Statistica & Applicazioni - Vol. XIV, n. 1, 2016, pp. 49-82

DECOMPOSITION BY SUBPOPULATIONS OF THE POINT AND THE SYNTHETIC GINI INEQUALITY INDEXES

Michele Zenga*

SUMMARY

Using the two-step approach proposed by Radaelli (2008, 2010) and by Zenga (2016) for the decomposition by subpopulations of the Zenga (2007) inequality index, this paper obtains the decomposition by k subpopulations of the point $G_h(Y)$ *and the synthetic* $G(Y) = \sum_{h=1}^{r} G_h(Y)\frac{n_{h.}}{N}$ *Gini (1914) inequality indexes: where N is the number of the households of the whole population and* $n_{h.}$ *is the frequency of the households with total income* $Y = y_h$*. In the first step, the point Gini index is decomposed in the sum of* $k \times k$ *contributions* C_{hlg} : $G_h(Y) = \sum_{l=1}^{k}\sum_{g=1}^{k} C_{hlg}$ *. In the second step, the synthetic Gini index is decomposed in the sum of* $k \times k$ *contributions* $C_{.lg} = \sum_{h=1}^{r} C_{hlg}\frac{n_{h.}}{N}$ *. From the reported decompositions, are obtained the decompositions* $G_h(Y) = \sum_{l=1}^{k} C_{hl.}$ *and* $G(Y) = \sum_{l=1}^{k} C_{.l.}$*, where* $C_{hl.} = \sum_{g=1}^{k} C_{hlg}$ *and* $C_{.l.} = \sum_{g=1}^{k} C_{.lg}$ *are the contributions of the subpopulation l to the point and to the synthetic Gini indexes, respectively. In addition, the within and the between parts of* $C_{hl.}$*,* $C_{.l.}$*,* $G_h(Y)$ *and of* $G(Y)$ *can be obtained by summing the respective contributions. In the within parts only incomes of the same subpopulations are compared, while in the between parts only incomes of different subpopulations are compared. The decompositions proposed in this paper are applied to the net disposable income of the 8151 Italian households partitioned in three macroregions, supplied by the 2012 Bank of Italy sample survey on household income and wealth. The decompositions of the point index* $G_{h(p)}(Y) = G_{(p)}(Y)$ *are illustrated for three values* $(0.10; 0.50; 0.95)$ *of the cumulative relative frequency p. For* $p = 0.95$ *some contributions are negative. Zenga and Valli (2016) have obtained similar results for the decompositions by subpopulations of the point and synthetic Bonferroni indexes.*

Keywords: *Gini Index, Point Inequality Index, Synthetic Inequality Index, Decomposition by Subpopulatios.*

1. INTRODUCTION

Radaelli (2010) provided a comprehensive survey of the literature on the decomposition by k subpopulations of the synthetic Gini (1914) inequality index. The decompositions considered by Radaelli are essentially based on the decomposition of the Gini mean difference with replacement $\Delta_R(Y)$. In particular, $\Delta_R(Y)$ is expressed as the following weighted mean of the k^2 Gini mean differences Δ_{lg} between the subpopulations l and g : $\Delta_R(Y) = \sum_{l=1}^{k}\sum_{g=1}^{k} \Delta_{lg}\, p_{.l}\, p_{.g}$, where the weights are given by the product of the corresponding subpopulation shares. From this basic decompo-

* Dipartimento di Statistica e Metodi Quantitativi - Università Milano-Bicocca - Piazza dell'Ateneo Nuovo, 1 - 20126 MILANO (e-mail: michele.zenga@unimib.it).

sition of $\Delta_R(Y)$ it derives that: $\Delta_R(Y) = \Delta_W(Y) + \Delta_B(Y)$, where $\Delta_W(Y) = \sum_{l=1}^{k} \Delta_{ll} p_{.l}^2$ and $\Delta_B(Y) = \sum_{l=1}^{k} \sum_{g \neq l} \Delta_{lg} p_{.l} p_{.g}$ are the within and the between components of $\Delta_R(Y)$, respectively. Finally, dividing both sides of the last decomposition of $\Delta_R(Y)$ by $2 \cdot M(Y)$ the following decomposition of the Gini concentration ratio $G(Y) = \frac{\Delta_R(Y)}{2 \cdot M(Y)}$ is obtained: $G(Y) = G_W(Y) + G_B(Y)$, where $G_W(Y) = \sum_{l=1}^{k} \frac{\Delta_{ll}}{2 \cdot M(Y)} p_{.l}^2$ and $G_B(Y) = \sum_{l=1}^{k} \sum_{g \neq l} \frac{\Delta_{lg}}{2 \cdot M(Y)} p_{.l} p_{.g}$ are the within and the between components of $G(Y)$, respectively. Recently, Ogwang (2014) using a ‘‘trick regression model’’ between the income ordered values $y_{(i)}$ and the weights $(2 \cdot i - N - 1)$ obtained the decomposition by subpopulations of $G(Y)$ into three components: within, between and interaction. Besides that, Ogwang decomposed $G(Y)$ in the sum of the contributions of each subpopulation.

Unfortunatly, computing $G(Y)$ through the mean difference the connections with the Lorenz curve are lost. The present paper mantains these connections because, first decomposes the point Gini index $G_{(i)}(Y) = 2(p_{(i)} - q_{(i)})$ by subpopulations, and second, using the relation $G(Y) = \frac{1}{N} \sum_{i=1}^{N} G_{(i)}(Y)$ the corrisponding decomposition of $G(Y)$ is derived. The next part of the paper is organized as follows. In the case of N values of Y arranged in non-decreasing order $0 \leq y_{(1)} \leq \ldots \leq y_{(i)} \leq \ldots \leq y_{(N)} > 0$, Section 2 provides the Gini relative inequality (concentration) $\rho\left(\frac{i}{N}\right)$ corresponding to the cumulative relative frequency $p_{(i)} = \frac{i}{N}$ and the Gini synthetic index: $\rho\left(\frac{i}{N}\right) = \frac{p_{(i)} - q_{(i)}}{p_{(i)}}$, and $q_{(i)}$ is the ordinate of the Lorenz curve. Then, the expressions of the Gini synthetic index related to the Lorenz curve and to the Gini mean difference (with and without replacement) are illustrated. Moreover, this section shows that the synthetic Gini index $G(Y)$ is also equal to the (simple) arithmetic mean $\frac{1}{N} \sum_{i=1}^{N} G_{(i)}(Y)$: where $G_{(i)}(Y) = 2(p_{(i)} - q_{(i)}) = \frac{M(Y) - \bar{M}_{(i)}(Y)}{M(Y)} \cdot 2p_{(i)}$, $\bar{M}_{(i)}(Y)$ is the mean of the i poorest population units and $M(Y)$ is the mean of Y. In addition, this section utilize $N = 10$ ordered values $y_{(i)}$ for the calculation of $G_{(i)}(Y)$ and of $G(Y)$. This example shows that $G_{(i)}(Y)$ may not be constant for units taking the same value of Y. This behaviour is not reasonable in the decomposition by subpopulations because units with the same value of Y may belong to different subpopulations. Section 3 overcomes this situation by substituing the values of $G_{(i)}(Y)$, corresponding to the $n_{h.}$ units with the same value y_h, with their arithmetic mean: $G_h(Y) = M\left[G_{(i)}(Y) | Y = y_h\right]$. Then, this section introduces the appropiate definitions and notation used in the decompositions by subpopulations of $G_h(Y)$ and $G(Y) = \sum_{h=1}^{r} G_h(Y) \frac{n_{h.}}{N}$. In particular this section introduces the basic $r \times k$ bivariate distribution of the N units according to: the r distinct values $\{0 \leq y_1 < ... < y_h < ... < y_r\}$ assumed by the variate Y and the k subpopulations. Moreover, this section provides: for each subpopulation l the lower means $\bar{M}_{hl}$ and the mean M_l, and for the whole population the lower means $\bar{M}_{h.}$ and the mean M. Section 4 achieves (i) the decomposition of $G_h(Y)$ in a sum of $k \times k$ contributions C_{hlg} and (ii) a $k \times k$ decomposition by subpopulations of the Gini synthetic index. In Section 4.1, these $k \times k$ additive decompositions are illustrated by a 6×3 bivariate distribution with $N = 10$ units, $k = 3$ subpopulations and $r = 6$ distinct values of Y. Section 5 provides an application to the net disposable income of the Italian households partitioned into three residence areas: North, Center, and South with Islands. The data are supplied by the 2012 Central Bank of Italy sample survey on household income and wealth (Bank of Italy, 2014). In particular this section points out that for the application at hand the values of the contributions C_{hlg} are mainly given by the product of the corresponding Bonferroni contributions V_{hlg} and $2p_{h.}$. More details on the contributions V_{hlg} are reported in Zenga and Valli (2016). Moreover, this section shows that if $M_l > M_g$, then there is an integer $h' \leq r$, such that $\left(M_g - \bar{M}_{hl}\right) < 0$, for $\forall(h \geq h')$. In this case the contributions V_{hlg} and C_{hlg} are negative. This important result explains, for example, why there are negative contributions to the Bonferroni point index (and to the Gini point index) when we evaluate, for $p = 0.95$, the contributions $V_{(0.95)13}$ and $C_{(0.95)13}$, where $l = 1$ is the North and $g = 3$ is the South. Section 6 is devoted to the conclusions and final remarks.

2. DEFINITIONS AND NOTATION

Let Y denote a non-negative variate, usually income, observed on the N units of a finite population. The N units can be partitioned according to some relevant characteristic, into k different subpopulations whose size is denoted by $n_{.g}$ $(g = 1, 2, \ldots, k)$. Let

$$0 \leq y_{(1)} \leq \ldots \leq y_{(i)} \leq \ldots \leq y_{(N)} > 0$$

be the N values of Y arranged in non-decreasing order. Let

$$Q_{(i)}(Y) = \sum_{t=1}^{i} y_{(t)}, (i = 1, ..., N)$$

be the income of the i poorest population units;

$$T = Q_{(N)}(Y) = \sum_{i=1}^{N} y_{(i)}\,;$$

$$M = M(Y) = T/N\,;$$

$$\bar{M}_{(i)}(Y) = \frac{Q_{(i)}(Y)}{i},\ (i = 1, ..., N).$$

According to Gini (1914) the relative inequality (concentration), corresponding to the relative frequency

$$p_{(i)} = \frac{i}{N}$$

is given by:

$$\rho\left(\frac{i}{N}\right) = \frac{p_{(i)} - q_{(i)}}{p_{(i)}}, \ (i = 1, ..., N),$$

where

$$q_{(i)} = \frac{Q_{(i)}(Y)}{T} = \frac{i \cdot \bar{M}_{(i)}(Y)}{N \cdot M(Y)}$$

is the ordinate of the Lorenz (1905) curve. Note that for $i = N$, $p_{(N)} = q_{(N)} = 1$, and $\frac{p_{(N)} - q_{(N)}}{p_{(N)}} = 0$.

The synthetic inequality index $\tilde{G}(Y)$ proposed by Gini (1914) is the weighted mean of $\frac{p_{(i)} - q_{(i)}}{p_{(i)}}$ with weights $p_{(i)}$:

$$\tilde{G}(Y) = \frac{1}{\sum_{i=1}^{N-1} p_{(i)}} \cdot \sum_{i=1}^{N-1} \frac{p_{(i)} - q_{(i)}}{p_{(i)}} \cdot p_{(i)} = \frac{2}{N-1} \cdot \sum_{i=1}^{N-1} \left(p_{(i)} - q_{(i)}\right). \qquad (1)$$

The value of $\tilde{G}(Y)$ is also provided by the ratio

$$\tilde{G}(Y) = \frac{\sum_{i=1}^{N} \frac{1}{N} \left\{ \frac{\left(p_{(i-1)} - q_{(i-1)}\right) + \left(p_{(i)} - q_{(i)}\right)}{2} \right\}}{\frac{1}{2} \cdot \frac{N-1}{N}}. \qquad (2)$$

The numerator of (2) is the concentration area CA and $\frac{1}{2} \cdot \frac{N-1}{N}$ is the value of CA in the case of maximum inequality. In (2), $p_{(0)} = q_{(0)} = 0$.

