Statistica & Applicazioni - Vol. XIV, n. 1, 2016, pp. 3-28

ON THE DECOMPOSITION BY SUBPOPULATIONS OF THE POINT AND SYNTHETIC BONFERRONI INEQUALITY MEASURES

Michele Zenga*
Igor Valli**

ABSTRACT

This paper, by using the ''two-step'' approach proposed in Radaelli (2008, 2010) and in Zenga (2016) for the decomposition of the Zenga (2007) index, obtains the decomposition of the Bonferroni (1930) inequality measure. In the first step the Bonferroni point measure $V_h(Y)$ is decomposed in a weighted mean of $k \times k$ relative differences between the mean $M_g(Y)$ of subpopulation g and the lower mean $\bar{M}_{h\ell}(Y)$ of the subpopulation ℓ; the weights are the product of their relative frequencies. From this decomposition, we obtain two decompositions of $V_h(Y)$ into the within and the between components, and into the sum of the k contributions of each subpopulation. In the second step, the decompositions of the Bonferroni point index are extended, in a simple way, to the Bonferroni synthetic measure $V(Y)$. We remark that the decomposition obtained in this paper is rather different from those proposed by Tarsitano (1990), and Bárcena-Martin and Silber (2013). Actually, they provide only a decomposition by subpopulations of the Bonferroni synthetic index.

Keywords: *Bonferroni Index, Point Inequality Index, Sguthetic Inequality Index, Decomposition by subpopulations.*

1. INTRODUCTION

Tarsitano (1990) decomposed the synthetic Bonferroni (1930) index in a sum of a within-group component (a weighted average of the subpopulations Bonferroni indexes with weights equal to the subpopulation income shares) and the accross component, a weighted average of the other subgroup characteristics. For more details on this point see Tarsitano (1990) and Valli (2016). Bárcena-Martin and Silber (2013) extended the matrix approach proposed by Silber (1989) for the decomposition of the synthetic Gini index, to the decomposition of the synthetic Bonferroni index. Recently, Zenga (2016) has used the two-step approach for the decomposition by subpopulations of the point and synthetic Zenga (2007) inequality indexes. This paper, by using this approach: obtains, in the first step, the decomposition of the Bonferroni point index, and extend, in the second step this decomposition to the synthetic Bonferroni index $V(Y)$. The paper is organized as follows. In the next section, some definitions and notation are introduced. In particular this section provides: the definition of

* Dipartimento di Statistica e Metodi Quantitativi - Università Milano-Bicocca - Piazza dell'Ateneo Nuovo, 1 - 20126 MILANO (e-mail: michele.zenga@unimib.it)
** Dipartimento Economia e Management - Università di Brescia - C.da S. Chiara 50 - 25122 BRESCIA (e-mail: igor.valli@gmail.com).

the lower mean and of the point $V_{(i)}(Y)$ and synthetic Bonferroni index $\tilde{V}(Y)$. Moreover, in this section, $N = 10$ ordered values $y_{(i)}$ are utilized for the calculation of the point index $V_{(i)}(Y)$. Section 3 extends the definitions of Section 2 to the case of the frequency distribution framework $\{(y_h, n_{h.}),\ h = 1, \dots, r\}$, where y_h are the different values assumed by the variate Y and $n_{h.}$ is the frequency of y_h. Moreover, this section illustrates the basic bivariate $r \times k$ distribution of the whole population partitioned into k subpopulations and provides the definitions of the lower mean in the whole population and in the subpopulations and the means in the whole population and in the subpopulations. In Section 4 different additive decompositions of $V_h(Y)$ and $V(Y)$ are obtained. In Section 4.1 the calculations for the decompositions of $V_h(Y)$ and $V(Y)$ are illustrated. Moreover, Section 4.2 and 4.3 obtain the contribution of each subpopulation to the point and synthetic indexes, and the decomposition of $V_h(Y)$ and $V(Y)$ into the within and the between components. 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). Finally, Section 6 is devoted to conclusions and final remarks.

2. DEFINITIONS AND NOTATION

Let Y denote a non-negative variate, usually income, observed on 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 $n_{.g}$, $g = 1, ..., k$. Let

$$0 \leq y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(i)} \leq \dots \leq y_{(N)} > 0$$

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

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

be the income of the i poorest population units;

$$T(Y) = Q_{(N)}(Y) = \sum_{i=1}^{N} y_{(i)}; \tag{2}$$

$$\bar{M}_{(i)}(Y) = \frac{Q_{(i)}(Y)}{i}, \tag{3}$$

be the lower mean $\{$the mean of $(y_{(1)}, y_{(2)}, \dots, y_{(i)})\}$;

$$M(Y) = \frac{T(Y)}{N} = \bar{M}_{(N)}(Y), \tag{4}$$

be the population mean.

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

$$V_{(i)}(Y) = \frac{M(Y) - \bar{M}_{(i)}(Y)}{M(Y)}; \qquad i = 1, \ldots, N \tag{5}$$

$$\tilde{V}(Y) = \frac{1}{N-1} \sum_{i=1}^{N-1} V_{(i)}(Y), \tag{6}$$

respectively. Note that $V_{(i)}(Y)$ is the relative variation of the lower mean $\bar{M}_{(i)}(Y)$ w.r.t. $M(Y)$, hence, $\tilde{V}(Y)$ is their (simple) arithmetic mean.

In case of maximum inequality $\{y_{(1)} = \ldots = y_{(N-1)} = 0, y_{(N)} > 0\}$, it is known that $\tilde{V}(Y) = 1$, for all $N \geq 2$. Note that $\tilde{V}(Y)$ does not discern among maximum inequality cases with different values of N. In the case of maximum inequality, it seems more reasonable that the value of an inequality index $C_N(Y)$, evaluated on N units, is such that:

(a) $C_N(Y)$ is an increasing and positive function of N;

(b) $\lim_{N\to\infty} C_N(Y) = 1$.

Now, multiplying, both sides of (6) by $\frac{N-1}{N}$, we have:

$$V'(Y) = \frac{N-1}{N} \cdot \tilde{V}(Y) = \frac{1}{N} \cdot \sum_{i=1}^{N-1} V_{(i)}(Y) = \frac{1}{N} \cdot \sum_{i=1}^{N} V_{(i)}(Y). \tag{7}$$

Thus, in case of maximum inequality, $V'(Y) = \frac{N-1}{N}$. Note that $\frac{N-1}{N}$ is an increasing and positive function of N such that $\lim_{N\to\infty} \frac{N-1}{N} = 1$.

The following $N = 10$ ordered values $\{y_{(i)} = 2, 2, 8, 24, 29, 37, 37, 37, 62, 62\}$ are utilized in Table 1 for the calculation of $V_{(i)}(Y)$ and $V'(Y)$. Table 1 shows that: $M(Y) = \bar{M}_{(N)}(Y) = 30$ and $V'(Y) = \frac{5{,}16}{10} = 0{,}5157$.

