ON THE DECOMPOSITION BY SUB POPULATIONS OF THE POINT $I_h(Y)$ AND SYNTHETIC $I(Y)$ INEQUALITY INDEXES

Michele Zenga

April 13, 2015

Dipartimento di Statistica e Metodi Quantitativi

Università degli Studi di Milano-Bicocca

Abstract

The Radaelli (2008) decomposition by k subpopulations of the Zenga (2007) point $I_h(Y) = \left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right] / \overset{+}{M}_{h.}(Y)$ index is based on the decomposition of the point uniformity measure $U_h(Y) = \bar{M}_{h.}(Y) / \overset{+}{M}_{h.}(Y)$. In this work, we first obtain, by the use of the relation between the mean value M of a mixture and the means M_l of the k subpopulations of the mixture, a $k \times k$ additive decomposition of the difference $\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)$: $\left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right] = \sum_{l=1}^{k} \sum_{g=1}^{k} \left[\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)\right] p(l|h) \cdot a(g|h)$; where $\overset{+}{M}_{hg}(Y)$ and $\bar{M}_{hl}(Y)$ are respectively the upper and the lower means of the subpopulation g and l, and $a(g|h)$ and $p(l|h)$ are their relative frequencies. Then, dividing both sides of the above reported decomposition we obtain a $k \times k$-additive decomposition of $I_h(Y)$. From this latter decomposition, with simple "aggregations" we obtain a k-additive decompositions of $I_h(Y)$, and the decomposition of $I_h(Y)$ into the within and the between components. 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. This application shows that the values of the conditional relative frequencies $a(g|h)$ and $p(l|h)$ help in the interpretation of the 3×3 -contributions

$B_{hlg}(Y) = \left[\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] p(l|h) \cdot a(g|h)$, as well as in the interpretation of the 3 contributions $B_{hl.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(l|h) \cdot$

1 Introduction

The point inequality index $I_h(Y)$, proposed by Zenga (2007), is based on the ratio between the arithmetic mean $\overset{-}{M}_{h.}(Y)$ of the lower group $(Y \leq y_h)$ and the arithmetic mean $\overset{+}{M}_{h.}(Y)$ of the upper group $(Y > y_h)$. The synthetic index $I(Y)$ is obtained by averaging $I_h(Y)$. The curves of $I_h(Y)$, evaluated on several income distributions, are U-shaped Zenga (2007, 2007b), Langel and Tillé (2012), Arcagni (2013), and Arcagni and Porro (2014). Polisicchio (2008) has shown that the truncated Pareto distribution with parameter inequality $\theta = 0.5$ has constant point inequality index. Porro (2011) has obtained the density of random variable with linear point inequality curve. Unnikrishnan Nair. N. et al (2012) have analyzed some property of the $I_h(Y)$ curve. Maffenini and Polisicchio (2014) have analyzed the effects of some transformations on the $I_h(Y)$curve, and they have also shown that the comparison of the empirical $I_h(Y)$ curve with the corresponding uniform inequality curve provides useful informations. Many inferential results on $I(Y)$, in the case of sampling from continuous model have been obtained by Greselin and Pasquazzi (2009), Greselin et al (2009, 2010, 2014). Moreover, in the case of complex sampling, Langel and Tillé (2012) obtained the variance of an estimator of $I(Y)$.

A very important characteristic of an inequality index is its suitability in the decompositions by sub-populations and by sources: Rao (1969), Meran (1975), Shorroks (1980), Lerman and Yitzhaki (1984, 1985), Zenga (1986), Bottiroli Civardi (1987), Zenga (1987), Tarsitano (1989), Deutsch-Silber (1999), Zenga(2001), Radaelli and Zenga (2005), Mussini (2013a, 2013b). The aim of all the above mentioned papers is the decomposition (by sources, by sub-populations) of synthetic inequality indexes: Gini, Bonferroni, Herfindahl,.... On the other hand, in the case of the Zenga (2007) index, the approaches proposed for the decompositions by sub-populations Radaelli (2008, 2010), and by sources Zenga et al.(2012) are such that:

- they obtain, first of all, "additive" decompositions of the point index $I_h(Y)$, and then
- by averaging these decompositions, they obtain the corresponding decompositions of $I(Y)$.

Recently, Zenga (2013) has used this two-step approach for the decomposition by sources of the Gini (1914) and the Bonferroni (1930) indexes too. In that paper it is shown (Lemma 1) that the relative contributions of the component X_j, $Y = \sum X_j$, to the Gini, Bonferroni and Zenga point indexes are equal.

Moreover, by the use of the two-step approach, the decompositions by subpopulations and by sources of the ζ Zenga (1984) index has been obtained respectively by Porro et al (2014) and by Arcagni et al (2014).

The first step of Radaelli's (2008, 2010) approach consists in decomposing first the uniformity point measure $U_h(Y) = 1 - I_h(Y) = \overset{-}{M}_{h.}(Y)/\overset{+}{M}_{h.}(Y)$, and later the inequality point measure $I_h(Y)$, while in this paper we decompose $I_h(Y)$ starting from the decomposition of the difference $\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)$. The paper is organized as follows. In the next section, some definitions and notation in the case of frequency distributions framework are introduced. In particular this section provides: the definitions of the lower and the upper means in the whole population and in the sub-populations, and of the point $I_h(Y)$ and synthetic $I(Y)$ indexes. In section 3 are obtained different additive decompositions for $I_h(Y)$ and $I(Y)$. Section 4 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 5 is devoted to the conclusions and final remarks.

2 Definitions and notation

Let Y denote a non-negative variate, usually income, observed on N units that can be partitioned, according to some relevant characteristic, into k different subpopulations whose numerousness is denoted by $n_{\cdot g}$ $(g = 1, \ldots, k)$. Let $\{0 \leq y_1 < \ldots < y_h < \ldots < y_r\}$ denote the set of the distinct values assumed by the variate Y on all the k subpopulations; it is possible to report the whole distribution as in Table 1: where n_{hg} denotes the frequency of the value y_h in the subpopulation g.

Table 1: Frequency distribution of the whole population partitioned into k subpopulations

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

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

$$\begin{aligned}
P_{h.} = P_{h.}(Y) &= \sum_{t=1}^{h} n_{t.} \quad ,h\text{=1,..., } r && (1)\\
S_{h.}(Y) &= y_h \cdot n_h \ \ , h = 1, \ldots, r && (2)\\
Q_{h.}(Y) = \sum_{t=1}^{h} S_{t.}(Y) &= \sum_{t=1}^{h} y_t \cdot n_{t.},\ h = 1, \ldots, , r && (3)\\
T = Q_{r.}(Y) = \sum_{h=1}^{r} S_{h.}(Y) &= \sum_{h=1}^{r} y_h \cdot n_{h.} > 0, && (4)\\
M = M(Y) &= T/N. && (5)
\end{aligned}$$

For the distribution $\{(y_h, n_{hg}) : h = 1, \ldots, r\}$ of the subpopulation g the analogous of $(1) - (5)$ are:

$$\begin{aligned}
P_{hg} = P_{hg}(Y) &= \sum_{t=1}^{h} n_{tg} \qquad , h = 1, \ldots, r && (6)\\
S_{hg}(Y) &= y_h \cdot n_{hg} \quad , h = 1, \ldots, r && (7)\\
Q_{hg}(Y) = \sum_{t=1}^{h} S_{tg}(Y) &= \sum_{t=1}^{h} y_t \cdot n_{tg} \qquad , h = 1, \ldots, r && (8)\\
T_g = Q_{rg}(Y) &= \sum_{h=1}^{r} y_h \cdot n_{hg} && (9)\\
M_g = M_g(Y) &= T_g/n_{.g} && (10)
\end{aligned}$$

At each y_h the whole population can split into two non overlapping groups: a lower group $\{(y_1, n_{1.}), \ldots, (y_h, n_{h.})\}$including the first $P_{h.}$ units and the corresponding upper group $\{(y_{h+1}, n_{h+1.}), \ldots, (y_r, n_{r.})\}$ including the remaining $N - P_{h.}$ units. Note that for $h = r$ the upper group is empty. Let

$$\bar{M}_{h.}(Y) = \frac{Q_{h.}(Y)}{P_{h.}}, \qquad h = 1, \ldots, r, \tag{11}$$

be the arithmetic mean (lower mean) in the lower group and

$$\overset{+}{M}_{h.}(Y) = \begin{cases} \frac{T - Q_{h.}(Y)}{N - P_{h.}}, & h = 1, \ldots, r-1 \\ \overset{+}{M}_{r-1.}(Y) & = y_r,\ h = r \end{cases}, \tag{12}$$

be the arithmetic mean (upper mean) in the upper group.

