Michał Gostkowski[1]
Department of Econometrics and Statistics
Warsaw University of Life Sciences - SGGW
Piotr Jałowiecki[2]
Department of Informatics
Warsaw University of Life Sciences - SGGW
Agnieszka Tekień[3]
Warsaw University of Life Sciences - SGGW

Regional Diversification of Expenditure Structure in Poland: a GCA Approach

Abstract. The paper presents the results of the analysis of household expenditure structures in Poland using Grade Correspondence Analysis (GCA). Studies have shown that in the analyzed period the structure of consumer expenditure changed. The share of food and non-alcoholic beverages expenditures decreased steadily, while the share of expenses related to maintenance and household equipment, culture and recreation, or restaurants and hotels steadily increased. Further studies have shown that the structure of consumer expenditures in Poland is diversified regionally. The greatest variation occurs between the north-eastern region voivodeships and the central region/south-west voivodeships.

Key words: Grade Correspondence Analysis, expenditure structure, clustering, overrepresentation maps, Grade Stat, Poland

Introduction

From the economic perspective, consumption is described as a process of satisfying human needs, which is closely related to socio-economic development. In addition, consumption plays an important role as a stimulator of economic growth and social development [Świetlik 2012]. The impact of consumption on the country's economic activity is quantified in the System of National Accounts as a component of the Gross Domestic Product (GDP) [Begg et al. 2007]. In Poland, the share of private consumption in the structure of GDP is around 60% [Świetlik 2012].

Household incomes are one of the most important factors determining the level and the structure of consumer spending [Dudek et al. 2012]. In recent years, consumption in Polish households was characterized by constant changes and fluctuations. The increase in household income led to spending increases on consumption goods like food and non-food items, while the share of food expenditures in total expenditures decreased [Badach 2012]. Besides income levels there are also other important non-economic factors like: demographic, social, cultural or psychological factors [Utzig 2013]. The main objective of this study was to identify changes in the structure of consumer expenditures in Polish households during the period from 1999 to 2012. The secondary objective was to verify

---

[1] MSc, e-mail: michal_gostkowski@sggw.pl
[2] PhD, e-mail: piotr_jalowiecki@sggw.pl
[3] MSc, e-mail: agnieszka.tekien@gmail.com

whether the structure of consumer expenditure in Polish households is regionally diversified.

Data

This study analyses the changes in the household expenditures structure in Poland over the period from 1999 to 2012. Data provided by the Central Statistical Office in Statistical Yearbooks of the Regions were used in the study. Expenditures refers to average monthly household expenditures per capita [4].

Household expenditures were divided into groups according to the classification of expenditures based on COICOP/HBS[5] [GUS 2011]. Twelve groups of expenditures were established: 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).

Methodology

The method used in the studies was Grade Correspondence Analysis (GCA) [Kowalczyk et al. 2004, Szczesny et al. 2012]. Grade Correspondence Analysis is an original method of data mining developed and supported at the Institute of Computer Science of the Polish Academy of Sciences. A wide variety of tools and software to support data analysis can be found at http://gradestat.ipipan.waw.pl/. An important feature of GCA is the fact that the GCA does not build a new synthetic measure but takes into account the original structure of the phenomenon.

To compare the structure of expenditures the concentration curve was used. The concentration distribution curve $\mathbf{q} = (q_1, ..., q_k)$ relative distribution $\mathbf{p} = (p_1, ..., p_2)$ is a polygonal chain in a unit square that connects the points $(0, 0)$, (p_1, q_1), $(p_1 + p_2, q_1 + q_2)$, ..., $(p_1 + ... + p_k, q_1 + ... + q_k) = (1, 1)$. The slope of the next segments of the polygonal chain to the x-axis determines the quotient q_i/p_i (the slope of the corresponding line). Quotients q_i/p_i can run freely. The concentration curve is obtained after moving coordinate vectors $\mathbf{q}$ and $\mathbf{p}$ so that successive quotients are non-decreasing – this is called the maximum concentration curve, whereas the corresponding concentration ratio is called the index of the maximum concentration ar_{max} [Binderman et al. 2010].