TABLE 1. - *Distribution of $N = 10$ units and calculation of $V_{(i)}(Y)$*

i	$y_{(i)}$	$Q_{(i)}(Y)$	$\bar{M}_{(i)}(Y)$	$V_{(i)}(Y)$
1	2	2	2.00	0.9333
2	2	4	2.00	0.9333
3	8	12	4.00	0.8667
4	24	36	9.00	0.70
5	29	65	13.00	0.5667
6	37	102	17.00	0.4333
7	37	139	19.8571	0.3381
8	37	176	22	0.2667
9	62	238	26.4444	0.1185
10	62	300	30	0.00
Total				5.1566

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

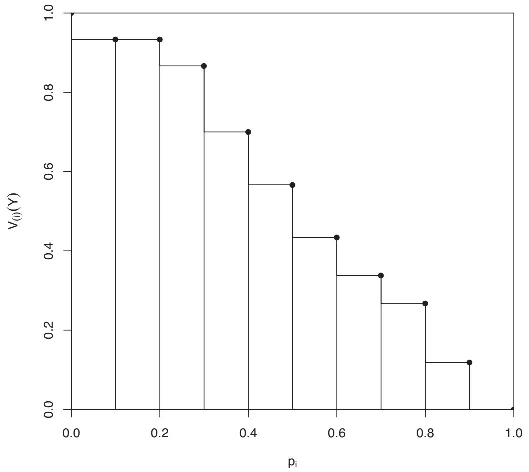

FIGURE 1. - *Graph of* $V_{(i)}(Y)$

3. DEFINITIONS AND NOTATION IN THE FREQUENCY DISTRIBUTION FRAMEWORK

The last column of Table 1 shows that $V_{(i)}(Y)$ may not be constant for units taking the same value of Y. This behavior of $V_{(i)}(Y)$ is not reasonable in the decomposition by subpopulation because units with the same value of Y may belong to different subpopulations.

We will overcome this situation by substituting the values of $V_{(i)}(Y)$ corresponding to units with the same value y_h of Y with the value $V_{(P_{h.})}(Y)$, where $P_{h.}$ is the number of units with $Y \leq y_h$. We introduce now the appropriate definitions and notation used in the decomposition by subpopulation proposed in this paper.

Let

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

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 shown 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*

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

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

$$P_{h}. = P_{h.}(Y) = \sum_{t=1}^{h} n_{t.}, \tag{8}$$

$$Q_{h.}(Y) = Q_{(P_{h}.)}(Y) = \sum_{t=1}^{h} y_t \cdot n_{t.}, \tag{9}$$

$$T(Y) = Q_{r.}(Y) = \sum_{h=1}^{r} y_h \cdot n_{h.} \tag{10}$$

$$\bar{M}_{h.}(Y) = \frac{Q_{h.}(Y)}{P_{h.}}. \tag{11}$$

Note that, $M(Y) = \bar{M}_{r.}(Y)$.

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

$$P_{hg} = P_{hg}(Y) = \sum_{t=1}^{h} n_{tg}, \tag{12}$$

$$Q_{hg}(Y) = Q_{(P_{hg})}(Y) = \sum_{t=1}^{h} y_t \cdot n_{tg}, \tag{13}$$

$$T_g(Y) = Q_{rg}(Y) = \sum_{h=1}^{r} y_h \cdot n_{hg}, \tag{14}$$

$$M_g(Y) = \frac{T_g}{n_{.g}} \tag{15}$$

$$o(g) = \min h : n_{hg} > 0 \tag{16}$$

$$\bar{M}_{hg}(Y) = \begin{cases} y_{o(g)} & \text{for } h < o(g) \\ \frac{Q_{hg}(Y)}{P_{hg}} & \text{for } h \geq o(g) \end{cases} \tag{17}$$

where $\bar{M}_{hg}(Y)$ in (17) denotes the mean of the first poorest P_{hg} units. Note that, from (17) and (15), follows $M_g(Y) = \bar{M}_{rg}(Y)$.

Note that, from (12) to (14), we can deduce the quantities defined in (8)-(11):

$$P_{h.} = \sum_{g=1}^{k} P_{hg}(Y) \tag{18}$$

$$Q_{h.} = \sum_{g=1}^{k} Q_{hg}(Y) \tag{19}$$

$$T(Y) = \sum_{g=1}^{k} T_g(Y) \tag{20}$$

We can now define the Bonferroni inequality measures in the frequency distribution framework. From (9) we have:

$$V'(Y) = \frac{1}{N}\sum_{i=1}^{N} V_{(i)}(Y)$$

$$= \frac{1}{N}\sum_{h=1}^{r} \sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} V_{(i)}(Y). \tag{21}$$

In order to assign same point inequality measure to units that have same value $Y = y_h$, we set, for $1 + P_{h.} - n_{h.} \leq i \leq P_{h.}$:

$$V_{(i)}(Y) = V_{(P_{h.})}(Y) = \frac{M(Y) - \bar{M}_{(P_{h.})}(Y)}{M(Y)} = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} = V_h(Y). \tag{22}$$

Hence, from (22), we approximate

$$\sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} V_{(i)}(Y)$$

with

$$\sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} V_{(P_{h.})}(Y) = \sum_{i=1+P_{h.}-n_{h.}}^{P_{h.}} V_h(Y) = V_h(Y) \cdot n_{h.} \tag{23}$$

and we define the synthetic inequality index $V(Y)$,

$$V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N}, \tag{24}$$

as weighted mean of the point inequality measures $V_h(Y)$, with weights $n_{h.}/N$.

Obviously, since $V_h(Y) = V_{(P_{h.})}(Y) \leq V_{(i)}(Y)$ for $1 + P_{h.} - n_{h.} \leq i \leq P_{h.}$, follows:

$$V(Y) \leq V'(Y).$$

The $N = 10$ values $y_{(i)}$, introduced in the previous section are utilized in Table 3 for the calculation of $\bar{M}_{h.}(Y)$, $V_h(Y)$ and $V(Y)$. Figure 4 reports the graphs of $V_{(i)}(Y)$ and $V_h(Y)$

TABLE 3. - *Calculation of the lower means, $V_h(Y)$ and $V(Y)$*

h	y_h	$Q_{(h.)}(Y)$	$\bar{M}_{h.}(Y)$	$V_h(Y)$
1	2	4	2	0.9333
2	8	12	4	0.8667
3	24	36	9	0.70
4	29	65	13	0.5667
5	37	176	22	0.2667
6	62	300	30	0.00
				$V(Y) = 0.48$

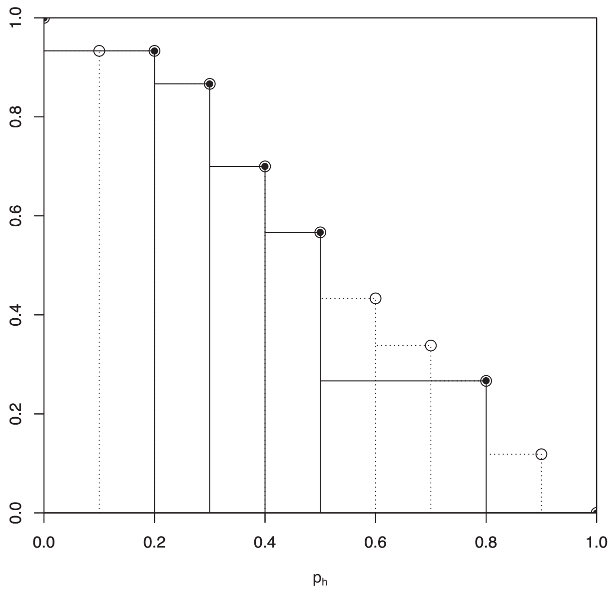

FIGURA 2. - *Graphs of $V_{(i)}(Y)$ (dotted line) and $V_h(Y)$ (solid line)*

4. DECOMPOSITION BY SUBPOPULATIONS OF THE POINT $V_h(Y)$ AND SYNTHETIC $V(Y)$ BONFERRONI'S INEQUALITY INDEXES

In this section we decompose by subpopulations the point $V_h(Y)$ and the synthetic $V(Y)$ Bonferroni indexes using the ‘‘two-step’’ approach, recently proposed by Zenga (2016) for the decomposition of the Zenga (2007) inequality indexes. In particular, in the first step, we decompose by subpopulations the Bonferroni point index $V_h(Y)$, then, putting this decomposition in the relation $V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N}$, we obtain the corresponding decomposition by subpopulations of $V(Y)$.

