Analyzing the health status of the population using ordinal data

Maria Livia ŞTEFĂNESCU[*]
Research Institute for Quality of Life Romanian Academy

Abstract

We intend to estimate the health status of the people using a Gini kind index GO for measuring the inequality and a polarization indicator PO too. The both indices were applied for ordinal health data which were selected from three national representative samplings designed in the period 2003-2010 in Romania. The results evaluate the evolution level of the polarization and inequality phenomena in the health domain.

Keywords*: ordinal data, polarization and inequality indices, health status, SAH data*

JEL classification : *I14, I19, C43, C19*

1. Introduction

The data in health domain are accessible under different forms. We mention here the nominal, ordinal and cardinal (for interval scale) health data. Obviously, the indices which are used to evaluate health status of the individuals or of the nations must be highly associated with the specific kind of data.

Our research is based on there national representative Romanian sampling surveys designed in the years 2003, 2006 and 2010. In the following we intend to analyze the response at the question *Q1*: "*How do you evaluate your status of health*". The question *Q1* was addressed to all the persons belonging to the selected samplings. At this question we specified five possibilities of answer, that is: *very bad* (code 1), *bad* (code 2), *satisfactory* (code 3), *good* (code 4), *very good* (code 5).

In this context the people answers concerning the subjective self-assessment of the Romanian health status are ordinal data. More precisely, we identify practically five categories C_1 - C_5 characterized by the answer codes 1-5. The individuals from a group C_k have often more difficulties concerning their health status in contrast with the persons belonging to the superior groups as C_{k+1}, C_{k+2}, etc.

Effectively, the primary data are characterized by the frequencies f_k, $1 \le k \le m$, where the natural number f_k, $f_k \in N$, represents the frequency to have individuals in the category C_k from the specified sampling. Shortly, we will designate by $\underline{f}$ the vector which includes all the frequencies f_k, $1 \le k \le m$, that is $\underline{f} = (f_1, f_2, f_3, \dots, f_m)$.

To simplify our presentation, we will use the notation $f_{k,+}$, $1 \le k \le m$, to designate the following expression

$$f_{k,+} = f_1 + f_2 + f_3 + \dots + f_k, \quad 1 \le k \le m \tag{1.1}$$

The size n of the sampling is just $f_{m,+}$.

More, by $\Delta_{m,n}$ we understand the subdomain of N^m having the form

$$\Delta_{m,n} = \{ \underline{f} \mid \underline{f} = (f_1, f_2, f_3, \dots, f_m),\ f \in N^m,\ f_1 + f_2 + f_3 + \dots + f_m = n \} \tag{1.2}$$

In the subsequent we intend to measure two distinct social phenomena, that is the polarization and also the inequality aspects in health. In practice, the both phenomena are closely related, but we have not a strictly dependence relation between them. More precisely, one of these phenomena can affect in part the other. But for effective real situations this multifaceted relation could have very distinct intensity degrees. A concrete evaluation concerning the inequality and the polarization health levels associated to the years 2003-2010 was given for Romania living in the rural environment.

[*] Corresponding author: livia.stefanescu@gmail.com

2. The polarization index *PO* for SAH data

In the literature, for nominal, ordinal or cardinal types of data, are known a lot of indicators to measure the degree of the polarization phenomenon. These measures depend effectively on the particular type of data. So, for the class of cardinal data we point out the papers: Esteban & Ray (1994), (2012), Duclos & Esteban & Ray (2004), Chakravarty & Majumder (2001), Chakravarty (2009), Rodriguez & Salas (2003), Zhang & Kanbur (2001), Wang & Tsui (2000), Foster & Wolfson (2010), Bossert & Schworm (2008), Deutsch & Silber & Yalonetzky (2013). Some polarization indices were also proposed for categorical data. We remark here the work of Permanyer & D'Ambrosio (2013). Also we mention especially the following references focused on ordinal data: Apouey (2007), (2010), Montalvo & Reynal-Querol (2005), Apouey & Silber (2013), Kobus (2014), Chakravarty & Maharaj (2012), Makdissi & Yazbecky (2014).

Strong related with a polarization measure are different indices of variation with could now be applied for ordinal data too. See, for example Berry & Mielke (1992), Blair & Lacy (1996), (2000).