Another popular expression of $\tilde{G}(Y)$ is given by the ratio

$$\tilde{G}(Y) = \frac{\Delta(Y)}{2 \cdot M(Y)}, \qquad (3)$$

where

$$\Delta(Y) = \frac{1}{N \cdot (N-1)} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} |y_{(i)} - y_{(j)}|$$

is the Gini mean difference (without replacement) of Y. Moreover, the numerator of $\Delta(Y)$ is also given by

$$\sum_{i=1}^{N} \sum_{j=1}^{N} |y_{(i)} - y_{(j)}| = 2 \cdot \sum_{i=1}^{N} y_{(i)} (2 \cdot i - N - 1). \qquad (4)$$

Computing $\tilde{G}(Y)$ through the mean difference, the connections with the Lorenz curve are lost, even if we use the expression (4).

Multiplying both sides of (3) by $(N-1)/N$ we have:

$$G(Y) = \frac{\Delta_R(Y)}{2 \cdot M(Y)},$$

where

$$G(Y) = \frac{N-1}{N} \cdot \tilde{G}(Y), \text{ and } \Delta_R(Y) = \frac{1}{N^2} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} |y_{(i)} - y_{(j)}|$$

is the mean difference with replacement of Y. Finally, multiplying both sides of (1) by $(N-1)/N$ it turns out that $G(Y)$ is also equal to the following (simple) arithmetic mean

$$G(Y) = \frac{N-1}{N} \cdot \frac{2}{N-1} \cdot \sum_{i=1}^{N-1} \left(p_{(i)} - q_{(i)}\right) = \frac{1}{N} \cdot \sum_{i=1}^{N} G_{(i)}(Y). \tag{5}$$

In (5)

$$G_{(i)}(Y) = 2 \cdot \left(p_{(i)} - q_{(i)}\right), \tag{6}$$

and $G_{(N)} = 0$.

The Bonferroni (1930) point and synthetic inequality measures are:

$$V_{(i)}(Y) = \frac{M(Y) - \bar{M}_{(i)}(Y)}{M(Y)} \text{ and } \tilde{V}(Y) = \frac{1}{N-1} \sum_{i=1}^{N-1} V_{(i)}(Y), \text{ respectively.}$$

De Vergottini (1940) showed that

$$\rho\left(\frac{i}{N}\right) = \frac{p_{(i)} - q_{(i)}}{p_{(i)}} = \frac{M(Y) - \bar{M}_{(i)}(Y)}{M(Y)} = V_{(i)}(Y). \tag{7}$$

Thus from (7), (6) and (5) we have:

$$G_{(i)}(Y) = 2 \cdot \left(p_{(i)} - q_{(i)}\right) = \frac{M(Y) - \bar{M}_{(i)}(Y)}{M(Y)} \cdot 2 \cdot p_{(i)}, \tag{8}$$

and

$$G(Y) = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{M(Y) - \bar{M}_{(i)}(Y)}{M(Y)} \cdot 2 \cdot p_{(i)}. \tag{9}$$

The following $N = 10$ ordered values $\left\{y_{(i)} : 2, 2, 8, 24, 29, 37, 37, 37, 62, 62\right\}$ are utilized in Table 1 for the calculation of $\rho\left(\frac{i}{N}\right)$ and of $G_{(i)}(Y)$. Moreover, from Table 1 we obtain: $M(Y) = \bar{M}_{(10)}(Y) = 30$ and $G(Y) = \frac{1}{10} \cdot 3.84 = 0.384$.

TABLE 1. - *Calculation of* $\rho\left(\frac{i}{N}\right)$ *and of* $G_{(i)}(Y)$

i	$y_{(i)}$	$Q_{(i)}(Y)$	$\bar{M}_{(i)}(Y)$	$p_{(i)}$	$q_{(i)}$	$p_{(i)}-q_{(i)}$	$\rho\left(\frac{i}{N}\right)$	$G_{(i)}(Y)$
1	2	2	2	0.1	0.006	0.093	0.933	0.186
2	2	4	2	0.2	0.013	0.186	0.933	0.373
3	8	12	4	0.3	0.04	0.260	0.866	0.520
4	24	36	9	0.4	0.12	0.28	0.7	0.560
5	29	65	13	0.5	0.216	0.283	0.566	0.566
6	37	102	17	0.6	0.34	0.26	0.433	0.520
7	37	139	19.85	0.7	0.463	0.236	0.337	0.473
8	37	176	22	0.8	0.586	0.213	0.266	0.426
9	62	238	26.44	0.9	0.793	0.106	0.119	0.213
10	62	300	30	1	1	0	0	0
							tot.	3.84

The value of $G(Y)=\frac{1}{N}\cdot\sum_{i=1}^{N}G_{(i)}(Y)$ can be interpreted as the sum of the areas of N rectangles, each with basis $\frac{1}{N}$ and height $G_{(i)}(Y)$. To draw the inequality diagram $G_{(i)}(Y)$, it is necessary, first of all, to obtain the N points of coordinates $\left(p_{(i)},G_{(i)}(Y)\right)$. Then we obtain N rectangles by the following procedure: the first rectangle has abscissas in the interval $[0,1/N]$ and ordinates in the interval $[0,G_{(1)}(Y)]$. The $i-th$ rectangle $(i=2,...,N)$ has abscissas in the interval $[\frac{i-1}{N},\frac{i}{N}]$ and ordinates in the interval $[0,G_{(i)}(Y)]$.

Figure 1 reports the graphs of $G_{(i)}(Y)$ and of $\rho\left(\frac{i}{N}\right)=V_{(i)}(Y)$.

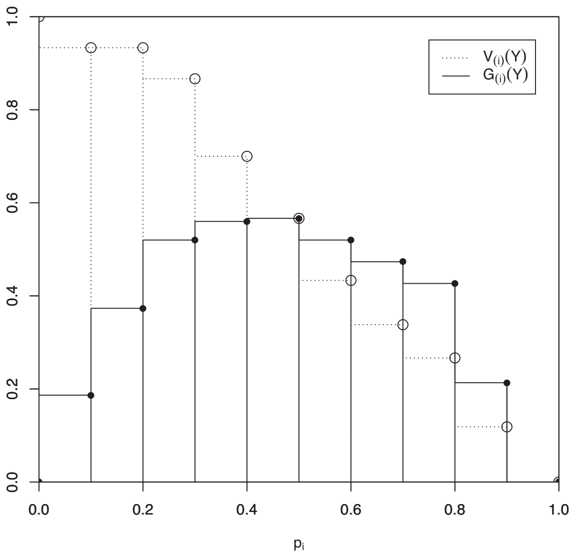

FIGURE 1. - *Graphs of* $G_{(i)}(Y)$ *and of* $\rho\left(\frac{i}{N}\right)=V_{(i)}(Y)$

3. DEFINITIONS AND NOTATION IN THE CASE OF FREQUENCY DISTRIBUTION FRAMEWORK

The last column of Table 1 shows that $G_{(i)}(Y)$ may not be constant for units taking the same value of Y. This behaviour of $G_{(i)}(Y)$ is not reasonable in the decomposition by subpopulations because units with the same value of Y may belong to different subpopulations. We will overcome this situation by substituting the values of $G_{(i)}(Y)$ corresponding to the units with the same value y_h of Y with their arithmetic mean: $G_h(Y) = M\left[G_{(i)}(Y)|Y = y_h\right]$.

We introduce now the appropriate definitions and notation used in the decomposition by subpopulations proposed in this paper for $G(Y)$.

Let,

$$\{0 \leq y_1 < ... < y_h < ... < y_r\}$$

denote the set of the r distinct values assumed by the variate Y over the k subpopulations. It is possible to report the whole $r \times k$ bivariate distribution of the N units as in Table 2, where: n_{hg} denotes the frequency of y_h in the subpopulation g, $n_{h.} = \sum_{g=1}^{k} n_{hg}$ is the frequency of y_h in the whole population, and $n_{.g} = \sum_{h=1}^{r} n_{hg}$ is the frequency of the subpopulation g.

TABLE 2. - *Bivariate $r \times k$ distribution of the whole population partitioned into k subpopulations*

	Subpopulation					
	1	$\dots$	g	$\dots$	k	tot
y_1	n_{11}	$\dots$	n_{1g}	$\dots$	n_{1k}	$n_{1.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_h	n_{h1}	$\dots$	n_{hg}	$\dots$	n_{hk}	$n_{h.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_r	n_{r1}	$\dots$	n_{rg}	$\dots$	n_{rk}	$n_{r.}$
tot	$n_{.1}$	$\dots$	$n_{.g}$	$\dots$	$n_{.k}$	N

Let us define, for the overall distribution $\{(y_h, n_{h.}) : h = 1, \dots, r\}$:

$$P_{h.} = P_{h.}(Y) = \sum_{t=1}^{h} n_{t.} \quad \text{, h=1,..., r} \tag{10}$$

$$Q_{h.}(Y) = \sum_{t=1}^{h} y_t \cdot n_{t.}, h = 1, \dots, , r \tag{11}$$

We remark that

$$Q_{h.}(Y) = Q_{(P_{h.})}(Y) = \sum_{i=1}^{P_{h.}} y_{(i)}, \tag{12}$$

and that

$$T = Q_{r.}(Y) = \sum_{h=1}^{r} y_h \cdot n_{h.} = Q_{(N)}(Y) > 0. \tag{13}$$

For the distribution $\{(y_h, n_{hg}) : h = 1, \ldots, r\}$ of the subpopulation g the analogous of (10), (11), and (13) are:

$$P_{hg} = P_{hg}(Y) = \sum_{t=1}^{h} n_{tg} \qquad h = 1, \ldots, r \tag{14}$$

$$Q_{hg}(Y) = \sum_{t=1}^{h} y_t \cdot n_{tg} \qquad h = 1, \ldots, r \tag{15}$$

$$T_g = Q_{rg}(Y) = \sum_{h=1}^{r} y_h \cdot n_{hg}. \tag{16}$$

Finally,

$$M_g = M_g(Y) = T_g / n_{.g}$$

denotes the mean of subpopulation g.

Let $\{(y_1, n_{1.}), \ldots, (y_h, n_{h.})\}$ be the group including the first $P_{h.}$ units (of the whole population). Let

$$\bar{M}_{h.}(Y) = \frac{Q_{h.}(Y)}{P_{h.}}, \qquad h = 1, \ldots, r, \tag{17}$$

be the arithmetic mean (lower mean) in the lower group: $\{Y \leq y_h\}$. Using (12) in (17) we obtain:

$$\bar{M}_{h.}(Y) = \bar{M}_{(P_{h.})}(Y) = \frac{1}{P_{h.}} \sum_{i=1}^{P_{h.}} y_{(i)} \tag{18}$$

For the distribution $\{(y_h, n_{hg}) : h = 1, \ldots, r\}$ of the subpopulation g let

$$y_{o(g)}, \text{ where } o(g) = \min \text{h} : \; n_{hg} > 0 \tag{19}$$

and define the lower mean $\bar{M}_{hg}(Y)$ as follows:

$$\bar{M}_{hg}(Y) = \begin{cases} y_{o(g)} & \text{for } h < o(g) \\ Q_{hg}(Y)/P_{hg} & \text{for } h \geq o(g). \end{cases} \tag{20}$$

Using the definitions and notation of this section we will now obtain a new expression for the numerator $\sum_{i=1}^{N} G_{(i)}(Y)$ of (5).