The mean $M(Y)$ is related to the k subpopulations means $M_g(Y)$ by the relation

$$M(Y) = \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N}, \tag{25}$$

and $\bar{M}_{h.}(Y)$ is related to the k subpopulation lower means $\bar{M}_{h\ell}(Y)$ by the relation

$$\bar{M}_{h.}(Y) = \sum_{e=1}^{k} \bar{M}_{h\ell}(Y) \cdot p(\ell|h), \tag{26}$$

where

$$p(\ell|h) = \frac{P_{h\ell}}{P_{h.}} \qquad h = 1, \ldots, r; \ell = 1, \ldots, k \tag{27}$$

is the relative frequency of the subpopulation ℓ in the lower group $\{Y \le y_h\}$. Note that $\sum_{\ell=1}^{k} p(\ell|h) = 1$, $\sum_{g=1}^{k} \frac{n_{.g}}{N} = 1$ and $\sum_{\ell=1}^{k} \sum_{g=1}^{k} p(\ell|h) \cdot \frac{n_{.g}}{N} = 1$.

Using (25) and (26), the following $k \times k$ additive decomposition of

$$[M(Y) - \bar{M}_{h.}(Y)]$$

is obtained:

$$\begin{aligned}[M(Y) - \bar{M}_{h.}(Y)] &= \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} \cdot \sum_{\ell=1}^{k} p(\ell|h) - \sum_{\ell=1}^{k} \bar{M}_{h\ell}(Y) \cdot p(\ell|h) \cdot \sum_{g=1}^{k} \frac{n_{.g}}{N} \\ &= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \left[M_g(Y) - \bar{M}_{h\ell}(Y)\right] \cdot p(\ell|h) \cdot \frac{n_{.g}}{N}. \end{aligned} \tag{28}$$

Now, dividing both sides of (28) by $M(Y)$, we obtain the following $k \times k$ additive decomposition of $V_h(Y)$:

$$V_h(Y) = \sum_{\ell=1}^{k} \sum_{g=1}^{k} V_{h\ell g}(Y) \tag{29}$$

where

$$V_{h\ell g}(Y) = \left[\frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)}\right] \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \tag{30}$$

is the contribution to $V_h(Y)$ that derives from the comparison of $\bar{M}_{h\ell}(Y)$ w.r.t. $M_g(Y)$.

Finally, putting (29) in (24), the following $k \times k$ additive decomposition of $V(Y)$ is obtained:

$$V(Y) = \sum_{h=1}^{r} \left\{ \sum_{\ell=1}^{k} \sum_{g=1}^{k} V_{h\ell g}(Y) \right\} \cdot \frac{n_{h.}}{N}$$

$$= \sum_{\ell=1}^{k} \sum_{g=1}^{k} V_{.\ell g}(Y), \tag{31}$$

where

$$V_{.\ell g}(Y) = \sum_{h=1}^{r} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{N} \tag{32}$$

is the weighted mean of $V_{h\ell g}(Y)$ with weights $\frac{n_{h.}}{N}$.

4.1 *Example*

The results of this section are illustrated by the 6×3 bivariate distribution, reported in Table 4, with $N = 10$ units, $k = 3$ subpopulations and $r = 6$ distinct values of Y. Table 5 reports all the values necessary for the decomposition (29) of $V_h(Y)$. Table 6 reports the calculation of the $k \times k$ values of $V_{1\ell g}(Y)$ and Table 7 reports the decomposition of all $V_h(Y)$, $h = 1, \ldots, r = 6$. Finally, Table 8 reports the decomposition of the synthetic index $V(Y)$.

TABLE 4. - *Joint frequencies n_{hg}, total frequencies $n_{.g}$ and $n_{h.}$, cumulative frequencies P_{hg} and $P_{h.}$ and cumulative incomes $Q_{hg} = Q_{hg}(Y)$ and $Q_{h.} = Q_{h.}(Y)$*

		g			Total	g			Total	g			Total
		1	2	3		1	2	3		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	0	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
$n_{.g}$		5	2	3									

TABLE 5. - *Lower means in the subpopulations and in the whole population and relative frequencies* $p(\ell|h) = \frac{P_{h\ell}}{P_{h.}}$

		g			$\bar{M}_{h.}$
		1	2	3	
$o(g)$		1	2	1	
h	y_h	$\bar{M}_{h1}$	$\bar{M}_{h2}$	$\bar{M}_{h3}$	
1	2	2	8	2	2
2	8	2	8	2	4
3	24	13	8	2	9
4	29	13	8	15.50	13
5	37	25	22.50	15.50	22
6	62	32.4	22.5	31.00	30
$\frac{n_{.g}}{N}$		0.50	0.20	0.30	

ℓ		
1	2	3
$\frac{P_{h1}}{P_{h.}}$	$\frac{P_{h2}}{P_{h.}}$	$\frac{P_{h3}}{P_{h.}}$
0.50	0.00	0.50
0.33	0.33	0.33
0.50	0.25	0.25
0.40	0.20	0.40
0.5	0.25	0.25
0.5	0.20	0.30

TABLE 6. - *Calculation of* $V_{1\ell g}(Y)$

$h = 1$			ℓ			$\frac{n_{.g}}{N}$
$M(Y) = 30$			1	2	3	
$V_{h\ell g}(Y)$			$\bar{M}_{h1} = 2$	$\bar{M}_{h2} = 8$	$\bar{M}_{h3} = 2$	
g	1	$M_1 = 32.40$	$[1.0133] \cdot 0.25 = 0.2533$	$[0.8133] \cdot 0 = 0,00$	$[1.0133] \cdot 0.25 = 0.2533$	0.50
	2	$M_2 = 22.50$	$[0.6833] \cdot 0.10 = 0.0683$	$[0.4833] \cdot 0 = 0.00$	$[0.6833] \cdot 0.10 = 0.0683$	0.20
	3	$M_3 = 31.00$	$[0.9667] \cdot 0.15 = 0.1450$	$[0.7667] \cdot 0 = 0.00$	$[0.9667] \cdot 0.15 = 0.1450$	0.30
$p(\ell|h)$			0.50	0.00	0.50	1.00

TABLE 7. - *Values of the contributions* $V_{h\ell g}(Y),\ h = 1, \ldots, 6$

$V_{1\ell g}(Y)$		ℓ		
		1	2	3
g	1	0.2533	0.00	0.2533
	2	0.0683	0.00	0.0683
	3	0.1450	0.00	0.1450
		$V_1(Y) = 0.9333$		

$V_{2\ell g}(Y)$		ℓ		
		1	2	3
g	1	0.1689	0.1356	0.1689
	2	0.0456	0.0322	0.0456
	3	0.0967	0.0767	0.0967
		$V_2(Y) = 0.8667$		

(*continue*)

$V_{3\ell g}(Y)$		ℓ = 1	2	3
g	1	0.1617	0.1017	0.1267
	2	0.0317	0.0242	0.0342
	3	0.0900	0.0575	0.0725
		$V_3(Y) = 0.70$		

