International Journal of Scientific Research and Management (IJSRM)
||Volume||06||Issue||03||Pages||M-2018-17-24||2018||
Website: www.ijsrm.in ISSN (e): 2321-3418
Index Copernicus value (2015): 57.47, (2016):93.67, DOI: 10.18535/ijsrm/v6i3.m01

Analysis of consumer behavior in a small size market entity: case study for Vlora District, Albania

Elmira Kushta[1], Dode Prenga[2], Fatmir Memaj[3]

[1]Department of Mathematics, Faculty of Technical Sciences, University "Ismail Qemali", Vlore;,
[2]Department of Physics, Faculty of Natural Sciences, University of Tirana University,
[3]Department of Informatics and Statistics, Faculty of Economy, University of Tirana

Abstract:
In standard econometric application all variables are analyzed statistically before being used in mathematical models. In this framework we considered non-stationary distribution as an starting procedure on the study of consumer behavior in a local market area whereof non-homogeneity of buyers and small size effect could be present. By evaluation of the degree of non-stationary of the actual state for particular variable as observed, we hope to be able to estimate and interpret the model outcomes. Assuming the non-stationary of variables as indicator of the overall stet itself, we argue that the state where observation were made is non-stationary too, and for that reason, models are expected to not fit well. In the other hand, by dropping the significance level in model fitting process we expect to count for this instability whereas the model remains valid. Herewith, the logistic model for consumer behavior in our system is applied and calculated using significance level 0.85-0.90. Under such limiting constraint assumption we identified the variables that mostly affected the proportion between expense categories and the characteristics of the expenses that mostly describe the market consumer behavior in the unity studied. We hope that methodically this procedure could be helpful for other similar market or socio-metric study as well.

Keywords: logistic regression, consumer behavior, distributions**.**

I. Remarks on consumer behavior as mixes of opinion and econometric aspects

Consumer behavior is a key element in understanding market dynamics and marketing itself. It refers to those actions and related activities of persons involved specifically in buying and using economic goods and services [1] .Therefore it has two parts to be considered in one, the opinion formation aspects which is more psychological and fulfilling the needs in a concrete environment which has more metrics and econometrics inside. The most prevalent model from this perspective is 'Utility Theory' which proposes that consumers make choices based on the expected outcomes of their decisions. Consumers are viewed as rational decision makers who are only concerned with self-interest [2]. Meanwhile, different approaches highlight parts of the wholly so each model became more psycho-metric or more econometric considering specific subclasses of predictors and so for indicators. As a rule one can speaks for a chain of factors which cause a specific decision, or a series of consumer conduct, and the conjecture between them could be complicated, or even complex. Here we just stop in the representative part of observed series for variables contouring the consumer profile; therefore a statistical analysis for stationary of the state is performed initially. In this aspect we refer to the notion of stability in mathematical sense as for example in Levy terms as in [3] or even physically as discussed in [5],[6]]. Having decision making as opinion formation in the focus, we refer to more technical view from socio-physics [12], [10], and [11], that assimilate psychometric and sociologic elements in network terms, and calculate the behavior as emergency property for typically complex system. Those finding tells that opinion systems behave mostly as complex and therefore decision making in consumer case has no exception. Moreover, non-stationary distributions are characteristic for such systems [12],[6], therefore testing this property will help in model analysis and outcome interpretations. Next, the observed quantities are result of particular relationships and interactions so this aspect is considered in the framework of the analysis for variables series data. In particular we consider here the model that presumes logistic probability density function to

describe the consumer decision outcome. General analysis as in [2], [9] and application as in [8] illustrate that the decision making process in buying is very complex. It happen than the outcome interacts in some way with next stage of purchasing and it is underlined that consumers experience post purchase dissonance at least to some extent at every purchase they make [13]. It is evident that this behavior could not be assimilated in a simple functional relationship. For our system in the study one should consider the fact that internal and temporal aspect of decision compromise assumption of continuity, homogeneity. But again, in statistical view, this could be managed by adjusting significances for assertion based on standard models. It is a large agreement that major models in consumer behavior are regression based where variables appears in logarithmic form as rule or in probability form leading to the logistic probability distribution function. Hence, herein we just analyze our concrete system with limited data consisting of monthly expenses, and try to identify the specifics of the regression models and logistic behavior, admitting that the significance could be tolerated inasmuch the statistics for observed variables measured the level of non-stationary. Working in concrete data for consumer expenses in an Albanian district, we performed statistical analysis for stationary of the observation, the factor analysis searching for less dimensional profile for consumer and finally we realized the regression for logistic approach.

