of possible arrangements of rows and columns is finite and equals $m!*k!$. After the GCA algorithm is over, elements located in both rows and columns are ordered in relation to suitable hidden variables. Strong correlation occurs, when there is a strong monotonic association between those two hidden variables. A very important quality of GCA is setting similar rows (columns) next to each other.

An important advantage of this method compared to commonly known methods of cluster analysis [Ostasiewicz 1999, Zeliaś 2000] is a simultaneous presentation of features and facilities in overrepresentation maps (Fig. 1). The idea is to show various structures in the background of the average structure. To create an overrepresentation map in the first place, based on the standardized table P, h_{ij} overrepresentation indexes are determined for each table cell:

$$h_{ij} = \frac{p_{ij}}{p_{i+} p_{+j}} \tag{3}$$

Overrepresentation index indicates the extent to which observed value differs from what would be expected from ideal proportionality distribution (i.e. when there exists no relationship between the rows and columns). For such a set of overrepresentation indexes a map showing the degree of data representation can be created (Fig. 1).

RESULTS AND DISCUSSION

The level and households expenditures structure strongly determines living situation of the households [Zalega 2011]. Tendency to consumption reflects in the level and structure of the consumer goods expenditures [Stanislawska and Majchrzak 2009], therefore evaluation of consumer expenditures development is valuable source of information for both the theorists, as well as practitioners. Based on information showed in the table it can be stated, that the consumer expenditures in Polish households systematically increased, reaching the amount of PLN 1,091.19 per person in 2015. After eliminating influence of the inflation, change of the expenditures is not as substantial anymore. It can be even noticed, that after the 2009, level of consumer expenditures stabilized on the level of approximately PLN 650 Polish per person, which could
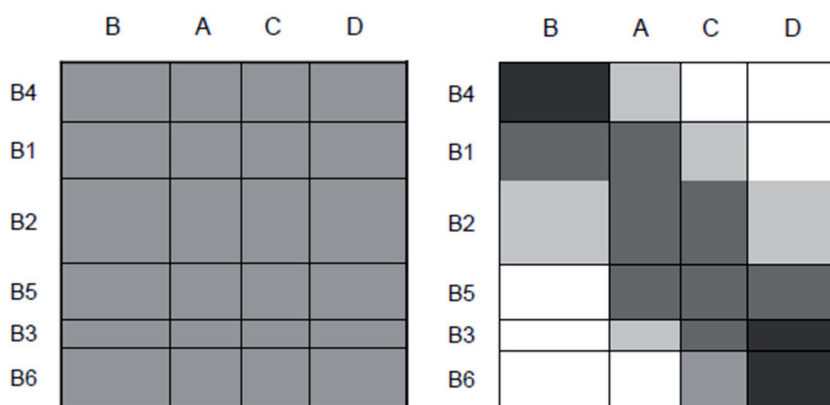
Fig. 1. Examples of overrepresentation maps in the absence of the relationship between the rows and columns (left) and otherwise (right)

Source: Own research.

be an effect of the economic slowdown. On the other hand, in 2012 re-growth of the average monthly consumer expenditures can be noticed.

Table. Average monthly expenditures for consumer goods and services per person in households and values of the consumer goods and services price indicator in relations to the year of 1999

Year	Nominal expenditure (PLN)	Real expenditure (PLN)	CPI
1999	530.15	530.15	1.000
2000	577.62	524.63	1.101
2001	585.72	504.26	1.162
2002	599.20	506.24	1.184
2003	648.74	543.75	1.193
2004	665.63	539.04	1.235
2005	660.67	524.02	1.261
2006	712.56	559.58	1.273
2007	755.58	578.89	1.305
2008	865.32	636.24	1.360
2009	913.86	649.21	1.408
2010	945.80	654.87	1.444
2011	971.83	645.16	1.506
2012	1 005.19	643.49	1.562
2013	1 061.70	673.61	1.576
2014	1 078.74	684.42	1.576
2015	1 091.19	698.60	1.562

Source: GUS [1999–2015].

In order to better illustrate changes in the average expenditures structure, an overrepresentation map showing expenditures structures in each period was used (Fig. 2). The determinant factors were average monthly expenditures (represented as structure) for consumer goods and services per person in Polish households.

