Using the median and the mean of the income to establish the poverty lines

Maria Livia ŞTEFĂNESCU*

Ştefan ŞTEFĂNESCU**

Abstract

One of the methods to estimate the poverty level inside a given population is based on how to define the poverty line values. Each person having his income under the poverty threshold will be considered to be poor. In the literature we distinguish at least three approaches: to evaluate the absolute poverty line, to find a relative poverty threshold depending on the main indicators of the income distribution in the analyzed community or to assume a subjective point of view.

The procedures for determining the relative poverty lines are based in practice on the mean or the median of the population income. To assure a concordance between the concrete estimation of several possible poverty thresholds and the poverty and inequality real phenomena we proposed to be verified three conditions. Finally, we also proved by examples that each of these restrictions are not obligatory satisfied by an arbitrary real positive data set.

The present study will be extended in the future to assure a theoretical support for estimating more exactly the relative poverty lines.

Keywords : relative poverty line, inequality, axioms, measurement, Gini index, income distribution, mean, median.

JEL classification: I32, C51, D63.

1. Measuring the inequality

To simplify our presentation we will consider in the next sections only the income of the persons in order to distinguish between the individuals. In reality the persons from a given population can be grouped using a multitude of criteria which are usually distinct from their monetary status. We must mention here the classifications based on the religion, education, age, gender, tradition, coefficient of intelligence, charisma, type of personality, individual skills, a variety of preferences, the acceptance of different social and cultural values, the possession of some goods, etc. The inequalities which appear inside a community define a complex phenomenon which takes into considerations all the attributes of a person.

The dispersion of the individual wage is often used in practice to measure the income inequality. Amartya Sen classified the inequality measures into objectives and normative ones. The normative indices are characterized by the effect on a social welfare redistribution. As an alternative way, the objective measures of inequality are based with priority on statistical methods to evaluate the effect of the income dispersion among the persons from the studied group.

For diminishing the social inequality is very important to emphasize the relevant causes of the income inequality. So, we distinguish in reality endogenous and exogenous causes too (for details see Jorge A. Charles-Coll [4]). In fact the income inequality is regarded as the cumulative effects of all endogenous and exogenous conditions which influence permanently, in an active way, the individuals, as well as the groups.

In [4] are analysed a lot of indicators designed to measure the inequality aspects from a population. The most well known coefficients are : relative mean deviation, variance, coefficient of variation, variance and standard deviation of the logarithm, income shares and income quantiles ratio, Theil and Pietra indices, Hoover indicator and Gini coefficient.

From all known inequality measures, Gini coefficient is the most popular index.

* Correspondence address : livia.stefanescu@gmail.com

** Research Institute for Quality of Life, Romanian Academy

Let X be the data set containing only nonnegative real numbers $x_1, x_2, x_3, \ldots, x_n$ which are not necessary distinct values. We sort all these values x_i, $0 \le i \le n$, in an ascending order such that $x_{(1)}$ and $x_{(n)}$ are the minimum, respectively the maximum of the x_i quantities. More precisely

$$x_{(1)} \le x_{(2)} \le x_{(3)} \le \ldots \le x_{(n-1)} \le x_{(n)}$$

To simplify our presentation we define the partial sums

$$S_k = x_{(0)} + x_{(1)} + x_{(2)} + x_{(3)} + \ldots + x_{(k)}, \ 0 \le k \le n$$

with the convention

$$x_{(0)} = 0$$

In this context, the Gini coefficient determined by the data set X, has the following form (more details in [12])

$$Gini(X) = 2 \sum_{k=1}^{n} A_k$$

where the areas A_k are given by the expressions

$$A_k = \big(\big((k-1)/n - S_{k-1}/S_n\big) + \big(k/n - S_k/S_n\big)\big)\big(k/n - (k-1)/n\big)/2$$

Concluding

$$Gini(X) = \frac{1}{n} \sum_{k=1}^{n} \big(\big((k-1)/n - S_{k-1}/S_n\big) + \big(k/n - S_k/S_n\big)\big) \tag{1.1}$$

The health status is a new other important characteristic of a person. Amartya Sen underlines in [10] that illness and health must figure too as a major condition regarding the quality of live of the persons. We emphasize here that the studies about the population health and health care are very important to propose an equitable public policy.