In order to measure the inequality between the lower group and the upper group, Zenga (2007) proposed the point index

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)}, \quad h = 1, \ldots, r\,. \tag{13}$$

Table 2: Joint frequencies n_{hg} and cell totals $S_{hg}(Y)$ of $N = 20$ units partitioned into $k = 3$ subpopulations and $r = 6$ distinct values of Y.

		Subpop. g			tot.	Sums $S_{hg}(Y)$ and $S_{h.}(Y)$			
h	y_h	1	2	3	$n_{h.}$	$S_{h1}(Y)$	$S_{h2}(Y)$	$S_{h3}(Y)$	$S_{h.}(Y)$
1	0	1	1	0	2	0	0	0	0
2	2	1	0	1	2	2	0	2	4
3	5	2	1	2	5	10	5	10	25
4	10	0	2	4	6	0	20	40	60
5	20.5	1	1	2	4	20.5	20.5	41	82
6	29	0	0	1	1	0	0	29	29
total $n_{.g}$		5	5	10	20	32.5	45.5	122	200

The synthetic Zenga's inequality measure $I(Y)$ is fournished by:

$$I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N} . \tag{14}$$

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

$$\begin{cases} y_{o(g)}, \text{ where } o(g) & = \min h : n_{hg} > 0 \\ y_{u(g)}, \text{ where } u(g) & = \max h : n_{hg} > 0, \end{cases} \tag{15}$$

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

$$\bar{M}_{hg}(Y) = \begin{cases} y_{o(g)} & \text{for } h < o(g) \\ Q_{hg}(Y)/P_{hg} & \text{for } h \geq o(g) \end{cases} ; \tag{16}$$

$$\overset{+}{M}_{hg}(Y) = \begin{cases} \frac{T_g(Y) - Q_{hg}(Y)}{n_{.g} - P_{hg}} & \text{for } h < u(g) \\ y_{u(g)} & \text{for } h \geq u(g) . \end{cases} \tag{17}$$

2.1 Example

The results of this paper are illustrated by a frequency distribution with: $N = 20$ units, $k = 3$ subpopulations , $r = 6$ distinct values of total income Y . Now, we illustrate the definitions and notation introduced in sec.2.

3 Decomposition by subpopulations of the point $I_h(Y)$ and the synthetic $I(Y)$ inequality indexes

First of all, in this section we decompose the difference $\left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right]$.
By the relations (11) and (12), for $h = 1, \ldots, r-1$, we have:

Table 3: Lower-group frequencies P_{hg} and $P_{h.}$, and upper-group frequencies $(n_{.g} - P_{hg})$ and $(N - P_{h.})$.

h	y_h	P_{h1}	P_{h2}	P_{h3}	$P_{h.}$	$n_{.1} - P_{h1}$	$n_{.2} - P_{h2}$	$n_{.3} - P_{h3}$	$N - P_{h.}$
1	0	1	1	0	2	4	4	10	18
2	2	2	1	1	4	3	4	9	16
3	5	4	2	3	9	1	3	7	11
4	10	4	4	7	15	1	1	3	5
5	20.5	5	5	9	19	0	0	1	1
6	29	5	5	10	20	0	0	0	0

Table 4: Lower-group incomes $Q_{hg}(Y)$ and $Q_{h.}(Y)$, and upper-group incomes $(T_g(Y) - Q_{hg}(Y))$ and $(T(Y) - Q_{h.}(Y))$.

h	y_h	$Q_{h1}(Y)$	$Q_{h2}(Y)$	$Q_{h3}(Y)$	$Q_{h.}(Y)$
1	0	0	0	0	0
2	2	2	0	2	4
3	5	12	5	12	29
4	10	12	25	52	89
5	20.5	32.5	45.5	93	171
6	29	32.5	45.5	122	200

h	y_h	$T_1(Y) - Q_{h1}(Y)$	$T_2(Y)$-$Q_{h2}(Y)$	$T_3(Y)$-$Q_{h3}(Y)$	$T(Y)$-$Q_{h.}(Y)$
1	0	32.5	45.5	122	200
2	2	30.5	45.5	120	196
3	5	20.5	40.5	110	171
4	10	20.5	20.5	70	111
5	20.5	0	0	29	29
6	29	0	0	0	0

Table 5: Lower and upper means of the subpopulations and of the whole population.

		Subpopulations			
$o(g)$		1	2	3	
h	y_h	$\overset{-}{M}_{h1}(Y)$	$\overset{-}{M}_{h2}(Y)$	$\overset{-}{M}_{h3}(Y)$	$\overset{-}{M}_{h.}(Y)$
1	0	0	0	2	0
2	2	1	0	2	1
3	5	3	2.5	4	3.22
4	10	3	6.25	7.42	5.93
5	20.5	6.5	9.1	10.33	9
6	29	6.5	9.1	12.2	10
		subpopulations			
u(g)		5	5	6	
h	y_h	$\overset{+}{M}_{h1}(Y)$	$\overset{+}{M}_{h2}(Y)$	$\overset{+}{M}_{h3}(Y)$	$\overset{+}{M}_{h.}(Y)$
1	0	8.125	11.375	12.2	11.11
2	2	10.16	11.375	13.3	12.25
3	5	20.5	13.5	15.71	15.54
4	10	20.5	20.5	23.3	22.2
5	20.5	20.5	20.5	29	29
6	29	20.5	20.5	29	29

Table 6: Points $I_h(Y)$ and synthetic $I(Y)$ inequality indexes.

h	$I_h(Y)$	$\frac{n_{h.}}{N}$	$I_h(Y)\cdot\frac{n_{h.}}{N}$
1	1.00	0.1	0.1
2	0.9184	0.1	0.0918
3	0.7927	0.25	0.1982
4	0.7327	0.3	0.2198
5	0.6897	0.2	0.1379
6	0.6552	0.05	0.0328
		1.0	$I(Y)$=0.7805

$$\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y) = \frac{T - Q_{h.}(Y)}{N - P_{h.}} - \frac{Q_{h.}(Y)}{P_{h.}}$$
$$= \frac{(T - Q_{h.}(Y)) \cdot P_{h.} - Q_{h.}(Y) \cdot (N - P_{h.})}{(N - P_{h.}) \cdot P_{h.}}.$$

Now, by the the relations $T = \sum_{g=1}^{k} T_g(Y)$, $Q_{h.}(Y) = \sum_{g=1}^{k} Q_{hg}(Y)$, $P_{h.} = \sum_{g=1}^{k} P_{hg}$, and $N = \sum_{g=1}^{k} n_{.g}$, the numerator of the latter expression can be written as:

$$\sum_{g=1}^{k} (T_g(Y) - Q_{hg}(Y)) \cdot \sum_{l=1}^{k} P_{hl} - \sum_{l=1}^{k} Q_{hl}(Y) \cdot \sum_{g=1}^{k} (n_{.g} - P_{hg})$$
$$= \sum_{g=1}^{k} \overset{+}{M}_{hg}(Y) \cdot (n_{.g} - P_{hg}) \cdot \sum_{l=1}^{k} P_{hl} - \sum_{l=1}^{k} \bar{M}_{hl} \cdot P_{hl} \cdot \sum_{g=1}^{k} (n_{.g} - P_{hg})$$
$$= \sum_{l=1}^{k} \sum_{g=1}^{k} \overset{+}{M}_{hg}(Y) \cdot (n_{.g} - P_{hg}) P_{hl} - \sum_{l=1}^{k} \sum_{g=1}^{k} \overset{-}{M}_{hl}(Y) \cdot (n_{.g} - P_{hg}) P_{hl}.$$

Thus, for $h = 1, \ldots, r-1$:

$$\left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right] = \sum_{l=1}^{k} \sum_{g=1}^{k} \left(\overset{+}{M}_{hg}(Y) - \overset{-}{M}_{hl}(Y)\right) \frac{P_{hl}}{P_{h.}} \cdot \frac{n_{.g} - P_{hg}}{N - P_{h.}}.$$

For $h = r$,

$$\overset{+}{M}_{r.}(Y) = y_r = \sum_{g=1}^{k} y_r \frac{n_{rg}}{n_{r.}} = \sum_{g=1}^{k} \overset{+}{M}_{rg}(Y) \cdot \frac{n_{rg}}{n_{r.}}.$$

.

