A JOINT DECOMPOSITION BY SOURCES AND BY SUBPOPULATIONS OF THE ZENGA-84 INEQUALITY INDEX

Francesca De Battisti*
Francesco Porro**

ABSTRACT

This paper provides a multi-decomposition (or joint decomposition) of the Zenga-84 inequality index, considering simultaneously different sources and subpopulations. The suggested procedure consists of two different stages: in the first one, each inequality pointwise measure is decomposed by sources and by subpopulations; in the second one, the decomposition of the synthetic index is obtained, by averaging the previously decomposed pointwise measures. An important feature of this multi-decomposition is that it allows to assess the portion of the global inequality related to each source in each subpopulation, since most of the decomposition procedures proposed in the literature are not able to achieve such goal. The proposed joint decomposition permits also to obtain, as particular cases, two "marginal" decompositions by sources and by subpopulations of the Zenga-84 index, already introduced in the literature. To show the usefulness of the described multi-decomposition, an application about the household consumption is provided. The real considered dataset comes from the Household Consumption Expenditure Survey (HCES) provided by ISTAT.

***Keywords*: Income inequality, Joint decomposition, Decomposition by subpopulations, Decomposition by sources, Zenga-84 inequality index**

1. INTRODUCTION

Zenga (1984) introduced the index ζ, based on the relationships between population quantiles and income quantiles, for evaluating inequality. This index owns the interesting property to be more susceptible than others to small changes in inequality; in particular, it has been proved that, within the most common degree of concentration, it is more sensitive than the Gini index (Zenga, 1984). Several papers in literature analyzed and applied the Zenga-84 inequality index ζ. Recently, Arcagni (2017) and Porro and Zenga (2020) introduced the decomposition by sources and by subpopulations, respectively. The aim of the paper is to provide a multi-decomposition (or joint decomposition) of the index ζ, following a procedure based on a two-step approach, introduced for the first time in the literature in Zenga et al. (2012). In particular, two different stages are implemented: in the first one, each inequality pointwise measure is decomposed

*Dipartimento di Economia, Management e Metodi Quantitativi - Universitá degli Studi di Milano - via Conservatorio, 7, 20122 MILANO (francesca.debattisti@unimi.it)
**Dipartimento di Statistica e Metodi Quantitativi - Universitá degli Studi di Milano-Bicocca - piazza dell'Ateneo Nuovo, 1, 20126 MILANO (francesco.porro1@unimib.it). (Corresponding author)

by sources and by subpopulations; in the second one, the decomposition of the synthetic index is obtained, by averaging the previously decomposed pointwise measures. The presented multi-decomposition allows to evaluate the portion of the global inequality related to each source in each subpopulation; this is a worthwhile result, since most of the decomposition procedures proposed in the literature are not able to achieve such goal. To highlight the interpretative power of the described multi-decomposition, an application about the household consumption is provided, by using data from the Household Consumption Expenditure Survey (2018) supplied by ISTAT.

2. METHODOLOGY

Firstly, it is useful to recall the definitions of the Zenga-84 inequality function $Z(p)$ and of the inequality index ζ, introduced in Zenga (1984). Here only the discrete case is reported, but the continous one straightforward follows.

DEFINITION 1 *Let $\{(y_h, n_{h.}); h = 1, 2, ...r\}$ be the non-negative, distinct, and ordered values assumed by the statistical variable Y (with $y_1 < y_2 < \cdots < y_r$) and their frequencies, observed on N units of a finite population: $N = \sum_{h=1}^{r} n_{h.}$. Let $M(Y)$ be the positive mean of Y, and let F and Q denote the distribution function and the first incomplete moment of Y, respectively:*

$$F(y) = \sum_{\{h\,:\,y_h \leq y\}} \frac{n_{h.}}{N}, \qquad Q(y) = \sum_{\{h\,:\,y_h \leq y\}} \frac{y_h n_{h.}}{N \cdot M(Y)}.$$

Then the Zenga-84 inequality function $Z(p)$ of Y is defined as the relative variation of $y_{(p)}$ with respect to $y^*_{(p)}$:

$$Z(p) = \frac{y^*_{(p)} - y_{(p)}}{y^*_{(p)}} = 1 - \frac{y_{(p)}}{y^*_{(p)}} \qquad p \in [0, 1] \tag{1}$$

where $y_{(p)}$ and $y^*_{(p)}$ are the generalized inverse functions of F and Q, respectively. It is a step function, and its support is finite. The inequality index ζ of Y is the area below the $Z(p)$ function, therefore it is the sum of the areas of NC rectangles:

$$\zeta = \sum_{t=1}^{NC} Z_t w_t \tag{2}$$

where w_t and Z_t denote the basis and the height of the t^{th} rectangle, respectively.