Ar concentration ratio is equal to twice the difference of two squares limited concentration curve and the diagonal system: located below and above the diagonal. Formally, this can be written as [Borkowski et al. 2009]:

---

[4] It should be stressed that data reliability depends on the accuracy of data, methods of data collection and data availability.

[5] COICOP-HBS is Classification of Individual Consumption by Purpose Adapted to the Needs of Household Budget Surveys.

$$ar = 2\int_0^1 (u - C(u))du, \tag{1}$$

where $C(u)$ is a function describing the concentration curve.

Ar_{max} concentration ratio and concentration curve are fundamental tools of Grade Correspondence Analysis. Grade Correspondence Analysis belongs to the exploratory data analysis (data mining) and is perfectly compatible with the variables measured on different scales [Ząbkowski and Szczesny 2012, Zmarzłowski and Koszela 2013]. GCA algorithm aims to get such set of rows and columns in a Table P where all indicators *ar* are equal to ar_{max}. It should be noted that maximizing *ar* per pair of variables reduces the value of *ar* for the other pairs of variables. Therefore GCA algorithm aims to ensure that the achieved indicators were as close as possible to ar_{max}. For this purpose the GCA reorders the rows and columns of the table with *m* rows and *k* columns to maximize the Spearman rank correlation coefficient $\rho*$ at each step:

$$\rho* = 3\sum_{i=1}^{m}\sum_{s=1}^{k}(p_{is}(2S_{row}(i)-1)(2S_{col}(s)-1)), \tag{2}$$

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

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

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

$p_{+t} = \sum_{t=1}^{m} p_{ts}.$

A very important property of the GCA is that the algorithm also sets similar rows (columns) close to each other. Similar rows (columns) always occupy contiguous space in optimal permutation. Moreover, maximization of the $\rho*$ ratio is equivalent to the maximum variation of these regressions. In other words, when setting the GCA data table, the rows and columns are always sorted by their similarity. Furthermore, the similarities are as varied between them as it is possible.

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

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

The overrepresentation index indicates the extent to which the observed value differs from that which would be expected from ideal proportionality distribution (i.e. when there exists no relationship between rows and columns). For such a set of overrepresentation

indicators a map showing a degree of data representation can be created. With a few shades of gray areas of underrepresentation, the ideal representation and overrepresentation of the data can be identified. In the paper the areas are identified as follows:

- value below 0.8 indicates a strong underrepresentation,
- value in the range of 0.8-0.98 determines there is poor underrepresentation,
- a value in the range of 0.98-1.02 is an ideal representation,
- value in the range of 1.02-1.2 determines there is poor overrepresentation,
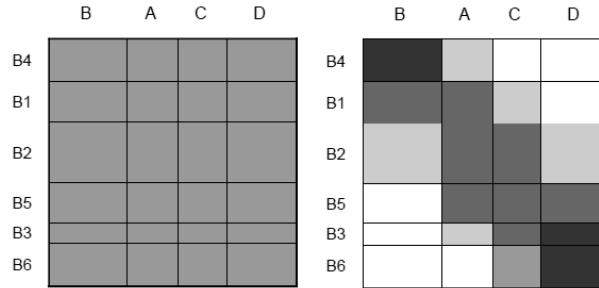- value greater than 1.2 means a strong overrepresentation.

Fig. 1. Examples of overrepresentation maps in the absence of the relationship between the rows and columns (left) and otherwise (right).

Source: Own research.

Results

In the first stage of the study using the method GCA, an overrepresentation map of Poland was created showing the structure of expenditures in the years 1999-2012 (Figure 2). Factors determining the division (or assignment) were average monthly household expenditures per capita.
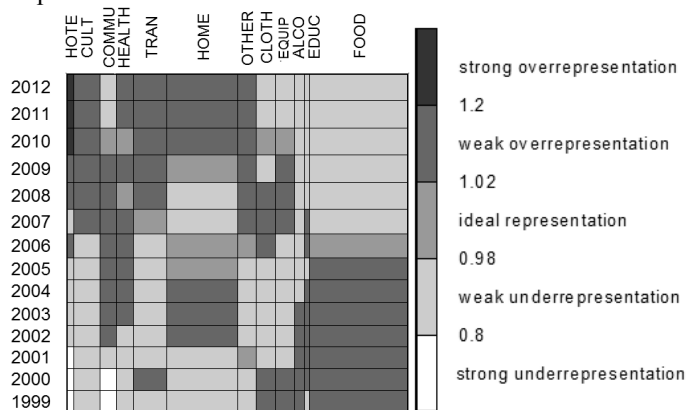

Fig. 2. Overrepresentation map for average monthly expenditure on consumer goods and services per person in Poland.