$V_{4\ell g}(Y)$		ℓ = 1	2	3
g	1	0.1293	0.0813	0.1127
	2	0.0253	0.0193	0.0187
	3	0.0720	0.0460	0.0620
		$V_4(Y) = 0.5667$		

$V_{5\ell g}(Y)$		ℓ = 1	2	3
g	1	0.0617	0.0412	0.0704
	2	-0.0083	0.00	0.0117
	3	0.0300	0.0212	0.0388
		$V_5(Y) = 0.2667$		

$V_{6\ell g}(Y)$		ℓ = 1	2	3
g	1	0.00	0.0330	0.0070
	2	-0.0330	0.00	-0.0170
	3	-0.0070	0.0170	0.00
		$V_6(Y) = 0.00$		

TABLE 8. - *Values of the contributions* $V_{.\ell g}(Y)$

$V_{.\ell g}(Y)$		ℓ = 1	2	3	
g	1	0.1152	0.0508	0.1140	
	2	0.0148	0.0076	0.0236	
	3	0.0625	0.0278	0.0637	
					$V(Y) = 0.48$

4.2 *Contributions of each subpopulations to the point* $V_h(Y)$ *and the synthetic* $V(Y)$ *indexes*

Starting from the $k \times k$ contributions in (29), it is possible to obtain other useful decompositions.

Let:

$$V_{h\ell.}(Y) = \sum_{g=1}^{k} V_{h\ell g}(Y). \tag{33}$$

Then, using (30) and (25) in (33), gives:

$$\begin{aligned} V_{h\ell.}(Y) &= \sum_{g=1}^{k} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \\ &= \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h). \end{aligned} \tag{34}$$

Note that $\frac{M(Y)-\bar{M}_{h\ell}(Y)}{M(Y)}$ is the relative variation of the lower mean $\bar{M}_{h\ell}(Y)$ w.r.t. the mean $M(Y)$.

Now, from (29), (33) and (34) the following k additive decompositions of $V_h(Y)$ is obtained:

$$V_h(Y) = \sum_{\ell=1}^{k}\sum_{g=1}^{k} V_{h\ell g}(Y) = \sum_{\ell=1}^{k} V_{h\ell.}(Y)$$

$$= \sum_{\ell=1}^{k} \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \tag{35}$$

In conclusion, $V_h(Y)$ is the sum of the k contributions $V_{h\ell.}(Y)$.

Equation (35) shows that the point index $V_h(Y)$ is the weighted mean of the k relative variations $\frac{M(Y)-\bar{M}_{h\ell}(Y)}{M(Y)}$ with weights $p(\ell|h)$. Thus, $V_{h\ell.}(Y)$ can be interpreted as the contribution of the subpopulation ℓ to the point inequality index $V_h(Y)$.

Finally, putting (35) in (24), the following k additive decomposition of $V(Y)$ is obtained

$$V(Y) = \sum_{h=1}^{r}\sum_{\ell=1}^{k} V_{h\ell.}(Y) \cdot \frac{n_{h\cdot}}{N} \tag{36}$$

$$= \sum_{\ell=1}^{k} V_{.\ell.}(Y) \tag{37}$$

where

$$V_{.\ell.}(Y) = \sum_{h=1}^{r} V_{h\ell.}(Y) \cdot \frac{n_{h.}}{N} \tag{38}$$

denotes the weighted mean of $V_{h\ell.}(Y)$ with weights $\frac{n_{h.}}{N}$. Thus, $V_{.\ell.}(Y)$ can be interpreted as the contribution of the subpopulation ℓ to the Bonferroni synthetic index $V(Y)$.

Note that $V_{.\ell.}(Y)$ can be also obtained from (31) by the sum:

$$V_{.\ell.}(Y) = \sum_{g=1}^{k} V_{.\ell g}(Y). \tag{39}$$

4.3 *Within and Between components of* $V_{h\ell.}(Y)$, $V_h(Y)$ *and* $V(Y)$

The contribution $V_{h\ell.}(Y)$ of the subpopulation ℓ to the point index $V_h(Y)$ can be split into a ‘‘within’’ and a ‘‘between’’ component.

From (33) we have:

$$V_{h\ell.}(Y) = \sum_{g=1}^{k} V_{h\ell g}(Y) = V_{h\ell W}(Y) + V_{h\ell B}(Y), \tag{40}$$

where,

$$\begin{aligned} V_{h\ell W}(Y) &= V_{h\ell\ell}(Y) \\ &= \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{N}, \end{aligned} \tag{41}$$

and

$$\begin{aligned} V_{h\ell B}(Y) &= \sum_{(g:g\neq\ell)}^{k} V_{h\ell g}(Y) \\ &= \sum_{(g:g\neq\ell)}^{k} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N}. \end{aligned} \tag{42}$$

In (41) the value of $V_{h\ell\ell}(Y)$ derives from comparisons of ''incomes'' of the same subpopulation ℓ; thus, $V_{h\ell W}(Y) = V_{h\ell\ell}(Y)$ can be interpreted as the ''within'' part of the contribution $V_{h\ell.}(Y)$. Vice versa, in (42) the value of $V_{h\ell B}(Y) = \sum_{(g:g\neq\ell)}^{k} V_{h\ell g}(Y)$ derives from the comparison of ''incomes'' of different subpopulations; thus, $V_{h\ell B}(Y)$ can be interpreted as the ''between'' part of $V_{h\ell.}(Y)$.

From (35) and (40), we obtain:

$$\begin{aligned} V_h(Y) &= \sum_{\ell=1}^{k} V_{h\ell.}(Y) = \sum_{\ell=1}^{k} [V_{h\ell W}(Y) + V_{h\ell B}(Y)] \\ &= V_{h.W}(Y) + V_{h.B}(Y); \end{aligned} \tag{43}$$

where,

$$V_{h.W}(Y) = \sum_{\ell=1}^{k} V_{h\ell W}(Y) \tag{44}$$

is the sum of the ''within'' parts of the contributions $V_{h\ell.}(Y)$ and can be interpreted as the within part of $V_h(Y)$ and,

$$V_{h.B}(Y) = \sum_{\ell=1}^{k} V_{h\ell B}(Y) \tag{45}$$

can be interpreted as the ''between'' part of the Bonferroni point index $V_h(Y)$.

Finally, putting the decomposition (43) in (24) gives the following decomposition of the Bonferroni synthetic index $V(Y)$:

$$V(Y) = V_{..W}(Y) + V_{..}B(Y), \tag{46}$$

where,

$$V_{..W}(Y) = \sum_{h=1}^{r} V_{h.W}(Y) \cdot \frac{n_{h.}}{N} \tag{47}$$

and

$$V_{..B}(Y) = \sum_{h=1}^{r} V_{h.B}(Y) \cdot \frac{n_{h.}}{N} \tag{48}$$

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

Using in (47) the relations (44) and (41), the following useful relation is obtained:

$$\begin{aligned} V_{..W}(Y) &= \sum_{h=1}^{r} \sum_{\ell=1}^{k} V_{h\ell W}(Y) \cdot \frac{n_{h.}}{N} \\ &= \sum_{\ell=1}^{k} \sum_{h=1}^{r} \left[\frac{M_{\ell}(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \right] \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{N} \cdot \frac{n_{h.}}{N}. \end{aligned} \tag{49}$$

Note that in (49) are involved only the relative variations of the lower means $\bar{M}_{h\ell}(Y)$ w.r.t. the means $M_{\ell}(Y)$ of the same subpopulation: by construction these relative variation $\frac{M_{\ell}(Y) - \bar{M}_{h\ell}(Y)}{M(Y)}$ are never negative.