In many recent papers are studied diverse techniques to transform a kind of data in a new type of data. In this context we underline the review of Van Doorslaer & Jones (2003). Therefore, an index used initially for a specific data can be modified to use for another data types.

Apouey (2007) proposed an one parameter class of polarization indices applied to ordinal SAH (self-assessed health) data. These types of indices depend on the probabilities of the ordinal categories. In the present paper we will express one of these Apouey polarization indicators in function of the frequency associated to every ordinal class C_1- C_m . More precisely, for any $\underline{f} \in \Delta_{m,n}$ the polarization index *PO* has the following expression :

$$PO(\underline{f}) = 1 - \frac{1}{m-1} \sum_{k=1}^{m-1} | 2 f_{k,+} / n - 1 | \tag{2.1}$$

Apouey (2007) established two main axioms which must be fulfilled compulsory by any polarization index applied to ordinal data. All his proposed indicators satisfy the both axioms and more these indices have good properties to measure the polarization phenomenon in health domain, Apouey (2007), (2010).

It is easy to show by a direct calculus that

Proposition 2.1. For any $\underline{f} \in \Delta_{m,n}$ we have always the inequalities

$$0 \le PO(\underline{f}) \le 1 \tag{2.2}$$

3. An inequality coefficient *GO* for ordinal data

For an arbitrary ordinal variable X characterized by the frequencies $\underline{f}$ is not able to operate correctly with the mean $\mu(X)$. Indeed, a score k attached to the ordinal category C_k of X is subjective. These scores k establish only the hierarchy of the classes C_k , $1 \le k \le m$. More, the value k is not often relevant when is used to characterize all the individuals belonging to the same class C_k . Any other set of real values $v_1 < v_2 < v_3 < \ldots < v_m$ could define the weights of the ordered groups C_k , $1 \le k \le m$.

The methodology to measure the inequality phenomena was intensively developed in the last 50 years. We mention only a bit from this multitude of references : Atkinson (1970), Chakravarty (2009), Duclos & Araar (2006), Haughton & Khandker (2009), Betti & Lemmi (2008), Foster & Seth & Lokshin & Sajaia (2013).

Gini coefficient $G(X)$ is the most popular index to evaluate the degree of inequality for a distribution of cardinal data X. This very known indicator was proposed by the famous Italian economist and statistician Corrado Gini at the beginning of the twenty century (Gini (1909a),(1909b)). We remind that Gini index $G(X)$ is based on the Lorenz curve where the mean of the cardinal variable X plays an essential role.

Since the mean $\mu(X)$ of an ordinal variable X has not a clear interpretation we can't apply correctly the classical Gini coefficient $G(X)$ to measure the inequality from X.

More indicators were proposed to evaluate inequality aspects in the case of ordinal data. We mention here some references regarding different approaches: Allison R. A., Foster J. E. (2004), Abul & Yalcin (2008), Madden (2010), Giudici & Raffinetti (2011).

Giudici & Raffinetti (2011) adapted the classical Gini coefficient $G(X)$ to any ordinal variable X.

More exactly, for all individuals belonging to the class C_k we associate the same rank r_k, $1 \le k \le m$. But, the rank r_k is modified in function of the ordinal distribution $\underline{f}$ which is analyzed (Giudici & Raffinetti (2011)). So

$$r_1 = 1 \quad , \quad r_k = r_{k-1} + f_{k-1} \text{ for any } 2 \le k \le m \tag{3.1}$$

The ordinal Gini index $GO(\underline{f})$ is based on the Lorenz curve defined by the points having the cartesian coordinates $(f_{k,+} / f_{m,+} , q_k / q_m)$, $0 \le k \le m$, where

$$q_k = \sum_{j=1}^{k} r_j f_j \quad , \quad 1 \le k \le m \tag{3.2}$$

with the convention $f_{0,+} = q_0 = 0$.