First of all, note that

$$\sum_{i=1}^{N} \left(p_{(i)} - q_{(i)}\right) = \sum_{h=1}^{r} \left\{ \sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} \left(p_{(i)} - q_{(i)}\right) \right\}. \tag{21}$$

Then:

$$\sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} p_{(i)} = \sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} \frac{i}{N}$$

is given by

$$\frac{1}{N}\left[n_{h.}\left(\frac{2P_{h.}-(n_{h.}-1)}{2}\right)\right] = \frac{n_{h.}}{N}\left[P_{h.}-\frac{n_{h.}-1}{2}\right], \tag{22}$$

and

$$\sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} q_{(i)} = \frac{(Q_{h.}-y_h\cdot n_{h.}+y_h)}{T}+$$

$$+\frac{(Q_{h.}-y_h\cdot n_{h.}+y_h\cdot 2)+\ldots+(Q_{h.}-y_h\cdot n_{h.}+y_h\cdot n_{h.})}{T}=$$

$$\frac{n_{h.}}{T}\left[Q_{h.}-y_h\cdot n_{h.}+y_h\cdot\left(1+\frac{n_{h.}}{2}\right)\right] = \frac{n_{h.}}{T}\left[Q_{h.}-y_h\cdot\left(\frac{n_{h.}-1}{2}\right)\right]. \tag{23}$$

Thus,

$$\sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}}\left(p_{(i)}-q_{(i)}\right) = n_{h.}\left[\frac{P_{h.}}{N}-\frac{n_{h.}-1}{2\cdot N}-\frac{Q_{h.}}{T}+\frac{y_h}{T}\cdot\left(\frac{n_{h.}-1}{2}\right)\right]=$$

$$= n_{h.}\left[(p_{h.}-q_{h.})-\frac{n_{h.}-1}{2\cdot N}\left(\frac{M-y_h}{M}\right)\right], \tag{24}$$

where

$$p_{h.}=\frac{P_{h.}}{N}\text{ and } q_{h.}=\frac{Q_{h.}}{T}. \tag{25}$$

Using (24) in (21) gives,

$$\sum_{i=1}^{N}\left(p_{(i)}-q_{(i)}\right) = \sum_{h=1}^{r}\left[(p_{h.}-q_{h.})-\frac{n_{h.}-1}{2\cdot N}\left(\frac{M-y_h}{M}\right)\right]\cdot n_{h.}. \tag{26}$$

In conclusion, from (6), (5) and (26) we obtain:

$$G(Y)=\sum_{h=1}^{r} G_h(Y)\cdot\frac{n_{h.}}{N}, \tag{27}$$

where

$$G_h(Y) = 2\cdot(p_{h.}-q_{h.})-\frac{n_{h.}-1}{N}\left(\frac{M-y_h}{M}\right) \tag{28}$$

is the arithmetic mean of the $n_{h.}$ values $G_{(i)}(Y)$ corresponding to the $n_{h.}$values $Y=y_h$.

The $N=10$ values $y_{(i)}$, introduced in Section 3, are utilized in Tables 3 and 4 for the calculation of $p_{h.}$, $q_{h.}$ and of $G_h(Y)$ and $G(Y)$, respectively.

TABLE 3. - *Calculation of:* $P_{h.}, Q_{h.}, \bar{M}_{h.}, p_{h.}$ *and* $q_{h.}$

h	y_h	$n_{h.}$	$P_{h.}$	$y_h \cdot n_{h.}$	$Q_{h.}$	$\bar{M}_{h.}$	$p_{h.}$	$q_{h.}$
1	2	2	2	4	4	2	0.20	0.013
2	8	1	3	8	12	4	0.3	0.04
3	24	1	4	24	36	9	0.40	0.12
4	29	1	5	29	65	13	0.5	0.216
5	37	3	8	111	176	22	0.8	0.586
6	62	2	10	124	300	30	1	1
		10		300				

TABLE 4. - *Calculation of* $G_h(Y)$ *and of* $G(Y)$

$a = 2 \cdot (p_{h.} - q_{h.})$	$b = \frac{n_{h.}-1}{N}$	$c = \frac{M-y_h}{M}$	$d = b \times c$	$G_h(Y) = a - d$	$G_h(Y) \cdot \frac{n_{h.}}{N}$
0.373	0.1	0.933	0.0933	0.28	0.056
0.52	0.0	0.733	0	0.52	0.052
0.56	0.0	0.2	0	0.56	0.056
0.566	0.0	0.03	0	0.566	0.0566
0.4266	0.2	-0.233	-0.0466	0.4733	0.142
0.0	0.1	-1.06	-0.106	0.1066	0.0213
					$G(Y) = 0.384$

In the frequency distribution framework, the Bonferroni point measures $/V_h(Y)$ is given by $V_h(Y) = \dfrac{M - \bar{M}_{h.}}{M}$. For detail see Zenga and Valli (2016, Sec. 3). Figure 2 reports the graphs of $G_h(Y)$ and $V_h(Y)$.

4. DECOMPOSITION BY SUBPOPULATIONS OF THE POINT $G_h(Y)$ AND THE SYNTHETIC $G(Y)$ INEQUALITY INDEXES

The approach followed in this paper for the decomposition by subpopulations of $G_h(Y)$ and of $G(Y)$ is similar to the two-step approach proposed in Zenga (2016) for the decomposition of the point and synthetic Zenga (2007) inequality indexes. To this purpose we need to modify formula (28) as follows.

From (7) and (18) we have:

$$\frac{p_{h.} - q_{h.}}{p_{h.}} = \frac{M - \bar{M}_{(P_{h.})}}{M} = \frac{M - \bar{M}_{h.}}{M}.$$

Thus:

$$2 \cdot (p_{h.} - q_{h.}) = \frac{M - \bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.}. \tag{29}$$

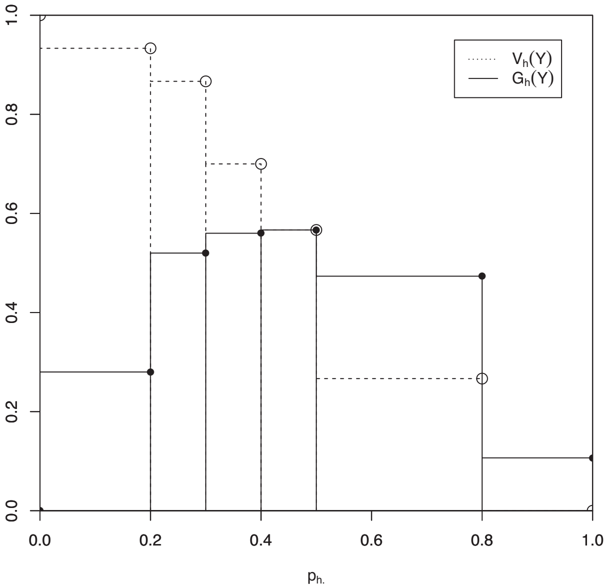

FIGURA 2. - *Graphs of $V_h(Y)$ and $G_h(Y)$*

Putting (29) in (28) and in (27) gives:

$$G_h(Y) = \frac{M - \bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.} - \frac{n_{h.} - 1}{N}\left(\frac{M - y_h}{M}\right), \tag{30}$$

$$G(Y) = \sum_{h=1}^{r}\left\{\frac{M - \bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.} - \frac{n_{h.} - 1}{N}\left(\frac{M - y_h}{M}\right)\right\} \cdot \frac{n_{h.}}{N}. \tag{31}$$

Now we will obtain ''first'' a $k \times k$ additive decomposition by subpopulations of $G_h(Y)$, and ''secondly'' by using this decomposition in (31) a $k \times k$ decomposition by subpopulations of the Gini synthetic index is obtained.

The mean M is related to the k means M_g by the relation

$$M = \sum_{g=1}^{k} M_g \cdot \frac{n_{.g}}{N}, \tag{32}$$

and $\bar{M}_{h.}$ is related to the k means $\bar{M}_{hl}$ by the relation

$$\bar{M}_{h.} = \sum_{l=1}^{k} \bar{M}_{hl} \cdot p(l|h), \tag{33}$$

where

$$p(l|h) = \frac{P_{hl}}{P_{h.}}, \; (h = 1, ..., r; l = 1, ..., k) \tag{34}$$

is the relative frequency of the subpopulation l in the lower group $\{Y \le y_h\}$. Note that $\sum_{l=1}^{k} p(l|h) = 1$.

Using (32) and (33) the following $k \times k$ additive decomposition of $[M - \bar{M}_{h.}]$ is obtained:

$$[M - \bar{M}_{h.}] = \sum_{g=1}^{k} M_g \cdot \frac{n_{.g}}{N} \sum_{l=1}^{k} p(l|h) - \sum_{l=1}^{k} \bar{M}_{hl} \cdot p(l|h) \sum_{g=1}^{k} \frac{n_{.g}}{N}$$
$$= \sum_{l=1}^{k} \sum_{g=1}^{k} \left[M_g - \bar{M}_{hl} \right] \frac{n_{.g}}{N} \cdot p(l|h). \tag{35}$$

The value y_h can be written as follows

$$y_h = \sum_{l=1}^{k} y_h \cdot f(l|h) \; , \tag{36}$$

where

$$f(l|h) = \frac{n_{hl}}{n_{h.}}, \;\; (h = 1, ..., r; l = 1, ..., k) \tag{37}$$

is the relative frequency of the subpopulation l in the group of the $n_{h.}$ units with $Y = y_{h.}$.

Thus, using (32) and (36) the following $k \times k$ additive decomposition of $[M - y_h]$ is obtained:

$$[M - y_h] = \sum_{g=1}^{k} M_g \cdot \frac{n_{.g}}{N} \sum_{l=1}^{k} f(l|h) - \sum_{l=1}^{k} y_h \cdot f(l|h) \sum_{g=1}^{k} \frac{n_{.g}}{N}$$
$$= \sum_{l=1}^{k} \sum_{g=1}^{k} \left[M_g - y_h \right] \frac{n_{.g}}{N} \cdot f(l|h). \tag{38}$$

Finally, substituting (35) and (38) in (30) the following $k \times k$ additive decomposition of $G_h(Y)$ is obtained:

$$G_h(Y) = \sum_{l=1}^{k} \sum_{g=1}^{k} C_{hlg}(Y), \tag{39}$$

where

$$C_{hlg}(Y) = \frac{M_g - \bar{M}_{hl}}{M} \cdot \frac{n_{.g}}{N} p(l|h) \cdot 2p_{h.} - \frac{M_g - y_h}{M} \cdot \frac{n_{.g}}{N} f(l|h) \cdot \frac{n_{h.} - 1}{N}, \tag{40}$$

is the contribution to $G_h(Y)$ that derives from the comparisons of $\bar{M}_{hl}$ and y_h w.r.t. $M_g(Y)$.

Now, putting (39) in (31) the following $k \times k$ additive decomposition of $G(Y)$ is obtained:

$$G(Y) = \sum_{h=1}^{r} \left\{ \sum_{l=1}^{k} \sum_{g=1}^{k} C_{hlg}(Y) \right\} \cdot \frac{n_{h.}}{N} = \sum_{l=1}^{k} \sum_{g=1}^{k} C_{.lg}(Y), \tag{41}$$

where

$$C_{.lg}(Y) = \sum_{h=1}^{r} C_{hlg}(Y) \cdot \frac{n_{h.}}{N}. \tag{42}$$

REMARK

Now we will obtain the expression (31) starting from formula (2).

Multiplying both sides of (2) by $(N-1)/N$ gives:

$$G(Y) = \frac{N-1}{N} \cdot \tilde{G}(Y) = \frac{N-1}{N} \cdot \frac{\sum_{i=1}^{N} \frac{1}{N} \left\{ \frac{\left(p_{(i-1)} - q_{(i-1)}\right) + \left(p_{(i)} - q_{(i)}\right)}{2} \right\}}{\frac{1}{2} \cdot \frac{N-1}{N}} = \frac{CA}{\frac{1}{2}}.$$

In the case of frequency distribution the value of the concentration area CA is given by:

$$CA = \sum_{h=1}^{r} \left\{ \frac{(p_{h-1.} - q_{h-1.}) + (p_{h.} - q_{h.})}{2} \right\} \cdot \frac{n_{h.}}{N}. \tag{43}$$

Putting the relations

$$p_{h-1.} = p_{h.} - \frac{n_{h.}}{N} \quad \text{and} \quad q_{h-1.} = q_{h.} - \frac{n_{h.} \cdot y_h}{M \cdot N}$$

in (43) gives:

$$CA = \sum_{h=1}^{r} \left\{ (p_{h.} - q_{h.}) - \frac{n_{h.}}{2 \cdot N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N}. \tag{44}$$

Using in (44) the popular property $\sum_{h=1}^{r} \left\{ \left(\frac{M-y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N} = 0$ gives:

$$CA = \sum_{h=1}^{r} \left\{ (p_{h.} - q_{h.}) - \frac{n_{h.}}{2 \cdot N} \left(\frac{M - y_h}{M} \right) + \frac{1}{2 \cdot N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N} =$$

$$= \sum_{h=1}^{r} \left\{ (p_{h.} - q_{h.}) - \frac{n_{h.} - 1}{2 \cdot N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N}. \tag{45}$$

Dividing both sides of (45) by $\frac{1}{2}$ gives

$$\frac{CA}{\frac{1}{2}} = \sum_{h=1}^{r} \left\{ 2 \cdot (p_{h.} - q_{h.}) - \frac{n_{h.} - 1}{N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N}.$$

Putting (29) in the latter expression the result is obtained:

$$G(Y) = \sum_{h=1}^{r} \left\{ \frac{M - \bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.} - \frac{n_{h.} - 1}{N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N}. \tag{31}$$

4.1 *Example*

The result of this paper are illustrated by the 6×3 bivariate distribution, reported in Table 5, with $N = 10$ units, $k = 3$ subpopulations and $r = 6$ distinct values of Y.