As detailed in Zenga (1991) and in Porro and Zenga (2020), the calculation of the index ζ can be simplified by using a procedure based on the cograduation table of Y, which makes explicit an association rule between two discrete variables. The computation of the cograduation table allows to obtain the number NC of cells with non-zero weight. Each one of them is associated with a value of the counter t (with $t = 1, 2, \ldots, NC$), a weight w_t, and two values y_t and

y_t^*, providing $Z_t = \frac{y_t^* - y_t}{y_t^*}$, which can be used to calculate the index ζ. The proposed joint decomposition procedure, which considers together sources and subpopulations, can be summarized in the following theorem.

THEOREM 1 (Joint decomposition) *Let Y be a statistical variable evaluated on a finite population of N units, partitioned into k subpopulations; and let Y be the sum of c sources $X_1, X_2, \ldots X_c$. Then the inequality index ζ can be decomposed as:*

$$\begin{aligned} \zeta &= \sum_{j=1}^{c}\sum_{l=1}^{k}\sum_{g=1}^{k}\sum_{t=1}^{NC} H_{tlg}(X_j)w_t = \sum_{j=1}^{c}\sum_{l=1}^{k}\sum_{t=1}^{NC} H_{tl\cdot}(X_j)w_t \\ &= \sum_{j=1}^{c}\sum_{l=1}^{k}\sum_{g=1}^{k} H_{\cdot lg}(X_j) = \sum_{j=1}^{c}\sum_{l=1}^{k} H_{\cdot l\cdot}(X_j), \textit{where} \end{aligned} \tag{3}$$

- *$H_{\cdot l\cdot}(X_j) = \sum_{g=1}^{k}\sum_{t=1}^{NC} H_{tlg}(X_j)w_t$ is the contribution due to the source X_j to the total inequality related to the subpopulation S_l;*
- *$H_{tl\cdot}(X_j) = \sum_{g=1}^{k} H_{tlg}(X_j) = \frac{x_{jt}^* - x_{jt}}{y_t^*} \cdot p(l|t)$ is the contribution due to the source X_j to the pointwise measure Z_t related to the subpopulation S_l;*
- *$H_{tlg}(X_j)$ is the contribution due to the source X_j to the pointwise measure Z_t related to the couple of the subpopulations S_l and S_g, defined as:*

$$H_{tlg}(X_j) = \frac{x_{jt}^* - x_{jt}}{y_t^*} \cdot a(g|t)p(l|t),$$

  where $a(g|t)$ is the relative frequency of y_t^* in the subpopulation S_g, and $p(l|t)$ is the relative frequency of y_t in the subpopulation S_l.

The result of the joint decomposition are c $(k \times k)$-matrices (one for each source), where the entries are the quantities $H_{tlg}(X_j)$. By considering one of these matrices, the sum of the values in the main diagonal provides the *Within* component to the inequality due to the corresponding source, while the sum of the remaining values provides the *Between* component. It is also worthwhile noting that the decompositions by sources and by subpopulations, already known in literature, can be easily obtained from the joint decomposition.