Using in (48) the relations (45) and (42) the following useful relation is obtained:

$$\begin{aligned} V_{..B} &= \sum_{h=1}^{r} \sum_{\ell=1}^{k} V_{h\ell B}(Y) \cdot \frac{n_{h.}}{N} = \sum_{h=1}^{r} \left[\sum_{\ell=1}^{k} \sum_{(g:g\neq\ell)}^{k} V_{h\ell g}(Y) \right] \cdot \frac{n_{h.}}{N} \\ &= \sum_{\ell=1}^{k} \sum_{g:g\neq\ell}^{k} \sum_{h=1}^{r} \left[\frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \right] \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \cdot \frac{n_{h.}}{N}. \end{aligned} \tag{50}$$

In (50) are involved only the relative variations of the lower means $\bar{M}_{h\ell}(Y)$ w.r.t. the means $M_g(Y)$, $g \neq \ell$, of different subpopulations. Unfortunatly, these relative variations $\frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)}$ can be negative. For more details on this point see: Section 5.2, Tables 16, 17 and 18, Section 6, Table 20.

5. APPLICATION

The data used in this application are supplied by the 2012 Central Bank of Italy sample survey on household income and wealth. This survey covers 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 8151 households have been partitioned according to their residence area: North $(g = 1)$, Center $(g = 2)$ and South with Islands $(g = 3)$. In all computations that follow we consider the weights $w_i > 0$ $(i = 1, 2, \ldots, 8151;\ W = \sum w_i = 8151)$ 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 semplicity's sake we will continue the use of the previous sections. Thus, to denote the sum of the weights of the $n_{h\ell}$ households of the subpopulation ℓ with total income $Y = y_h$ we will use $n_{h\ell}$ instead of $w_{h\ell}$. Note that the frequency distribution of the total income Y has $r = 7287$ different values.

5.1 *Aggregate characteristics in three Italian macro-regions*

Table 9 reports for the total income Y of each geographic area: the arithmetic mean, the median, the synthetic index $V_{.\ell}(Y)$, the sum of the weights $n_{.\ell}(= w_{.\ell})$ and the relative weights $\frac{n_{.\ell}}{N}(= \frac{w_{.\ell}}{W})$. The synthetic inequality index $V_{.\ell}(Y)$ of the subpopulation ℓ is given by:

$$V_{.\ell}(Y) = \sum_{h=1}^{r} V_{h\ell}(Y) \cdot \frac{n_{h\ell}}{n_{.\ell}} = \sum_{h=1}^{r} \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M_\ell(Y)} \cdot \frac{n_{h\ell}}{n_{.\ell}}, \tag{51}$$

where

$$V_{h\ell}(Y) = \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M_\ell(Y)} \tag{52}$$

is the point inequality index of the subpopulation ℓ and $n_{h\ell}$ is the sum of the weights of the households in the subpopulation ℓ with total income $Y = y_h$. Table 9 shows that the mean value 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 the inequality of the whole population is a little bit greater than the one of the North. The synthetic inequality index $V(Y) = 0.4795$ means that in the whole population, on average, the lower mean is equal to the $(1 - 0.4795) \cdot 100 \simeq 52\%$ of the mean $M(Y)$.

In other words, from $V(Y) = \sum_{h=1}^{7287} \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} \cdot \frac{n_{h.}}{N} = 0.4795$ it derives that:

$$(1 - 0.4795) = \sum_{h=1}^{7287} \frac{\bar{M}_{h.}(Y)}{M(Y)} \cdot \frac{n_{h.}}{N},$$

and consequently we can say that on average (w.r.t. the weights $\frac{n_{h.}}{N}$), the lower mean $\bar{M}_{h.}(Y)$ is the 52% of the mean $M(Y)$.

Figure 3 displays the graphs of the point inequality measures for the whole population, while the graphs of the North, the Center and the South are reported in Figure 4. For the subpopulation ℓ the abscissas and the ordinates are given respectively by

$$\left(p_{h\ell} = \frac{P_{h\ell}}{n_{.\ell}}, \quad V_{h(p_{h\ell})}(Y) = V_{h\ell}(Y) \right),$$

$\forall h = 1, \ldots, r$.

For the whole population the abscissas and the ordinates are given respectively by

$$\left(p_{h.} = \frac{P_{h.}}{N}, \quad V_{h(p_{h.})}(Y) = V_h(Y)\right),$$

$\forall h = 1, 2 \ldots, r$.

TABLE 9. - *Some aggregate characteristics for geographic area*

	Subpopulation			Italy
	North	*Center*	*South*	
$n_{.\ell}$	3971.949	1537.372	2641.679	$8151 = N\,(= W)$
$n_{.\ell}/N$	0.4873	0.1886	0.3241	1.00
Median	27527.57	29824.24	19123.67	24590.10
Mean	33543.17	34000.09	23517.86	30380.22
$V_{.\ell}(Y)$	0.4740	0.4421	0.4695	$0.4795 = V(Y)$

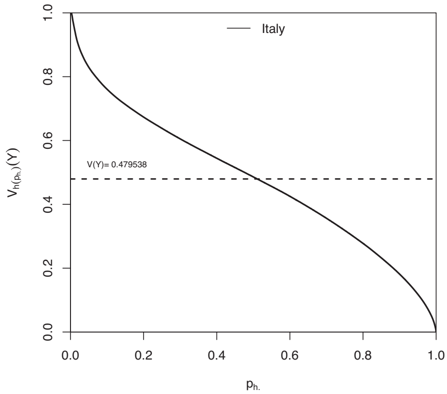

FIGURE 3. - *Graph of the Bonferroni point measure* $V_{h(ph.)}(Y)$

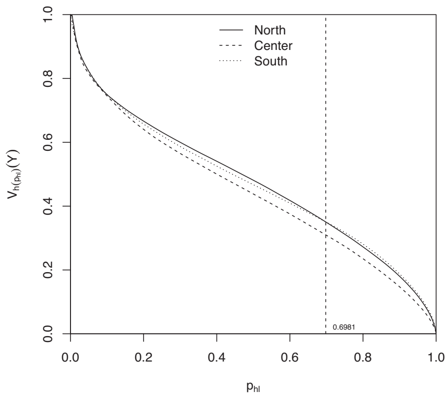

FIGURE 4. - *Graphs of the Bonferroni point measures* $V_{h(ph\ell)\ell}(Y)$, $\ell = 1, 2, 3$

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 $V_{h(p_{h.})}(Y)$ for three values of $p_{h.}$, and the decompositions of the synthetic index $V(Y) = 0.4795$. For $p_{h.}$ we have chosen the following values:

- $p_{h.} = 0.10$; $V_{h(0.10)}(Y) = 0.7607$ compares the income mean of the poorest 10% of households with the income mean $M(Y)$.
- $p_{h.} = 0.50$; $V_{h(0.50)}(Y) = 0.4859$ compares the income mean of the households with $Y \leq Med(Y)$ with $M(Y)$.
- $p_{h.} = 0.95$; $V_{h(0.95)}(Y) = 0.1181$ compares the income mean of the lower group that is the 95% of the whole population with $M(Y)$.