In conclusion, for each $h = 1, \ldots, r$:

$$\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y) = \sum_{l=1}^{k} \sum_{g=1}^{k} \left[\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)\right] p(l|h) \cdot \mathrm{a}(g|h), \tag{18}$$

where, the relative frequency $p(g/h)$ is given by

$$p(g/h) = \frac{P_{hg}}{P_{h.}} \tag{19}$$

and the relative frequency $a(g/h)$ is given by

$$a(g/h) = \begin{cases} \frac{n_{.g} - P_{hg}}{N - P_{h.}}, \text{ for} & h = 1, \ldots, r-1 \\ \frac{n_{rg}}{n_{r.}} & \text{for } h = r \quad . \end{cases} \tag{20}$$

Note that: $\sum_{l=1}^{k} p(l|h) = \sum_{g=1}^{k} a(g|h) = \sum_{l=1}^{k}\sum_{g=1}^{k} p(l|h) \cdot a(g|h) = 1$.

In conclusion, the difference $\left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right]$ is the weighted mean of the $k \cdot k$ differences $\left[\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)\right]$ with weights $p(l|h) \cdot \mathrm{a}(g|h)$.

The decomposition (18) can also be obtained by the use of the following relations:

$$\overset{+}{M}_{h.}(Y) = \sum_{g=1}^{k} \overset{+}{M}_{hg}(Y) \cdot a(g/h), \ h = 1, \ldots, r; \tag{21}$$

$$\bar{M}_{h.}(Y) = \sum_{g=1}^{k} \bar{M}_{hg}(Y) \cdot p(g|h), \ h = 1, \ldots, r. \tag{22}$$

It is useful to remark that the expression (18) has the structure of the Gini mean difference Δ_{hgl} between the two distributions:

$$\left\{\left[\overset{+}{M}_{hg}(Y), a(g|h)\right] : g = 1, \ldots, k\right\} \text{ and } \left\{\left[\bar{M}_{hl}(Y), p(l|h)\right] : l = 1, \ldots, k\right\}.$$

Now, to obtain the first decomposition by subpopulations of $I_h(Y)$, we divide both sides of (18) by $\overset{+}{M}_{h.}(Y)$. Thus:

$$I_h(Y) = \sum_{l=1}^{k}\sum_{g=1}^{k} B_{hlg}(Y) = B_{h..}(Y), \tag{23}$$

where:

$$B_{hlg}(Y) = \left[\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] p(l|h) \cdot a(g|h) \tag{24}$$

is the contribution to the point index $I_h(Y)$ that derives from the comparison of the lower mean $\bar{M}_{hl}(Y)$ with the upper mean $\overset{+}{M}_{hg}(Y)$.

Obviously, $I_h(Y)$ may be interpreted as the weighted mean of the k^2 "relative differences" $\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}$.

Now let,

$$B_{hl.}(Y) = \sum_{g=1}^{k} B_{hlg}(Y) = \sum_{g=1}^{k}\left[\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] p(l|h) \cdot a(g|h).$$

Then, after some steps we obtain:

$$B_{hl.}(Y)=\left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]\cdot p\left(l|h\right)\cdot \tag{25}$$

In other words, $B_{hl.}(Y)$ is equal to the product of the relative variation of $\bar{M}_{hl}(Y)$ w.r.t $\overset{+}{M}_{h.}(Y)$ and the relative frequency $p\left(l/h\right)=\frac{P_{hl}}{P_{h.}}$. Thus, $B_{hl.}(Y)$ can be interpreted as the contribution of the subpopulation l to the point inequality index $I_h(Y)$.

Finally, from (23) and (25) we obtain the following decomposition for $I_h(Y)$:

$$I_h(Y)=B_{h..}(Y)=\sum_{l=1}^{k}B_{hl.}(Y)=\sum_{l=1}^{k}\left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]\cdot p\left(l|h\right)\cdot \tag{26}$$

Formula (26) shows that, the point index $I_h(Y)$ is the weighted mean of the k relative variations $\left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]$ with weights $p\left(l/h\right)=\frac{P_{hl}}{P_{h.}}$.

Finally, $B_{hl.}(Y)$ can be split into a within and a between component.

$$\begin{aligned}
B_{hl.}(Y) &= \sum_{g=1}^{k}B_{hlg}(Y)=B_{hll}(Y)+\textstyle\sum_{(g:g\neq l)}B_{hlg}(Y)\\
&= p(l\mid h)\left[\frac{\overset{+}{M}_{hl}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}a\left(l|h\right)+\textstyle\sum_{(g:g\neq l)}\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}a\left(g|h\right)\right]
\end{aligned}$$

Note that, $\sum_{(g:g\neq l)}^{k}a\left(g|h\right)=\left(1-a(l\mid h)\right)$. Thus,

$$B_{hl.}(Y)=p(l\mid h)\left[\frac{\overset{+}{M}_{hl}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]a\left(l|h\right)+$$

$$+p(l|h)\left[\sum_{(g:g\neq l)}\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\cdot\frac{a\left(g|h\right)}{\left(1-a(l\mid h)\right)}\right]\left(1-a(l|h)\right) \tag{27}$$

The comparison of (25) with (27) shows that the relative variation of $\bar{M}_{hl}(Y)$ w.r.t $\overset{+}{M}_{h.}(Y)$ is the weighted mean of:

- the ratio $\left[\frac{\overset{+}{M}_{hl}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]$ with weight $a(l\mid h)$, and

- the weithed mean $\left[\sum_{(g:g\neq l)}\left(\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right)\cdot\frac{a(g|h)}{(1-a(l|h))}\right]$ with weight $(1-a(l|h))$.

In the expression of B_{hll} the ratio $\frac{\overset{+}{M}_{hl}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}$ derives from the comparison of the lower mean and the upper mean of the same subpopulation l. Thus,

$$B_{hll}(Y) = p(l|h)\cdot\left[\frac{\overset{+}{M}_{hl}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]\cdot a(l|h) = B_{hlW}(Y) \tag{28}$$

can be interpreted as the within part of the contribution $B_{hl.}(Y)$. Viceversa, in the expression of the weighted mean $\sum_{(g:g\neq l)}\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\cdot\frac{a(g|h)}{(1-a(l|h))}$, the ratios $\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}$ derive by comparing lower means and upper means of different subpopulations. Consequently,

$$\sum_{(g:g\neq l)} B_{hlg}(Y) = p(l\mid h)\sum_{(g:g\neq l)}\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}a(g|h) = B_{hlB}(Y) \tag{29}$$

can be interpreted as the between part of $B_{hl.}(Y)$.
In conclusion,

$$B_{hl.}(Y) = B_{hlW}(Y) + B_{hlB}(Y). \tag{30}$$

From (26) and (30) we obtain:

$$\begin{aligned} I_h(Y) = \sum_{l=1}^{k} B_{hl.}(Y) &= \sum_{l=1}^{k} B_{hlW}(Y) + \sum_{l=1}^{k} B_{hlB}(Y) \\ &= B_{h.W}(Y) + B_{h.B}(Y), \end{aligned} \tag{31}$$

where:

$$B_{h.W}(Y) = \sum_{l=1}^{k} B_{hlW}(Y) \tag{32}$$

can be interpreted as the within contribution of all the subpopulations to $I_h(Y)$, and

$$B_{h.B}(Y) = \sum_{l=1}^{k} B_{hlB}(Y) = I_h(Y) - B_{h.W}(Y) \tag{33}$$

can be interpreted as the between contribution of all the subpopulations to $I_h(Y)$.