3. AN APPLICATION TO HOUSEHOLD CONSUMPTION EXPENDITURE

The HCES survey in 2018 involves 18342 households and detects all the expenses incurred by resident families to purchase goods and services intended for household consumption or to make gifts. The dataset is available online at www.istat.it/it/archivio/180356. The estimate of mean monthly consumption expenditure is 2569.52€. Half of households spend montly more than 2146.95€. The expenditures are grouped in different categories; in this study only two

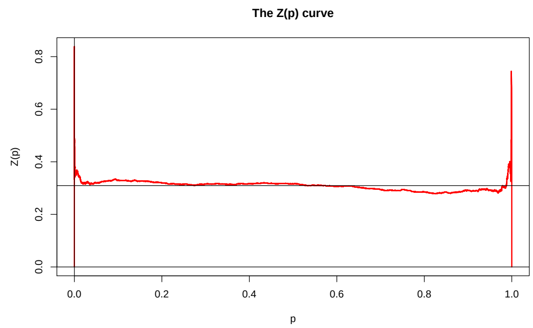

FIGURE 1: The $Z(p)$ curve of the *Household Consumption Expenditure* Y.

sources are considered: *Food and non-alcoholic beverages* and *Non-Food*, which represent the 18.2% and the 81.8% on the total expenditure, respectively. The survey sample has been divided in three territorial subpopulations: *North, Center, South and Islands*; for these ones the expenditure are 51.2%, 21.4% and 27.4% of the total amount, respectively. The survey sample consists of 8476 (46.2%) household of subpopulation *North*, 3632 (19.8%) of *Center*, and 6234 (34%) of *South and Islands.* The aim of this analysis is to evaluate the disparity level by Zenga-84 inequality index and, subsequently, to decompose its value jointly by sources (*Food, Non-Food*) and by territorial subpopulations (*North, Center, South and Islands*). The calculation of the cograduation matrix provides a number $NC = 36683$ of cells with non-zero weight. The value of Zenga-84 inequality index for the *Household consumption expenditure* related to all the considered families is equal to $\zeta = 0.3094$. This value is not high and it denotes a discrete total level of inequality. In Figure 1 the $Z(p)$ curve of the variable Y is drawn: it can be remarked that its behaviour is quite similar to an horizontal line, meaning that the distribution of the *Household Consumption Expenditure* is quite lognormal, mostly in the central values of the support. By applying the joint decomposition proposed, we obtain two (as the number of considered sources) (3×3)-matrices, each of them representing the contribution to total inequality due to each source. They are all stored in Table 1. In each matrix, the sum of values in one column represents the contribution due to such source to the total inequality related to the corresponding subpopulation. For both the sources (*Food* and *Non-Food*), the most relevant subpopulation is *North* (with contributions equal to 0.0245 and 0.1173, respectively), then *South and Islands* follows (0.0102 and 0.0962), and finally *Center* (0.0092 and 0.0520). From the joint decomposition, the decomposition by subpopulations (proposed in Porro and Zenga, 2020) can be easily achieved. By summing the two matrices in Table 1 we obtain the decomposition by subpopulations matrix in Table 2, with the values of $Z_{\cdot 1 \cdot} = 0.1418$ (45.8%), $Z_{\cdot 2 \cdot} = 0.0612$ (19.8%) and $Z_{\cdot 3 \cdot} = 0.1064$ (34.4%). They show that the highest contribution to total inequality of *Household Consumption Expenditure* is from the subpopulation

TABLE 1: The (3×3)-decomposition matrices for the two *Household Consumption Expenditure* sources: *Food* (X_1), and *Non-Food* (X_2).

Food (X_1)	*North*	*Center*	*South and Islands*
North	0.0089	0.0034	0.0017
Center	0.0047	0.0016	0.0016
South and Islands	0.0109	0.0042	0.0069
Total	*0.0245* *(55.8%)*	*0.0092* *(21%)*	*0.0102* *(23.2%)*

Non-Food (X_2)	*North*	*Center*	*South and Islands*
North	0.0650	0.0286	0.0499
Center	0.0265	0.0115	0.0202
South and Islands	0.0258	0.0119	0.0261
Total	*0.1173* *(44.2%)*	*0.0520* *(19.6%)*	*0.0962* *(36.2%)*

TABLE 2: The decomposition by subpopulations matrix for the total inequality of the *Household Consumption Expenditure* (Y).