For this reason, to establish correctly the level of inequality which is present among the individuals of a given community, we often apply the concentration index of health (Erreygers, [5], [6]). Koolman and Van Doorslaer used also in [7] the concentration index, intending to measure the relative inequality of health in Europe. It is very well known the relationship between the concentration indicator, the coefficient of variation and a type of correlation index (see [7]).

The literature underlines a relevant link between the socio-economic status x of a person and his health status $h(p)$, p being the rank associated to the quantity x which belongs to the set values X ([14], [15], [7]). More precisely, if X are the wages of all individuals from a given population, then $p = F(x)$, where $F(x)$ is the cumulative distribution function of the income variable X.

In this context the health achievement indices defined by Wagstaff et all. ([14], [15], [16]) have the expression

$$W(X;r) = \int_0^1 r(1-p)^{r-1} h(p)\, dp \tag{1.2}$$

where r is regarded as a health inequality aversion parameter.

The Wagstaff's health index $W(X;2)$, resulted for an aversion of risk $r = 2$, is just the standard health concentration index. This particular variant of the indicator $W(X;r)$ is widely used in practice for measuring the health inequality.

Makdissi and Yazbeck proposed in [9] a more general class of health achievement and inequality indices based on an axiomatic approach. The accepted health indices must verify five axioms, that is : monotonicity, pure healh transfer, socioeconomic health transfer, pure health trasfer sensitivity and socioeconomic health transfer sensitivity (more details in [9]).

2. The poverty lines

The poverty reduction is the main goal of all development policies. But in the literature there is little agreement about a consensus regarding the poverty definition. Concerning the poverty measures there are at least four theoretical approaches. Consequently, we mention here the implementation of very different principles as follows: the monetary point of view, take into consideration the human potential, social exclusion or the participatory activity ([8]).

The monetary approach is the most commonly used in practice. In fact this method measures the underperformance in income or consumption of the individuals. For the subsequent we will measure the poverty basing only on the income X of the persons.

As a result, any person which has his income x under the poverty threshold π will be considered poor. The value of this threshold π is usually known in socio-economic studies as the "poverty line".

The choice of a concrete method which defines the poverty threshold is extremely important to understand correctly the poverty phenomenon. Some basic criteria to estimate the poverty lines are presented in [3]. Many studies mentioned by the literature link the poverty line with the minimal labor wage, the social security assistance or with the minimal pension for retirement and age (see [3]).

Concerning the establishment of the poverty lines we identify in practice three basic approaches : an absolute poverty line, a relative poverty threshold or a subjective one. The subjective score cumulates the people's perception about the minimum necessary for a standard family budget to assure a decent life. The absolute poverty threshold takes into consideration the lowest amount of money necessary to satisfy the basic needs of the individuals from a given community. The income or the expenditures of the persons are also used to determine the relative poverty upper limit. More exactly, the relative poverty lines π_1 and π_2 are computed as a proportion from the mean or median income X of the whole population.

Preserving the previous notations, the arithmetic average $Mean(X)$ of all values x_i, $1 \le i \le n$, from the data set X has the expression

$$Mean(X) = (x_1 + x_2 + x_3 + ... + x_n)/n = (x_{(1)} + x_{(2)} + x_{(3)} + ... + x_{(n)})/n \qquad (2.1)$$

The median indicator $Mdn(X)$ associated to the set x_i is computed by the formula

$$Mdn(X) = \begin{cases} (x_{(m)} + x_{(m+1)})/2 & ; \quad when \;\; n \;\; is \; even \\ x_{(m+1)} & ; \quad when \;\; n \;\; is \; odd \end{cases} \qquad (2.2)$$

where m is the integer part of the real value $n/2$.

Respecting these notations in the literature are proposed

$$\pi_1 = \alpha\, Mean(X) \qquad \pi_2 = \beta\, Mdn(X) \qquad (2.3)$$

with given proportion coefficients $0 < \alpha, \beta < 1$ (see for example Berthoud's approach, [2]).