From (14) and the decompositions (23), (26), and (31) we obtain, after some steps, the following decompositions for $I(Y)$.

$$\begin{aligned} I(Y) &= \sum_{h=1}^{r}\sum_{l=1}^{k}\sum_{g=1}^{k} B_{hlg}(Y)\frac{n_{h.}}{N} = \sum_{l=1}^{k}\sum_{g=1}^{k}\sum_{h=1}^{r} B_{hlg}(Y)\frac{n_{h.}}{N} = \\ &= \sum_{l=1}^{k}\sum_{g=1}^{k} B_{.lg}(Y); \\ &= \sum_{l=1}^{k}\left\{B_{.ll}(Y) + \sum_{(g:g\neq l)} B_{.lg}(Y)\right\} = \sum_{l=1}^{k} B_{.l.}(Y) \\ &= \sum_{l=1}^{k}\{B_{.lW}(Y) + B_{.lB}(Y)\}; \\ &= B_{..W}(Y) + B_{..B}(Y) = B_{...}(Y). \end{aligned} \tag{34}$$

In these latter decompositions:

$$B_{.lg} = \sum_{h=1}^{r} B_{hlg}(Y)\frac{n_{h.}}{N} \tag{35}$$

is the weighted mean of the r contributions $B_{hlg}(Y)$ with weights $n_{h.}/N$;

$$B_{.l.}(Y) = \sum_{g=1}^{k} B_{.lg}(Y), \tag{36}$$

is the contribution of the subpopulation l, to the synthetic index $I(Y)$;

$$B_{.lW}(Y) = B_{.ll}(Y) \tag{37}$$

and

$$B_{.lB}(Y) = \sum_{(g:g\neq l)} B_{.lg}(Y) \tag{38}$$

can be interpreted respectively as the within and the beetween contribution of subpopulation l to $I(Y)$;

$$B_{..W}(Y) = \sum_{l=1}^{k} B_{.lW}(Y), \tag{39}$$

and

Table 8: Calculus of the contributions $B_{2lg}(Y) = \left[\frac{\overset{+}{M}_{2g}(Y)-\bar{M}_{2l}(Y)}{\overset{+}{M}_{2.}(Y)}\right] p(l|2) \cdot a(g|2)$

			l			$a(g\vert 2)$
			1	2	3	
$\overset{+}{M}_{2.}(Y) = 12.25$			$\bar{M}_{21}(Y) = 1$	$\bar{M}_{22}(Y) = 0$	$\bar{M}_{23}(Y) = 2$	
g	1	$\overset{+}{M}_{21}(Y)$ $=10.16$	$[0.7479]\cdot 0.0937$ $= 0.07002$	$[0.8295]\cdot 0.0469$ $= 0.03888$	$[0.6662]\cdot 0.0468$ $= 0.03123$	0.1875
	2	$\overset{+}{M}_{22}(Y)$ $=11.375$	$[0.8469]\cdot 0.125$ $= 0.1059$	$[0.9286]\cdot 0.0625$ $= 0.058$	$[0.7653]\cdot 0.0625$ $= 0.0478$	0.25
	3	$\overset{+}{M}_{23}(Y)$ $= 13.33$	$[1.0068]\cdot 0.2813$ $= 0.2832$	$[1.0884]\cdot 0.1406$ $= 0.1531$	$[0.9252]\cdot 0.1406$ $= 0.1301$	0.5625
$p(l\vert 2)$			0.5	0.25	0.25	1.000

$$B_{..B}(Y) = \sum_{l=1}^{k} B_{.lB}(Y) \tag{40}$$

can be interpreted respectively as the within and the between components of the synthetic index $I(Y)$.

3.1 Numerical illustration of the decomposition by subpopulations of the point $I_h(Y)$ and the synthetic $I(Y)$ inequality indexes

In this section, for the example introduced in (2.1), we show the decompositions by subpopulatios of $I_2(Y) = 0.9184$ and of $I(Y) = 0.7805$.

Table 7: Relative frequencies:$p(l|h) = \frac{P_{hl}}{P_{h.}} \quad \forall(h)$; $a(g|h)=\frac{n_{.g}-P_{hg}}{N-P_{h.}}$ for $h = 1,\ldots,r-1$, and $a(g|h)=\frac{n_{rg}}{n_{r.}}$, for $h = r$

	l				g			
	1	2	3	tot	1	2	3	tot
h	$p(1\vert h)$	$p(2\vert h)$	$p(3\vert h)$	1	$a(1\vert h)$	$a(2\vert h)$	$a(3\vert h)$	1
1	0.5	0.5	0.0	1	0.222	0.222	0.5555	1
2	0.5	0.25	0.25	1	0.1875	0.25	0.5625	1
3	0.444	0.222	0.333	1	0.0909	0.2727	0.6363	1
4	0.266	0.266	0.466	1	0.2	0.2	0.6	1
5	0.2631	0.2631	0.4737	1	0.0	0.0	1	1
6	0.25	0.25	0.5	1	0.0	0.0	1	1

Table 9: Decompositions of $I_2(Y) = 0.9184$ into the contributions: $B_{2lg}(Y)$; $B_{2l.}(Y)$;$B_{2lW}(Y)$, $B_{2lB}(Y)$;$B_{2.W}(Y)$,$B_{2.B}(Y)$.

$B_{2lg}(Y)$		l			
		1	2	3	
g	1	0.070153	0.038903	0.03125	
	2	0.105867	0.058035	0.047831	
	3	0.283163	0.15306	0.130102	
$B_{2l.}(Y)$		0.459183	0.25	0.209183	$0.9184 = I_2(Y)$
$B_{2lW}(Y)$		0.0702	0.0580	0.1308	$0.2582 = B_{2.W}(Y)$
$B_{2lB}(Y)$		0.38903	0.19196	0.07908	$0.6601 = B_{2.B}(Y)$

Table 10: Decompositions of $I(Y) = 0.7852$ into the contributions:$B_{.lg}(Y)$; $B_{.l.}(Y)$;$B_{.lW}(Y)$, $B_{.lB}(Y)$;$B_{..W}(Y)$,$B_{..B}(Y)$.

$B_{.lg}(Y)$		l			
		1	2	3	
g	1	0.0391239	0.0281335	0.0276524	
	2	0.0550422	0.0381701	0.0351485	
	3	0.2111432	0.1574888	0.1886142	
$B_{.l.}(Y)$		0.305309	0.223792	0.251415	$0.7805 = I(Y)$
$B_{.lW}(Y)$		0.0391239	0.0381701	0.1886142	$0.265908 = B_{..W}(Y)$
$B_{.lB}(Y)$		0.266185	0.185622	0.062801	$0.514608 = B_{..B}(Y)$

From tables 9 we may obtain many important informations. For example we note that the contribution of the subpopulation 1 to $I_2(Y)$ is $B_{21.}(Y) = 0.45918$. This value is equal to, see formula (25), the product of :

- the relative variation $\frac{\overset{+}{M}_{2.}(Y) - \bar{M}_{21}(Y)}{\overset{+}{M}_{2.}(Y)} = \frac{12.25-1}{12.25} = 0.918367$ and
- the relative frequency $p(1|2) = \frac{P_{21}}{P_{2.}} = \frac{2}{4} = 0.5$.

4 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 $w_i > 0$ ($i = 1, 2, \ldots, 8151; W = \sum w_i = 8151$) 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). We remark that the frequency distribution of the total income Y has $r = 7287$ different values.

4.1 Aggregate characteristic in three Italian macro-regions

Table 11 reports for the total income Y of each geographic area: the arithmetic mean, the median, and the synthetic index $I_{.l}(Y)$, the sum of the weights $W_{.l}$, and the relative weights $W_{.l}/W$. The synthetic inequality index of the subpopulation l is given by:

$$I_{.l}(Y) = \sum_{h=1}^{r} I_{hl}(Y) \cdot \frac{W_{hl}}{W_{.l}} = \sum_{h=1}^{r} \frac{\overset{+}{M}_{hl}(Y) - \overset{-}{M}_{hl}(Y)}{\overset{+}{M}_{hl}(Y)} \cdot \frac{W_{hl}}{W_{.l}},$$

where,

$$I_{hl}(Y) = \frac{\overset{+}{M}_{hl}(Y) - \overset{-}{M}_{hl}(Y)}{\overset{+}{M}_{hl}(Y)}$$

is the point inequality index of the subpopulation l and W_{hl} is the sum of the weights of the households of subpopulation l with total income $Y = y_h$.

Table 11 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 $I(Y)$ =0.70014 means that in the whole population, on average, the lower mean is equal to the (1-0.70014)·100 $\simeq$ 30% of the upper mean.