With these notations we define (Giudici & Raffinetti (2011))

$$GO(\underline{f}) = 1 - \sum_{k=1}^{m} (q_k / q_m + q_{k-1} / q_m)(f_{k,+} / f_{m,+} - f_{k-1,+} / f_{m,+}) \tag{3.3}$$

After a straightforward computations we deduce too

Proposition 3.1. For any $\underline{f} \in \Delta_{m,n}$ the following inequalities are true

$$0 \le GO(\underline{f}) \le 1 \tag{3.4}$$

4. Some differences between *PO* and *GO* indicators

Therefore, the indicators $PO(\underline{f})$ and $GO(\underline{f})$ can be used successfully to measure the polarization degree, respectively the inequality level for an arbitrary distribution of frequencies $\underline{f}$ which characterize an ordinal variable X. We established that the both coefficients vary in the interval [0 , 1].

But, in practice, between the polarization and the inequality phenomena there is a complex relation of dependence. For this reason, from very closed values of the polarization index *PO* is possible to often obtain very different values of the inequality coefficient *GO*.

For the subsequent we will prove this assertion taking into consideration the frequency distributions $\underline{f}$, $\underline{f} \in \Delta_{3,500}$, of sixteen ordinal variable X precised in *Table 4.1* .

In *Graphic 4.2* are represented the points j, $1 \le j \le 16$, having the cartesian coordinates $(PO(X_j) , GO(X_j))$, where X_j are the ordinal variables with the frequency distributions $\underline{f}$ from *Table 4.1*. The scatter of the points j from *Graphic 4.2* suggests us that there is not a simple dependence relation between the polarization and inequality aspects of the variables X_j. More, none of the following inequalities $PO(X) < GO(X)$ or $PO(X) > GO(X)$ are always true for any ordinal variable X (see *Graphic 4.2*).

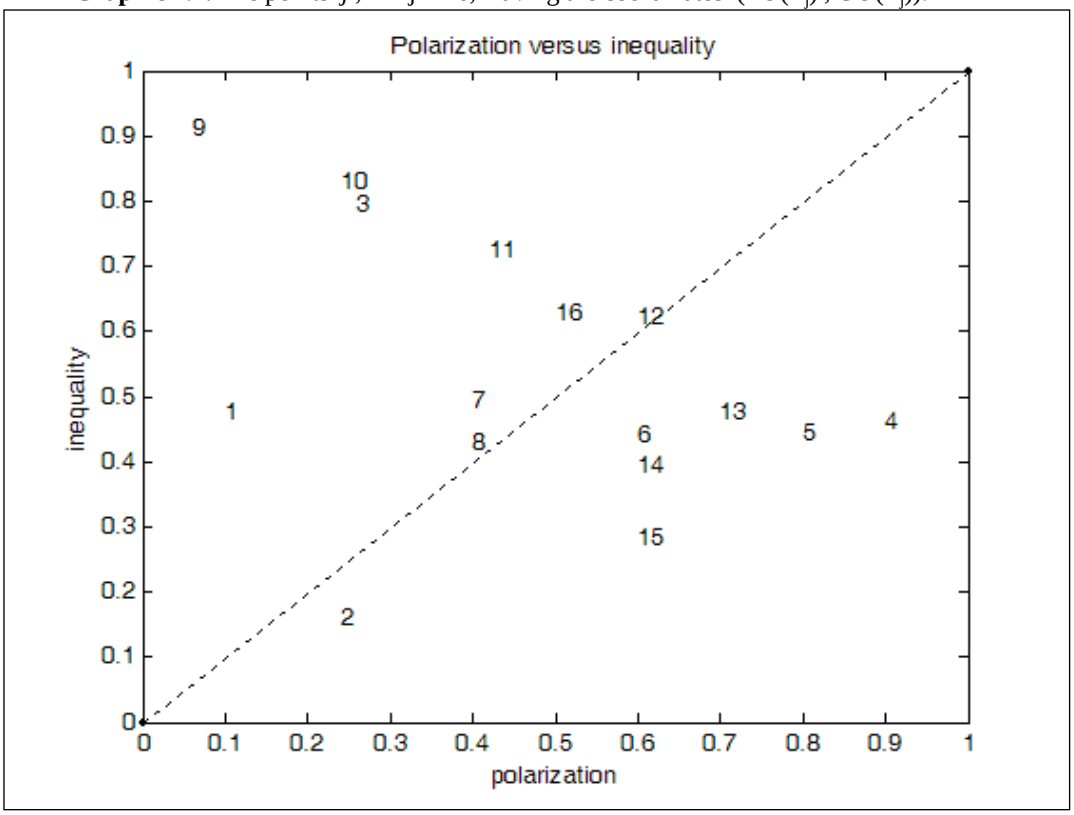
Having in mind this conclusion, for an accurate interpretation of the health status of a given population, we recommend to use together the two indicators *PO* and *GO*.