Source: own research based on Statistical Yearbook of the Regions 1999-2012.

As was mentioned earlier, one of the advantages of the GCA algorithm is setting similar columns (rows) close together. Analyzing the overrepresentation map (presented on Figure 2) it can be seen that on the left side of the array were expenditures related to restaurants and hotels, culture and recreation and communication. It can be argued that these are expenses that occur at higher levels of income and can be described as luxury goods [Gostkowski & et al. 2014]. On the other side there are expenses related to food and non-alcoholic beverages, education and alcoholic beverages, tobacco and narcotics. It is quite remarkable that the structure of food and non-alcoholic beverages consumption is similar to the structure of education as well as to alcoholic beverages, tobacco and narcotics, however expenditures on alcoholic beverages, tobacco and drug-specific products may be underestimated by households, which may be reflected in the results. It also should be mentioned that the education group includes only education tuition fees in schools and kindergartens; all the expenses such as purchasing textbooks and school supplies are in the recreation and culture group.

It can also be concluded that the overrepresentation map (Figure 2) can indicate the years 2005-2006 as a key moment of change in the expenditure structure.

Additionally, in order to illustrate the changes in the structure of expenditures in individual voivodeships overrepresentation maps for four deliberately selected periods were created: 1999, 2004, 2008 and 2012 (Figure 3-6).

Based on the information presented in the graphs (Figures 3-6, Table 1) it can be stated that the structure of Polish household expenditures in the analysed period has changed. The largest share of expenditures was allocated to food and non-alcoholic beverages, but from period to period it has systematically decreased. On this basis, the First Engel's Law, which states that an increase in consumer income leads to a decrease in percentage of food expenditures among total consumer expenditures has been confirmed.
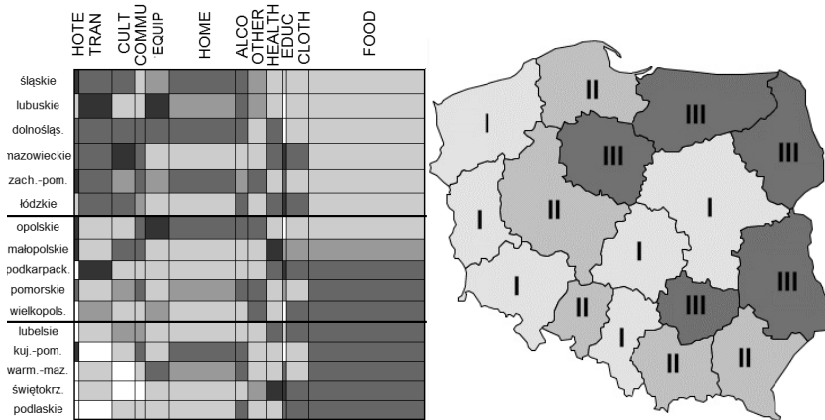

Fig. 3. Overrepresentation map for average monthly expenditure on consumer goods and services per person in each voivodeship in 1999 (left) and Polish territorial map showing the distribution (right).

Source: own research based on Statistical Yearbook of the Regions 1999.