Figure 1 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

$$p_{hl} = \frac{P_{hl}}{W_{.l}} \text{ and } I_{(p_{hl})l}(Y) = I_{hl}(Y), \forall h = 1, \ldots, r, \text{where } P_{hl} = \sum_{t=1}^{h} W_{tl},$$

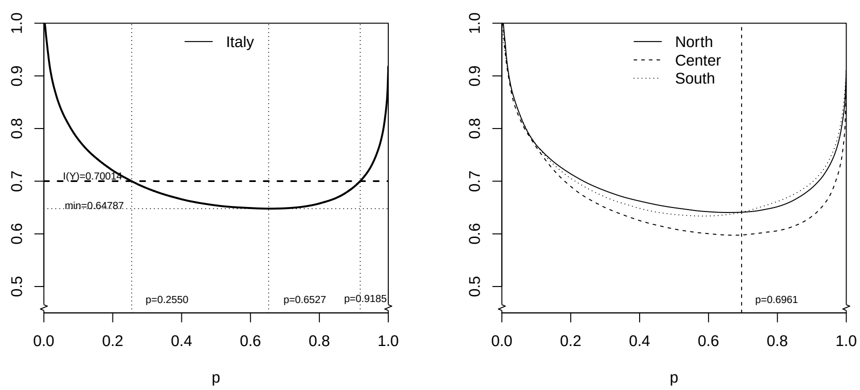
while for the whole population the abscissas and the ordinates are given respectively by

$$p_{h.} = \frac{P_{h.}}{W} \text{ and } I_{(p_{h.})}(Y) = I_h(Y), \forall h = 1, \ldots, r \text{ where } P_{h.} = \sum_{l=1}^{3} P_{hl}.$$

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
$I_{.l}(Y)$	0.6949	0.6592	0.6919	$0.70014 = I(Y)$
$W.l$	3971.949	1537.372	2641.679	$8151 = W$
$W.l/\mathrm{W}$	0.48729	0.18861	0.32409	1.00000

Figure 1: Graphs of the point measure for geographic areas



4.2 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 $I_{h(p)}(Y) = I_{(p)}(Y)$ for three values of p, and the decompositions of the synthetic index $I(Y) = 0.70014$. For p we have chosen the following values:

- $p = 0.10$, because $I_{(0.10)}(Y) = 0.7793$ compares the income mean of the poorest 10% households with the income mean of the other 90% households;

- $p = 0.50$, because $I_{(0.50)}(Y) = 0.654$ compares the income mean of the households with $Y \leqq Median(Y)$ with the mean income of the households with $Y > Median(Y)$;

- $p = 0.95$, because $I_{(0.95)}(Y) = 0.7282$ compares the income mean of the lower group that is the 95% of the whole population with the income mean of the reachest 5% of the households.

Table 13: Upper and lower means in the geographic areas: $p = 0.10$; $h = 460$; $y_h = 10600$

$p = 0.10; h = 460$	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	7335.80
Total$=W_{.l}$	3971.95	1537.37	2641.68	8151
Conditional relative frequencies				
$p\,(l\|h)$	0.3383	0.1399	0.5218	1.0000
$a\,(g\|h)$	0.50385	0.19403	0.3021	1.0000
Upper and Lower means				
$\overset{-}{M}_{hl}\,(Y)$	7091.45	7554.05	7310.03	7270.21
$\overset{+}{M}_{hl}\,(Y)$	35516.80	36118.46	26628.84	32948.33

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

Table 12: Cumulative frequency and quantile values for three values of p of the total income Y for the whole country

p	$h\,(p)$	$P_{h(p).}$	$P_{h(p).}$/W	$y_{h(p)}$
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 $I_{(0.10)}\,(Y)$=0.7793. These decompositions are shown in Table 14.

Table 14: Decompositions of $I_{(0.1)}\,(Y) = 0.7793$ into the contributions: $B_{(0.1)lg}(Y)$; $B_{(0.1)l.}(Y)$;$B_{(0.1)lW}\,(Y)$, $B_{(0.1)lB}\,(Y)$;$B_{(0.1).W}\,(Y)$,$B_{(0.1).B}\,(Y)$.

$B_{(0.1)lg}(Y)$		l			
		1	2	3	
g	1	0.1471	0.0596	0.2251	
	2	0.0578	0.0235	0.0885	
	3	0.0606	0.0245	0.0924	
$B_{(0.1)l.}(Y)$		0.2655	0.1078	0.4061	0.7793= $I_{(0.1)}\,(Y)$
$B_{(0.1)lW}(Y)$		0.1471	0.0235	0.0924	0.2630=$B_{(0.1).W}\,(Y)$
$B_{(0.1)lB}\,(Y)$		0.1184	0.0843	0.3137	0.5164=$B_{(0.1).B}\,(Y)$

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

$$= \left[\frac{\overset{+}{M}_{h1}(Y)-\bar{M}_{h3}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p\left(3|h\right) \cdot a\left(1|h\right) = [0.8569] \cdot 0.5218 \cdot 0.50385 = 0.22501.$$

This result depends from the difference of the lower mean of the South and the upper mean of the North, and from their relative weights :$p\left(3|h\right)$ and $a\left(1|h\right)$. Let us consider now the contribution

$$B_{(0.1)13}(Y) = \left[\frac{\overset{+}{M}_{h3}(Y)-\bar{M}_{h1}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p\left(1|h\right) \cdot a\left(3|h\right) = [0.5930] \cdot 0.3383 \cdot 0.3021 = 0.0606.$$

In this latter case the difference between the upper mean and the lower mean 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 the point index $I_{(0.10)}\left(Y\right) = 0.7793$ into the three contribution $B_{(0.1)l.}(Y)$ of each macro region:

$$B_{(0.1)1.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{h1}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p\left(1|h\right) = 0.7848 \cdot 0.3383 = 0.2655;$$

$$B_{(0.1)2.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{h2}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p\left(2|h\right) = 0.7707 \cdot 0.1399 = 0.1078;$$

$$B_{(0.1)3.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{h3}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p\left(3|h\right) = 0.7781 \cdot 0.5218 = 0.4061.$$

These values show that the relative variations of the lower means of the three macro-regions w.r.t the upper mean of the whole population are similar, while their relative weights $p\left(l|h\right)$ 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 $Y \leq y_{h(0.10)} = 10600$ Euro is the 52.18% of the er" of the corresponding households of the whole lower group. This explains why the greatest contribution to the point index $I_{(0.10)}\left(Y\right) = 0.7793$ 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 necessary for the four decompositions of $I_{(0.5)}\left(Y\right) = 0.6540$ which are reported in Table 16

.

Table 15: Upper and lower means in the geographic areas: $p = 0.50$; $h = 3064$; $y_h = 24590.1$

$p = 0.50; h = 3064$	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$=W_{.l}$	3971.95	1537.37	2641.68	8151
Conditional relative frequencies				
$p(l\|h)$	0.3383	0.1399	0.5218	1.0000
$a(g\|h)$	0.50385	0.19403	0.3021	1.0000
Upper and Lower means				
$\overset{-}{M}_{hl}(Y)$	160413.3	16449.9	14972.1	15618.2
$\overset{+}{M}_{hl}(Y)$	46796.9	44542.29	41435.5	45143.3

Table 16: Decompositions of $I_{(0.5)}(Y) = 0.6540$ into the contributions: $B_{(0.5)lg}(Y)$; $B_{(0.5)l.}(Y)$;$B_{(0.5)lW}(Y)$, $B_{(0.5)lB}(Y)$;$B_{(0.5).W}(Y)$,$B_{(0.5).B}(Y)$.

$B_{(0.5)lg}(Y)$		l			
		1	2	3	
g	1	0.1588	0.0528	0.1717	
	2	0.0625	0.0207	0.0677	
	3	0.0495	0.0164	0.0538	
$B_{(0.5)l.}(Y)$		0.2707	0.0900	0.2933	0.6540$= I_{(0.5)}(Y)$
$B_{(0.5)lW}(Y)$		0.1588	0.0207	0.0538	0.2335 $=B_{(0.5).W}(Y)$
$B_{(0.5)lB}(Y)$		0.1120	0.0692	0.2394	0.4206$=B_{(0.5).B}(Y)$

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

$$B_{(0.5)1.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h1}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(1|h) = 0.6453 \cdot 0.4196 = 0.2707;$$

$$B_{(0.5)2.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h2}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(2|h) = 0.6356 \cdot 0.1416 = 0.0900;$$

$$B_{(0.5)3.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h3}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(3|h) = 0.6683 \cdot 0.4389 = 0.2933.$$

Finally Table 17 reports all the values necessary for the decompositions of $I_{(0.95)}(Y) = 0.7282$ which are reported in Table 18.