Table 4.1. The frequency distributions f of the ordinal variables X_j

(m=3 , n=500).

j	f_1	f_2	f_3	$PO(X_j)$	$GO(X_j)$
1	25	450	25	0.10	0.4775
2	20	80	400	0.24	0.1611
3	400	70	30	0.26	0.7975
4	225	50	225	0.90	0.4622
5	200	100	200	0.80	0.4472
6	150	200	150	0.60	0.4428
7	50	300	150	0.40	0.4963
8	150	300	50	0.40	0.4301
9	480	10	10	0.06	0.9131
10	420	40	40	0.24	0.8318
11	360	70	70	0.42	0.7258
12	300	100	100	0.60	0.6241
13	200	150	150	0.70	0.4789
14	100	200	200	0.60	0.3975
15	100	100	300	0.60	0.2837
16	300	150	50	0.50	0.6287

Graphic 4.2. The points j , $1 \le j \le 16$, having the coordinates $(PO(X_j)$, $GO(X_j))$.



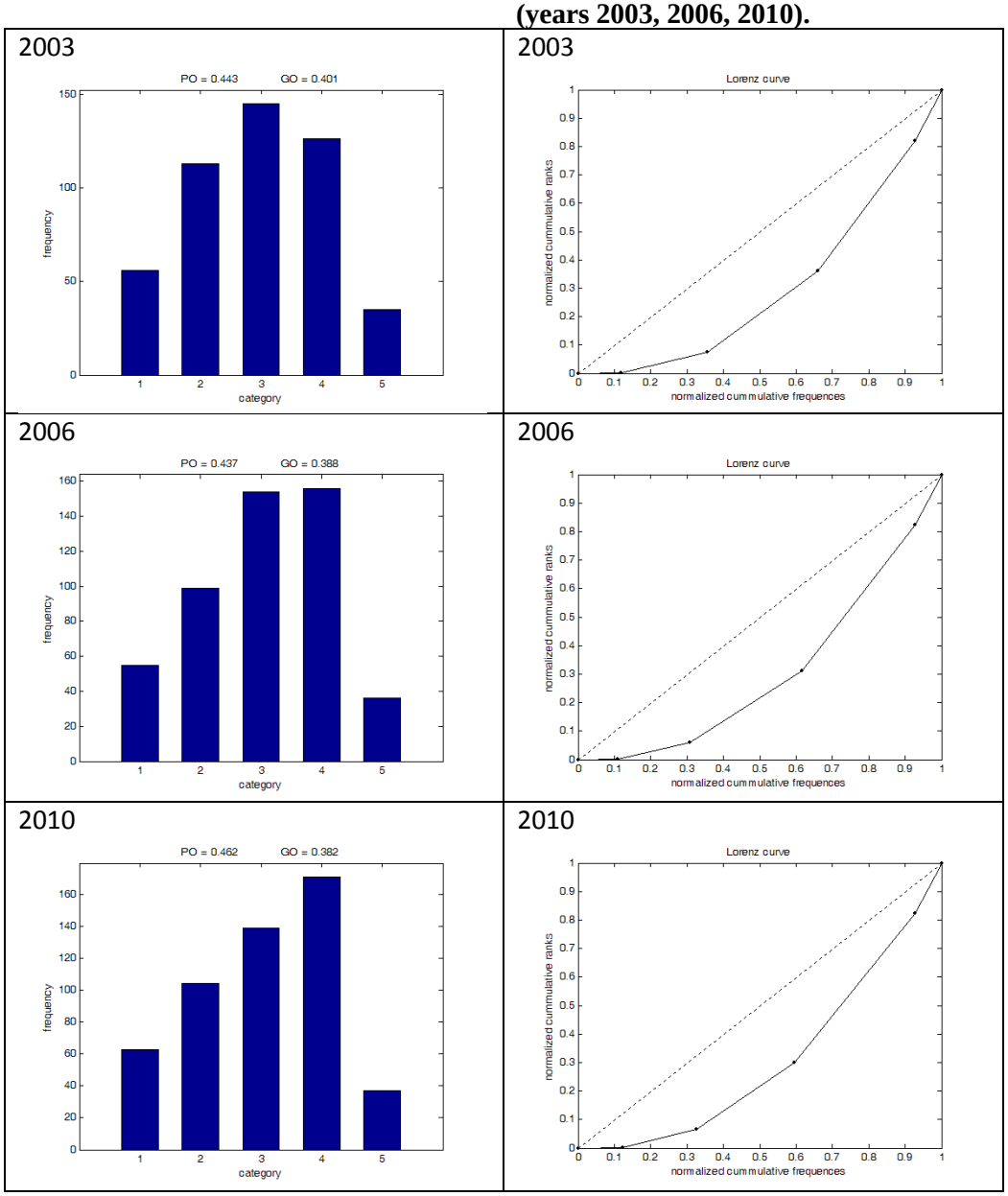
5. An application

We will analyze the evolution of the health status for Romanian living in rural, during the years 2003-2010. In this context we used three representative samplings concerning the quality of life of the Romanian people. The sampling surveys were designed at Research Institute for Quality of Life, Romanian Academy, in the years 2003, 2006 and 2010.

