Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 2, 187 – 198
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n2p187

ON THE DECOMPOSITION OF THE SCHUTZ COEFFICIENT: AN EXACT APPROACH WITH AN APPLICATION

Elsayed A. Habib*

Department of Statistics and Mathematics, Benha University, Egypt
&
Management and Marketing Department, University of Bahrain, Kingdom of Bahrain

Received 23 March 2011; Accepted 13 February 2012
Available online 14 October 2012

Abstract: *Decomposing inequality indices across household groups are useful in estimating the contribution of each component to total inequality. Decomposing relative inequality index for the Schutz coefficient (S-coefficient) is not simple since the functional form of inequality indices is not additively separable in incomes. In this article, the decomposition of Schultz coefficient across sub-groups is derived where it has a general form of decomposition as between-groups term, within-groups term and error term. Moreover, the error analysis is used to obtain the exact decomposition for the Schutz coefficient by dividing the error term to within-groups and between-groups terms. The final two main component types that we obtained are the exclusive within-groups and between-groups terms. Several examples are given that illustrate the benefits of the proposed method.*

Keywords: *Lorenz curve, mean absolute deviation, Pietra ratio, Robin Hood index.*

1. Introduction

The Lorenz curve is a tool used to represent income distributions; it tells us which proportion of total income is in the hands of a given percentage of population. However, instead of ending with income shares, the Lorenz curve relates the cumulative proportion of income to the cumulative proportion of individuals; see, for example, [5], [12] and [13].

* E-mail: shabib40@gmail.com

The Gini index is the most commonly applied inequality measure in the Literature, probably because of its link with Lorenz curve which gives an intuitive and graphical representation of inequality, where it is defined as the ratio of the area between the Lorenz curve and the equality line to the area below the equality line. Its main application has been in the measurement of inequalities in income and wealth, but it has also a long history in other areas; see, for example [9], [12], [7], and [1].

The Schutz coefficient (S-coefficient), also known as the Robin Hood index and Pietra ratio, has link with the Lorenz curve which is equivalent to the maximum vertical distance between the Lorenz curve and the equal line of incomes. Its value approximates the share of total income that has to be transferred from households above the mean to those below the mean to achieve equality in the distribution of incomes; see, [11].

In this article, the decomposition of S-coefficient across household groups is studied where it has a general form of decomposition as between-groups term, within-groups term and error term. The error analysis is proposed by dividing the error term to within-groups error and between-groups error and adds these errors to between-groups and within-groups terms to obtain the perfect decomposition for the S-coefficient.

In Section 2 the S-coefficient is defined in terms of the ratio between the mean absolute deviation and the population mean as well as in terms of "overs" and "unders". Some desirable features for S-coefficient index are studied in Section 3. Decomposition of S-coefficient by household groups is presented in Section 4. An exact approach of decomposition is derived in Section 5. An empirical study is given in Section 6. Section 7 is devoted to conclusion.

2. Schutz coefficient (S-coefficient)

2.1 Population S-coefficient

Let a vector of income X (a positive random variable) from a continuous distribution with cumulative distribution function (cdf) $F(x) = F_X = F$, density function $f(x)$, $x(F) = F^{-1}(x)$ quantile function and let $X_{i:n}$ denote the corresponding order statistics for a general distribution function $F(x)$. One of the most widely used measures for the extent of inequality is the Lorenz Curve. Let the population mean:

$$\mu = \int x f(x) dx = \int_0^1 x(F) dF(x)$$

is assumed to be finite and positive. The cumulative distribution function:

$$F(x) = \int_0^x f(x) dx$$

represents the proportion of persons with income less than or equal to x then the Lorenz curve is defined as:

$$L(x) = \frac{\int_0^x xf(x)dx}{\mu} = \frac{\int_0^F x(F)dF(x)}{\mu} = L(F)$$

represents the proportionate share of these persons in the aggregate income of all persons; see, for example, [6]. Note that, $L(x)$ exists only if μ exists.