Table 17: Upper and lower means in the geographic areas: $p = 0.95$; $h = 6841$;$y_h = 68819.2$

$p = 0.95; h = 6841$	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$=W_{.l}$	3971.95	1537.37	2641.68	8151
Conditional relative frequencies				
$p(l\|h)$	0.4792	0.1864	0.3344	1.0000
$a(g\|h)$	0.6411	0.2308	0.1281	1.0000
Upper and Lower means				
$\overset{-}{M}_{hl}(Y)$	28856.94	20401.45	21819.47	26791.44
$\overset{+}{M}_{hl}(Y)$	100096.67	89231.78	107775.02	98572.91

Table 18: Decompositions of $I_{(0.95)}(Y) = 0.7282$ into the contributions $B_{(0.95)lg}(Y)$; $B_{(0.95)l.}(Y)$; $B_{(0.95)lW}(Y)$, $B_{(0.95)lB}(Y)$; $B_{(0.95).W}(Y)$, $B_{(0.95).B}(Y)$.

$B_{(0.95)lg}(Y)$		l = 1	l = 2	l = 3	
g	1	0.2220	0.0845	0.1703	
	2	0.0677	0.0257	0.0528	
	3	0.0491	0.0187	0.0373	
$B_{(0.95)l.}(Y)$		0.3389	0.1289	0.2604	$0.7282= I_{(0.95)}(Y)$
$B_{(0.95)lW}(Y)$		0.2220	0.0257	0.0373	$0.2850=B_{(0.95).W}(Y)$
$B_{(0.95)lB}(Y)$		0.1169	0.1032	0.2231	$0.4432=B_{(0.95).B}(Y)$

Table 18 shows that the North has the greatest contribution to the the point index $I_{(0.95)}(Y) = 0.7282$, although the relative variation of the lower mean of the South w.r.t the upper mean of the whole population is greater than the one of the North. This happens because the relative weight of the South $(0.3344)$is smaller than the one of the North (0.4792):

$$B_{(0.95)1.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{h1}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(1|h) = 0.7072 \cdot 0.4792 = 0.3389;$$

$$B_{(0.95)2.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{h2}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(2|h) = 0.6916 \cdot 0.1864 = 0.1289;$$

$$B_{(0.95)3.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{h3}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(3|h) = 0.7786 \cdot 0.3344 = 0.2604.$$

Finally Table 19 reports the decompositions of the synthetic index $I(Y) = 0.7001$. *It is useful to remember that the contributions* $B_{.lg}(Y)$ *reported in this Table are the weighted arithmetic means of the corrisponding contributions* $B_{hlg}(Y)$ *with weights* $W_{h.}/W$:

$$B_{.lg}(Y) = \sum_{h=1}^{r} B_{hlg}(Y) \cdot \frac{W_{h.}}{W}.$$

This Table confirms that the two greatest contributions $B_{.lg}(Y)$ *are* $B_{.31}(Y) = 0.1815$ *and* $B_{.11}(Y) = 0.1719$.

Table 19: Decompositions of $I(Y) = 0.7001$ into the contributions:$B_{.lg}(Y)$; $B_{.l.}(Y)$;$B_{.lW}(Y)$, $B_{.lB}(Y)$;$B_{..W}(Y)$,$B_{..B}(Y)$.

$B_{.lg}(Y)$		l			
		1	2	3	
g	1	0.1719	0.0595	0.1815	
	2	0.0637	0.0219	0.0687	
	3	0.0545	0.0187	0.0598	
$B_{.l.}(Y)$		0.2901	0.1001	0.3100	$0.7001 = I(Y)$
$B_{.lW}(Y)$		0.1719	0.0219	0.0598	$0.2536 = B_{..W}(Y)$
$B_{.lB}(Y)$		0.1182	0.0782	0.2502	$0.4465 = B_{..B}(Y)$

Table 20 reports for the three macro-regions their:

- *relative contributions to the point indexes*

$$\beta_{hl.}(Y) = \frac{B_{hl.}(Y)}{I_h(Y)} = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)} \cdot p(l|h),$$

- *relative contributions to the synthetic index*

$$\beta_{.l.}(Y) = \frac{B_{.l.}(Y)}{I(Y)},$$

- *relative weights* $W_{.l}/W$.

Table 20: Relative contributions $\beta_{(0.1)l.}(Y)$, $\beta_{(0.5)l.}(Y)$, $\beta_{(0.95)l.}(Y)$, $\beta_{.l.}(Y)$

	l			Total
	1	2	3	
$\beta_{(0.1)l.}(Y)$	*0.3407*	*0.1383*	*0.5210*	*1.0000*
$\beta_{(0.5)l.}(Y)$	*0.4140*	*0.1376*	*0.4485*	*1.0000*
$\beta_{(0.95)l.}(Y)$	*0.4654*	*0.1770*	*0.3576*	*1.0000*
$\beta_{.l.}(Y)$	*0.4143*	*0.1429*	*0.4428*	*1.0000*
$W.l/W$	*0.4873*	*0.1886*	*0.3241*	*1.0000*

We note that, the relative contribution of the North increases for increasing values of p*; viceversa for the South there is an opposite relation. The contributions to the synthetic inequality index of these two regions are similar. Comparing the relative contribution to the inequalitry of the macro-regions with their "demopgraphic" weights we can assert that the South is a region that increases the income inequality while the contrary happens for the North and the Center.*

We end this section observing that for the whole population the within component is the 36.22% of the synthetic index.

5 Conclusions and final remarks

The Zenga point inequality index $I_h(Y)$ is the relative variation $I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)}$ of the lower mean $\overset{-}{M}_{h.}(Y)$ w.r.t. the upper mean $\overset{+}{M}_{h.}(Y)$. The synthetic index is given by $I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N}$. In order to decompose by subpopulations the Zenga (2007) point and synthetic inequality indexes, Radaelli (2008, 2010) obtains, first, the following decomposition of the point

$$\begin{aligned} \text{uniformity measure } U_h(Y) &= (1 - I_h(Y)) = \frac{\overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} \\ &= \sum_{l=1}^{k}\sum_{g=1}^{k} \frac{\overset{-}{M}_{hl}(Y)}{\overset{+}{M}_{hg}(Y)} \cdot \frac{P_{hl}}{P_{h.}} \cdot \frac{T_g - Q_{hg}(Y)}{T - Q_{h.}}. \end{aligned}$$

Then he splits the point uniformity measure in the within and between components:

$$U_h(Y) = U_{hW}(Y) + U_{hB}(Y),$$

where

$$U_{hW}(Y) = \sum_{l=1}^{k} \frac{\overset{-}{M}_{hl}(Y)}{\overset{+}{M}_{hl}(Y)} \cdot \frac{P_{hl}}{P_{h.}} \cdot \frac{T_l - Q_{hl}(Y)}{T - Q_{h.}}$$

is the within part and

$$U_{hB}(Y) = \sum_{l=1}^{k} \sum_{(g:g\neq l)}^{k} \frac{\overset{-}{M}_{hl}(Y)}{\overset{+}{M}_{hg}(Y)} \cdot \frac{P_{hl}}{P_{h.}} \cdot \frac{T_g - Q_{hg}(Y)}{T - Q_{h.}}$$

is the between part. Finally, after some steps Radaelli obtains the corresponding decompositions of $I_h(Y)$, and of $I(Y)$.

We remark now, that Zenga M.M,Radaelli P., Zenga Ma. (2012) obtain the decomposition by sources of $I_h(Y)$ starting, viceversa, from the decomposition of the difference $\left[\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)\right]$.