The frequency distributions f at the question *Q1* where illustrated in *Graphic 5.1* together with the associated Lorenz curves obtained after the Giudici & Raffinetti (2011) methodology.

From *Graphic 5.1* we remark that the studied distributions and their Lorenz curves are very similar. For this reason is very difficult to evaluate the progress of the Romanian health status in the period 2003-2010.

Graphic 5.1. The frequency distribution and the Lorenz curve at the question *Q1* (years 2003, 2006, 2010).



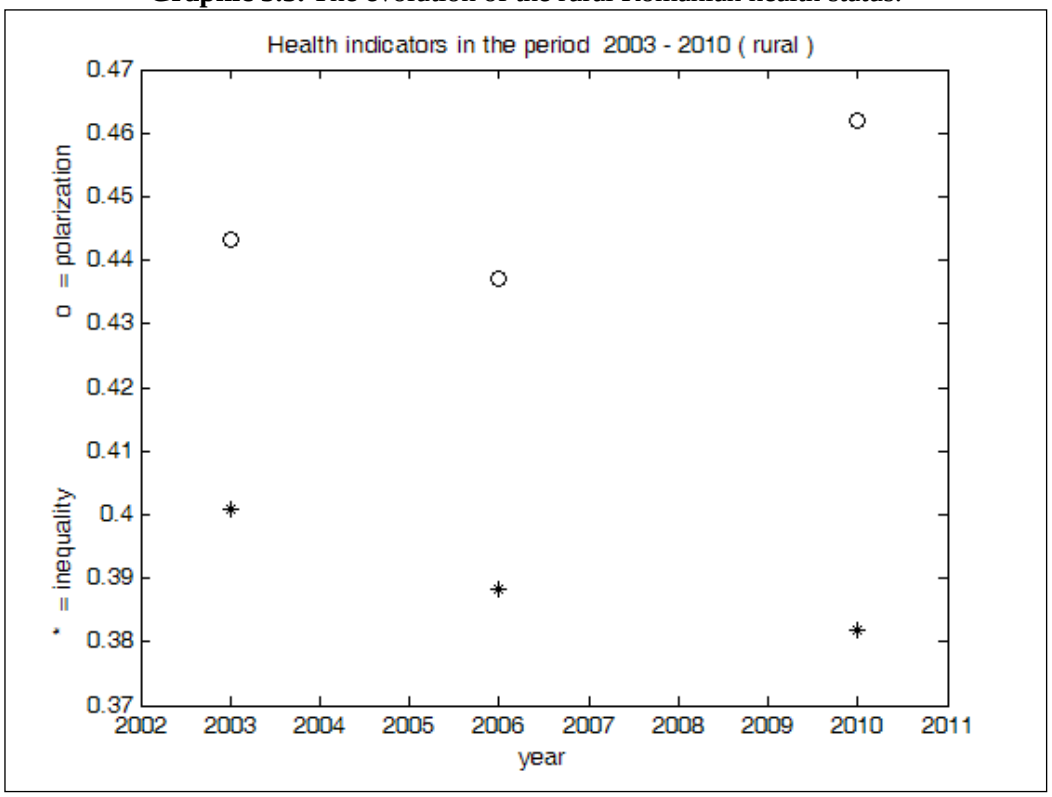
By applying the polarization PO and inequality GO indices to the selected samplings having the volume n we obtained the values mentioned in *Table 5.2*. *Graphic 5.3* suggests the evolution of the rural Romanian health