The S-coefficient in terms of Lorenz curve is the **maximum vertical distance** between the Lorenz curve or the cumulative portion of the total income held below a certain income percentile, and the perfect equality Line, that is the 45 degree line of equal incomes. Figure 1 below show the S-coefficient and the Lorenz curve. Therefore, S-coefficient is given by:

$$S = DP = Max.[F(x) - L(F(x))] = F(\mu) - L(F(\mu)) = \int_0^\mu xf(x)(1 - \frac{1}{\mu})dx = \frac{MAD}{2\mu}$$

See; [8]. The mean absolute deviation (MAD) about population mean μ is defined as

$$MAD = E|X - \mu| = \int |x - \mu| f(x)dx$$

This equation expresses the S-coefficient as a ratio between the mean absolute deviation and the twice of the population mean.

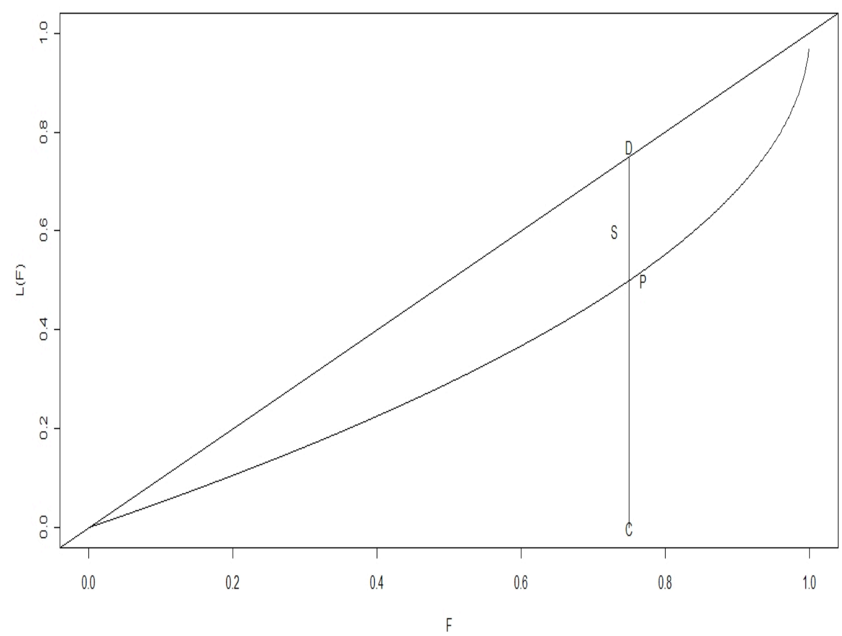

Figure 1. Schultz coefficient and Lorenz curve.

2.2 S-coefficient in terms of "overs" and "unders"

Another important equivalent form for the population S-coefficient can be written as the gap between the individual's income x_i and the population mean income μ. The MAD can be rewritten as:

$$MAD = E|X-\mu| = 2\int_{\mu}^{\infty}(x-\mu)f(x)dx = -2\int_{-\infty}^{\mu}(\mu-x)f(x)dx$$

Therefore, the S-coefficient is:

$$S = E(D_i)/\mu$$

where:

$$D_i = \begin{cases}(X_i-\mu), & X_i > \mu \\ 0 & X_i \le \mu\end{cases}$$

Note that the values of D_i represent a person's income to be more than the population mean μ ("overs") and $E(D_i)$ represents the expected amount of money which have to be transferred from households above the mean to those below the mean to achieve equality. Moreover, in the same manner we may write S-coefficient as:

$$S = -E(\overline{D}_i)/\mu$$

where:

$$\overline{D}_i = \begin{cases}(\mu - X_i), & X_i > \mu \\ 0 & X_i \le \mu\end{cases}$$

the values of $\overline{D}_i$ represent a person's need of money to achieve equality ("unders") and $E(\overline{D}_i)$ represent the expected need of money to achieve equality.

Example

The Pareto distribution with shape and scale parameters; k and b is:

$$f(x;b,k) = kb^k x^{-(k+1)}, \quad b \le x < \infty$$

with mean $\mu = kb/(k-1)$. The MAD can be obtained as:

$$MAD = 2\int_{\mu}^{\infty}(x-\mu)kb^k x^{-(k+1)}dx$$

Hence,

$$MAD = \frac{2b}{k-1}\left(\frac{k}{k-1}\right)^{-(k-1)}$$

Therefore:

$$S = \frac{MAD}{2\mu} = \frac{1}{k}\left(\frac{k}{k-1}\right)^{-(k-1)} = \frac{1}{k-1}\left(\frac{k-1}{k}\right)^{k}$$

This depends on the scale parameter k.