The subsequent expenditure groups characterized by the highest shares were housing, water, electricity, gas and other fuels and those related to transport. Moreover, furnishings, household equipment and routine maintenance expenditures in the final period of the study

were overrepresented, which can be explained by a systematic price increase in various energy sources (Figure 2, Table 1). Additionally, a systematic increase in the share of expenditures on recreation and culture, restaurants and hotels is noticed. Similar changes were observed in the work of Kuśmierczyk and Piskiewicz [2012] about changes in the expenditure structure in European Union countries. The lowest shares in total expenditure are alcoholic beverages, tobacco and narcotics and those related to education (Figure 2).
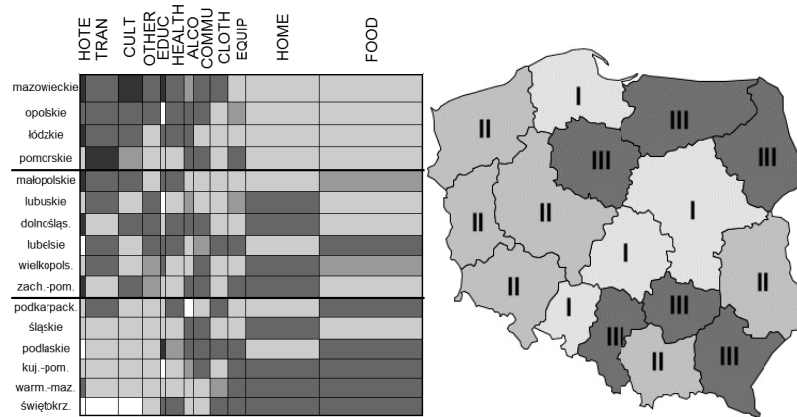

Fig. 4. Overrepresentation map for average monthly expenditure on consumer goods and services per person in 2004 (left) and Polish territorial map showing the distribution (right).

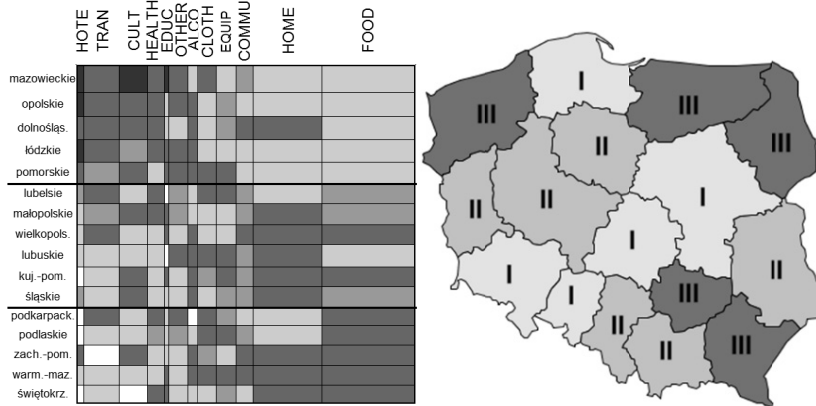Source: own research based on Statistical Yearbook of the Regions 2004.

Fig. 5. Overrepresentation map for average monthly expenditure on consumer goods and services per person in 2008 (left) and Polish territorial map showing the distribution (right).

Source: own research based on Statistical Yearbook of the Regions 2008.

Further study leads to the conclusion that the differentiating factor in the expenditure patterns of individual voivodeships is the income situation of households. Typically, the poorest households spend most of their income for basic needs (expenditures on food and non-alcoholic beverages and solid counter charges like energy, water, etc.), resulting in overrepresentation compared to other expenses groups [Dudek et al. 2012].
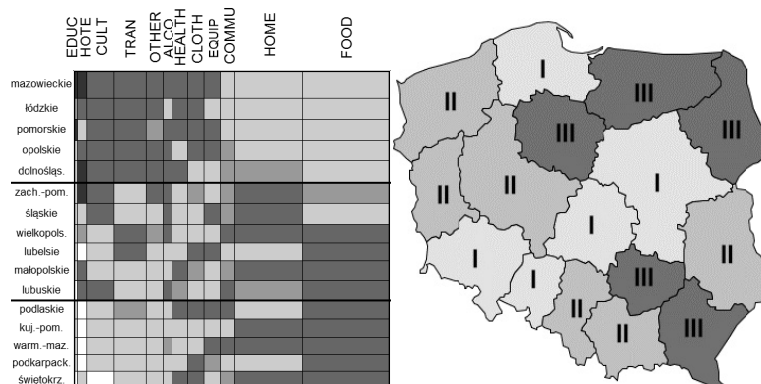
Fig. 6. Overrepresentation map for average monthly expenditure on consumer goods and services per person in 2012 (left) and Polish territorial map showing the distribution (right).