II. Profile of consumer and prediction factors

We realized the survey in Vlore district, a seaside city in south west of Albania. The city has no specific characteristics in the sense of economical level, ethnic cultures etc., one could assume with no doubt as practically representative for all urban area of the country, hence no additional variables are expected to interfere in the system [14]. Our data for response variables consist in the amount of monthly expenses in national currency Lek (ALL) for each family for 12 typical commodities and services paid in-between the period considered. In predictor variables we recorded values of seven most significant individual characteristics expecting to affect the consumer behavior [14]. As usually applied [8], [9], [2], most of them are categorical format but appropriate units has been discussed during the modeling process. We selected as potential predictors family size, the education level, employment status, gender of the most frequent buyer for family, etc. First we elaborate data to better categorize them based on general assumption, theoretical issue and tested them step by step. Next we perform factorial analysis to observe any possible reduction in the indicators as routinely suggested. In a preliminary stage, we realized a mixed view on the responses of system by changing the nature of variables from numerical to categorical. The two types are used in regression and of we assign as more natural the one that offer a better model fit based on the measurement for the system. In the table below we represented such actions.

Table 1:Predictor Variables

	Predictor Variables		Value	
X1				Set
	Variable	Type	set I	II
X2	Family type	categorical	1-5	1-5
X3	Education level	categorical	1-5	1-4
X4	Age	categorical	1-6	1-4
X5	Employment Status	categorical	1-3	1-2
X6	Income Type	categorical	1-3	1-9
X7	Gender	categorical	1-2	1-2
X8	Total Budged	numerical	Real	Real

Indicators or response variables have been considered in two approaches. First we use the proportion of the expenses for each type of expenditure in the role of the probability that consumer would spend its own budget in this type of commodity. Next, we change them to the categorical representation assuming that the behavior of the consumer is driven from the decision to pay in a specific range for specific goods needed. To avoid over detailed profile of the consumer, the data gathered from the inquiry were reorganized to form 4 distinct variables respectively the expenses for basic goods and services, common expenditure, life quality expenses and luxury expenses. This reduction is supported by factorial analysis too..

III. Stationary of the variables

Reorganizing the data i frequencies and getting distribution from optimized histograms as recommended in [7], we observe that data series for indicator variables are non-stable distribution in the Levy terms. For this, we make use of q-Gaussians properties as introduced in [3], [5] applying a step

by step procedure of fitting. Generally speaking processes governing such systems are expected to be a mixed form of multiplicative and additive one as in other similar systems treated in [12] so Gaussian and Lognormal forms are expected to be present in standard econometric parlance , whereas correlated processes give rise to q-distribution as follow [3]

$$Gausian_q(x) = \alpha\left(1-\beta(1-q)(x-\mu)^2\right)^{\frac{1}{1-q}}$$

$$Lognormal_q(x) \sim \frac{1}{x^q}\left(1-\beta(1-q)\left(\frac{x^{1-q}-1}{1-q}\right)-\mu\right)^{2}\Bigg)^{\frac{1}{1-q}} \quad (2)$$

which in the limit q→1 reproduce the Gaussians and lognormal respectively [3],[5]. According to such detailed view this will happen when additive or multiplicative properties turn to be dominant as detailed for other systems in [5] and acknowledged as Tsallis statistics. In this approach if the distribution is stationary and otherwise it is not. Moreover, in the interval distribution has infinite variance whereas for q in the interval the distribution still exists but the variance is indefinite [3]. Above q=3 there is no distribution. Remember that those (Tsallis) conditions in q correspond to the Levy stability as seen from the relationship [3]. Letting aside the nature of processes governing such system, but using conclusions for non-stationary states and practical estimation of it as briefed above, we apply directly relations (1) to fit distributions of data series for variables. Considering the fact that number of individuals interviewed (number of valid observation) is small (N=350), the histogram optimization has been considered with great care according to Scot and Freidman-Diaconics rule as discussed in [19]. We applied a tuning technique as proposed in [18] around the bin size found in standard procedures. We observe that q parameter is obtained in the range [1.3-2.9] and all variables except one have q>5/3 and therefore distributions are non-stationary. An immediate finding is that the average values of variables do not represent the series if strictly statistically speaking. For some of them, the statistical variance is infinite or even indefinite according to [3], [5]. But the most important is the finding that some of them have not distribution at all as the 1 parameter is 3 or more. We excluded this variable as highly disturbed. From the other side, we obtained that many variables have indefinite variance and therefore those are in-appropriate to be included in models because they miss an important statistical characterizes the second moment. In the Figure 1 we showed those distributions in log-log graph to better emphasize the visual differences of the distributions.