2.3 Sample S-coefficient

For a sample or population of size, n, an income distribution is $x_1, x_2, \ldots, x_n$ of nonnegative values and their order statistics are, $x_{1:n}, \ldots, x_{n:n}$. We define the following nonparametric estimators for S-coefficient using data.

1. Based on the mean absolute deviation:

$$\hat{S} = \frac{mad}{2\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{2n\bar{x}}$$

where $\bar{x}$ $\bar{x}$ is the mean and n n the number of observations.

2. Based on the ranking of the income in ascending order:

$$\hat{S} = \frac{\sum_{i=v}^{n} (x_{i:n} - \bar{x})}{n\bar{x}} = \frac{-\sum_{i=1}^{\dot{v}} (x_{i:n} - \bar{x})}{n\bar{x}}$$

where $x_{i:n}$ the order income, and $v = \left[\sum_{i=1}^{n} I\left(x_i \le \bar{x}\right)\right] + 1$, is the number of values less than or equal to the mean plus 1, $\dot{v} = \sum_{i=1}^{n} I(x_i < \bar{x})$, is the number of values less than mean and I indicator function, 1 if true and 0 else.

3. Based on the expected money transferred from rich to poor:

$$\hat{S} = \frac{\sum_{i=1}^{n} d_{i:n}}{n\bar{x}} = \frac{\bar{d}}{\bar{x}}$$

Where

$$d_i = \begin{cases} x_i - \bar{x} & x_i > \bar{x} \\ 0 & x_i \le \bar{x} \end{cases}$$

Note that the index has the same value if we used $\tilde{d}_i$.

3. Desirable features

There are some desirable features that make an index a good measure of inequality; see, for example, [2]. To see if the S-coefficient satisfies these features, namely:

- *Range of S*. When all incomes are equal, the numerator of S is equal to zero, as any difference between any income and the mean is zero. When all incomes are zero but the last is not, S has a maximum value at 1, we have $m = \frac{(x_n - \bar{x})/n}{x_n/n} = 1 - \frac{\bar{x}}{x_n} \approx 1$ where $\bar{x} <<< x_n$.

- *Mean independence (scale invariant).* This means that if all incomes were multiplied by factor a, S remains unchanged where both the numerator and the denominator increases with the factor a.
- *S is not translation invariant:* when all incomes of the original income distribution are added (subtracted) the same amount, the numerator of S remains unchanged, while the denominator increases (decreases) by an amount which is equal to the original addition (subtraction) times the number of observations. Therefore S decreases (increases).
- *Population size independence:* requires inequality measures to be invariant to replications of the population; merging two identical distributions should not alter inequality. For any scalar, $k > 0$, $S(x) = S(x[k])$, where $x[k]$ is a concatenation of the vector x $\mathbf{x}$, k $\mathbf{k}$ time, where the mean absolute deviation and the mean will not change.
- *Symmetry.* When two persons swap their income, S will not change as the value of d_i will not change and we sum for all individuals.
- *Pigou-Dalton Transfer sensitivity.* When income is redistributed from richer to poorer, S decreases as numerator $d_i = 0$ (decreases). The opposite holds true for redistributions from poorer to richer individuals. It is worth noting that S react to redistribution only for transfers across the mean. In other words, if the change in the same side, S will not change.
- *Decomposability.* This means that inequality may be broken down by population groups or income sources or in other dimensions. This requires overall inequality to be related consistently to constituent parts of the distribution, such as population sub-groups. See the details in Section 4.

4. Decomposition

Decomposing inequality indices across household groups or income sources are useful in estimating the contribution of each component to total inequality. This can help policy makers draw efficient policies to reduce disparities in the distribution of incomes using targeting tools. Decomposing relative inequality indices, such as the Gini coefficient, are not a simple procedure since, in many cases, the functional form of inequality indices are not additively separable in incomes; see, for example, [10] and [14]. More importantly, for some of the indices on which this decomposition can be performed, the interpretation of the decomposition components is often not well founded; see, for example, [3] and [4]. This section aims to obtain the decomposition of S-coefficient by sub-groups. The decomposition seems to have an error term. We propose a method by analyzing the error term to within-groups and between-groups terms. The final two main component types that we obtained are the exclusive within-groups and between-groups.