Table 6 reports all the values necessary for the decomposition (39) of $G_h(Y)$.

For the calculations of $C_{hlg}(Y)$ is useful the following relation:

$$C_{hlg}(Y) = V_{hlg}(Y) \cdot 2p_{h.} - A_{hlg}(Y) \cdot \frac{n_{h.} - 1}{N}, \tag{46}$$

where

$$V_{hlg}(Y) = \frac{M_g - \bar{M}_{hl}}{M} \cdot \frac{n_{.g}}{N} \cdot p(l|h), \tag{47}$$

and

$$A_{hlg}(Y) = \frac{M_g - y_h}{M} \cdot \frac{n_{.g}}{N} \cdot f(l|h). \tag{48}$$

TABLE 5. - *Joint frequencies n_{hg}, total frequencies $n_{.g}$ and $n_{h.}$, cumulative frequencies P_{hg} and $P_{h.}$, and cumulative incomes Q_{hg} and $Q_{h.}$*

		g			tot	g			tot	g			tot
		1	2	3		1	2	2		1	2	3	
h	y_h	n_{h1}	n_{h2}	n_{h3}	$n_{h.}$	P_{h1}	P_{h2}	P_{h3}	$P_{h.}$	Q_{h1}	Q_{h2}	Q_{h3}	$Q_{h.}$
1	2	1	0	1	2	1	0	1	2	2	-	2	4
2	8	0	1	0	1	1	1	1	3	2	8	2	12
3	24	1	0	0	1	2	1	1	4	26	8	2	36
4	29	0	0	1	1	2	1	2	5	26	8	31	65
5	37	2	1	0	3	4	2	2	8	100	45	31	176
6	62	1	0	1	2	5	2	3	10	162	45	93	300
total $n_{.g}$		5	2	3	10								

TABLE 6. - *Lower means $\bar{M}_{hg}$ and $\bar{M}_{h.}$, and relative frequencies $p(l|h) = \frac{P_{hl}}{P_{h.}}$ and $f(l|h) = \frac{n_{hl}}{n_{h.}}$*

		g			$\bar{M}_{h.}$	l			l		
		1	2	3		1	2	3	1	2	3
$o(g)$		1	2	1							
h	y_h	$\bar{M}_{h1}$	$\bar{M}_{h2}$	$\bar{M}_{h3}$		$\frac{P_{h1}}{P_{h.}}$	$\frac{P_{h2}}{P_{h.}}$	$\frac{P_{h3}}{P_{h.}}$	$\frac{n_{h1}}{n_{h.}}$	$\frac{n_{h2}}{n_{h.}}$	$\frac{n_{h3}}{n_{h.}}$
1	2	2	8	2	2	0.5	0	0.5	0.5	0	0.5
2	8	2	8	2	4	0.33	0.33	0.33	0	1	0
3	24	13	8	2	9	0.5	0.25	0.25	1	0	0
4	29	13	8	15.5	13	0.4	0.2	0.4	0	0	1
5	37	25	22.5	15.5	22	0.50	0.25	0.25	0.66	0.33	0
6	62	32.4	22.5	31	30	0.5	0.2	0.3	0.5	0	0.5

Tables 7, 8 and 9 show the calculations for $V_{1lg}(Y)$, $A_{1lg}(Y)$ and $C_{1lg}(Y)$, respectively. Moreover, Table 10 reports the contributions $C_{hlg}(Y)$, $(h = 1, ..., 6)$, and $C_{.lg}(Y)$.

TABLE 7. - *Calculus of* $V_{1lg}(Y) = \left[\frac{M_g - \bar{M}_{1l}}{M}\right] \cdot \frac{n_{.g}}{N} \cdot p(l|1)$

$h = 1$ $M(Y) = 30$			l			$\frac{n_{.g}}{N}$
			1	2	3	
			$\bar{M}_{11}(Y) = 2$	$\bar{M}_{12}(Y) = 8$	$\bar{M}_{13}(Y) = 2$	
g	1	$M_1(Y)$ =32.4	$[1.013] \cdot 0.25 = 0.2533$	$[0.] \cdot 0. = 0.$	$[1.013] \cdot 0.25 = 0.2533$	0.5
	2	$M_2(Y)$ =22.5	$[0.683] \cdot 0.1 = 0.0683$	$[0.] \cdot 0.0 = 0.$	$[0.683] \cdot 0.1 = 0.0683$	0.2
	3	$M_3(Y)$ = 31	$[0.966] \cdot 0.15 = 0.145$	$[1.0884] \cdot 0.0 = 0.0$	$[0.966] \cdot 0.15 = 0.145$	0.3
$p(l\|1)$			0.5	0.0	0.5	1.000

TABLE 8. - *Calculus of* $A_{1lg}(Y) = \frac{M_g - y_1}{M} \cdot \frac{n_{.g}}{N} \cdot f(l|1)$

$h = 1$ $M(Y) = 30$			l			$\frac{n_{.g}}{N}$
			1	2	3	
			$y_1 = 2$	$y_1 = 2$	$y_1 = 2$	
g	1	$M_1(Y)$ =32.4	$[1.013] \cdot 0.25 = 0.253$	$[0.] \cdot 0. = 0.$	$[1.013] \cdot 0.25 = 0.253$	0.5
	2	$M_2(Y)$ =22.5	$[0.683] \cdot 0.1 = 0.0683$	$[0.] \cdot 0.0 = 0.$	$[0.683] \cdot 0.1 = 0.0683$	0.2
	3	$M_3(Y)$ = 31	$[0.966] \cdot 0.15 = 0.145$	$[1.0884] \cdot 0.0 = 0.0$	$[0.966] \cdot 0.15 = 0.145$	0.3
$f(l\|1)$			0.5	0.0	0.5	1.000

TABLE 9. - *calculus of* $C_{1lg}(Y) = V_{1lg}(Y) \cdot 2p_{1.} - A_{1lg}(Y) \cdot \frac{n_{1.}-1}{N}$

$2 \cdot p_{1.} = 0.4$ $\frac{n_{1.}-1}{N} = 0.1$		l		
		1	2	3
g	1	$0.253 \cdot 0.4 - 0.253 \cdot 0.1 = 0.076$	$0 - 0 = 0$	$0.253 \cdot 0.4 - 0.253 \cdot 0.1 = 0.076$
	2	$0.0683 \cdot 0.4 - 0.0683 \cdot 0.1 = 0.0205$	$0 - 0 = 0$	$0.0683 \cdot 0.4 - 0.0683 \cdot 0.1 = 0.0205$
	3	$0.145 \cdot 0.4 - 0.145 \cdot 0.1 = 0.0435$	$0 - 0 = 0$	$0.145 \cdot 0.4 - 0.145 \cdot 0.1 = 0.0435$

TABLE 10. - *Values of the contributions* $C_{hlg}(Y), (h = 1, ..., 6)$, *and of* $C_{.lg}(Y)$

C_{1lg}		l = 1	2	3
g	1	0.076	/	0.076
	2	0.0205	/	0.0205
	3	0.0435	/	0.0435

$G_1 = 0.28$

C_{2lg}		l = 1	2	3
g	1	0.1013	0.0813	0.1013
	2	0.0273	0.0193	0.0273
	3	0.058	0.046	0.058

$G_2 = 0.52$

C_{3lg}		l = 1	2	3
g	1	0.1293	0.0813	0.1013
	2	0.0253	0.0193	0.0273
	3	0.072	0.046	0.058

$G_3 = 0.56$

C_{4lg}		l = 1	2	3
g	1	0.1293	0.0813	0.1127
	2	0.0253	0.0193	0.0187
	3	0.072	0.046	0.062

$G_4 = 0.5666$

C_{5lg}		l = 1	2	3
g	1	0.1089	0.0711	0.1127
	2	-0.0004	0.0064	0.0187
	3	0.056	0.038	0.062

$G_5 = 0.4733$

C_{6lg}		l = 1	2	3
g	1	0.0247	0.066	0.0387
	2	-0.0528	/	-0.0208
	3	0.0015	0.034	0.0155

$G_6 = 0.1066$

$C_{.lg}$		l = 1	2	3
g	1	0.0888	0.0589	0.0883
	2	0.0012	0.0077	0.0129
	3	0.046	0.032	0.0482

$0.384 = G(Y)$

4.2 *Contributions of each subpopulation to the point and the synthetic Gini indexes*

Let

$$C_{hl.}(Y) = \sum_{g=1}^{k} C_{hlg}(Y). \tag{49}$$

Putting (40) in (49) gives:

$$C_{hl.}(Y) = \sum_{g=1}^{k} C_{hlg}(Y) =$$

$$= \sum_{g=1}^{k} \left[\frac{M_g - \bar{M}_{hl}}{M} \cdot 2p_{h.} \cdot p(l|h) - \frac{M_g - y_h}{M} \cdot \frac{n_{h.} - 1}{N} \cdot f(l|h) \right] \cdot \frac{n_{.g}}{N}.$$

Using in this latter expression the relation (32) gives:

$$C_{hl.}(Y) = \frac{M - \bar{M}_{hl}}{M} \cdot 2p_{h.} \cdot p(l|h) - \frac{M - y_h}{M} \cdot \frac{n_{h.} - 1}{N} \cdot f(l|h). \tag{50}$$

We remark that in (50)

$$\frac{M - \bar{M}_{hl}}{M} \cdot 2p_{h.}$$

is the weighted (weight $2p_{h.}$) relative variation of $\bar{M}_{hl}$ w.r.t. M, and

$$\frac{M - y_h}{M} \cdot \frac{n_{h.} - 1}{N}$$

is the weighted (weight $\frac{n_{h.}-1}{N}$) relative variation of y_h w.r.t. M.

Finally, from (39), (49) and (50) the following k additive decomposition of $G_h(Y)$ is obtained:

$$G_h(Y) = \sum_{l=1}^{k} \sum_{g=1}^{k} C_{hlg}(Y) = \sum_{l=1}^{k} C_{hl.} =$$

$$= \sum_{l=1}^{k} \left[\frac{M - \bar{M}_{hl}}{M} \cdot 2p_{h.} \cdot p(l|h) - \frac{M - y_h}{M} \cdot \frac{n_{h.} - 1}{N} \cdot f(l|h) \right]. \tag{51}$$

Note that, if h is kept fixed, the value of $C_{hl.}(Y)$ depends on the values of $\bar{M}_{hl}$, $p(l|h)$ and $f(l|h)$ related to the subpopulation l. Thus, $C_{hl.}(Y)$ can be interpreted as the contribution of the subpopulation l to the point index $G_h(Y)$.

Now, putting (51) in (27) the following k additive decomposition of $G(Y)$ is obtained:

$$G(Y) = \sum_{h=1}^{r} G_h(Y) \cdot \frac{n_{h.}}{N} = \sum_{h=1}^{r} \left\{ \sum_{l=1}^{k} C_{hl.}(Y) \right\} \cdot \frac{n_{h.}}{N} =$$

$$\sum_{l=1}^{k} \sum_{h=1}^{r} C_{hl.}(Y) \cdot \frac{n_{h.}}{N} = \sum_{l=1}^{k} C_{.l.}(Y), \tag{52}$$

where

$$C_{.l.}(Y) = \sum_{h=1}^{r} C_{hl.}(Y) \cdot \frac{n_{h.}}{N} \tag{53}$$

is the contribution of the subpopulation l to $G(Y)$. Note that $C_{.l.}(Y)$ can also be obtained from (41) by the sum:

$$C_{.l.}(Y) = \sum_{g=1}^{k} C_{.lg}(Y). \tag{54}$$

4.3 *Within and Between components of $C_{hl.}(Y)$, $G_h(Y)$ and $G(Y)$*

The contribution $C_{hl.}(Y)$, of the subpopulation l to the point index $G_h(Y)$, can be split into a within and a between component.