Berthoud ([2]) mentioned the weights $\alpha = 0.5$ and $\beta = 0.6$ which were applied to year 1998 to determine the relative poverty lines.

In addition, the selected values of the coefficients α and β assured for year 1998 a very good approximation $\pi_1 \approx \pi_2$, that is

$$\alpha\, Mean(X) \approx \beta\, Mdn(X)$$

So, using a new other data set X we obtain different values for $Mean(X)$ and $Mdn(X)$ parameters which characterize the distribution of the income X. For this reason we could work in this case with another ratio α / β,

$$\alpha / \beta = Mdn(X) / Mean(X) \tag{2.4}$$

It is clear that for a same socio-economic analysis the proportion α / β does not have to fluctuate very much when is taken into consideration another subgroup of the population. So, the ratio α / β must be comparatively unchanged.

In practice, for every country is defined its own relative poverty line. So, to compare the poverty phenomenon from more countries is necessary to calibrate all these poverty thresholds ([2], [1]). For this reason is indispensable to establish an unitary methodology to estimate comparable cross-nationality poverty indicators ([1]).

To work with realistic poverty thresholds we should establish a set of axioms which must be verified by the poverty lines. In this context we suggest to consult a practical study about the reconsideration of the poverty lines in Malaysia ([13]).

We saw that the relative poverty threshold could be taken as a proportion from the population mean. In [11] is discussed an empirical method to locate the mean $Mean(X)$ of a random variable X depending on the mode of its probability density function. The form of the p.d.f. tails performs many times an important role to establish the real position of the $Mean(X)$ indicator.

3. Restrictions for poverty and inequality indices

Now we intend to propose some axioms which must be validated by all relative poverty lines.

In section 1 we talked about a large variety of indices for measuring the degree of inequality which exist inside the data set X. From all these indicators we will choose Gini's coefficient $Gini(X)$ to analyze the disproportions of the wages of X (formula 1.1).

We remind that both the mean and the median distribution parameter were applied together in section 2 to characterize the relative "poverty threshold" π for the elements of the data set X.

Taking into consideration the relation (2.4) with the EU coefficients $\alpha = 0.5$ and $\beta = 0.6$ (Berthoud [2]) we deduce $Mdn(X) / Mean(X) = \alpha / \beta < 1$. Therefore we will require to be satisfied the following condition (at least for EU countries):

C1. For any data sets X and Y we have $Mdn(X) < Mean(X)$.

In practice is important to study the concordance of the functions $Mean(X)$, $Mdn(X)$ and $Gini(X)$ concerning the monotony property. For this reason we will associate to all data sets X and Y the symmetric indicators :

$$\Delta(X,Y) = (Mdn(X) - Mdn(Y))(Mean(X) - Mean(Y)) \tag{3.1}$$

$$\delta(X,Y) = (Mdn(X) - Mdn(Y))(Gini(X) - Gini(Y)) \tag{3.2}$$

Our approach uses the values of the indices $Mdn(X)$ and $Mean(X)$ to define the relative poverty thresholds which characterize the poor people from the population X. In this context, to assure a good comparison between the poverty levels of two distinct populations X and Y, we imposed the restriction:

C2. $\Delta(X,Y) \geq 0$ for any populations X and Y.

Frequently we relate in practice the poverty phenomenon with the inequality aspects. More precisely, increasing (diminishing) the number of poor people inside the community X is often grown (decreased) the degree of inequality existing between the individuals of X. Generally, in many sociological studies, the relative poverty line is based on the median coefficient. Having in mind all these aspects we must certify the following conditional demand:

C3. For two arbitrary data sets X and Y we must have $\delta(X,Y) \geq 0$.

4. Some examples

In the subsequent we will consider seven data sets *X1-X7* , every variable X having the values x_i , $1 \leq i \leq n$, which belong to the open interval $(0 , 1)$. For this reason we always have the inequalities

$$0 < Mean(X) < 1 \ , \quad 0 < Mdn(X) < 1$$

Table 4.1 contains the concrete data sets *X1-X7* with the same volume $n = 21$.

In *Table 4.2* are listed the values of the indicators *Mean(X), Mdn(X), Gini(X)* and the ratio α / β which characterize the sets *X1-X7*.