4.1 Decomposing by population sub-groups

Inequality may stem from different groups or sectors of population with different intensities (workers and pensioners; rural and urban households; households with and without children). Avery important features of inequality measures are decomposability, i.e. the possibility of calculating the contribution of each group to total insulate. In its most general form, decomposability of inequality measures requires a consistent relation between overall inequality and its parts. When dealing with decomposability, we must be able to distinguish between within inequality (W) and between inequalities (B). The within inequality element captures the inequality due to variability of income within each group, while the between inequality captures the inequality due to variability of income across different groups. The most general form decomposition of any inequality index (I) generates a within element (I_W), a between element (I_B) and a residual term (E):

$$I = I_W + I_B + E$$

Gini index has this general form; see, [6]. While Theil index is perfectly decomposable without a residual term ($E = 0$); see, for example, [14], therefore, their economic interpretation is straightforward.

Our approach to decompose the S-coefficient seems to have the general form with residual. Assume we have G different groups from a population with individuals in each group, n_g, $g = 1,...,G$ and their income is x_{gi}, $i = 1,2,...,n_g$. We define:

$$d_{gi} = \begin{cases} x_{gi} - \bar{x} & x_{gi} > \bar{x} \\ 0 & else \end{cases}$$

As overall variation, $\bar{x} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} x_{gi} / n$,

$$y_{gi} = \begin{cases} x_{gi} - \bar{x}_g & x_{gi} > \bar{x}_g \\ 0 & else \end{cases}$$

The within-groups variation, $\bar{x}_g = \sum_{i=1}^{n_g} x_{gi} / n_g$, $g = 1,2,...,G$ and

$$z_g = \begin{cases} \bar{x}_g - \bar{x} & \bar{x}_g > \bar{x} \\ 0 & else \end{cases}$$

the between-groups variation.

Therefore, d_{gi} represents a person income comparing with the grand mean $\bar{x}$, and y_{gi} represents a person income comparing with the mean income inside the group, $\bar{x}_g$ while z_g represents the mean income for each group comparing with the grand mean. Based on this explanation, we may write S-coefficient as:

$$\hat{S} = \frac{\sum_{g=1}^{G} \sum_{i=1}^{n_g} d_{gi}}{n\bar{x}} = \frac{\sum_{g=1}^{G} \sum_{i=1}^{n_g} (y_{gi} + z_g + d_{gi} - y_{gi} - z_g)}{n\bar{x}}$$

First, we start by within-groups term:

$$W_1 = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g} y_{gi}}{n\bar{x}} = \sum_{g=1}^{G}\frac{n_g}{n}\frac{\bar{x}_g}{\bar{x}}\sum_{i=1}^{n_g}\frac{y_{gi}}{n_g\bar{x}_g} = \sum_{g=1}^{G} P_g O_g \hat{S}_g$$

$P_g = n_g / n$ is the population share, $O_g = \bar{x}_g / \bar{x}$ is the mean income share, and $S_g = \sum_{i=1}^{n_g} y_{gi} / (n_g \bar{x}_g)$ is the S-coefficient within- groups. For the between-groups term:

$$B_1 = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g} z_g}{n\bar{x}}$$

With algebraic manipulation we obtain:

$$B_1 = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g} z_g}{n\bar{x}} = \frac{\bar{z}}{\bar{x}_G}\frac{\bar{x}_G}{\bar{x}}\sum_{g=1}^{G}\frac{n_g}{n}\frac{z_g}{\bar{z}} = S_b C \sum_{g=1}^{G} P_g T_g$$

$\bar{x}_G = \sum_{g=1}^{G}\bar{x}_g / G, \bar{z} = \sum_{g=1}^{G} z_g / G, S_b = \bar{z} / \bar{x}_G$ is S-coefficient between-groups, $P_g = n_g / n$ is the population share, $C = \bar{x}_G / \bar{x}$ is the grand mean of between-groups share and $T_g = z_g / \bar{z}$ is the between-groups variation share. The remaining term (error term) can now be written as:

$$E = \frac{\sum\sum(d_{gi} - y_{gi} - z_g)}{n\bar{x}} = \frac{n\bar{d} - \sum_{g=1}^{G} n_g \bar{y}_g - \sum_{g=1}^{G} n_g z_g}{n\bar{x}} = \frac{\bar{d}}{\bar{x}} - \sum_{g=1}^{G}\frac{n_g}{n}\left[\frac{\bar{y}_g}{\bar{x}} + \frac{z_g}{\bar{x}}\right]$$

The decomposition of S-coefficient is:

$$\hat{S} = W_1 + B_1 + E$$

The error term can be considered as

1. The necessary term to maintain the identity.
2. The overestimation of the total variation created by within-groups term plus between-groups term; it could be separate to within-groups error and between-groups error and added to each term to obtain perfect decomposition.