Table 10 reports for these three values of $p_{h.}$ the corresponding values of $h(p_{h.})$, $P_{h(p_{h.})}$, $P_{h(p_{h.})}/N$ and of $y_{h(p_{h.})}$. Note that:

$$h(p_{h.}) = \min h : \frac{P_{h(p_{h.}).}}{N} \geq p_{h.}.$$

TABLE 10. - *Cumulative frequencies and quantiles for some values of $p_{h.} = P(Y \leq y_h)$*

$p_{h.}$	$h(p_{h.})$	$P_{h(p_{h.}).}$	$P_{h(p_{h.}).}/N$	y_h
$p_{h.} = 0.10$	460	815.20	0.10	10600.00
$p_{h.} = 0.50$	3064	4075.65	0.50	24590.10
$p_{h.} = 0.95$	6841	7743.48	0.95	68819.23
$p_{h.} = 1.00$	7287	8151.00	1.00	368689.70

Table 11 reports all the values needed for the decompositions of $V_{h(0.10)}(Y) = 0.7607$. These decompositions are shown in Table 12.

The greatest contribution $V_{h(0.10)\ell g}(Y)$ is $V_{h(0.10)31}(Y)$:

$$V_{h(0.1)31}(Y) = \frac{M_1(Y) - \bar{M}_{h(0.1)3}(Y)}{M(Y)} \cdot \frac{n_{.1}}{N} \cdot p(3|460)$$

$$= \frac{33543.17 - 7310.03}{30380.22} \cdot 0.4873 \cdot 0.5218$$

$$= 0.8635 \cdot 0.4873 \cdot 0.5218 = 0.2196.$$

TABLE 11. - *Means and lower means in the subpopultaion ℓ, frequencies $P_{h\ell}$ and $P_{h.}$, and relative frequencies $\frac{n_{.\ell}}{N}$ and $p(\ell|h)$; $h = 460, y_h = 10600.00$*

$p_{h.} = 0,10$ $h = 460$	North $\ell = 1$	Center $\ell = 2$	South $\ell = 3$	**Italy**
	Frequencies $P_{h\ell}, P_{h.}$ and $n_{.\ell}$			
$Y \leq 10600.00$	275.78	114.01	425.40	815.20
$Y > 10600.00$	3696.16	1423.36	2216.28	7335.80
Total $= n_{.\ell}$	3971.95	1537.37	2641.68	8151
Relative frequencies				
$n_{.\ell}/N$	0.4873	0.1886	0.3241	1.00
$p(\ell\|h)$	0.3383	0.1399	0.5218	1.00
Means				
$M_\ell(Y)$	33543.17	34000.09	23517.86	30380.22
$\bar{M}_{h\ell}(Y)$	7091.45	7554.05	7310.03	7270.21

TABLE 12. - *Decomposition of $V_{h(0.1)}(Y) = V_{460}(Y) = 0.7607$ in $V_{460\ell g}(Y), V_{460\ell.}(Y)$, $V_{460\ell W}(Y)$, $V_{460\ell B}(Y)$, $V_{460.W}(Y)$, $V_{460.B}(Y)$*

$V_{460\ell g}(Y)$	ℓ			
g	1	2	3	
1	0.1435	0.0583	0.2196	
2	0.0565	0.0230	0.0865	
3	0.0593	0.0238	0.0902	
$V_{460\ell.}(Y)$	0.2593	0.1051	0.3963	$0.7607 = V_{460}(Y)$
$V_{460\ell W}(Y)$	0.1435	0.0230	0.0902	$0.2567 = V_{460.W}(Y)$
$V_{460\ell B}(Y)$	0.1158	0.0821	0.3060	$0.5040 = V_{460.B}(Y)$

This result depends from the difference of the lower mean of the South and the mean of the North, and from their relative weights: $p(\ell|h)$ and $\frac{n_{.g}}{N}$. Conversely, for the contribution $V_{(0.10)13}(Y)$ we have:

$$\begin{aligned} V_{h(0.1)13}(Y) &= \frac{M_3(Y) - \bar{M}_{h(0.1)1}(Y)}{M(Y)} \cdot \frac{n_{.3}}{N} \cdot p(1|460) \\ &= \frac{23517.86 - 7091.45}{30380.22} \cdot 0.3241 \cdot 0.3383 \\ &= 0.5407 \cdot 0.3241 \cdot 0.3383 = 0.0593. \end{aligned}$$

In this way the difference between $V_{h(0.10)13}(Y)$ and $V_{h(0.10)31}(Y)$ is explained by the different comparison between means and the different weights.

Let us consider now the decomposition of the point index $V_{h(0.10)}(Y) = 0.7607$ into the three contributions $V_{h(0.10)\ell.}(Y)$ of each macro regions:

$$\begin{aligned}
V_{h(0.1)1.}(Y) &= \frac{M(Y) - \bar{M}_{h(0.1)1.}(Y)}{M(Y)} \cdot p(1|460) \\
&= \frac{30380.22 - 7091.45}{30380.22} \cdot 0.3383 \\
&= 0.7666 \cdot 0.3383 = 0.2593; \\
V_{h(0.1)2.}(Y) &= \frac{M(Y) - \bar{M}_{h(0.1)2.}(Y)}{M(Y)} \cdot p(2|460) \\
&= \frac{30380.22 - 7554.05}{30380.22} \cdot 0.1399 \\
&= 0.7513 \cdot 0.1399 = 0.1051; \\
V_{h(0.1)3.}(Y) &= \frac{M(Y) - \bar{M}_{h(0.1)3.}(Y)}{M(Y)} \cdot p(3|460) \\
&= \frac{30380.22 - 7310.03}{30380.22} \cdot 0.5218 \\
&= 0.7594 \cdot 0.5218 = 0.3963.
\end{aligned}$$

These values show that the relative variations of the lower means of the three macroregions w.r.t. the mean of the whole population are similar, while their relative weights $p(\ell|h)$ are very different. This explain why the greatest contribution to the point index $V_{h(0.10)}(Y) = 0.7607$ comes from the South.

Table 13 reports all the values needed for the decompositions of $V_{h(0.50)}(Y) = 0.4859$ reported in Table 14.

TABLE 13. - *Means $M_\ell(Y)$ and related lower means, frequencies $P_{h\ell}$ and $P_{h.}$, and relative frequencies $\frac{n_{.\ell}}{N}$ and $p(\ell|h)$; $h = 3064$, $y_h = 24590.10$*

$p_{h.} = 0.50$ $h = 3064$	North $\ell = 1$	Center $\ell = 2$	South $\ell = 3$	**Italy**
	Frequencies $P_{h\ell}$, $P_{h.}$ and $n_{.\ell}$			
$Y \leq 24590.10$	1710.11	576.93	1788.61	4075.65
$Y > 24590.10$	2261.8	960.44	853.07	4075.4
Total $= n_{.\ell}$	3971.95	1537.37	2641.68	8151
Relative frequencies				
$n_{.\ell}/N$	0.4873	0.1886	0.3241	1.00
$p(\ell\|h)$	0.4196	0.1416	0.4389	1.00
Means				
$M_\ell(Y)$	33543.17	34000.09	23517.86	30380.22
$\bar{M}_{h\ell}(Y)$	16013.33	16449.92	14972.15	15618.20

TABLE 14. - *Decomposition of* $V_{h(0.5)}(Y) = V_{3064}(Y) = 0.4859$ *in* $V_{3064\ell g}(Y)$, $V_{3064\ell.}(Y)$, $V_{3064\ell W}(Y)$, $V_{3064\ell B}(Y)$, $V_{3064.W}(Y)$, $V_{3064.B}(Y)$

$V_{3064\ell g}(Y)$	ℓ			
g	1	2	3	
1	0.1180	0.0388	0.1307	
2	0.0469	0.0154	0.0518	
3	0.0336	0.0107	0.0400	
$V_{3064\ell.}(Y)$	0.1984	0.0649	0.2226	$0.4859 = V_{3064}(Y)$
$V_{3064\ell W}(Y)$	0.1180	0.0154	0.0400	$0.1734 = V_{3064.W}(Y)$
$V_{3064\ell B}(Y)$	0.0804	0.0495	0.1826	$0.3125 = V_{3064.B}(Y)$

Finally, Table 15 reports all the values needed for the decompositions of $V_{h(0.95)}(Y) = 0.1181$ which are reported in Table 16.