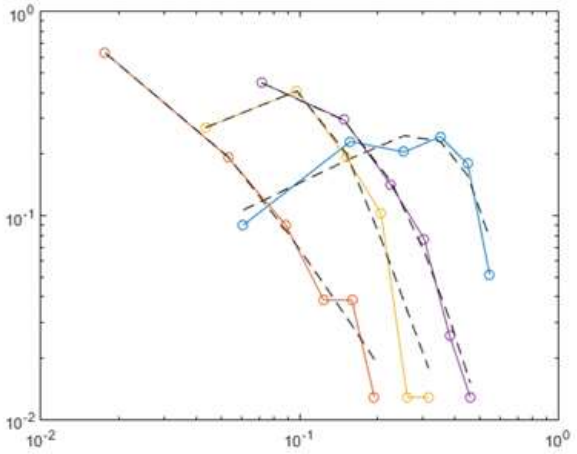

Figure 1: Log-log representation of q-Gaussians fitted to the empiric distribution of variables values. Variables 1-5.

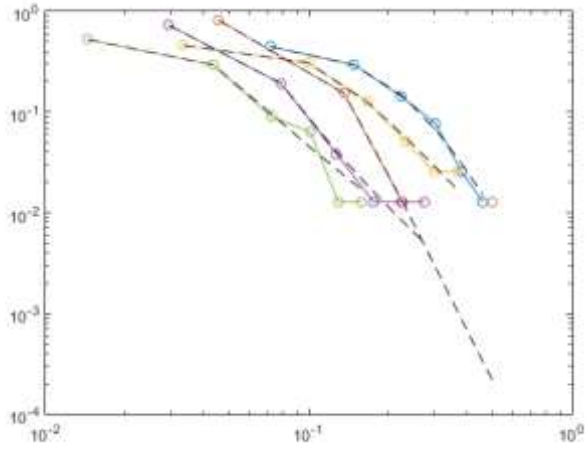

Figure 2: Log-log representation of q-Gaussians fitted to the empiric distribution of variables values, variables 6-11.

Moreover we observe that if one select subclasses belonging to a fixed categorical value, the number of such sampler become smaller than the minimum size requested to proceed with statistical analysis. Therefore considering a specific variable for example "luxury expenses" we should filter families with many children that have no more than one parent employed; families that are not having their incomes limited to the pensions or other social care etc., that by definition are not subject of discussion what to do on luxury expenses because they have predefined choice, no budget in this disposal. Doing so the sampler size became abnormally small to work with, hence we should work with all values of cause variables included. We have a preliminary result that either non stable per se, or non-stable caused from non-optimal observation the series are non-stationary or the average values are not representative.

Table 2 Evidences for the stability of the distribution for predictor variables and FA analysis

	Qstat	E[Y]	Q-variance

Y1	1.7265	0.2834	Infinite
Y2	2.9705	0.0088	indefinite
Y3	1.7351	0.0891	indefinite
Y4	1.3925	0.0353	finite
Y5	1.4295	0.0568	finite
Y6	1.7756	0.0423	Infinite
Y7	1.9281	0.0402	Infinite
Y8	2.0907	0.022	Indefinite
Y9	2.0836	0.104	indefinite
Y10	2.0403	0.1142	indefinite
Y11	1.8831	0.005	Infinite
Y12	1.9333	0.005	Infinite

Therefore, the set of observations of expenditures for specific goods or services and their proportion fail to represent the profile of consumers system analyzed (or generalized consumer characteristics. This result told us that either these are in-appropriate for the system or the measurements are far from the homogenized consumer medium. Considering that some of variables are less stationary than others, there is logic move to check which of them could be removed or transformed to make the model fits better. Therefore factorial analysis has been performed to check the idea of reducing analytically the number of variables. We identify that from 12 variables of the profile 5 of them express more than 92% of variance and have the eigenvalue higher than