From (49) we have:

$$C_{hl.}(Y) = \sum_{g=1}^{k} C_{hlg}(Y) = C_{hlW}(Y) + C_{hlB}(Y); \tag{55}$$

where:

$$C_{hlW}(Y) = C_{hll}(Y) =$$

$$= \left[\frac{M_l - \bar{M}_{hl}}{M} \cdot p(l|h) \cdot 2p_{h.} - \frac{M_l - y_h}{M} \cdot f(l|h) \cdot \frac{n_{h.} - 1}{N} \right] \cdot \frac{n_{.l}}{N}, \tag{56}$$

and

$$C_{hlB}(Y) = \sum_{(g:g \neq l)}^{k} C_{hlg}(Y) =$$

$$= \sum_{(g:g \neq l)}^{k} \left[\frac{M_g - \bar{M}_{hl}}{M} \cdot 2p_{h.} \cdot p(l|h) - \frac{M_g - y_h}{M} \cdot \frac{n_{h.} - 1}{N} \cdot f(l|h) \right] \cdot \frac{n_{.g}}{N}. \tag{57}$$

In (56) the value of $C_{hll}(Y)$ derives from comparisons of ‘‘incomes’’ of the same subpopulation l; thus $C_{hlW}(Y) = C_{hll}(Y)$ can be interpreted as the within part of the contribution $C_{hl.}(Y)$. Viceversa, in (57) the values of $C_{hlg}(Y)$ derive from comparisons of ‘‘incomes’’ of different subpopulations; thus $C_{hlB}(Y) = \sum_{(g:g \neq l)}^{k} C_{hlg}(Y)$ can be interpreted as the between part of $C_{hl.}(Y)$.

From (51) and (55) we obtain:

$$G_h(Y) = \sum_{l=1}^{k} C_{hl.}(Y) = \sum_{l=1}^{k} [C_{hlW}(Y) + C_{hlB}(Y)] =$$

$$= C_{h.W}(Y) + C_{h.B}(Y); \tag{58}$$

where

$$C_{h.W}(Y) = \sum_{l=1}^{k} C_{hlW}(Y) \tag{59}$$

is the sum of the within part of the contributions $C_{hl.}(Y)$ and can be interpreted as the within part of $G_h(Y)$, analogously

$$C_{h.B}(Y) = \sum_{l=1}^{k} C_{hlB}(Y) \tag{60}$$

can be interpreted as the between part of the Gini point index $G_h(Y)$.

Putting the decomposition (58) in (27) gives the following decomposition of the synthetic Gini index:

$$G(Y) = C_{..W}(Y) + C_{..B}(Y); \tag{61}$$

where

$$C_{..W}(Y) = \sum_{h=1}^{r} C_{h.W}(Y) \cdot \frac{n_{h.}}{N} \tag{62}$$

and

$$C_{..B}(Y) = \sum_{h=1}^{r} C_{h.B}(Y) \cdot \frac{n_{h.}}{N} \tag{63}$$

are the within and between components of $G(Y)$, respectively.

Putting in (62) the relations (59) and (56), (62) can be re-formulated as:

$$C_{..W}(Y) = \sum_{h=1}^{r} \sum_{l=1}^{k} C_{hlW}(Y) \cdot \frac{n_{h.}}{N} =$$

$$= \sum_{l=1}^{k} \sum_{h=1}^{r} \left[\frac{M_l - \bar{M}_{hl}}{M} \cdot p(l|h) \cdot 2p_{h.} - \frac{M_l - y_h}{M} \cdot f(l|h) \cdot \frac{n_{h.} - 1}{N} \right] \frac{n_{.l}}{N} \frac{n_{h.}}{N}. \tag{64}$$

Finally, using in (63) the relations (60) and (57), relation (63) can be re-formulated as:

$$C_{..B}(Y) = \sum_{h=1}^{r} \sum_{l=1}^{k} C_{hlB}(Y) \cdot \frac{n_{h.}}{N} = \sum_{h=1}^{r} \left[\sum_{l=1}^{k} \sum_{(g:g\neq l)}^{k} C_{hlg}(Y) \right] \frac{n_{h.}}{N} =$$

$$= \sum_{l=1}^{k} \sum_{g\neq l}^{k} \sum_{h=1}^{r} \left[\frac{M_g - \bar{M}_{hl}}{M} 2p_{h.}p(l|h) - \frac{M_g - y_h}{M} \frac{n_{h.} - 1}{N} f(l|h) \right] \frac{n_{.g}}{N} \frac{n_{h.}}{N}. \tag{65}$$

5. APPLICATION

The data used in this application are supplied by the 2012 Central Bank of Italy sample survey on household income and wealth (Bank of Italy, 2014). This survey covers $N = 8151$ households.

In this paper we deal with the household net disposable income Y, that is the sum of: the payroll income X_1, the pensions and net transfers X_2, the net self employment income X_3, and the property incomes X_4. The $N = 8151$ households have been partitioned according to their residence area: North (1), Center (2) and South with Islands (3). In all computations that follow we consider the weights $w_i > 0$ $(i = 1, 2, \ldots, 8151;\ W = \sum w_i = 8151 = N)$ supplied by the Central Bank of Italy for each household; these weights are defined as the inverse of household's probability of inclusion in the sample (for further details see Banca d'Italia, 2014). Now we remark that, in the following sections we will not use the notation related to the weights w_i, but for simplicity's sake we will continue the use of the notation of the previous sections. Thus, to denote the sum of the weights of the n_{hl} households of the subpopulation l with total income $Y = y_h$ we will use n_{hl} instead of w_{hl}. Note that the frequency distribution of the total income Y has $r = 7287$ different values, 6995 have frequency $n_{h.} = 1$.

5.1 *Aggregate characteristic in three Italian macro-regions*

Table 11 reports for the total income Y of each geographic area: the arithmetic mean, the median, the synthetic Gini index $G_{.l}(Y)$, the sum of the weights $n_{.l} (= w_{.l})$ and the relative weights $n_{.l}/N$. The synthetic Gini index $G_{.l}(Y)$ of the subpopulation l is given by

$$G_{.l}(Y) = \sum_{h=1}^{r} G_{hl}(Y) \cdot \frac{n_{hl}}{n_{.l}},$$

where

$$G_{hl}(Y) = \frac{M_l - \bar{M}_{hl}}{M_l} \cdot 2 \cdot p_{hl} - \frac{M_l - y_h}{M_l} \cdot \frac{n_{hl} - 1}{n_{.l}},$$

is the Gini point index of the subpopulation l and $p_{hl} = \frac{P_{hl}}{n_{.l}}$ is its relative cumulative frequency.

Table 11 informs that: the mean values of the South is very far from the means of the other two macro italian regions, the North has the greatest inequality, while the Center has the lowest one, and that the inequality of the whole population is a little bit greater than the one of the North. The synthetic inequality index $G(Y)$ =0.3561 means that in the whole population, on (weighted) average, the lower mean is equal to the (1-0.3561)·100 $\simeq$ 64.39% of the mean $M(Y)$.

TABLE 11. - *Some aggregate characteristics for geographic areas*

	North	Center	South	Italy
Mean	33543.17	34000.09	23517.86	30380.22
Median	27527.57	29824.24	19123.67	24590.10
$G_{.l}(Y)$	0.3513	0.3159	0.3497	$0.3561 = G(Y)$
$n_{.l}$	3971.949	1537.372	2641.679	$8151 = N = W$
$n_{.l}/N$	0.48729	0.18861	0.32409	1.00000

Figure 3 displays the graphs of the point inequality measures for: *a*) the whole population; *b*) the North, the Center and the South. For the subpopulation l the abscissas and the ordinates are given respectively by p_{hl} and $G_{(p_{hl})l}(Y) = G_{hl}(Y), \forall h = 1, \ldots, r$, while for the whole population the abscissas and the ordinates are given respectively by $p_{h.} = \dfrac{P_{h.}}{N}$ and $G_{(p_{h.})}(Y) = G_h(Y), \forall h = 1, \ldots, r$.

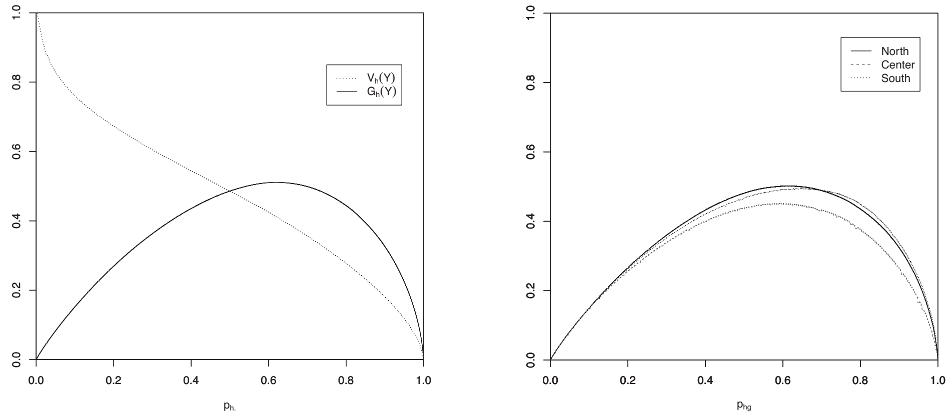

FIGURE 3. - *Graphs of the point measures for geographic areas*

5.2 *Decomposition by geographical areas of the point and synthetic inequality indexes of the whole country*

In this section we illustrate the decompositions of the point measure $G_{h(p)}(Y) = G_{(p)}(Y)$ for three values of p, and the decompositions of the synthetic index $G(Y) = 0.3561$. For p we have chosen the following values:

- $p = 0.10$; $G_{(0.10)}(Y) = 0.1522$ ‘‘compares’’ the income mean of the poorest 10% households with the income mean $M(Y)$ of the total households.
- $p = 0.50$; $G_{(0.50)}(Y) = 0.4859$ ‘‘compares’’ the income mean of the households with $Y \leq Median(Y)$ with $M(Y)$.
- $p = 0.95$; $G_{(0.95)}(Y) = 0.2247$ ‘‘compares’’ the income mean of the lower group that is the 95% of the whole population with $M(Y)$.

Table 12 reports for these three values of p the corresponding values of $h(p)$, $P_{h(p).}$, $P_{h(p).}/N$, and of $y_{h(p)}$; note that $h(p) = min\left(h : \frac{P_{h.}}{N} \geqq p\right)$.

TABLE 12. - *Cumulative frequency and quantile values for three values of p of the total income Y for the whole country*

p	$h(p)$	$P_{h(p).}$	$P_{h(p).}/N$	$y_{h(p)}$
0.10	460	815.20	0.10001	10600.00
0.50	3064	4075.65	0.50002	24590.10
0.95	6841	7743.48	0.95000	68819.23
1.00	7287	8151.00	1.00000	368689.7

Table 13 reports all the values necessary for the decompositions of $G_{(0.10)}(Y) = 0.1522$. These decompositions are shown in Table 14.

TABLE 13. - *Means $M_l(Y)$ and $\bar{M}_{hl}(Y)$, relative frequencies $n_{.l}/N, p(l|h)$ and $f(l|h)$ in the geographic areas: $p = 0.10$; $h = 460$; $y_h = 10600$; $n_{h.} = 0.5181$*

$p = 0.10; h = 460$	North $l = 1$	Center $l = 2$	South $l = 3$	Italy
$Y \leq 10600$	275.78	114.01	425.40	815.2
$Y > 10600$	3696.16	1423.36	2216.28	7335.8
Total=$n_{.l}$	3971.95	1537.37	2641.68	8151
		Relative frequencies		
$n_{.l}/N$	0.48729	0.18861	0.32409	1.0000
$p(l\|h)$	0.3383	0.1399	0.5218	1.0000
$f(l\|h)$	0.0	0.0	1.0	1.0000
		Means		
$\bar{M}_{hl}(Y)$	7091.45	7554.05	7310.03	7270.21
$M_l(Y)$	33543.17	34000.09	23517.86	30380.22

TABLE 14. - *Decompositions of $G_{(0.1)}(Y) = 0.1522$ into the contributions: $C_{(0.1)lg}(Y); C_{(0.1)l.}(Y); C_{(0.1)lW}(Y), C_{(0.1)lB}(Y); C_{(0.1).W}(Y), C_{(0.1).B}(Y)$*

$C_{(0.1)lg}(Y)$		l			
		1	2	3	
g	1	0.0287	0.0117	0.0439	
	2	0.0113	0.0046	0.0173	
	3	0.0119	0.0048	0.0181	
$C_{(0.1)l.}(Y)$		0.0519	0.0210	0.0793	$0.1522 = G_{(0.1)}(Y)$
$C_{(0.1)lW}(Y)$		0.0287	0.0046	0.0181	$0.0514 = C_{(0.1).W}(Y)$
$C_{(0.1)lB}(Y)$		0.0232	0.0165	0.0612	$0.1008 = C_{(0.1).B}(Y)$

For the interpretation of the values of the contributions $C_{hlg}(Y)$ it is useful to report the relations (46), (47) and (48) introduced in Section 4.1.

$$C_{hlg}(Y) = V_{hlg}(Y) \cdot 2p_{h.} - A_{hlg}(Y) \cdot \frac{n_{h.} - 1}{N}, \tag{46}$$

where

$$V_{hlg}(Y) = \frac{M_g - \bar{M}_{hl}}{M} \cdot \frac{n_{.g}}{N} \cdot p(l|h), \tag{47}$$

and

$$A_{hlg}(Y) = \frac{M_g - y_h}{M} \cdot \frac{n_{.g}}{N} \cdot f(l|h). \tag{48}$$

First of all we point out that in the present application there are $N = 8151$ households, consequently the value of $C_{hlg}(Y)$ is fundamentally given by the first term of the right side of (46). In addition, if for $Y = y_h$ there is only one Household , it follows that two of the corresponding relative frequencies $f(l|h)$ are equal to zero and the remaining is equal to one. Thus, two colums of the contributions $C_{hlg}(Y)$ are exactly equal to the first term $V_{hlg}(Y) \cdot 2p_{h.}$ of the right side of (46).We remark that, in the application at hand, this happens for the $\frac{6995}{7287} \cdot 100 = 95.993\%$ of the different values of the total income Y.