5. Proposed method for perfect decomposition

Since the error term captures information between-groups and within-groups effect at the same time, they should be adjusted for error. We suggest that the error term will be analyzed to two terms; within-groups term and between-groups term as follows:

$$E_{WB} = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g}(d_{gi} - y_{gi} - z_g)_W}{n\bar{x}} + \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g}(d_{gi} - y_{gi} - z_g)_B}{n\bar{x}}$$

This term can be written as:

$$E_{WB} = \frac{\sum_{i=1}^{n} E_{wi}}{n\bar{x}} + \frac{\sum_{i=1}^{n} E_{bi}}{n\bar{x}} = E_W + E_B$$

Where E_W captures within-groups error and will be added to W_1 and E_B captures between-groups error and will be added to B_1 to obtain the perfect decomposition for S-coefficient as:

$$\hat{S} = \left[\sum_{g=1}^{G} P_g O_g S_g + E_W\right] + \left[\hat{S}_b C \sum_{g=1}^{G} P_g T_g + E_B\right]$$

The perfect decomposition is:

$$\hat{S} = W + B$$

Where the within-groups inequality is:

$$W = \left[\sum_{g=1}^{G} P_g O_g S_g + E_W\right]$$

and the between-groups inequality is:

$$B = \left[\hat{S}_b C \sum_{g=1}^{G} P_g T_g + E_B\right]$$

The separation of the error term (E) to within-groups and between-groups errors could be based on a proportionate of each error to the total error.

Example

Suppose we have two groups every group consists of two individuals. The data are given in Table 1. The S-coefficient is $\frac{11}{(4)(8)} = 0.34375$. The total variation which we want to capture is 11 while the within-groups plus between-groups variation capture (11+4=15), i.e., they overestimate the total variation by 4 units (error term).

From Table 1 the error term is divided equally between with-groups and between-groups errors (-0.0625-0.0625=-0.125) which should be added to each inequality. Therefore, the within contribution is:

$$W = \frac{(2)(6)(3)}{(4)(8)(2)(6)} + \frac{(2)(10)(8)}{(4)(8)(2)(10)} - .0625 = 0.34375 - 0.0625 = 0.28125$$

and the between contribution is

$$B=\frac{(1)(8)}{(8)(8)}\left[\frac{(2)(0)}{(4)(1)}+\frac{(2)(2)}{(4)(1)}\right]-0.0625=0.125-0.0625=0.0625$$

Hence,

$$\hat{S}=0.28125+0.0625=\frac{11}{32}=0.34375$$

Table 1. Error term analysis to within-groups and between-groups errors.

g	x_{gi}	Total d_{gi}	Within y_{gi}	Between z_{gi}	E	E_W	E_B
	variation analysis				Error analysis		
1	**3**	0	0	0	0	0	0
	9	1	3	0	-2	-2	0
	$\bar{x}_1=6$	$n_1=2$					
2	**2**	0	0	2	-2	0	-2
	18	10	8	2	0	0	0
To.	32	11	11	4	-4	-2	-2
	$\bar{x}_2=10$	$n_2=2$					
	$\bar{x}=8$	$n=4$					
	Error	$E_{WB}=\frac{4}{4(8)}=0.125$		$E_W=\frac{2}{4(8)}=.0625$		$E_B=\frac{2}{4(8)}=.0625$	

6. An application

This section presents the results of the estimation of the S-coefficient from labor force survey conducted on non-Bahraini in 2008. Table 2 shows the distribution of the wage for the male and female is pertaining to a sample of 8571 individuals (8154 male and 417 female). These data are obtained from labor force survey conducted in 2008.

From table 2, the S-coefficient is 0.509. The contribution of within-groups is 0.495 and between-groups is 0.014. This indicates that the disparities in the distribution of the data return to within-groups. This will help policy makers draw efficient policies to reduce disparities in the distribution of incomes within each group. Indeed, it may think in S-coefficient not only an alternative to the popular Gini index, but rather, a far more natural and meaningful quantitative tool for the measurement of egalitarianism and consequently for the measurement of statistical heterogeneity.