TABLE 15. - *Means* $M_\ell(Y)$ *and related lower means, frequencies* $P_{h\ell}$ *and* $P_{h.}$, *and relative frequencies* $\frac{n_{.\ell}}{N}$ *and* $p(\ell|h)$; $h = 6841$, $y_h = 68819.23$

$p_{h.} = 0.95$ $h = 6841$	North $\ell = 1$	Center $\ell = 2$	South $\ell = 3$	**Italy**
	Frequencies $P_{h\ell}$, $P_{h.}$ and $n_{.\ell}$			
$Y \leq 68819.20$	3710.67	1443.33	2589.48	7743.48
$Y > 68819.20$	261.28	94.04	52.20	407.52
Total $= n_{.\ell}$	3971.95	1537.37	2641.68	8151
Relative frequencies				
$n_{.\ell}/N$	0.4873	0.1886	0.3241	1.00
$p(\ell\|h)$	0.4792	0.1864	0.3344	1.00
Means				
$M_\ell(Y)$	33543.17	34000.09	23517.86	30380.22
$\bar{M}_{h\ell}(Y)$	28856.94	30401.45	21819.47	26791.44

TABLE 16. - *Decomposition of* $V_{h(0.95)}(Y) = V_{6841}(Y) = 0.1181$ *in* $V_{6841\ell g}(Y)$, $V_{6841\ell.}(Y)$, $V_{6841\ell W}(Y)$, $V_{6841\ell B}(Y)$, $V_{6841.W}(Y)$, $V_{6841.B}(Y)$

$V_{6841\ell g}(Y)$	ℓ			
g	1	2	3	
1	0.0360	0.0094	0.0629	
2	0.0153	0.0042	0.0253	
3	-0.0273	-0.0137	0.0061	
$V_{6841\ell.}(Y)$	0.0240	-0.0001	0.0942	$0.1181 = V_{6841}(Y)$
$V_{6841\ell W}(Y)$	0.0360	0.0042	0.0061	$0.0462 = V_{6841.W}(Y)$
$V_{6841\ell B}(Y)$	-0.0120	-0.0043	0.0882	$0.0719 = V_{6841.B}(Y)$

Table 15 shows that the greatest contribution to $V_{6841}(Y)$ is given by $V_{h(0.95)31}(Y) = 0.0629$:

$$V_{h(0.95)31}(Y) = \frac{M_1(Y) - \bar{M}_{h(0.95)3}(Y)}{M(Y)} \cdot \frac{n_{.1}}{n} \cdot p(3|6841)$$

$$= \frac{33543.17 - 21819.47}{30380.22} \cdot 0.4873 \cdot 0.3344$$

$$= 0.3859 \cdot 0.4873 \cdot 0.3344 = 0.0629.$$

About the contribution $V_{h(0.95)13}(Y) = -0.0273$ we have:

$$V_{h(0.95)13}(Y) = \frac{M_3(Y) - \bar{M}_{h(0.95)1}(Y)}{M(Y)} \cdot \frac{n_{.3}}{n} \cdot p(1|6841)$$

$$= \frac{23517.86 - 28856.94}{30380.22} \cdot 0.3241 \cdot 0.4792$$

$$= -0.1757 \cdot 0.3241 \cdot 0.4792 = -0.0273.$$

About the contributions $V_{h(0.95)\ell.}(Y)$ of each subpopulation to the point index $V_{h(0.95)}(Y)$, we have:

$$V_{h(0.95)1.}(Y) = \frac{M(Y) - \bar{M}_{h(0.95)1}(Y)}{M(Y)} \cdot p(1|6841)$$

$$= \frac{30380.22 - 28856.94}{30380.22} \cdot 0.4792$$

$$= 0.0501 \cdot 0.4792 = 0.0240;$$

$$V_{h(0.95)2.}(Y) = \frac{M(Y) - \bar{M}_{h(0.95)2}(Y)}{M(Y)} \cdot p(2|6841)$$

$$= \frac{30380.22 - 30401.45}{30380.22} \cdot 0.1864$$

$$= -0.0007 \cdot 0.1864 = -0.0001;$$

$$V_{h(0.95)3.}(Y) = \frac{M(Y) - \bar{M}_{h(0.95)3}(Y)}{M(Y)} \cdot p(3|6841)$$

$$= \frac{30380.22 - 21819.47}{30380.22} \cdot 0.3344$$

$$= 0.2818 \cdot 0.3344 = 0.0942.$$

From Table 16 we note that:

- the comparison between the income mean of the South $M_3(Y)$ and the income mean of the household with $Y \leq y_{6841} = 68819.23$ in the North produces a negative contribution ($V_{(0.95)13}(Y) = -0.0273$) to the point inequality measure $V_{6841}(Y)$;
- the contribution of the Center to $V_{6841}(Y)$ is ''null''.

From Table 16 we note also that the South has the greatest contribution to the point index $V_{h(0.95)}(Y) = 0.1181$.

Table 17 reports the decompositions of the synthetic index $V(Y) = 0.4795$.

TABLE 17. - *Decomposition of the synthetic index $V(Y) = 0.4795$ in $V_{.\ell g}(Y)$, $V_{.\ell .}(Y)$, $V_{.\ell W}(Y)$, $V_{.\ell B}(Y)$, $V_{..W}(Y)$, $V_{..B}(Y)$*

$V_{.\ell g}(Y)$	ℓ			
g	1	2	3	
1	0.1114	0.0364	0.1366	
2	0.0443	0.0145	0.0541	
3	0.0289	0.0084	0.0449	
$V_{.\ell .}(Y)$	0.1847	0.0593	0.2355	$0.4795 = V(Y)$
$V_{.\ell W}(Y)$	0.1114	0.0145	0.0449	$0.1708 = V_{..W}(Y)$
$V_{.\ell B}(Y)$	0.0733	0.0448	0.1907	$0.3088 = V_{..B}(Y)$

It is useful to remember that the contribution $V_{.\ell g}(Y)$ reported in this table are the weighted means of the corresponding contribution $V_{h\ell g}(Y)$ with weights $n_{h.}/N$:

$$V_{.\ell g}(Y) = \sum_{h=1}^{r} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{N}.$$

This table confirms that the two greatest contributions $V_{.\ell g}(Y)$ are $V_{.31}(Y) = 0.1366$ and $V_{.11}(Y) = 0.1114$.