The greatest contributions $C_{(0.1)lg}(Y)$ is $C_{(0.1)31}(Y)$:

$$= \left[\frac{M_1 - \bar{M}_{h3}}{M}\right] \cdot \frac{n_{.1}}{N} \cdot p(3|h) \cdot 2 \cdot p_{h.} - \left[\frac{M_1 - y_h}{M}\right] \cdot \frac{n_{.1}}{N} \cdot f(3|h) \cdot \frac{n_{h.} - 1}{N} =$$

$$= [0.86349] \cdot 0.48729 \cdot 0.5218 \cdot 0.2 - [0.7552] \cdot 0.48729 \cdot 1 \cdot (-0.00005912) =$$

$$= 0.04391156 + 0.36800 \cdot 0.00005912 = 0.04391156 + 0.000021756 = 0.0439333.$$

This result depends fundamentally from the relative difference between the mean of the North and the lower mean of the South, from their relative weights $\frac{n_{.1}}{N}$ and $p(3|h)$, and from two times the value of $p_{h.} = 0.1$.

Let us consider now the contribution $C_{(0.1)13}(Y)$:

$$= \left[\frac{M_3 - \bar{M}_{h1}(Y)}{M}\right] \cdot \frac{n_{.3}}{N} \cdot p(1|h) \cdot 2 \cdot p_{h.} - \left[\frac{M_3 - y_h}{M}\right] \cdot \frac{n_{.3}}{N} \cdot f(1|h) \cdot \frac{n_{h.} - 1}{N}.$$

Table 13 informs that $f(1|h) = 0$. Consequently, $C_{(0.1)13}(Y)$ is equal to:

$$\left[\frac{M_3 - \bar{M}_{h1}(Y)}{M}\right] \cdot \frac{n_{.3}}{N} \cdot p(1|h) \cdot 2 \cdot p_{h.} =$$

$$= [0.54069] \cdot 0.32409 \cdot 0.3383 \cdot 0.2 = 0.011856.$$

In this latter case the relative difference between the mean of the South and the lower mean of the North and the relative weights are smaller than the corresponding values of the previous case. This clarifies the remarkable difference between the two contributions analized.

Let us consider now the decomposition of $G_{(0.10)}(Y) = 0.1522$ into the three contribution $C_{(0.1)l.}(Y)$ of each macro region:

$$C_{hlg}(Y) = V_{hlg}(Y) \cdot 2p_{h.} - A_{hlg}(Y) \cdot \frac{n_{h.} - 1}{N}, \tag{46}$$

where

$$V_{hlg}(Y) = \frac{M_g - \bar{M}_{hl}}{M} \cdot \frac{n_{.g}}{N} \cdot p(l|h), \tag{47}$$

and

$$A_{hlg}(Y) = \frac{M_g - y_h}{M} \cdot \frac{n_{.g}}{N} \cdot f(l|h). \tag{48}$$

$$C_{(0.1)1.}(Y) = \frac{M - \bar{M}_{h1}}{M} \cdot p(1|h) \cdot 2p_{h.} - \frac{M - y_h}{M} \cdot f(1|h) \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.76657 \cdot 0.3383 \cdot 0.2 - \frac{M - y_h}{M} \cdot 0 \cdot \frac{n_{h.} - 1}{N} = 0.05187;$$

$$C_{(0.1)2.}(Y) = \frac{M - \bar{M}_{h2}}{M} \cdot p(2|h) \cdot 2p_{h.} - \frac{M - y_h}{M} \cdot f(2|h) \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.75135 \cdot 0.1399 \cdot 0.2 - \frac{M - y_h}{M} \cdot 0 \cdot \frac{n_{h.} - 1}{N} = 0.02102;$$

$$C_{(0.1)3.}(Y) = \frac{M - \bar{M}_{h3}}{M} \cdot p(3|h) \cdot 2p_{h.} - \frac{M - y_h}{M} \cdot f(3|h) \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.75938 \cdot 0.5218 \cdot 0.2 + 0.65109 \cdot 1 \cdot 0.0000591 =$$

$$= 0.0792488 + 0.0000385 = 0.07929.$$

These values show that the relative variations of the lower means of the three macro-regions w.r.t the mean of the whole population are similar, while their relative weights $p(l|h)$ are very different. This explains why there are so remarkable differences among these three contributions. In particular we note that ‘‘the number’’ of the households of the South with $Y \le y_{h(0.10)} = 10600$ € is the 52.18% of the ‘‘number’’ of the corresponding households of the whole lower group. This explains why the greatest contribution to the point index $G_{(0.10)}(Y) = 0.1522$ comes from the South. Many other interesting informations can be obtained from the other decompositions reported in Table 14.

Table 15 reports all the values needed to calculate the four decompositions of $G_{(0.5)}(Y) = 0.4859$ which are reported in Table 16.

TABLE 15. - *Means $M_l(Y)$ and $\bar{M}_{hl}(Y)$, relative frequencies $n_{.l}/N, p(l|h)$ and $f(l|h)$ in the geographic areas: $p = 0.50; h = 3064$; $y_h = 24590.1; n_{h.} = 1.3504$*

$p = 0.50; h = 3064$	North $l = 1$	Center $l = 2$	South $l = 3$	Italy
$Y \leq 24590.1$	1710.1	576.9	1788.6	4075.6
$Y > 24590.1$	2261.8	960.44	853.07	4075.4
Total=$n_{.l}$	3971.95	1537.37	2641.68	8151
Relative frequencies				
$n_{.l}/N$	0.48729	0.18861	0.32409	1.0000
$p(l\|h)$	0.4196	0.1415	0.4388	1.0000
$f(l\|h)$	0.0	1.0	0.0	1.0000
Means				
$\bar{M}_{hl}(Y)$	16013.3	16449.9	14972.1	15618.2
$M_l(Y)$	33543.17	34000.09	23517.86	30380.22

TABLE 16. - *Decompositions of $G_{(0.5)}(Y) = 0.4859$ into the contributions: $C_{(0.5)lg}(Y)$; $C_{(0.5)l.}(Y)$; $C_{(0.5)lW}(Y), C_{(0.5)lB}(Y)$; $C_{(0.5).W}(Y), C_{(0.5).B}(Y)$*

$C_{(0.5)lg}(Y)$		l = 1	2	3	
g	1	0.1180	0.0388	0.1307	
	2	0.0469	0.0154	0.0518	
	3	0.0336	0.0107	0.0400	
$C_{(0.5)l.}(Y)$		0.1984	0.0649	0.2226	$0.4859 = G_{(0.5)}(Y)$
$C_{(0.5)lW}(Y)$		0.1180	0.0154	0.0400	$0.1734 = C_{(0.5).W}(Y)$
$C_{(0.5)lB}(Y)$		0.0804	0.0495	0.1826	$0.3125 = C_{(0.5).B}(Y)$

Table 16 shows that the contribution of the North $C_{(0.5)1.}(Y)$ is very near to the one of the South $C_{(0.5)3.}(Y)$. This happens because the relative variations of the lower means of these two macro-regions are similar as well as their relative weights:

$$C_{(0.5)1.}(Y) = \frac{M - \bar{M}_{h1}}{M} \cdot p(1|h) \cdot 2 \cdot 0.5 - \frac{M - y_h}{M} \cdot 0 \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.472904 \cdot 0.4196 \cdot 1 = 0.1984;$$

$$C_{(0.5)2.}(Y) = \frac{M - \bar{M}_{h2}}{M} \cdot p(2|h) \cdot 1 - \frac{M - y_h}{M} \cdot 1 \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.45853 \cdot 0.1415 - 0.190588 \cdot 0.0000429 = 0.06488 - 0.000008193 = 0.06487;$$

$$C_{(0.5)3.}(Y) = \frac{M - \bar{M}_{h3}(Y)}{M} \cdot p(3|h) \cdot 1 - \frac{M - y_h}{M} \cdot 0 \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.507176 \cdot 0.4389 = 0.2226.$$

Table 17 reports all the values necessary for the decompositions of $G_{(0.95)}(Y) = 0.2247$ which are reported in Table 18. This latter Table shows that the contributions $C_{(0.95)13}(Y)$, $C_{(0.95)23}(Y)$, $C_{(0.95)2.}(Y)$, $C_{(0.95)1B}(Y)$ and $C_{(0.95)2B}(Y)$ are negative. Relations (46) and (47) ''basically'' show that: if $\bar{M}_{hl} > M_g$ and $p(l|h) > 0$, then the corresponding contribution $C_{hlg}(Y)$ is negative. Recently, Zenga and Valli (2016, Section 5.2) have obtained similar results for the decomposition by subpopulations of the point and the synthetic Bonferroni inequality measures.

$$C_{(0.95)13}(Y) = \frac{M_3 - \bar{M}_{h1}(Y)}{M} \cdot \frac{n_{.3}}{N} \cdot p(1|h) \cdot 2 \cdot 0.95 +$$

$$- \frac{M_3 - y_h}{M} \cdot \frac{n_{.3}}{N} \cdot f(1|h) \cdot \frac{n_{h.} - 1}{N} =$$

$$= -0.1757 \cdot 0.3241 \cdot 0.4792 \cdot 1.9 + 1.49114 \cdot 1 \cdot \frac{1.509}{8151} =$$

$$= -0.05185 + 0.000089 = -0.05176.$$

TABLE 17. - *Means $M_l(Y)$ and $\bar{M}_{hl}(Y)$, relative frequencies $n_{.l}/N, p(l|h)$ and $f(l|h)$ in the geographic areas: $p = 0.95; h = 6841$; $y_h = 68819.2$; $n_{h.} = 2.5090$*

$p = 0.95; h = 6841$	North $l = 1$	Center $l = 2$	South $l = 3$	Italy
$Y \leq 68819.2$	3710.67	1443.33	2589.48	7743.48
$Y > 68819.2$	261.28	94.04	52.20	407.52
Total$= n_{.l}$	3971.95	1537.37	2641.68	8151
Relative frequencies				
$n_{.l}/N$	0.48729	0.18861	0.32409	1.0000
$p(l\|h)$	0.4792	0.1864	0.3344	1.0000
$f(l\|h)$	1.0000	0.0000	0.0000	1.0000
Means				
$\bar{M}_{hl}(Y)$	28856.94	30401.45	21819.47	26791.44
$M_l(Y)$	33543.17	34000.09	23517.86	30380.22

$$C_{(0.95)1.}(Y) = \frac{M - \bar{M}_{h1}}{M} \cdot p(1|h) \cdot 2 \cdot 0.95 - \frac{M - y_h}{M} \cdot 1 \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.05014 \cdot 0.4792 \cdot 1.9 + 1,26526 \cdot 1 \cdot 0.000185 = 0.04565 + 0.0002342 = 0.0459.$$

$$C_{(0.95)2.}(Y) = \frac{M - \bar{M}_{h2}}{M} \cdot p(2|h) \cdot 2 \cdot 0.95 - \frac{M - y_h}{M} \cdot 0 \cdot \frac{n_{h.} - 1}{N} =$$

$$= -0.0006988 \cdot 0.1864 \cdot 1.9 = -0.000247.$$

$$C_{(0.95)3.}(Y) = \frac{M - \bar{M}_{h3}}{M} \cdot p(3|h) \cdot 2 \cdot 0.95 - \frac{M - y_h}{M} \cdot 0 \cdot \frac{n_{h.} - 1}{N} =$$

$$= 0.28178696 \cdot 0.3344 \cdot 1.9 = 0.17903.$$

TABLE 18. - *Decompositions of* $G_{(0.95)}(Y) = 0.2247$ *into the contributions* $C_{(0.95)lg}(Y)$; $C_{(0.95)l.}(Y)$; $C_{(0.95)lW}(Y)$, $C_{(0.95)lB}(Y)$; $C_{(0.95).W}(Y)$, $C_{(0.95).B}(Y)$