When analyzing the overrepresentation map (Fig. 2), it can be stated, that the structure of average expenditures in Polish households underwent substantial changes. The biggest share in the budget in the analysed period had food and non-alcoholic beverages expenditures (FOOD), housing, water, electricity, gas

and other fuels (HOME) and transport expenditures (TRAN) (column width). The darker hue shows overrepresentation and points out, that share of the consumer expenditures in the analyzed period has set over the average level of the expenditures calculated for the whole researched period. Taking that into account it can be stated, that in the period of 1999–2005 food and non-alcoholic expenditures (FOOD) were overrepresented – which means, that share of the expenditures in this period formed above the average level of the expenditures' share. To summarize – food and non-alcoholic beverages (FOOD) in the period of 1999–2005 had the biggest share in the expenditures of households.
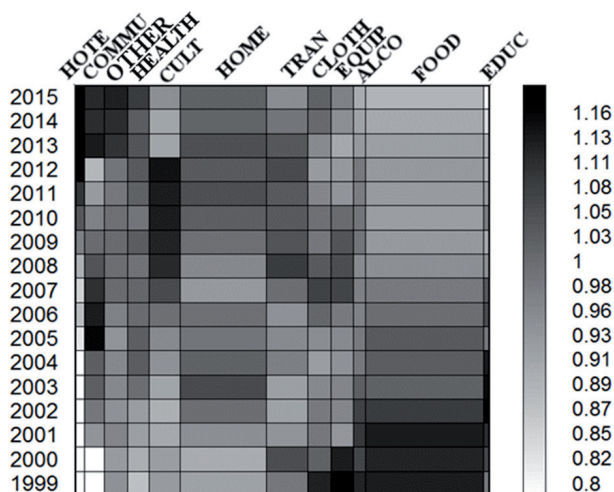
Fig. 2. Overrepresentation map for average monthly expenditure on consumer goods and services per person in Poland

Source: GUS [1999–2015].

Nevertheless, at the end of the analysed period an increase in the housing, water, electricity, gas and other fuels (HOME) and transport expenditures (TRAN) can be noticed. At the same time a decrease in share of the food and non-alcoholic beverages (FOOD) can be noticed, which can mean, that food and non-alcoholic beverages at the end of the analyzed period had no longer such a big share in the household expenditures structure as at the beginning of the analyzed period. An increase in housing, water, electricity, gas and other fuels expenditure (HOME) can

be explained with systematic increase of the various housing, water, electricity, gas and other fuels prices [Stolarska 2009, Świetlik 2014]. An increase of the expenditures connected with transport can be an effect of an increase of the people's mobility thanks to freedom of mobility ensured after Poland joined the EU [Hoszman 2013].

Additionally it can be noticed, that at the end of the analyzed period, there has been a systematic increase of the meaning of the expenditures for restaurants and hotels (HOTE), communication (COMMU), miscellaneous goods and services (OTHER) and health expenditures (HEALTH). What is more, during the period of 2013–2015 an underrepresentation of the expenditures for culture and recreation (CULT) can be noticed. This change can be caused by transferring expenditures for Internet services from the recreation and culture (CULT) group to the communication (COMMU) group, which in the period of 2013–2015 were overrepresented. A part of the observed changes was confirmed in the works of Kuśmierczyk and Piskiewicz [2012], and Piekut [2015], where an analysis of the change of the expenditures structure in chosen European Countries was described.

In the next stage of the research the overrepresentation map for the average expenditures for consumer goods and services in chosen socio-economic groups in 2015 was made (Fig. 3).
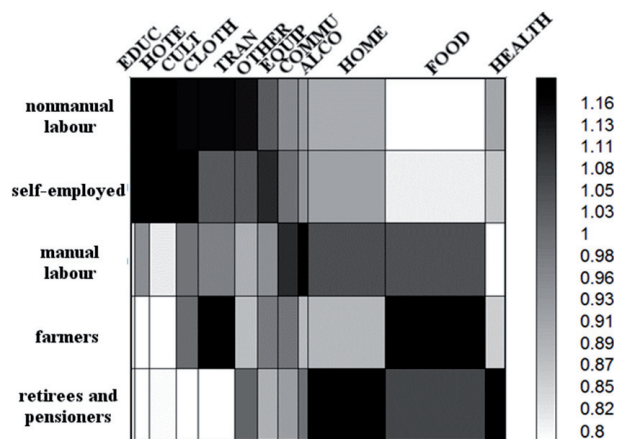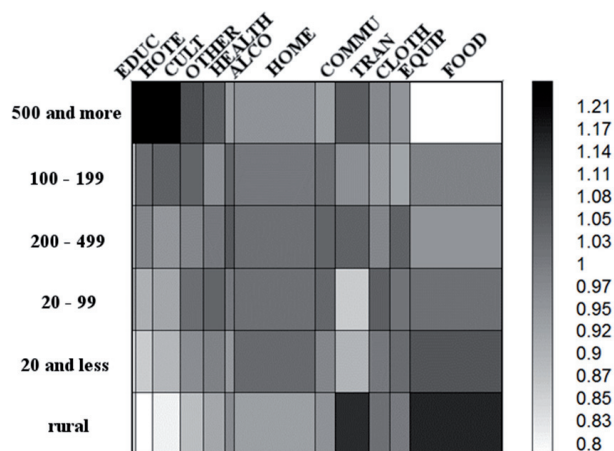