In the present paper the decomposition by subpopulations of $I_h(Y)$ is also obtained starting from the decomposition of the difference $\left[\overset{+}{M}_{h.}(Y)-\overset{-}{M}_{h.}(Y)\right]$. This latter decomposition is achieved by the use of the popular relation between the mean M of a mixture and the means M_l of the k subpopulations of the mixture: $M=\sum_{l=1}^{k} M_l \cdot p(l)$, where $p(l)$ is the relative frequency of the subpopulation l . In particular, for the decomposition of $\left[\overset{+}{M}_{h.}(Y)-\overset{-}{M}_{h.}(Y)\right]$, we have:

- one mixture for the lower group $\{(y_1,n_{1.}),\ldots,(y_h,n_{h.})\}$ that is composed of the k subpopulations $\{(y_1,n_{1l}),\ldots,(y_h,n_{hl});\, l=1,\ldots,k\}$, and
- one mixture for the upper group $\{(y_{h+1},n_{h+1.}),\ldots,(y_r,n_{r.})\}$ that is composed of the k subpopulations $\{(y_{h+1},n_{h+1l}),\ldots,(y_r,n_{rl});\, l=1,\ldots,k\}$.

Each subpopulation of the lower group includes $P_{hl}=\sum_{t=1}^{h} n_{tl}$ units, and the (whole) lower group includes $P_{h.}=\sum_{l=1}^{k} P_{hl}$ units. Thus, $\overset{-}{M}_{h.}(Y)=\sum_{l=1}^{k} \overset{-}{M}_{hl}\cdot p(l|h)$, where $\overset{-}{M}_{hl}$ is the lower mean of the subpopulation l and p($l|h$)=$P_{hl}/P_{h.}$ is its relative frequency . Analogously each subpopulation of the upper group includes $n_{.l}-P_{hl}$ units, and the (whole) upper group includes $N-P_{h.}$ units. Thus, $\overset{+}{M}_{h.}(Y)=\sum_{l=1}^{k} \overset{+}{M}_{hl}(Y)\cdot a(l|h)$, where $\overset{+}{M}_{hl}(Y)$ is the upper mean of the subpopulation l and $a(l|h)=(n_{.l}-P_{hl})/(N-P_{h.})$ is its relative frequency.

Using the above reported "representations" for $\overset{+}{M}_{h.}(Y)$ and $\overset{-}{M}_{h.}(Y)$, in Sec.3 with simple algebra, we have obtained the following decomposition

$$\left[\overset{+}{M}_{h.}(Y)-\overset{-}{M}_{h.}(Y)\right]=\sum_{l=1}^{k}\sum_{g=1}^{k}\left[\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)\right]p(l|h)\cdot \mathrm{a}(g|h). \qquad (18)$$

We point out that the expression (18) has the structure of the Gini mean difference Δ_{hgl} between the two distributions:

$$\left\{\left[\overset{+}{M}_{hg}(Y),a(g|h)\right]:\ g=1,\ldots,k\right\} \text{ and } \left\{\left[\overset{-}{M}_{hl}(Y),p(l|h)\right]:\ l=1,\ldots,k\right\}.$$

Then, dividing both sides of (18) by $\overset{+}{M}_{h.}(Y)$, we have achieved the following $k\times k$ additive decomposition of $I_h(Y)$:

$$\begin{aligned} I_h(Y) &= \sum_{l=1}^{k}\sum_{g=1}^{k}\left[\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]p(l|h)\cdot a(g|h), \qquad (23)\\ &= \sum_{l=1}^{k}\sum_{g=1}^{k} B_{hlg}(Y). \end{aligned}$$

Note that in the decomposition (23), the contributions $B_{hlg}(Y)$ have the same denominator, while this is not the case of the Radaelli (2008, 2010) decomposition. From $\sum_{l=1}^{k}\sum_{g=1}^{k} p(l|h)\cdot a(g|h) = 1$, it derives also that $I_h(Y)$ can be interpreted as the weighted mean of the relative differences $\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}$ with weights given by the product of the relative frequencies $p(l|h)\cdot a(g|h)$.

It is worth to remark that, starting from the $k\times k$ decomposition (23), we have obtained the following k additive decomposition of $I_h(Y)$:

$$\begin{aligned} I_h(Y) &= \sum_{l=1}^{k} p(l|h)\left\{\sum_{g=1}^{k}\left[\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]\cdot a(g|h)\right\}, \\ &= \sum_{l=1}^{k}\left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]\cdot p(l|h) = \sum_{l=1}^{k} B_{hl.}(Y), \quad (26) \end{aligned}$$

where $B_{hl.}(Y) = \sum_{g=1}^{k} B_{hlg}(Y)$. Note that $B_{hl.}(Y)$ is equal to the product of the relative variation of $\bar{M}_{hl}(Y)$ w.r.t $\overset{+}{M}_{h.}(Y)$ and the relative frequency $p(l|h) = \frac{P_{hl}}{P_{h.}}$. In other words, formula (26) shows that, the point index $I_h(Y)$ is the weighted mean of the k relative variations $\left[\frac{\overset{+}{M}_{h.}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]$ with weights $p(l|h) = \frac{P_{hl}}{P_{h.}}$. Thus, $B_{hl.}(Y)$ can be interpreted as the contribution of the subpopulation l to the point inequality index $I_h(Y)$.

It is interesting to point out that each contribution $B_{hl.}(Y) = \sum_{g=1}^{k} B_{hlg}(Y) = \sum_{g=1}^{k}\left[\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] p(l|h)\cdot a(g|h)$ can be split into the following:

whithin component

$$B_{hll}(Y) = p(l|h)\cdot\left[\frac{\overset{+}{M}_{hl}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right]\cdot a(l|h) = B_{hlW}(Y) \text{ and}$$

between component

$$\sum_{(g:g\neq l)} B_{hlg}(Y) = p(l|h)\sum_{(g:g\neq l)}\frac{\overset{+}{M}_{hg}(Y)-\bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)} a(g|h) = B_{hlB}(Y)$$

Consequently, the within and the between component of the point index $I_h(Y)$ are given by.

$$\begin{aligned} I_h(Y) = \sum_{l=1}^{k} B_{hl.}(Y) &= \sum_{l=1}^{k} B_{hlW}(Y) + \sum_{l=1}^{k} B_{hlB}(Y) \\ &= B_{h.W}(Y) + B_{h.B}(Y), \quad (31) \end{aligned}$$

where $B_{h.W}(Y) = \sum_{l=1}^{k} B_{hlW}(Y)$ can be interpreted as the within contribution of all the subpopulations to $I_h(Y)$, and $B_{h.B}(Y) = \sum_{l=1}^{k} B_{hlB}(Y)$ can be interpreted as the between contribution of all the subpopulations to $I_h(Y)$.

Finally, the decompositions of the synthetic index are obtained putting the decompositions of $I_h(Y)$ into the expression $I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N}$. Thus,

$$\begin{aligned}
I(Y) &= \sum_{h=1}^{r} \left\{ \sum_{l=1}^{k} \sum_{g=1}^{k} \left[\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)} \right] p(l|h) \cdot a(g|h) \right\} \cdot \frac{n_{h.}}{N} \\
&= \sum_{l=1}^{k} \sum_{g=1}^{k} \left\{ \sum_{h=1}^{r} \left[\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)} \right] p(l|h) \cdot a(g|h) \cdot \frac{n_{h.}}{N} \right\} \\
&= \sum_{l=1}^{k} \sum_{g=1}^{k} \left\{ \sum_{h=1}^{r} B_{hlg}(Y) \cdot \frac{n_{h.}}{N} \right\} = \sum_{l=1}^{k} \sum_{g=1}^{k} B_{.lg}(Y) \qquad (34) \\
&= \sum_{l=1}^{k} \left\{ B_{.ll}(Y) + \sum_{(g:g \neq l)} B_{.lg}(Y) \right\} = \sum_{l=1}^{k} B_{.l.}(Y) \\
&= \sum_{l=1}^{k} \left\{ B_{.lW}(Y) + B_{.lB}(Y) \right\}; \\
&= B_{..W}(Y) + B_{..B}(Y) = B_{...}(Y).
\end{aligned}$$