Source: own research based on Statistical Yearbook of the Regions 2012.

Based on the information presented above it can be concluded that the structure of consumer spending is regionally diverse (Figure 6). With the use of the GCCA[6] algorithm, three focus areas (groups) were established containing regions characterized by similar expenditure structures. In the third analyzed focus group **(**Group III) were the eastern and northern voivodeships (Podlaskie, Kujawsko-Pomorskie, Warmińsko-Mazurskie, Podkarpackie, Świętokrzyskie). They are primarily characterized by expenditure overrepresentation in food and non-alcoholic beverages, housing, water, electricity, gas and other fuels. All of them are expenses related to satisfying basic needs. Three voivodeships (Warmińsko-Mazurskie, Podlaskie and Świętokrzyskie) belonged to focus Group III in all the analyzed periods. The other two provinces in earlier periods changed their affiliation.

Table 1. Growth rate of expenditures on particular groups (2012/1999) and the average annual pace rate of expenditures in each group in Poland.

Expenditure group	Growth ratio	Average annual rate of change
food and non-alcoholic beverages	0,99	-0,11%
alcoholic beverages, tobacco and narcotics	1,05	0,36%
clothing and footwear	0,99	-0,09%
housing, water, electricity, gas and other fuels	1,36	2,36%
furnishings, household equipment and routine maintenance of the house	0,93	-0,53%
health	1,42	2,75%
transport	1,29	1,97%
communication	1,76	4,44%
culture and recreation	1,47	3,03%
education	1,09	0,63%
restaurants and hotels	2,52	7,37%
miscellaneous goods and services	1,26	1,77%

Source: own research based on Statistical Yearbook of the Regions 1999-2012.

---

[6] GCCA (Grade Correspondence Cluster Analysis) is the GCA with posterior clustering

In the second area were voivodeships from north-western and southern regions (ie. Lubuskie, Wielkopolskie, Zachodniopomorskie, Małopolskie and Śląskie). These voivodeships were characterized by a very diversified structure of consumer spending. In the first group were voivodeships from the central and south-western regions (ie. Mazowieckie, Łódzkie, Dolnośląskie, Opolskie) and the Pomorskie voivodeship). The income situation of households in this concentration is better when compared to other clusters, because there is overrepresentation of expenditures for expenses unrelated to the satisfaction of basic needs (eg. culture and recreation, restaurants, and hotels).

The division obtained in the analysis may reflect regional differences in consumption patterns mainly determined by the economic situation of the voivodeships. In order to confirm this finding we made a suitable map (Figure 7). Analyzing the map of Poland (Figure 7). it can be stated that provinces qualified to the first focus area are characterized by a relatively higher level of GDP *per capita*. In contrast, in the third focus area were voivodeships where the values of GDP *per capita* are lower. Exceptions are the Lublin and Opole voivodeships, which despite the relatively lower values of GDP *per capita* were selected to the second focus area.
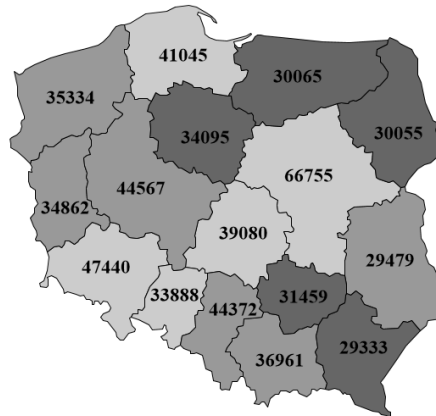

Fig. 7. The map of Poland with applied divisions, created using the GCCA for the year 2012. The values in the graph represent the level of GDP per capita in 2012 in current prices.

Source: own research based on Gross Domestic Product Regional Accounts in 2012.

Summary and Conclusions

Based on the analysis of results presented in the paper, it can be concluded that in the period from 1999 to 2012. the Polish household expenditure structure changed significantly. The largest share of expenditures went for food and non-alcoholic beverages, but over the years, share as a percent of total expenditures for this group shares are in continuous decline. On this basis, it can be stated that the First of Engel's Law has been confirmed. Further studies confirmed the growing share of expenses related to the maintenance and household equipment - which may be the result of rising prices. In addition, studies have shown a growing share of expenditures for culture and recreation, as well as for restaurants and hotels. This may mean that the situation of Polish households

during the period has improved. Obviously this is a preliminary conclusion and requires further study for confirmation.