Table 2. Wage distribution for non-Bahraini in 2008.

Class	Range (BD)	# of Male	# of Female	Total
1	Below 99	5760	164	5924
2	100-149	765	53	818
3	150-199	267	30	297
4	200-249	146	26	172
5	250-299	405	24	429
6	300-349	183	31	214
7	350-399	81	12	93
8	400-499	43	10	53
9	450-499	85	8	93
10	500-599	52	12	64
11	600-699	41	9	50
12	700-799	31	7	38
13	800-899	27	5	32
14	900-999	23	3	26
15	1000-1499	84	12	96
16	1500 and over	161	11	172
	Total	8154	417	8571
	Mean wage	194	299	199

6. Conclusion

Avery important feature of inequality measures is decomposability, i.e. the possibility of calculating the contribution of each group to total insulates. The S-coefficient is contained several illuminating insights. Graphically it is represented as the vertical distance between the line of equality and generalized Lorenz curve. The decomposability based on S-coefficient presents a much greater challenge, but the analysis of error term of the S-coefficient showed clearly how it was derived the exact decomposability by analyzing error to inclusive between and within errors that added to between and within groups.

Moreover, the representation of S-coefficient in terms of “over” and “under”-unlike other measures of inequalities-such as Gini coefficient- reflected important information about the expected value to be transferred from households above the mean to those below the mean to achieve equality. In other words it reflected more information about the upper and lower tails of the distribution. Still many things need to be explored in S-coefficient, for example, connection of S-coefficient to the variance-mean ratio, entropy maximization and analysis of variance.

Acknowledgment

The author is greatly grateful to associate editor and two reviewers for careful reading, valuable comments and highly constructive suggestions that have led to clarity, better presentation and improvements in the manuscript.

References

[1]. Barrett, G.F., and Donald, S.G. (2009). Statistical inference with generalized Gini indices of inequality, poverty, and welfare. *Journal of Business and Economic Statistics*, **20**, 1-17.

[2]. Cowell, F.A. (1995). *Measuring Inquality*. (2nd edition). Harvester Wheatsheaf, Hamel Hempstead.

[3]. Dagum, C. (2008). Inequality decomposition, directional economic distance, metric distance, and Gini dissimilarity between income distributions. In *Advances on Income Inequality and Concentration Measures,* Eds., Betti, G.,Lemmi, A., Routledge, London, pp. 131-160.

[4]. Deutsch, J., Silber, J. (2008). On the Shapley value and the decomposition of inequality by population subgroups with special emphasis on the Gini index. In *Advances on Income Inequality and Concentration Measures*, Eds., Betti, G., Lemmi, A., Routledge, London, pp. 161-178.

[5]. Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Ecnometrica*, **39**, 1037-1039.

[6]. Griffiths, W. (2008). *On Dagum's decomposition of the Gini coefficient*. URL:http//www.economics.unimelb.edu.au.

[7]. Kikuchi, T. (2007). An analysis of impacts of development on Gini inequality using grouped and individual observations: Examples from the 1998 Vietnamese Household Expenditure Data. *Journal of Asian Economics*, **18**, 537-552.

[8]. Lambert, P. J. and Aronson, J. (2001). Inequality decomposition analysis and the Gini coefficient revisited. *Economic Journal*, **103**, 1221–27.

[9]. Milanovic, B. (2002). True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *Economic Journal*, **112**, 51-92.

[10]. Sastry, D.V.S., Kelkar, U.R. (1994). Note on the decomposition of the Gini inequality. *Review of Economics and Statistics*, **76**, 584-586.

[11]. Schutz, R. (1951). On measurement of income inequality. *The American Economic Review*, **41**, 107-122.

[12]. Sen, A. (1997). *On Economic Inquality*. Clarendon Press, Oxford, UK.

[13]. Shahateet, M.I. (2006). How serious regional economic inequality in Jordan? Evidence from two national household surveys. *American Journal of Applied Sciences*, **3**, 1735-1744.

[14]. Shorrocks, A. F. (1984). Inequality decomposition by population subgroups. *Econometrica*, **52**, 1369–1385.