$C_{(0.95)lg}(Y)$		l			
		1	2	3	
g	1	0.0685	0.0178	0.1195	
	2	0.0291	0.0079	0.0480	
	3	-0.0518	-0.0260	0.0115	
$C_{(0.95)l.}(Y)$		0.0459	-0.0002	0.1790	$0.2247 = G_{(0.95)}(Y)$
$C_{(0.95)lW}(Y)$		0.0685	0.0079	0.0115	$0.0879 = C_{(0.95).W}(Y)$
$C_{(0.95)lB}(Y)$		-0.0227	-0.0082	0.1675	$0.1367 = C_{(0.95).B}(Y)$

Table 19 reports the decompositions of the synthetic index $G(Y) = 0.3561$. It is useful to remember that the contributions $C_{.lg}(Y)$ reported in this Table are the weighted arithmetic means of the corresponding contributions $C_{hlg}(Y)$ with weights $n_{h.}/N$:

$$C_{.lg}(Y) = \sum_{h=1}^{r} C_{hlg}(Y) \cdot \frac{n_{h.}}{N}.$$

This Table confirms that the two greatest contributions $C_{.lg}(Y)$ are $C_{.31}(Y) = 0.1067$ and $C_{.11}(Y) = 0.089$.

TABLE 19. - *Decompositions of* $G(Y) = 0.3561$ *into the contributions:* $C_{.lg}(Y)$; $C_{.l.}(Y)$; $C_{.lW}(Y), C_{.lB}(Y)$; $C_{..W}(Y)$, $C_{..B}(Y)$

$C_{.lg}(Y)$		l			
		1	2	3	
g	1	0.0890	0.0286	0.1067	
	2	0.0357	0.0115	0.0424	
	3	0.0118	0.0019	0.0285	
$C_{.l.}(Y)$		0.1364	0.0421	0.1776	$0.3561 = G(Y)$
$C_{.lW}(Y)$		0.0890	0.0115	0.0285	$0.1291 = C_{..W}(Y)$
$C_{.lB}(Y)$		0.0475	0.0305	0.1491	$0.2271 = C_{..B}(Y)$

Table 20 reports for the three macro-regions their:

- relative contributions to the point indexes

$$\theta_{hl.}(Y) = \frac{C_{hl.}(Y)}{G_h(Y)} = \frac{(M - \bar{M}_{hl}) \cdot p(l|h) \cdot 2p_{h.} - (M - y_h) \cdot f(l|h) \cdot \frac{n_{h.}-1}{N}}{(M - \bar{M}_{h.}) \cdot 2p_{h.} - (M - y_h) \frac{n_{h.}-1}{N}},$$

- relative contributions to the synthetic index

$$\theta_{.l.}(Y) = \frac{C_{.l.}(Y)}{G(Y)},$$

- relative weights $n_{.l}/N$.

TABLE 20. - *Relative contributions* $\theta_{(0.1)l.}(Y)$, $\theta_{(0.5)l.}(Y)$, $\theta_{(0.95)l.}(Y)$, $\theta_{.l.}(Y)$

	l			Total
	1	2	3	
$\theta_{(0.1)l.}(Y)$	0.3410	0.1380	0.5210	1.0000
$\theta_{(0.5)l.}(Y)$	0.4083	0.1336	0.4581	1.0000
$\theta_{(0.95)l.}(Y)$	0.20427	-0.0011	0.7966	1.0000
$\theta_{.l.}(Y)$	0.3830	0.1182	0.4987	1.0000
$n_{.l}/N$	0.4873	0.1886	0.3241	1.0000

The values reported in Table 20 are very close to the corresponding ones obtained by Zenga and Valli (2016, Sections 4.2 and 5.2) for the decompositions of the Bonferroni inequality measure. This happens because, in the present application: the values of $C_{hl.}(Y)$ are very close to $\frac{M-\bar{M}_{hl}}{M} \cdot p(l|h) \cdot 2p_{h.}$ and the values of $G_h(Y)$are very close to $\frac{M-\bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.}$. Then the values of $\theta_{hl.}(Y)$ are very close to the ones of the ratio

$$\frac{\frac{M-\bar{M}_{hl}}{M} \cdot p(l|h) \cdot 2p_{h.}}{\frac{M-\bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.}} = \frac{M - \bar{M}_{hl}}{M - \bar{M}_{h.}} \cdot p(l|h) = \nu_{hl.}(Y),$$

where

$$\nu_{hl.}(Y) = \frac{V_{hl.}(Y)}{V_h(Y)} = \frac{\frac{M-\bar{M}_{hl}}{M} \cdot p(l|h)}{\frac{M-\bar{M}_{h.}}{M}} = \frac{M - \bar{M}_{hl}}{M - \bar{M}_{h.}} \cdot p(l|h)$$

is the relative contribution of the subpopulation l to the Bonferroni point inequality idex $V_h(Y)$.

Table 21 reports the shares of the within components of $C_{hl.}(Y)$, $G_h(Y)$ and $G(Y)$. We note that the shares of the within components of the whole population and of the North increase for increasing values of p; viceversa for the South there is an opposite relation. We end this section observing that for the whole population the within component is the 36.25% of the synthetic index.

TABLE 21. - *Shares of the within components of the contributions $C_{hl.}(Y)$ and of the point and the synthetic Gini indexes*

	l			Italy
	1	2	3	
$\frac{C_{(0.1)lW}(Y)}{C_{(0.1)l.}(Y)}$	0.5529	0.2190	0.2283	$0.3377 = \frac{C_{(0.1).W}(Y)}{G_{(0.1)}(Y)}$
$\frac{C_{(0.5)lW}(Y)}{C_{(0.5)l.}(Y)}$	0.5947	0.2373	0.1796	$0.3569 = \frac{C_{(0.5).W}(Y)}{G_{(0.5)}(Y)}$
$\frac{C_{(0.95)lW}(Y)}{C_{(0.95)l.}(Y)}$	1.4923	-39.5	0.0642	$0.3912 = \frac{C_{(0.95).W}(Y)}{G_{(0.95)}(Y)}$
$\frac{C_{.lW}(Y)}{C_{.l.}(Y)}$	0.6525	0.2731	0.1604	$0.3625 = \frac{C(Y)}{G(Y)}$

6. FINAL REMARKS AND CONCLUSIONS

This paper, by using the two-step approach proposed in Zenga (2016) for the decomposition by subpopulations of the Zenga (2007) inequality index, obtains the decomposition by subpopulations of the point and synthetic Gini (1914) indexes. Recently, through this approach Zenga and Valli (2016) obtained the decomposition by subpopulations of the Bonferroni (1930) point and synthetic indexes. Moreover, Porro and Zenga (2014) by using this methodology obtained the decomposition by subpopulations of the point and synthetic Zenga (1984) inequality measure, too. We point out that this approach is also used in: Zenga, Radaelli and Zenga (2012) and in Zenga (2013) for the decomposition by sources of the Gini, Bonferroni and Zenga (2007) indexes; Arcagni and Zenga (2014) and in Arcagni (2016) for the decomposition by sources of the Zenga (1984) index; Zenga (2015) for the joint decomposition by subpopulation and sources of the Zenga (2007) index.

In the mentioned decompositions by subpopulations, an important role is played by the bivariate $r \times k$ distribution of the N households of the whole population according to: the r distinct values $\{0 \leq y_1 < \dots < y_h < \dots < y_r\}$ of the variate Y and the k distinct subpopulations. In this bivariate distribution: n_{hg} denotes the frequency of y_h in the subpopulation g; $n_{h.}$ is the frequency of y_h in the whole population; $n_{.g}$ is the size of the subpopulation g; $P_{h.} = \sum_{t=1}^{h} n_{t.}$ and $Q_{h.} = \sum_{t=1}^{h} y_t \cdot n_{t.}$ are respectively the cumulative frequencies and the cumulative incomes in the whole population; $P_{hl} = \sum_{t=1}^{h} n_{tl}$ and $Q_{hl} = \sum_{t=1}^{h} y_t \cdot n_{tl}$ are respectively the cumulative frequencies and the cumulative incomes in the subpopulation l; $T = Q_{r.}(Y)$ and $M = T/N$ are the total income and the arithmetic mean of the whole population, respectively; $T_g = Q_{rg}(Y)$ and $M_g = T_g/n_{.g}$ are the total income and the arithmetic mean of the subpopulation g, respectively; $p_{h.} = P_{h.}/N$ and $q_{h.} = Q_{h.}/T$ are the coordinates of the Lorenz curve; $\bar{M}_{h.} = Q_{h.}/P_{h.}$ is the arithmetic mean (lower mean) in the lower group $\{Y \leq y_h\}$.

The value of the synthetic Gini index can be provided by the ratio

$$G(Y) = \frac{CA}{\frac{1}{2}},$$

where CA is the concentration area. In the case of frequency distribution framework CA and $G(Y)$, can be evaluated – see Section 4 – by the following expressions:

$$CA = \sum_{h=1}^{r} \left\{ (p_{h.} - q_{h.}) - \frac{n_{h.} - 1}{2 \cdot N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N} =$$

$$\sum_{h=1}^{r} \left\{ \frac{M - \bar{M}_{h.}}{M} \cdot p_{h.} - \frac{n_{h.} - 1}{2 \cdot N} \left(\frac{M - y_h}{M} \right) \right\} \cdot \frac{n_{h.}}{N};$$

$$G(Y) = \sum_{h=1}^{r} \{G_h(Y)\} \cdot \frac{n_{h.}}{N}, \tag{31}$$

where the Gini point measure $G_h(Y)$ is provided by

$$G_h(Y) = \frac{M - \bar{M}_{h.}}{M} \cdot 2 \cdot p_{h.} - \frac{n_{h.} - 1}{N} \left(\frac{M - y_h}{M} \right). \tag{30}$$

In the first step the decomposition of $G_h(Y)$ is obtained starting from the two following decompositions:

$$[M - \bar{M}_{h.}] = \sum_{l=1}^{k} \sum_{g=1}^{k} [M_g - \bar{M}_{hl}] \frac{n_{.g}}{N} \cdot p(l|h), \tag{35}$$

and

$$[M - y_h] = \sum_{l=1}^{k} \sum_{g=1}^{k} [M_g - y_h] \frac{n_{.g}}{N} \cdot f(l|h). \tag{38}$$

In (35) and (38): $\bar{M}_{hl}$ is the lower mean in the subpopulation l, $p(l|h) = \frac{P_{hl}}{P_{h.}}$ is the relative frequency of subpopulation l in the lower group $\left(Y \leq y_h\right)$, and $f(l|h) = \frac{n_{hl}}{n_{h.}}$ is the relative frequency of subpopulation l in the group of $n_{h.}$ units with $Y = y_h$.

It is worth to remark that the decompositions (35) and (38) are achieved, with simple algebra, by the use of the relations:

$M = \sum_{g=1}^{k} M_g \cdot \frac{n_{.g}}{N}$, $\bar{M}_{h.} = \sum_{l=1}^{k} \bar{M}_{hl} \cdot p(l|h)$ and $y_h = \sum_{l=1}^{k} y_h \cdot f(l|h)$.

Finally, putting (35) and (38) in (30) the following $k \times k$ additive decomposition of the Gini point inequality index is obtained:

$$G_h(Y) = \sum_{l=1}^{k} \sum_{g=1}^{k} C_{hlg}(Y), \tag{39}$$

where

$$C_{hlg}(Y) = \frac{M_g - \bar{M}_{hl}}{M} \cdot \frac{n_{.g}}{N} p(l|h) \cdot 2p_{h.} - \frac{M_g - y_h}{M} \cdot \frac{n_{.g}}{N} f(l|h) \cdot \frac{n_{h.} - 1}{N}, \tag{40}$$

is the contribution to $G_h(Y)$ that derives from the comparisons of $\bar{M}_{hl}$ and y_h w.r.t. $M_g(Y)$.