In addition to changes in the structure of expenditures the analysis showed that in Poland there are regional differences in the structure of consumer expenditures. Differences in the structure of consumer expenditure are most visible between voivodeships from the first and third focus areas. The third focus group includes the voivodeships where expenditures related to the satisfying of basic needs (food and non-alcoholic beverages and solid counter charges) were overrepresented. On the other hand, the first focus area includes voivodeships characterized by a structure of expenditures with overrepresentation of expenditures related rather to higher needs (eg. culture and recreation, restaurants, hotels). Further studies have shown that the structure of expenditures in individual voivodeships may be determined by the economic situation of the voivodeship (Figure 7).

The study also showed that Grade Data Analysis (GCA) and its extension (GCCA) are proper tools for the analysis of structural changes in consumer expenditure. This method can be useful for identifying expenditure patterns, identification of trends and, as a consequence, in planning marketing campaigns and estimating the GDP structure.

References

Badach E. [2012]: Zastosowanie metod statystycznej analizy wielowymiarowej do badania struktury wydatków gospodarstw domowych. *Metody Ilościowe w Badaniach Ekonomicznych* XIII(1). Warszawa.

Begg D., Fischer S., Dornbush R. [2007]: Makroekonomia. Polskie Wydawnictwo Ekonomiczne, Warszawa.

Binderman Z., Borkowski B., Szczesny W. [2010]: Metody wizualizacji danych w analizie zmian poziomu i profilu konsumpcji w krajach UE. *Roczniki Nauk Rolniczych* SERIA G, T. 97(3).

Borkowski B., Dudek H., Szczesny W. [2009]: Spatial Differentiation of Food Production Structure and Consumption Profile in Enlarged European Union. *Journal of Economic Asymmetric* 5(2).

Dudek H., Koszela G., Landmesser J. [2012]: Wpływ sytuacji dochodowej na strukturę wydatków gospodarstw domowych. Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, *Ekonomika i Organizacja Gospodarki Żywnościowej* 97.

Główny Urząd Statystyczny [2011]: Metodologia badania budżetów gospodarstw domowych. Warszawa.

Gostkowski M., Gajowniczek K., Jałowiecki P. [2014]: Elastyczność dochodowa popytu na poszczególne grupy dóbr konsumpcyjnych w Polsce w latach 1999-2008. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu* XVI(3).

Kowalczyk T., Pleszczeńska E., Ruland F. [2004]: Grade Models and methods for data analysis. Springer, Berlin.

Kuśmierczyk K., Piskiewicz L. [2012]: Konsumpcja w Polsce na tle pozostałych krajów Unii Europejskiej. *Konsumpcja i Rozwój* 2.

Ostasiewicz W. [1999]: Statystyczne Metody Analizy Danych. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.

Świetlik K. [2012]: Konsumpcja jako czynnik wzrostu PKB w Polsce. *Konsumpcja i Rozwój* 1.

Szczesny W., Kowalczyk T., Wolińska-Welcz A., Wiech M., Dunicz-Sokolowska A., Grabowska G., Pleszczyńska E., [2012]: Models and Methods of Grade Data Analysis: Recent Developments, Institute of Computer Science Polish Academy of Sciences, Warszawa

Utzig M. [2013]: Regionalne zróżnicowanie konsumpcji i dochodów gospodarstw domowych w Polsce. Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu XV(2).

Ząbkowski T., Szczesny W. [2012]: The survey of the offer attractiveness of Internet access via Grade Correspondence Analysis. *Quantitative Methods in Economics* XIII(3).

Zeliaś A. [2000]: Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym. Akademia Ekonomiczna, Kraków.

Zmarzłowski K., Koszela G. [2013]: Czynniki różnicujące spożycie wyrobów spirytusowych w polskich gospodarstwach domowych w latach 1999-2008. *Metody Ilościowe w Badaniach Ekonomicznych* XIV(3).