Table 4.1 **The data sets *X1-X7*.**

***X1* :**						
0.2332	0.0794	0.0082	0.0683	0.9942	0.0852	0.8446
0.6496	0.0271	0.0525	0.8907	0.0810	0.7903	0.1047
0.8255	0.3399	0.0436	0.9569	0.9914	0.1028	0.7525
***X2* :**						
0.7620	0.9886	0.9806	0.2988	0.9357	0.9492	0.9946
0.9155	0.9681	0.6408	0.0022	0.2161	0.9786	0.1358
0.9712	0.2207	0.0092	0.7013	0.1516	0.9142	0.8019
***X3* :**						
0.1592	0.7845	0.7974	0.6231	0.1423	0.7782	0.6810
0.5910	0.0920	0.0837	0.0178	0.9963	0.0402	0.1319
0.1712	0.9778	0.8741	0.9804	0.0706	0.1340	0.9388
***X4* :**						
0.7649	0.8084	0.7062	0.7037	0.0525	0.1401	0.8132
0.1971	0.5504	0.8323	0.0191	0.6570	0.6229	0.9089
0.0818	0.4858	0.0980	0.6403	0.6664	0.2556	0.0301
***X5* :**						
0.9140	0.0169	0.0410	0.1140	0.7097	0.5125	0.9660
0.1539	0.2550	0.1086	0.3585	0.5579	0.1673	0.9905
0.9838	0.9393	0.8057	0.5147	0.4281	0.0242	0.7740
***X6* :**						
0.9299	0.0442	0.9902	0.9427	0.1680	0.0519	0.9169
0.9746	0.1406	0.9185	0.9327	0.0979	0.7015	0.3001
0.9083	0.8628	0.0196	0.8414	0.2232	0.8555	0.0211
***X7* :**						
0.5541	0.2498	0.9642	0.1116	0.7625	0.7570	0.4609
0.8994	0.9012	0.4822	0.5389	0.2056	0.7568	0.6112
0.0669	0.2622	0.8433	0.9843	0.5792	0.2476	0.2129

***Table 4.2.* The statistical characteristics of the data sets *X1-X7*.**

Data set	*Mean*(*X*)	*Mdn*(*X*)	*Gini*(*X*)	α / β
X1	0.42483	0.23317	0.49392	0.54885
X2	0.64461	0.80188	0.30690	1.24398
X3	0.47932	0.59102	0.43065	1.23304
X4	0.47784	0.62293	0.35888	1.30364
X5	0.49217	0.51246	0.40472	1.04123
X6	0.56389	0.84139	0.37683	1.49212
X7	0.54533	0.55409	0.30372	1.01606

Now we intend to distinguish between the sets *X1-X7* using the criteria ***C1-C3***.

First, only the data set *X1* validates the condition ***C1***. All the other variables, that is *X2-X7*, do not verify the restriction ***C1***. From the point of view of the requirement ***C1*** the variable *X1* plays a distinct role.

In an opposite way, most of the pairs of variables (X_j, X_k) accomplished the constraint ***C2***, We enumerate here the cases : *(X1 , X2) ; (X1 , X3) ; (X1 , X4) ; (X1 , X5) ; (X1 , X6) ; (X1 , X7) ; (X2 , X3) ; (X2 , X4) ; (X2 , X5) ; (X2 , X7) ; (X3 , X6) ; (X4 , X6) ; (X5 , X6) ; (X5 , X7) ; (X6 , X7) . As alternative, the pairs (X2 , X6) ; (X3 , X4) ; (X3 , X5) ; (X3 , X7) ; (X4 , X5) ; (X4 , X7)* do not verify the axiom ***C2***.

The condition **C3** is satisfied by the pairs of variables : (X2 , X6) ; (X2 , X7) ; (X3 , X5) ; (X3 , X7) ; (X4 , X6) ; (X4 , X7) ; (X6 , X7) . But we do not include here the pairs of data sets : (X1 , X2) ; (X1 , X3) ; (X1 , X4) ; (X1 , X5) ; (X1 , X6) ; (X1 , X7) ; (X2 , X3) ; (X2 , X4) ; (X2 , X5) ; (X3 , X4) ; (X3 , X6) ; (X4 , X5) ; (X5 , X6) ; (X5 , X7).