status in the period 2003-2010. We remark an easy decreasing of the inequality and an enough consistent increasing of the polarization aspects.

Table 5.2. Synthesis results regarding the selected samplings.

year	2003	2006	2010
n	475	500	514
PO	0.443	0.437	0.462
GO	0.401	0.388	0.382

Graphic 5.3. The evolution of the rural Romanian health status.



6. Partial conclusions

This proposed methodology, based on the polarization and inequality indices for ordinal data, was applied to evaluate health status for rural Romanian people in the period 2003-2010.

The indices *PO* and *GO* measure two distinct aspects of the reality, that is the polarization and the inequality phenomena. The two coefficients *PO* and *GO* vary inside the interval [0, 1]. Considering sixteen possible answer distributions we proved that the polarization and inequality situations can be often close related but not identical. In reality, the increase of the polarization level into a community do not compulsory involve the grow of the inequality degree inside that population (see Graphic 4.2). For a precise interpretation of the evolution for the population health status we recommend to use together the both indicators *PO* and *GO*.

For the rural Romanian communities we have a stable decrease of the *GO* inequality coefficient in the period 2003-2010. But the behavior of the *PO* polarization indicator is different. So, after a light decreasing of the *PO* values it results finally an enough consistent increase of the polarization (Graphic 5.3).

To apply correctly our proposed approach is necessary to study the properties of the indices *PO* and *GO* and in addition, to precise clearly the concrete cases when the both indicators can act in the same direction. It is essential to use more indices to measure distinct aspects of a complex reality. Primary, our option regards an index to identify a positive evolution of a concrete situation. From complementary studies we must also establish some reference distributions considered as equilibrium circumstances for the society.

In the future, using the same kind of processing, we intend to compare the self- assessed health answers at the question *Q1* of the people which is divided in more groups. So, the individual health data must be analysed in contrast for different age categories, taking also into consideration the gender of the persons and their domicile, the household income, families with more children, the unemployed people, individuals with disabilities or other deprivated groups.