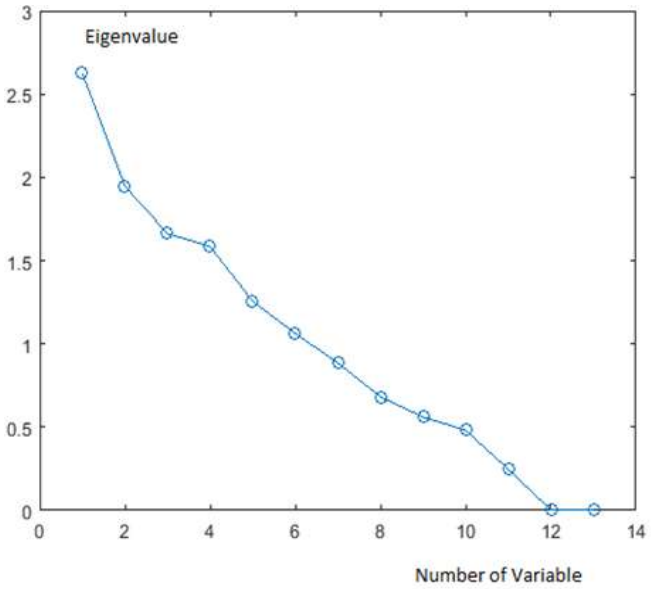

Figure 3: eigenvalues plots versus number of variables

Based on standard assumption of factorial analysis we hypothesize that our system of consumer profile should have no more than 5 variables. But considering the fact that all of actual series are in highly non-stable state the real set of independent variables is expected to be even lower. Using those qualitative arguments we propose to select 2 or 3 variables as indicators for consumer behavior in our system.

Table 3: Predictor Variables

	Initial variable. Value Real/Proportion. Categorical	Representative variables. Real/ Proportional	Proposed Variable. Real Proportional
{Y}	Expenses for:		
Y1	Alimentary goods		
Y2	Clothes	Basic	
Y3	Subsistence	expenditure	
Y4	Alcoholic drinks and cigarettes		
Y5	Health*		
Y6	Transport	Extra	Common
Y7	Communication	expenditures	expenditure
Y8	Culture and safety expenses	Qualitative	Quality life
Y9	Education	Life	and luxury
Y10	Other services	Expenditure	expenditures
Y12	Family expenses		
Y13	Luxury goods		
Y14	Restaurant expenses	Luxury Expenditure	

We conclude that in our system, the most significant expenses contributing in the behavior of the consumers are normalized expenditures in “basic commodities and services”, in “some custom goods and services”, expenditures in “goods and services related to the quality of life” and “luxury expenses”. In this insight more detailed expenses consist in over detailed behavior and hence not matching good models. By estimation of most characteristics set of indicator variables we went more inside the relationship between factors and indicators. In this case we reconsider the variables as the exhibition of consumer behavior not only as result of direct measurement of particular activity. This last has as final result the decision that in substance consist in an agreement for what to do if two or more alternatives exist. The numerical value or even the rapport of any expenses to the total budget does not report a clear decision or clearly a decision of the consumer.

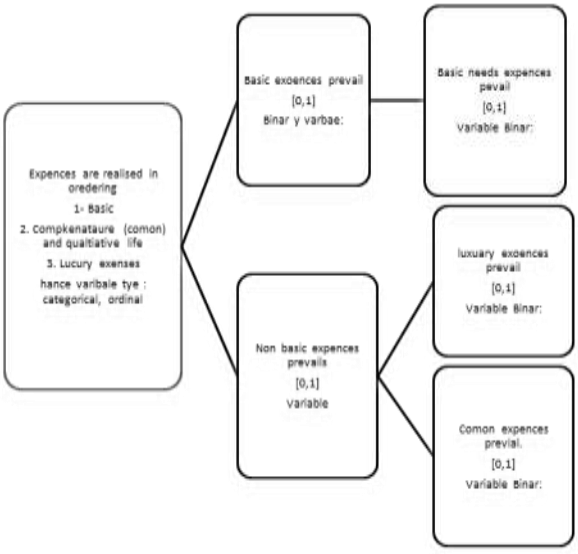

Figure 4: The use of categorical variable for consumer profile

Our analysis is extended in a broader representation of consumer activity and his decision. The rapport between each category tells us which decision has been frontrunner after a reasonable rivalry between them. In this stage we avoid even behavioral classes and models because we are interested in final decision and cause that affect it, but surely one can discuss about them when interpreting results.