Fig. 3. Overrepresentation map for average monthly expenditure per capita in households by socio-economic groups in 2015 in Poland

Source: Own research based on GUS [2015].

Analysis of the overrepresentation map (Fig. 3) showed, that substantial differences in the average structure of the expenditures for consumer goods and services were noticed between each socio-economic group. The most similar consumer expenditures structures have households of employees in non-manual labour and households of the self-employed. For those two types of households strong overrepresentation of such expenditures groups as education (EDUC), hotels and restaurant (HOTE) or recreation and culture (CULT) may be noticed. It can be stated, that these are expenditures occurring when the revenue level is higher. Nevertheless, food and non-alcoholic beverages expenditures were strongly underrepresented. Different expenditures structure can be noticed in the households of employees in manual labour, farmers and households of retirees and pensioners. Within this group of households expenditures on education (EDUC), hotels and restaurant (HOTE) or recreation and culture (CULT) were strongly underrepresented. What is more, in this type of group of household strong overrepresentation of the food and non-alcoholic beverages (FOOD) expenditures can be observed – especially substantial for the farmers' households. Moreover, farmers' households show strong overrepresentation of the transport expenditures (TRAN). Conducted researched also proved, that alcohol, tobacco products and narcotics expenditures were severely higher in the group of households of employees in manual labour than in the other socio-economic groups.

Further analysis of the average consumer expenditures was performed with taking into account grouping households by class of locality (Fig. 4).

Conducted research showed, that there are no substantial differences in the average structure of the expenditures for consumer goods and services between each of the locality classes. The greatest differences in the average structure of the expenditures for consumer goods and services can be observed between households in cities with more than 500,000 citizens and households located in the rural areas – what arises from natural conditions. Additionally, households located in the rural areas showed overrepresentation of the expenditures for transport (TRAN) – what was acknowledged in the previous example – and expenditures for food and non-alcoholic beverages

Fig. 4. Overrepresentation map for average monthly expenditure per capita in households by class of locality in 2015 in Poland

Source: Own research based on GUS [2015].

(FOOD). By contrast, very similar average structure of the consumer expenditures can be shown for other household classes. Without a doubt, their common feature is higher expenditure on housing, electricity, gas and other fuels (HOME).

CONCLUDING REMARKS

Conducted research showed, that in the analyzed period average expenditures on consumer goods and services were systematically growing to PLN 1,092 per person. Additionally, the largest share in the average expenditure on consumer goods and services is expenditures for food and non-alcoholic beverages. What is more, by the end of the analysed period it can be noticed that expenditures for restaurants and hotels (HOTE), communication (COMMU) and health (HEALTH) were overrepresented. The most similar structures of the average expenditures for consumer goods and services had households of employees in nonmanual labour and households of the self-employed. Moreover, the most different structures were the structures of households in cities with more than 500,000 residents and households located in rural areas.

Conducted research showed, that the GCA is a useful tool when analyzing changes in the average structure of expenditures for consumer goods and services. Use of this method allows analysis of the structures with usage of so-called overrepresentation maps, which deliver additional information about nature of the analyzed phenomenon.

REFERENCES

Borkowski, B., Dudek, H., Szczesny, W. (2008). Spatial Differentiation of Food Production Structure and Consumption Profile in Enlarged European Union. Journal of Economic Asymmetric, 5 (2), 145–156.

Borkowski, B., Szczesny, W. (2002). Metody taksonomiczne w badaniach przestrzennego zróżnicowania rolnictwa. Roczniki Nauk Rolniczych PAN, Seria G, 89 (2), 11–20.

Borkowski, B., Szczesny, W. (2005). Metody wizualizacji danych wielowymiarowych jako narzędzie syntezy informacji. Roczniki Naukowe SERiA, 7 (5), 11–18.

GUS (2011). Metodologia badania budżetów gospodarstw domowych. Zeszyty Metodyczne i Klasyfikacje. Zakład Wydawnictw Statystycznych, Warszawa.