References

[1] Abul Naga R., Yalcin T. (2008). Inequality measurement for ordered response health data. *Journal of Health Economics* 27, 1614-1625.
[2] Allison R. A., Foster J. E. (2004). Measuring health inequality using qualitative data. *Journal of Health Economics* 23, 505-524.
[3] Apouey Bénédicte, (2007). Measuring health polarization with self-assessed health data. *Health Economics*, 16 (9), 875–894.
[4] Apouey Bénédicte. (2010). On measuring and explaining socioeconomic polarization in health with an application to French data. *Review of Income and Wealth*, *56*(1), 141-170.
[5] Apouey Bénédicte, Silber Jacques, (2013). Inequality and bi-polarization in socioeconomic status and health: Ordinal approaches. PSE Working Papers, no. 2013-30. https://halshs.archives-ouvertes.fr/halshs-00850014.
[6] Atkinson A. B., (1970). On the measurement of inequality. *Journal of Economic Theory* 2, 244-263.
[7] Berry K.J, Mielke P.W., (1992). Indices of ordinal variation, *Perceptual and Motor Skills*, 74, 576-578.
[8] Betti Gianni, Lemmi Achille (eds.), (2008). Advances on income inequality and concentration measures. Routledge – Taylor & Francis Group, London.
[9] Blair, J., Lacy M.G., (1996). Measures of variation for ordinal data as functions of the cumulative distribution. *Perceptual & Motor Skills*, *82*(2), 411-418.
[10] Blair J, Lacy MG., (2000). Statistics of ordinal variation. *Sociological Methods and Research*, 28(251), 251-280.
[11] Bossert W., Schworm W., (2008). A Class of Two-Group Polarization Measures. *Journal of Public Economic Theory* **10**, 1169-1187.
[12] Chakravarty S.R., (2009). Inequality, polarization and poverty. *Advances in Distributional Analysis*. Springer, New York.
[13] Chakravarty S.R, Maharaj B., (2012), Ethnic polarization orderings and indices. *Journal of Economic Interaction and Coordination*. 7(1), 99-123.
[14] Chakravarty S.R, Majumder A., (2001). Inequality, polarization and welfare: theory and applications. *Australian Economic Papers* **40**, 1-13.
[15] Deutsch J., Silber J., Yalonetzky G., (2013). On relative bi-polarization and the middle class in Latin America. A look at the first decade of the 21st century. Paper presented at the XX Encuentro de Economía Pública, Sevilla, Spain.
[16] Duclos J-Y., Esteban J., Ray D. , (2004). Polarization: Concepts, measurement, estimation. *Econometrica* **72**, 1737-1772.
[17] Duclos Jean-Yves, Araar Abdelkrim, (2006). Poverty and equity: measurement, policy and estimation with DAD. Kluwer Academic Publishers, Boston.
[18] Esteban J., Ray D., (1994). On the measurement of polarization" *Econometrica* **62**, 819-852.
[19] Esteban J., Ray D., (2012). Comparing polarization measures. In Garfinkel M R, Skaperdas S (Eds), The Oxford Handbook of the Economics of Peace and Conflict, Chapter 7, Oxford University Press, New York.
[20] Foster J., Wolfson M.C., (2010). Polarization and the decline of the middle class: Canada and the US, *Journal of Economic Inequality*, 8, 247-273.
[21] Foster James, Seth Suman, Lokshin Michael, Sajaia Zurab, (2013). A unified approach to measuring poverty and inequality - Theory and Practice. Streamlined analysis with ADePT software. The World Bank, Washington D.C.
[22] Gini C., (1909a). Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza. Giornale degli Economisti, 38, 27–83.
[23] Gini C., (1909b). Concentration and dependency ratios (in Italian). English translation in Rivista di Politica Economica, 87 (1997), 769–789.
[24] Giudici Paolo, Raffinetti Emanuela, (2011). A Gini concentration quality measure for ordinal variables. *Quaderni del Dipartimento di Economia, Statistica e Diritto*, Uuniversità di Pavia, Serie Statistica, no. 1/2011 (April 7, 2011), 18 pag.
[25] Haughton Jonathan, Khandker Shahidur R., (2009). Handbook on Poverty and Inequality. The World Bank, Washington D.C.
[26] Kobus Martyna, (2014). On the measurement of polarization for ordinal data. ECINEQ 2014 – 325, March 2014.
[27] Madden D., (2010). Ordinal and cardinal measures of health inequality: an empirical comparison. *Health Economics Letters* 19, 243-250.
[28] Makdissi Paul, Yazbecky Myra, (2014). Robust Wagstaff orderings of distributions of self-reported health status. University of Ottawa (Canada), University of Queensland (Australia), Report September 2014, 39 pages.
[29] Montalvo J., Reynal-Querol M., (2005). Ethnic polarization, potential conflict and civil wars. *American Economic Review*, **95**, 796-815.
[30] Permanyer Inaki, D'Ambrosio Conchita, (2013). Measuring social polarization with ordinal and categorical data. *Journal of Public Economic Theory*, 11 pag.
[31] Rodriguez J., Salas R., (2003). Extended bi-polarization and inequality measures. *Research on Economic Inequality*, **9**, 69–83.
[32] Zhang X., Kanbur R., (2001). What difference do polarization measures make? An application to China. *Journal of Development Studies*, **37**, 85-98.
[33] Van Doorslaer E., Jones A.M., (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics* 22, 61-87.
[34] Wang Y.Q, Tsui K.Y., (2000). Polarization orderings and new classes of polarization indices. Journal of Public Economic Theory, 2, 349-363.
[35] ***** Data bases "Diagnosis for Quality of Life", Research Institute for Quality of Life, Romanian Academy, (2003, 2006, 2010).