Expenditure (Y)	*North*	*Center*	*South and Islands*	
North	0.0739	0.0320	0.0516	
Center	0.0312	0.0131	0.0218	
South and Islands	0.0367	0.0161	0.0330	
Total	*0.1418* *(45.8%)*	*0.0612* *(19.8%)*	*0.1064* *(34.4%)*	$\zeta = 0.3094$

North, followed by *South and Islands* and finally by *Center*. Another important decomposition is the classical one, in *Between* and *Within* components. By summing the values on the main diagonal of the previous matrix (see Table 2), we obtain the *Within* component; referring to the remaining cells, we calculate the *Between* component. These quantities are equal to $\zeta_W = 0.12$ and $\zeta_B = 0.1894$, and they represent the 38.8% and the 61.2% of total inequality, respectively. The contribution to total inequality due to comparisons of units in different subpopulations is higher than the contribution based on the comparisons of values related to the same subpopulation: such result is not surprising, since the means of the *Household Consumption Expenditure* are very different in the three subpopulations. Summing the nine entries in each matrix of Table 1, the decomposition by sources (proposed in Arcagni, 2017) can be achieved, which allows to identify the contribution to total inequality due to each source. The values are stored in the last column of Table 3. It is worth remarking that the highest contribution to the inequality is largely due to the source *Non-Food*:

TABLE 3: The joint decomposition and the decomposition by sources matrix for the total inequality.

Source	*North*	*Center*	*South and Islands*	*Contribution to ζ: $H.(X_j)$*
Food (X_1)	0.0245 (17.3%)	0.0092 (15%)	0.0102 (9.6%)	*0.0439 (14.2%)*
Non-Food (X_2)	0.1173 (82.7%)	0.0520 (85%)	0.0962 (90.4%)	*0.2655 (85.8%)*
				$\zeta =$ *0.3094*

this note was expected, since such source represents most part of the variable Y (81.8% against the 18.2% of *Food*). It is important to remark that the joint decomposition allows to highlight some aspects that would otherwise remain hidden. In particular, the last rows of matrices in Table 1 and Table 2 show that the relative contributions to the total inequality of the subpopulations for the *Expenditure* Y and for *Non-Food* source are both similar to the partition in subpopulations of the sample, while those corresponding to source *Food* are quite different, denoting that the inequality is more due to subpopulation *North* for the source *Food* (55.8%) than for the source *Non-Food* (44.2%) and for the *Expenditure* Y (45.8%). The results of the joint decomposition, stored in Table 3, allow to state that the two sources count very dissimilarly in the subpopulations. The most relevant point is for subpopulaton *South and Islands*, where the source *Non-Food* contributes for the 90.4% against a 85.8% at the national level.

4. CONCLUSIONS

In this paper we introduce a new joint decomposition of Zenga-84 inequality index ζ, which allows to split the contribution related to each source among the subpopulations and to evaluate the contribution to each source due to each subpopulation. This result is original and informative since it makes possible to identify the contribution to the overall inequality of each source due to each subpopulation. The application on the household consumption, which considers two sources (*Food and non-alcoholic beverages* and *Non-Food*) and three subpopulations (*North*, *Center*, *South and Islands*), highlights that the relative contributions to the total inequality of the subpopulations for the source *Food* are quite different from the partition in subpopulations of the sample, denoting that the inequality is more due to subpopulation *North* for the source *Food* than for the source *Non-Food*. Furthermore, for subpopulaton *South and Islands* the source *Non-Food* contributes to the inequality more than at the national level. These results are due to the joint decomposition; only this kind of approach allows to state that the two sources count very differently in the subpopulations.

REFERENCES

Arcagni A. (2017). On the decomposition by sources of the Zenga 1984 point and synthetic inequality indexes. *Statistical Methods & Applications*, **26**(1), 113-133.

Porro F., Zenga M. (2020) Decomposition by subpopulations of the Zenga-84 inequality curve and the related index ζ: an application to 2014 Bank of Italy survey. *Statistical Methods & Applications*, doi: 10.1007/s10260-019-00459-9.

Zenga M. (1984). Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito. *Giornale degli Economisti e Annali di Economia*, **43**(5-6), 301-326.

Zenga M. (1991). Impiego delle tabelle di cograduazione per la determinazione dell'indice puntuale di concentrazione $Z(p)$. *Statistica Applicata*, **3**, 283-291.

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