In these latter decompositions: $B_{.lg}(Y)$ is the weighted mean of the r contributions $B_{hlg}(Y)$ with weights $\frac{n_{h.}}{N}$; $B_{.l.}(Y) = \sum_{g=1}^{k} B_{.lg}(Y)$ is the contribution of the subpopulation l to $I(Y)$, and $B_{.lW} = \sum_{h=1}^{r} B_{hll}(Y) \cdot \frac{n_{h.}}{N}$ and $B_{.lB}(Y) = \sum_{(g:g \neq l)} \left\{ \sum_{h=1}^{r} B_{hlg}(Y) \cdot \frac{n_{h.}}{N} \right\}$ are respectively its within and between parts. Finally,

$$B_{..W}(Y) = \sum_{l=1}^{k} \left\{ \sum_{h=1}^{r} \left[\frac{\overset{+}{M}_{hl}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)} \right] p(l|h) \cdot a(l|h) \cdot \frac{n_{h.}}{N} \right\} \qquad (39)$$

and

$$B_{..B}(Y) = \sum_{l=1}^{k} \sum_{(g:g \neq l)} \left\{ \sum_{h=1}^{r} \left[\frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)} \right] p(l|h) \cdot a(g|h) \cdot \frac{n_{h.}}{N} \right\} \qquad (40)$$

are the within and the between part of the synthetic index $I(Y)$, respectively. Note that in $B_{..W}(Y)$ only comparisons between upper and lower means of the same subpopulations are involved, while in $B_{..B}(Y) = I(Y)$ only comparisons between upper and lower means of different subpopulations are involved.

Concluding, starting from the decomposition of $I(Y)$ into the $k \times k$ contributions $B_{.lg}(Y)$ we obtain its decomposition into a within and a between term just separating the contributions $B_{.lg}(Y)$ evaluated within the same population from the ones regarding different subpopulations. This approach, that does not depend on an *a priori* definition of the within or the between term, gives the within and between component of $I(Y)$ in a natural way preserving the structure of the index itself (see Radaelli P.2008, 2010).

The application of Section 4 shows that there is "strong" dependence of the household net disposable income Y from the Italian regional areas. In fact, the mean value of the South is very far from the mean values of the other two Italian macroregions. Besides that, this dependence exerts a strong influence on the values of the conditional relative frequencies $p(l|h) = \frac{P_{hl}}{P_{h.}}$ of the lower groups and on the values of the relative conditional frequencies $a(l|h) = (n_{.l} - P_{hl}) / (N - P_{h.})$ of the upper groups too. The values of these conditional relative frequencies helps in the interpretation of the contributions $B_{hlg}(Y) = p(l|h) \cdot \left[\frac{\overset{+}{M}_{hl}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot a(g|h)$, and of the contributions $B_{hl.}(Y) = \left[\frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{hl}(Y)}{\overset{+}{M}_{h.}(Y)}\right] \cdot p(l|h)$.

References

Arcagni A., Zenga M.M. (2013). Application of Zenga's distribution to a panel survey on household incomes of European member states. Statistica § Applicazioni, XI(1), 1-24.

Arcagni A., Porro F. (2014). The graphical representation of inequality. Revista Colombiana de estadistica , 37(2), 419-436.

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

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.

Bottiroli-Civardi M. (1987). Proposte di scomposizione dell'indice del Gini. In Zenga M. (a cura di), La distribuzione personale del reddito: problemi di formazione, di ripartizione e di misurazione,Vita e Pensiero,Milano.

DeVergottini M. (1940). Sul Significato di Alcuni Indici di Concentrazione. Giornale degli Economisti e Annali di Economia , 2. 317-347.

Deutsch J. Silber J. (1999). Inequality Decomposition by Population Subgroups and the Analysis of Interdistributional Inequality. In Silber J., Sen A. (eds), Handbook of Income Inequality Measurement, 363-403. Kluer Accademic Press.

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.

Greselin F., Pasquazzi L. (2009). Asymptotic confidence intervals for a new inequality measure. Comunications in Statistics. Simulation and Computation, 38(8), 1742-1756.

Greselin F., Pasquazzi L., Zitikis R. (2010). Zenga's new index of economic inequality, its estimation, and analysis of incomes in Italy. Journal of Probality and Statistics, 2010,1-26.

Greselin F., Pasquazzi L., Zitikis R. (2013). Contrasting the Gini and Zenga indices of economic inequality. Journal of Applied Statistics.

Greselin F., Pasquazzi L., Zitikis R. (2014). Heavy tailed capital incomes: Zenga index, statistical inference, and ECHP data analysis. Extremes: Volume 17, Issue 1 (2014), Page 127-155

Langel M., Tillé Y. (2012). Inference by linearization for the Zenga inequality index: A compararison with the Gini index. Metrika, 75(8), 1093-1110.

Lerman R., Yitzhaki S. (1984). A note on the calculation and the interpretation of the Gini index. Economic Letters, 15, 363-368.

Lerman R., Yitzhaki S. (1985). Income inequality effects by income source: a new approach and applications to the United States. The Review of Economics and Statistics, 67, 151-156.

Maffenini W., Polisicchio M. (2014). The potential of the $I_{(p)}$ inequality curve in the analysis of empirical distributions. Statistica § Applicazioni, XII(1), 63-85.

Mehran F. (1975). A statistical analysis of income inequality based on a decomposition of the Gini index. In Proceedings of the 40th Session of the International Statistical Institute, Vol. 4,pp. 145-150. Warsaw: International Statistical Institute.

Mussini M., Zenga M. (2013). A longitudinal decomposition of Zenga's new inequality index. Statistica § Applicazioni, XI(1), 63-77.

Mussini M. (2013). A matrix approach to the Gini index decomposition by subgroup and by income sources. Applied Economics. Taylor & Francis.

Piesch W. (2005). Bonferroni-Index und De Vergottini-Index. Diskussionsbeitrage aus dem Institut fur Volkswirtschaftslehre der Universitat Hohenheim. Nr.259/2005.

Polisicchio M. (2008). The continuous random variable with uniform point inequality measure $I\left(p\right)$. Statistica § Applicazioni, VI(2), 137-151.

Porro F. (2011). The distribution model with linear inequality curve $I(p)$. Statistica § Applicazioni, IX(1), 47-61.

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., Zenga M. (2002). Decomposition of the Mean Difference of the Sum of K Variables. Abstract of contributed papers. Fourth International Conference on the L_1-Norm and related Methods- University of Nuechatel, Switzerland.

Radaelli P., Zenga M. (2005). On the decomposition of the Gini mean difference and concentration ratio. Statistica § Applicazioni, III(2), 5-24.

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.

Rao V.(1969). Two decomposition of concentration ratio. Journal of the Royal Statistical Society. Series A, 132, 418-425.

Shorrocks A. (1980). The Class of Additively Decomposable Inequality Measures. Econometrica , 48(3), 613-626.

Tarsitano A. (1989). The Bonferroni index income inequality. In Dagum C. and Zenga M. (Eds). Income and Wealth Distribution, Inequality and Poverty-Studies in Contemporary Economics, Springer-Verlag, Berlin.

Unnikrishnan Nair N., Muraleedharan Nair K.R.,Sreelakshmi N. (20012). Some properties of the new Zenga curve. Statistica § Applicazioni, X(1), 43-52.

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. (1986). On the normalization, the invariance to the population replication and the additive decomposition in the $\left\{I^{(\alpha)}\right\}$ class of inequality measures.Metron,X LIV(1-4), 131-151.

Zenga M. (1987). Effetti della normalizzazione sul principio della somiglianza e sulla scomponibilità degli indici di concentrazione.In Zenga M. (a cura di), La distribuzione personale del reddito: problemi di formazione, di ripartizione e di misurazione,Vita e Pensiero,Milano.

Zenga M. (2001). A multiplicative decomposition of Herfindahl concentration measure.Metron, LIX(1-2), 3-10.

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.M. (2007b). Applications of a new inequality curve and inequality index based on ratios between lower and upper arithmetic means. In M.I.Gomes , D. Pestana and P.Silva (eds), Bulletin of the International Statistical Institute, Volume Proceeding on CD. Lisbon, Portugal.

Zenga M.M., Radaelli P., Zenga Ma. (2012). Decomposition of Zenga's Inequality Index by Sources. Statistica § Applicazioni, X(1), 3-31.

Zenga M. (2013). Decomposition by Sources of the Gini, Bonferroni and Zenga Inequality Indexes. Statistica § Applicazioni, XI(2), 131-158.