Concluding, the restrictions ***C1-C3*** are satisfied only by a part of the real data X . A lot of imagined data sets X do not verify necessary all these axioms.

5. Concluding remarks

The relative poverty lines were defined in practice by using a proportion from the mean or from the median of the wages for the individuals of a known population.

Having in mind a possible link between the poverty and the inequality phenomena and to ensure a concordance among different poverty thresholds we proposed three axioms ***C1-C3***. These axioms must be verified by all income data sets X .

Finally we proved that an arbitrary real data set do not satisfy necessary the conditions ***C1-C3***. Therefore, more elaborate theoretical studies about the concrete selection of the relative poverty lines are needed to be developed in the future.

References

[1] Richard Anker , *A new methodology for estimating internationally comparable poverty lines and living wage rates*. Working Paper no. 72, Policy Integration Department, Statistical Development and Analysis Group, International Labour Office, Geneva, July 2005.

[2] Richard Berthoud, *Calibrating a cross European poverty lines*. Institute for Social and Economic Research, University of Essex, No. 2012-02, January 2012.

[3] Bogdan Bogdanov, Kentaro Nakajima, Susan Kuivalainen, Pasi Moisio, Yuka Takeda, *Methods for estimating the poverty lines: Four country case studies.* International Labour Organization, ILO Decent Work Technical Support Team and Country Office for Eastern Europe and Central Asia, 2012.

[4] Jorge A. Charles-Coll, *Understanding income inequality: concept, causes and measurement*. International Journal of Economics and Management Sciences, vol.1, no.3, 2011, 17-28.

[5] G. Erreygers, G. (2006), *Beyond the health concentration index: An Atkinson alternative for the measurement of the socioeconomic inequality of health.* University of Antwerp, Faculty of Applied Economics, Research Paper no. 29 (2006).

[6] G. Erreygers, *Correcting the concentration index*, Journal of Health Economics, 28 (2009), 504-515.

[7] Xander Koolman, Eddy van Doorslaer, *On the interpretation of a concentration index of inequality*. Working Paper #4 (August 2003), ECuity II Project (“Economic determinants of the distribution of health and health care in Europe”), European Community’s Biomed II programme (contract BMH4-CT98-3352).

[8] Caterina Ruggeri Laderchi, Ruhi Saith, Frances Stewart, *Does it matter that we don't agree on the definition of poverty? A comparison of four approaches*. Queen Elizabeth House, University of Oxford, Working Paper Number 107, May 2003.

[9] Paul Makdissi, Myra Yazbeck, *Avoiding blindness to health status: A new class of health achievement and inequality indices*, Working Paper #1207E, Department of Economics, Faculty of Social Sciences, University of Ottawa, May 2012.

[10] Amartya Sen, *Why heath equity*? Health Economics, 11 (2002), 659–666.

[11] Maria Livia Ştefănescu, Ştefan Ştefănescu, *On an empiric method to locate the mean of a random variable depending on the mode of its probability density function*. Studii şi Cercetări de Calcul Economic şi Cibernetică Economică, vol.35, no.4 (2001), 19-29 (in Romanian).

[12] Kuan Xu, *How has the literature on Gini’s index evolved in the past 80 years?* Dalhousie University, Department of Economics, Halifax, Nova Scotia, Canada, December 2, 2004, 41 pages.

[13] Ragayah Haji Mat Zin, *Understanding the formulation of the revised poverty line in Malaysia*. Akademika 70 (January 2007), 21-39.

[14] A. Wagstaff, E. van Doorslaer, P. Paci, *Equity in the finance and delivery of health care: some tentative cross-country comparisons*, Oxford Review of Economic Policy, 5 (1989), 89-112.

[15] A Wagstaff, P. Paci, E. van Doorslaer, *On the measurement of inequalities in health*, Social Science and Medicine, 33 (1991), 545-557.

[16] A. Wagstaff, *Inequality aversion, health inequalities and health achievement*, Journal of Health Economics, 21 (2002), 627-641.