IV. Identification of the causes and relationship with indicators using logistic regression

Logistic regression is a standardized model in this case as routinely recommended and used [9], [8] aside others as discussed in [4]. So, we proceed with calculation of logit function of binary valued new variables that assigned the dominance of one expense to the other. In Figure 4 we show the logistic regression for the variable "spending in common or basic goods and services prevail " measured in binary values {0,1}. As seen in the graph, only central part of the expected sigmoid is seen, hence we argue that there are missing variables in the set of consumer profile. We show even the predicted margins from the regressions in two different set of predictors, and as it seen on the graph removing variables increase the error (red points on the figure) but the model continue to be robust. Again the central part of sigmoid remains the same. Next we extended the set of predictor variable by adding the amount of budget for each individual by a simple assumption that it can force people to make choices. This contribute to span the graph in full sigmoid. Therefore the model is considered adequate and we conclude that consumer behavior variable "Living Expenses prevail to the other" which we call reduced profile is found to have log-it function in the form

$$\pi = \log\left[\frac{\Pr ob(Y=1|X)}{1-\Pr ob(Y=0|X)}\right] = \alpha + \sum_{i=1}^{n}\beta_i X_i \quad (1)$$

and the regression of the right hand side of (3) give

$$\begin{aligned}\pi &= 4.304 - \\ &- 0.275 * FamilyType + 0.707 * EducationLevel \\ &+ 0.367 * AgeGroup - 0.454 * Employment\,Type \\ &+ 3.3*10^{-5} * Budget + \varepsilon\end{aligned} \quad (2)$$

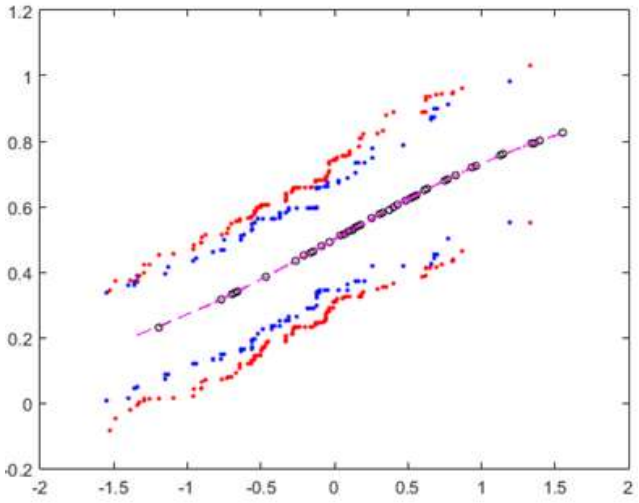

Figure 5: Probability versus log it regressed

In relation (4) variables "Family Type" has 5 value for one to five members respectively; "Education Level" has 4 values starting from elementary level to postgraduate level; "AgeGroup" for the householder (or most frequent buyer on the market) has 3 values respectively for young, mature or above age 60; "Employment" has 3 values respectively for being employed in state agency or institution (3), private company (2) and having no job at all or being retired (1).

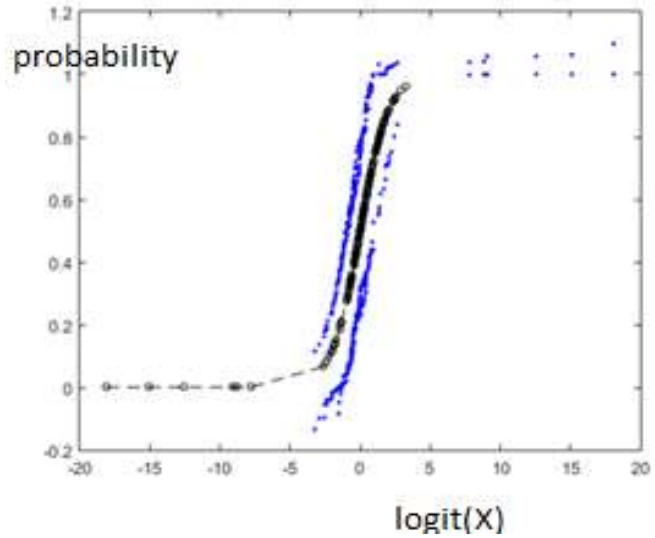

Figure 6: Regression for the full set e variable including budget

All characteristics above are attributes of the most

frequently used to manage the expenses as stated in the survey.The statistics for each parameter has p-value above 0.85 an only two have value above 0.95, but based on the arguments provided we accept this regression and acknowledge the description (4) as real behavior of the reduced consumer profile.