GUS (1999–2015). Budżety gospodarstw domowych [Household Budget Survey]. Zakład Wydawnictw Statystycznych, Warszawa.

Gostkowski, M., Jałowiecki, M., Tekień, A. (2015). Regional Diversification of Expenditure Structure in Poland: A GCA Approach. Zeszyty Naukowe SGGW w Warszawie – Problemy Rolnictwa Światowego, 15 (4), 71–79.

Hoszman, A. (2013). Konsumpcja usług transportowych na tle przemian gospodarczych w Polsce w latach 2003––2012. Konsumpcja i Rozwój, 2 (5), 63–77.

Koszela, G. (2016). Wykorzystanie gradacyjnej analizy danych do klasyfikacji podregionów pod względem struktury agrarnej. Wiadomości Statystyczne, 6, 10–30.

Koszela, G., Szczesny, W. (2017). Changes in Food Waste Level in The EU Countries. Acta Scientiarum Polonorum Oeconomia, 16 (1), 43–52.

Kowalczyk, T., Pleszczeńska, E., Ruland, F. (2004). Grade models and methods for data analysis. Springer, Berlin.

Kukuła, K. (2000). Metoda unitaryzacji zerowej. PWN, Warszawa.

Kuśmierczyk, K., Piskiewicz, L. (2012). Konsumpcja w Polsce na tle pozostałych krajów Unii Europejskiej. Konsumpcja i Rozwój, 2 (3), 78–93.

Malina, A. (2004). Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarski Polski według województw. Monografie 162. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

Olejniczuk-Merta, A. (2016). Innovations and consumption versus raising the quality of life or society. Acta Scientiarum Polonorum Oeconomia, 15 (3), 91–100.

Ostasiewicz, W. (1999). Statystyczne metody analizy danych. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.

Piekut, M. (2015). The Rich North-West, The Poor Middle-East – Consumption in EU Households. Comparative Economic Research, 18 (1), 43–63.

Schwartz, B. (2013). The Paradox of choice. Why More is Less. PWN, Warszawa.

Stanisławska, J., Majchrzak, A. (2009). Preferencje konsumpcyjne w zakresie artykułów żywnościowych w gospodarstwach domowych rolników na tle pozostałych grup społeczno-ekonomicznych ludności. Roczniki Naukowe SERiA, 11 (3), 319–324.

Stolarska, A. (2009). Tendencje i zróżnicowanie regionalne poziomu oraz struktury konsumpcji w Polsce w latach 2000–2007. Roczniki Naukowe SERiA, 11 (3), 336––341.

Świetlik, K. (2014). Dylematy rozwoju konsumpcji żywności w warunkach niestabilnej koniunktury. Konsumpcja i Rozwój, 3, 16–30.

Zalega, T. (2011). Spożycie żywności w gospodarstwach domowych z osobami bezrobotnymi w województwie mazowieckim. Zeszyty Naukowe SGGW w Warszawie Ekonomika i Organizacja Gospodarki Żywnościowej, 93, 119–135.

Ząbkowski, T., Szczesny, W. (2012). Badanie atrakcyjności oferty dostępu do internetu za pomocą analizy gradacyjnej. Metody Ilościowe w Badaniach Ekonomicznych, 13 (3), 276–287.

Zeliaś, A. (2000). Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym. Wydawnictwo Akademii Ekonomicznej, Kraków.

MAPY NADREPREZENTACJI JAKO NARZĘDZIE DO ANALIZY STRUKTURY WYDATKÓW KONSUMPCYJNYCH

STRESZCZENIE

W pracy podjęto próbę wykorzystania narzędzi gradacyjnej analizy danych do analizy struktury wydatków konsumpcyjnych. Materiałem badawczym były dane Głównego Urzędu Statystycznego dotyczące wydatków gospodarstw domowych w Polsce w latach 1999–2015. Gradacyjna analiza danych jest zaliczana do metod wielowymiarowej analizy danych i można ją wykorzystywać jako technikę wizualizacji danych umożliwiającą klarowne pokazanie ich współzależności. W tym celu wykorzystano tzw. mapy nadreprezentacji. Analizę przeprowadzono z uwzględnieniem czasu, klasy miejscowości zamieszkania oraz grupy społeczno-ekonomicznej. Przeprowadzone badania wskazały kierunek zmian w przeciętnej strukturze wydatków konsumpcyjnych.

Słowa kluczowe: gradacyjna analiza danych, GCA, struktura, wydatki na dobra i usługi konsumpcyjne