Finally, Table 18 reports for the three macro regions their:

- relative contributions to the point indexes

$$\nu_{h\ell .}(Y) = \frac{V_{h\ell .}(Y)}{V_h(Y)} = \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y) - \bar{M}_{h.}(Y)} \cdot p(\ell|h) \quad (h = 1, ..., r-1) \tag{53}$$

- relative contributions to the sythetic index

$$\nu_{.\ell .}(Y) = \frac{V_{.\ell .}(Y)}{V(Y)} \tag{54}$$

- relative weights $\frac{n_{.\ell}}{N}$

From Table 18, we note that the South is the region that shows the greatest contribution to the income inequality, while the contrary happens for the Center.

We conclude this section observing that for the whole population, the within component is the $0.1708/0.4795 = 35.62\%$ of the synthetic index.

TABLE 18. - *Relative contributions* $\nu_{460\ell.}(Y)$, $\nu_{3064\ell.}(Y)$, $\nu_{6841\ell.}(Y)$, $\nu_{.\ell.}(Y)$

	ℓ			Tot.
	Nord	Centro	Sud	
$\nu_{460\ell.}(Y)$	0.3409	0.1381	0.5209	1.00
$\nu_{3064\ell.}(Y)$	0.4084	0.1336	0.4581	1.00
$\nu_{6841\ell.}(Y)$	0.2034	-0.0011	0.7977	1.00
$\nu_{.\ell.}(Y)$	0.3851	0.1237	0.4912	1.00
$n_{.\ell}/n$	0.4873	0.1886	0.3241	1.00

6. CONCLUSION AND FINAL REMARKS

This paper, by using the ‘‘two-step’’ approach recently proposed in Zenga (2016) for the decomposition by subpopulations of the Zenga (2007) index, obtains the decomposition by subpopulations of the Bonferroni (1930) index. This ‘‘two-step’’ approach is also used in Zenga, Radaelli and Zenga (2012) and in Zenga (2013) for the decomposition by sources of the Gini (1914), the Bonferroni (1930) and the Zenga (2007) indexes. Moreover, this approach is used in Arcagni (2016) and in Porro and Zenga (2014) for the decompositions by sources and by subpopulations of the Zenga (1984) index, respectively. We remark that the decompositons obtained in this paper are rather different from those proposed by Tarsitano (1990) and by Bárcena-Martin and Silber (2013). Actually, they provide only the decomposition by subpopulations of the Bonferroni synthetic index. More details on these latter decompositions are reported in Valli (2016).

It is worthwhile to point out that the main result of the present proposal is the decomposition of the Bonferroni point measure

$$V_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} \tag{22}$$

into the sum of $k \times k$ contributions $V_{h\ell g}(Y)$ as follows:

$$V_h(Y) = \sum_{\ell=1}^{k} \sum_{g=1}^{k} V_{h\ell g}(Y), \tag{29}$$

where, as defined in (30),

$$V_{h\ell g}(Y) = \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N}$$

is the contribution to $V_h(Y)$ that derives from the comparison of the lower mean

$\bar{M}_{h\ell}(Y)$ of the subpopulation ℓ w.r.t. the mean $M_g(Y)$ of subpopulation g. In (30), $p(\ell|h)$ and $\frac{n_{.g}}{N}$ are the relative frequencies of the subpopulation ℓ in the lower group $\{Y \leq y_h\}$ and of the subpopulation g in the whole population, respectively.

From this $k \times k$ decomposition of $V_h(Y)$, with simple aggregation, the within and the between components and the contribution of each subpopulation to $V_h(Y)$ are obtained.

Formula (30) shows that if $\bar{M}_{h\ell}(Y) > M_g(Y)$ and $p(\ell|h) > 0$, then the corresponding contribution $V_{h\ell g}(Y)$ is negative. This characteristic is illustrated in Tables 7 and 16. Moreover, if $M_g(Y) < M_\ell(Y)$, then $V_{h\ell g}(Y) < 0$ for some $h = 1, \ldots, r$. Note that it is also possible to have negative contributions $V_{.\ell g}(Y)$ to the synthetic Bonferroni index. This case is illustrated by Tables 19 and 20.

TABLE 19. - *Bivariate* 6×3 *distribution*

h	y_h	n_{h1}	n_{h2}	n_{h3}	$n_{h.}$
1	2	0	1	1	2
2	8	0	1	0	1
3	24	0	0	1	1
4	29	1	0	0	1
5	37	2	0	1	3
6	62	2	0	0	2
$n_{.g}$		5	2	3	$10 = N$

TABLE 20. - *Contributions* $V_{.\ell g}(Y)$ *and* $V_{.\ell.}(Y)$ *from the bivariate distribution in Table 19*

$V_{.\ell g}(Y)$		ℓ			
		1	2	3	
g	1	0.0262	0.2553	0.2152	
	2	-0.0521	0.0020	-0.0205	
	3	-0.0410	0.0625	0.0325	
$V_{.\ell.}(Y)$		-0.0669	0.3197	0.2272	$V(Y) = 0.48$

Moreover, it is also possible to have negative values for the contribution

$$V_{h\ell.}(Y) = \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h)$$

of the subpopulation ℓ to the point index $V_h(Y)$. In the application reported in this paper:

- for $h > 7055$, $\bar{M}_{h1}(Y) > M(Y)$, and $V_{h1}(Y) < 0$;
- for $h > 6850$, $\bar{M}_{h2}(Y) > M(Y)$, and $V_{h2}(Y) < 0$.

We point out that it is also possible to have negative contributions $V_{.\ell.}(Y)$ of the subpopulation ℓ to the synthetic index $V(Y)$ as illustrated in Table 20. The fact that some contributions are negative may cause some difficulties for the interpretation.

This cannot happen for the Zenga (2007) index. In fact, by costruction (See Zenga, 2016), the contributions

$$B_{h\ell g}(Y) = \frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot a(g|h) \cdot p(\ell|h) \tag{55}$$

to the Zenga (2007) point inequality index

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

are never negative. In (55) $\overset{+}{M}_{h.}(Y)$ is the mean in the upper group $\{Y > y_h\}$ of the whole population, $\overset{+}{M}_{hg}(Y)$ 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. (2016). On the decomposition by sources of the Zenga 1984 point and synthetic inequality indexes. *Statistical Methods & Applications*, **26(1)**, 113-133.

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

Bárcena-Martin E., Silber J. (2013). On the generalization and decomposition of the Bonferroni index. *Social Choice and Welfare*, **41(4)**, 863-787.

Bonferroni C. E. (1930). *Elementi di statistica generale*. Litografia Felice Gili, Torino.

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

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*, **6(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.

Silber J. (1989). Factor components, population subgroups and the computation of the Gini index of inequality. *The Review of Economics and Statistics*, **71(1)**, 107-115.

Tarsitano A. (1990). The Bonferroni index of inequality. In *Income and Wealth Distribution, Inequality and Poverty*, (pp. 228-242). Springer-Verlag, Berlin.

Valli I. (2016). Scomposizione per sottopopolazioni dell'indice di Bonferroni. Phd Thesis. http://hdl.handle.net/10281/114249.

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

Zenga M. (2007). Inequality curve and inequality index based on the ratios between lower and upper means. *Statistica & Applicazioni*, **5(1)**, 3-27.

Zenga M.M., Radaelli P., Zenga Ma. (2012). Decomposition of Zenga's inequality index by sources. *Statistica & Applicazioni*, **10(1)**, 3-31.

Zenga M. (2013). Decomposition by sources of the Gini, Bonferroni and Zenga 2007 inequality indexes. *Statistica & Applicazioni*, **11(2)**, 133-16.

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