V. Conclusions

Using analysis of stationary for observed variables data as extra statistical tools has improved the study of specific econometric systems in the preliminary stage of model calculation end regression. It helped in filtering high disturbed data series that seem to not be drawn from a known distribution and hence they state is undefined. Next, knowing the level of stationary for actual state that characterized the variable we were able to judge on the expected range of significance in model fitting at least qualitatively, and motivate the model fitting in somewhat lower level of confidence but remaining valid. In the application to our concrete system in the limitation of small size sampler, we realized the study and obtained some characteristics. We obtained that the profile of the consumer was described by the binary variable that evaluates the dominance of basic needs expenses. The variables that define the predictors set are the size of the family, the nature of employment, education, the age group of householder and the budget. We observe that the category that weights mostly in the decision to spend in basic needs goods is the education level (highest score for highest level) , the second is the employment type, third is the age group etc. The budget has fine tune coefficient but its effect is important in the decision of the consumer. We expect that some of those finding can be extended to a larger system size and the methodology could be effective elsewhere.

References

[1] Jisana T. K. Consumer behaviour models: an overview. Volume 1, Issue 5 (May, 2014)

[2] Jeff Bray. Consumer Behaviour Theory: Approaches and Models . http://eprints.bournemouth.ac.uk

[3] Sabir Umarov, Constantino Tsallis, Murray Gell-Mann, Stanly Steinberg. Generalization of symmetric -stable Lévy distributions for q>1. Journal of mathematical physics 51, 033502 2010

[4] Steiger, J.H. (1990), "Structural model evaluation and modification," Multivariate Behavioral Research, 25, 214-12.

[5] Constantino Tsallis Computational applications of non-extensive statistical mechanics. Journal of Computational and Applied Mathematics 227 (2009) pp 51-58.

[6] G.P. Pavlos, M.N. Xenakis,L.P. Karakatsanis, A.C. Iliopoulos, A.E.G. Pavlos D.V. Sarafopoulos. Universality of Tsallis Non-Extensive Statistics and Fractal Dynamics for Complex Systems. Chaotic Modeling and Simulation (CMSIM) 2: 395-447, 2012.

[7] Glen Meden, Kun He. Selecting the Number of Bins in a Histogram: A Decision Theoretic Approach. Journal of Statistical Planning and Inference, Volume 61 (1997), 59-69.

[8] Kadri G Yilmaz , Sedat Belbag. Prediction of Consumer Behavior Regarding Purchasing Remanufactured Products: A Logistics Regression Model. International Journal of Business and Social Research Volume 06, Issue 02, 2016

[9] Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. Statistics in Medicine, 22, 1433–1446.

[10]G. Deffuant, D.Neau, F.Amblard , G.Weisbuch, “Mixing beliefs among interacting agents”, adv. Compl. Sys. 3(1-4), 87-98. (2000).

[11]Rainer Hegselmann. Opinion Dynamics and bounded confidence models, analysis and simulation.. Journal of Artifical Societies and Social Simulation (JASSS) vol.5, no. 3, 2002

[12]Claudio Castellano, Santo Fortunato,Vittorio Loreto.: Statistical physics of social dynamics. Rev. Mod. Phys. 81, 591-646 . April-June 2009

[13]Maria-Cristiana Munthiu. The buying decision process and types of buying decision behaviour

[14]Sibiu Alma Mater University Journals. Series A. Economic Sciences – Volume 2, no. 4, December / 2009

[15]Horn, J. L. A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

[16]F. Mema. Marketing research in helping decision making” . “Ekonomia dhe Biznesi” Nr. 2 (22) 2006

[17]F. Mema. Statistical methods in acknowledgment of consumer behavior" International Review of Science, Innovation and New technology, Vol. 1, Nr 12, February, 2015

[18]D. Prenga, M. Ifti. Complexity Methods Used in the Study of Some Real Systems with Weak

Characteristic Properties. AIP Conf. Proc. 1722, 080006 (2016)
[19] Scott, David W. Multivariate Density Estimation and Visualization Papers Humboldt-Universitet Berlin, Center for Applied Statistics and Economics (CASE), no. 2004,16

Author(s) Profile

Elmira Kushta, PhD student received the M.S. degrees in Mathematics from the Faculty of Natural Sciences, University of Tirana in 2010. She started as lecturer of statistics in the Faculty of Technical Sciences, University of Vlore. Now she is working on statistical and operational researches and simulation for consumer behavior and marketing.

Dode Prenga, Physicist, graduated in 1992 at the Faculty of Natural Sciences is associate professor is lecturer of nonlinear dynamics in the Faculty of Natural Sciences, University of Tirana.

Fatmir Memaj, Statistician, graduated in 1984, is Professor of Statistics, Demography and Research Methods on Applied Mathematics in the Faculty of Economy, University of Tirana.