In the second step, putting (39) in (31) the following $k \times k$ additive decomposition of $G(Y)$ is obtained:

$$G(Y) = \sum_{h=1}^{r} \left\{ \sum_{l=1}^{k} \sum_{g=1}^{k} C_{hlg}(Y) \right\} \cdot \frac{n_{h.}}{N} = \sum_{l=1}^{k} \sum_{g=1}^{k} C_{.lg}(Y), \tag{41}$$

where

$$C_{.lg}(Y) = \sum_{h=1}^{r} C_{hlg}(Y) \cdot \frac{n_{h.}}{N}. \tag{42}$$

It is important to remark that, starting from the decompositions (39) and (41), it is possible to obtain others interesting decompositions.

In particular, from (39) we obtain:

$$G_h(Y) = \sum_{l=1}^{k} \left\{ \sum_{g=1}^{k} C_{hlg}(Y) \right\} = \sum_{l=1}^{k} C_{hl.}, \tag{51}$$

where

$$C_{hl.} = \frac{M - \bar{M}_{hl}}{M} \cdot 2p_{h.} \cdot p(l|h) - \frac{M - y_h}{M} \cdot \frac{n_{h.} - 1}{N} \cdot f(l|h).$$

Note that, if h is kept fixed, the value of $C_{hl.}(Y)$ depends on the values of $\bar{M}_{hl}$, $p(l|h)$ and $f(l|h)$ related to the subpopulation l. Thus, $C_{hl.}(Y)$ can be interpreted as the contribution of the subpopulation l to the point index $G_h(Y)$. Moreover, $C_{hl.}$ can be split as follows.

$$C_{hl.} = \sum_{g=1}^{k} C_{hlg} = C_{hll} + \sum_{(g:g \neq l)}^{k} C_{hlg} = C_{hlW} + C_{hlB}; \tag{55}$$

where

$$C_{hlW} = \left[\frac{M_l - \bar{M}_{hl}}{M} \cdot p(l|h) \cdot 2p_{h.} - \frac{M_l - y_h}{M} \cdot f(l|h) \cdot \frac{n_{h.} - 1}{N} \right] \cdot \frac{n_{.l}}{N}, \tag{56}$$

and

$$C_{hlB} = \sum_{(g:g \neq l)}^{k} \left[\frac{M_g - \bar{M}_{hl}}{M} \cdot 2p_{h.} \cdot p(l|h) - \frac{M_g - y_h}{M} \cdot \frac{n_{h.} - 1}{N} \cdot f(l|h) \right] \cdot \frac{n_{.g}}{N} \tag{57}$$

are the within and the between part of $C_{hl.}$, respectively.

Putting (55) in (51) gives:

$$G_h(Y) = \sum_{l=1}^{k} [C_{hlW}(Y) + C_{hlB}(Y)] = C_{h.W}(Y) + C_{h.B}(Y); \tag{58}$$

where

$$C_{h.W} = 2p_{h.} \sum_{l=1}^{k} \frac{M_l - \bar{M}_{hl}}{M} \cdot \frac{n_{.l}}{N} p(l|h) - \frac{n_{h.} - 1}{N} \sum_{l=1}^{k} \frac{M_l - y_h}{M} \cdot \frac{n_{.l}}{N} f(l|h), \tag{59}$$

and

$$C_{h.B} = 2p_{h.}\sum_{l=1}^{k}\sum_{g\neq l}^{k}\frac{M_g - \bar{M}_{hl}}{M}\cdot\frac{n_{.g}}{N}p(l|h) - \frac{n_{h.}-1}{N}\sum_{l=1}^{k}\sum_{g\neq l}^{k}\frac{M_g - y_h}{M}\cdot\frac{n_{.g}}{N}f(l|h) \quad (60)$$

are the within and the between part of $G_h(Y)$, respectively.

Finally, putting the decomposition (58) in (27) the following decomposition of G(Y) is obtained:

$$G(Y) = C_{..W} + C_{..B}; \quad (61)$$

where:

$$C_{..W} = \sum_{l=1}^{k}\frac{n_{.l}}{N}\sum_{h=1}^{r}\frac{M_l - \bar{M}_{hl}}{M}p(l|h)2p_{h.}\frac{n_{h.}}{N} + \\ - \sum_{l=1}^{k}\frac{n_{.l}}{N}\sum_{h=1}^{r}\frac{M_l - y_h}{M}f(l|h)\frac{n_{h.}-1}{N}\frac{n_{h.}}{N} \quad (64)$$

and

$$C_{..B} = \sum_{l=1}^{k}\sum_{g\neq l}\frac{n_{.g}}{N}\sum_{h=1}^{r}\frac{M_g - \bar{M}_{hl}}{M}p(l|h)2p_{h.}\frac{n_{h.}}{N} + \\ - \sum_{l=1}^{k}\sum_{g\neq l}\frac{n_{.g}}{N}\sum_{h=1}^{r}\frac{M_g - y_h}{M}f(l|h)\frac{n_{h.}-1}{N}\frac{n_{h.}}{N} \quad (65)$$

are the within and the between parts of $G(Y)$, respectively.

We remark that the value of $C_{..W}$ derives from comparisons of ‘‘incomes’’ of the same subpopulations, while the value of $C_{..B}$ derives from comparisons of ‘‘incomes’’ of different subpopulations.

The theoretical results of this paper are applied to the 2012 Bank of Italy sample survey on household income and wealth. Section 5.1 shows that there is ‘‘strong’’ dependence of the household net disposable income Y from the Italian regional areas. In fact, the mean of the South is very far from the means of the other two Italian macro-regions, the North has the greatest inequality, while the Center has the lowest one, and that the inequality of the whole population is a little bit greater than the one of the North. Moreover, Section 5.2 illustrates the decompositions of the synthetic index $G(Y)$ and of the point index $G_{h(p)}(Y) = G_{(p)}(Y)$ for three values $(0.10; 0.50; 0.95)$ of the cumulative relative frequency p. For the interpretetion of the contributions C_{hlg} are useful the following relations (46), (47) and (48) introduced in Section 4.1.

$$C_{hlg}(Y) = V_{hlg}(Y)\cdot 2p_{h.} - A_{hlg}(Y)\cdot\frac{n_{h.}-1}{N}, \quad (46)$$

where

$$V_{hlg}(Y) = \frac{M_g - \bar{M}_{hl}}{M}\cdot\frac{n_{.g}}{N}\cdot p(l|h), \quad (47)$$

and

$$A_{hlg}(Y) = \frac{M_g - y_h}{M}\cdot\frac{n_{.g}}{N}\cdot f(l|h). \quad (48)$$

Now we point out that in the present application there are $N = 8151$ households, consequently the value of C_{hlg} is fundamentally given by the first term of the right side of (46). In addition, if for $Y = y_h$ there is only one household, it follows that two of the corresponding relative frequencies $f(l|h)$ are equal to zero and the remaining is equal to one. Thus, two colums of the contributions C_{hlg} are exactly equal to the first term $V_{hlg}(Y) \cdot 2p_{h.}$ of the right side of (46). In the application at hand, this happens for the $\frac{6995}{7287} \cdot 100 = 95.993\%$ of the different values of the total income Y. Note that Zenga and Valli (2016) obtained for the Bonferroni point index

$$V_h(Y) = \frac{M - \bar{M}_{h.}}{M}$$

the following $k \times k$ decomposion by subpopulations

$$V_h(Y) = \sum_l \sum_g V_{hlg}(Y) = \sum_l \sum_g \frac{M_g - \bar{M}_{hl}}{M} \cdot \frac{n_{.g}}{N} \cdot p(l|h).$$

In other words, in the present application the values of the contributions C_{hlg} are fundamentally given by the product of the corresponding Bonferroni contributions V_{hlg} and $2p_{h.}$.

It is worthwhile to remark that in Table 17 the contributions $C_{(0.95)13}$, $C_{(0.95)23}$, $C_{(0.95)2.}$, $C_{(0.95)1B}$, and $C_{(0.95)2B}$ are negative. Relations (46) and (47) ''basically'' show that: if $\bar{M}_{hl} > M_g$ and $p(l|h) > 0$, then the corresponding contribution C_{hlg} is negative. Note that: the lower mean $\bar{M}_{hl}$ is a non decreasing function in h and $\bar{M}_{rl} = M_l$. Now, let us suppose that $M_l > M_g$. Then, there is an integer $h' \leq r$, such that $\left(M_g - \bar{M}_{hl}\right) < 0$, for $\forall(h \geq h')$. This result explain, for example, why there are negative contributions to the Bonferroni point index (and to the Gini point index) when we compare, for $p = 0.95$, the lower mean of the North with the mean of the South. For more details on this point see Zenga and Valli (2016).

The fact that some contributions are negative may cause some difficulties for the interpretation. This cannot happens for the Zenga (2007) inequality index. In fact, by constrution, see Zenga (2016), the contributions

$$B_{hlg}(Y) = \frac{\overset{+}{M}_{hg} - \bar{M}_{hl}}{\overset{+}{M}_{h.}} \cdot a(g|h) \cdot p(l|h) \tag{66}$$

to the Zenga (2007) point inequality index

$$I_h(Y) = \frac{\overset{+}{M}_{h.} - \bar{M}_{h.}}{\overset{+}{M}_{h.}} \tag{67}$$

are never negative. In (66) $\overset{+}{M}_{h.}$ is the mean of the upper group $\{Y > y_h\}$ of the whole population, $\overset{+}{M}_{hg}$ is the mean in the upper group of the subpopulation g and $a(g|h)$ is the relative frequency of the subpopulation g in the upper group.

REFERENCES

Arcagni A., Zenga M. (2014). The decomposition by sources of the ζ inequality index. *Proceedings of SIS 2014*, Cagliari.

Arcagni A. (2016). On the decomposition by sources of the Zenga 1984 point and synthetic inequality indexes. *Statistical Methods and Applications*, DOI:10.1007/s10260-016-0360-0.

Banca d'Italia (2014). Household income and wealth in 2012. Supplements to the Statistical Bulletin - Sample Survey, 5, Year XXIV.

Bonferroni C. (1930). Elementi di statistica generale. Seeber, Firenze.

De Vergottini M. (1940). Sul significato di alcuni indici di concentrazione. *Giornale degli Economisti e Annali di Economia*, **2**, 317-347.

Gini C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. In *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*. Anno Accademico 1913-1914, Tomo LXXIII - Parte Seconda.

Lorenz M. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**, 209-219.

Ogwang T. (2014). A convenient method of decomposing the Gini index by population subgroups. *Journal of Official Statistics*, **30(1)**, 91-105.

Porro F., Zenga M. (2014). The decomposition by subgroups of the inequality curve $Z(p)$ and the inequality index ζ. *Proceedings of SIS 2014*, Cagliari.

Radaelli P., (2008). A subgroups decomposition of Zenga's uniformity and inequality indexes. *Statistica & Applicazioni*, **VI(2)**, 117-136.

Radaelli P., (2010). On the decomposition by subgroups of the Gini index and Zenga's uniformity and inequality indexes. *International Statistical Review*, **78(1)**, 81-101.

Zenga M. (1984). Proposta per un indice di concentrazione basato sui rapporti tra quantili di popolazione e quantili di reddito. *Giornale degli Economisti e Annali di Economia*, **43**, 301-326.

Zenga M.M. (2007). Inequality curve and inequality index based on the ratios between lower and upper means. *Statistica & Applicazioni*, **V(1)**, 3-27.

Zenga M. (2013). Decomposition by sources of the Gini, Bonferroni and Zenga inequality indexes. *Statistica & Applicazioni*, **XI(2)**, 131-158.

Zenga M. (2015). Joint decomposition by subpopulations and sources of the Zenga inequality index $I(Y)$. *Statistica & Applicazioni*, **XIII(2)**, 163-195.

Zenga M. (2016). On the decomposition by subpopulations of the point and the synthetic Zenga (2007) inequality indexes. *Metron*, **74(3)**, pp. 375-405.

Zenga M.M., Radaelli P., Zenga Ma. (2012). Decomposition of Zenga's inequality index by sources. *Statistica & Applicazioni*, **X(1)**, 3-31.

Zenga M., Valli I. (2016). On the decomposition by subpopulations of the point and synthetic Bonferroni inequality measures. *Statistica & Applicazioni,* **XIV(1)**, 3-28.