**Statistical Methods & Applications manuscript No.**
(will be inserted by the editor)

# A multi-decomposition of Zenga-84 inequality index: an application to the disparity in $CO_2$ emissions in European countries

**Francesca De Battisti, Francesco Porro**

Received: date / Accepted: date

**Abstract** The monitoring of $CO_2$ emissions has become a sensitive topic of discussion in the last years. The engagement of the protocol of Kyoto, and the subsequent activities that the different countries have carried out to reduce the $CO_2$ emissions, are factors which push the topic into the spotlight. An interesting issue regards how the disparities of such emissions can be analyzed by sources and by subpopulations. In this paper an innovative procedure to jointly decompose the disparity by sources and by subpopulations is proposed. The assessment of the inequality is determined by the Zenga-84 index. This new methodology is applied to the analysis of the per capita $CO_2$ emission disparities for European countries, by simultaneously considering their sources (coal, oil, natural gas, and other) and the membership of the country to OECD.

**Keywords** $CO_2$ emission · decomposition by sources · decomposition by subpopulations · inequality · joint decomposition · Zenga-84 inequality index

## 1 Introduction

Many tools have been proposed in the literature for the inequality evaluation and its analysis. Among them, a relevant role is played by the inequality synthetic indexes: they have the capability to summarize the inequality level in a single number. Among all the characteristics and properties they can have, a very interesting peculiarity is their chance to be decomposed. The most

Francesca De Battisti
francesca.debattisti@unimi.it
Dipartimento di Economia, Management e Metodi Quantitativi
Università degli Studi di Milano, Milan, Italy

Francesco Porro
francesco.porro@unige.it
Dipartimento di Matematica
Università degli Studi di Genova, Genoa, Italy

famous kinds of decompositions for an inequality index are two: by sources and by subpopulations.

The decomposition by subpopulations (or by subgroups) basically consists of considering the observed population divided into - say $k$- exhaustive and disjoint groups, and its aim is to assess a level of inequality to each subpopulation, or at least to obtain an aggregate value for the inequality *into* the subpopulations (Within component), and one related to the inequality *across* them (Between component): for the statement of the topic, see for example (Shorrocks, 1984). Instead, the decomposition by sources (or by factors) can be performed when the target variable is the sum of - say $c$- other variables, called sources, and its aim is to evaluate the contribution to the total inequality due to each source: see for more details Shorrocks (1982).

Several proposals in literature deal with these well-known decompositions. The papers of Bourguignon (1979), and Shorrocks (1980, 1982, 1984) have had a great influence on the researches regarding the decomposition by subpopulations of the Theil indexes and the generalized entropy indexes. The two Dagum papers (Dagum, 1997*a*,*b*) have had a great effect on a lot of researches concerning the decomposition by subpopulations of the Gini concentratio ratio. Several approaches have been proposed to decompose inequality indexes by sources. For instance, Rao (1969), Lerman and Yitzhaki (1984, 1985) and Radaelli and Zenga (2005) proposed different methods to decompose the Gini index by sources. Some other papers on the decompositions of inequality indexes are Mehran (1975), Mookherjee and Shorrocks (1982), Mussard (2004), and more recently, Radaelli (2010), Frosini (2012), Ebert (2010), Fiori and Porro (2020), Arnold and Sarabia (2018), Porro and Zenga (2021).

Recently, Arcagni (2017), and Porro and Zenga (2020) have proposed the decomposition by sources and by subpopulations of Zenga-84 inequality index, respectively.

Many papers dealing with the decompositions of inequality indexes follow the approach proposed in Shorrocks (1980, 1982, 1984). In these three papers some constrict hypotheses about the decomposition procedure are assumed, in analogy to the classical variance decomposition. The result is a very restricted class of "decomposable" inequality measures, which excludes many largely-used inequality indexes as the Gini coefficient and the Bonferroni index, just to mention two of them. To overcome this issue, many authors have followed a different approach.

In the recent years Zenga and Valli (2017, 2018) have proposed a joint decomposition, by sources and by subpopulations simultaneously, for the Bonferroni inequality measures and for the point and synthetic Gini indexes, respectively.

In this paper we propose an innovative joint decomposition by sources and by subpopulations of the Zenga-84 inequality index, based on the relationships between population quantiles and income quantiles, introduced in Zenga (1984).

The same author proposed a new inequality index some years later, with different properties (Zenga, 2007). Nevertheless, we decided to use the Zenga-

84 index, because in the literature there is no joint decomposition for such index. The aim of this paper is therefore to fill this gap. We highlight that an exaustive comparison with decomposition procedures of other indexes is a very interesting topic, but it is out of scope of this paper. In the literature, it is well-known that the Zenga-84 index satisfies all the usual requirements of an inequality indicator and it has shown to be a valid alternative to the most common inequality indexes. We refer to Berti and Rigo (1995), Zenga (1990, 1993), Porro and Zenga (2020) and references therein for a more detailed revision of its properties, for all the proofs, and for a review of its impact in the literature. Here it is worth to mention just some papers: Belzunce et al. (1995), Fernandez Morales et al. (1996), Aly and Hervas (1999), Bertoli-Barsotti (2001), Kleiber and Kotz (2003), Berti and Rigo (2006), Jedrzejczak (2013), Jedrzejczak and Kubacki (2013), Arnold (2015), Jedrzejczak (2015), and Arnold and Sarabia (2018).

Among the properties of the Zenga-84 index $\zeta$, here we want to recall that it has been empirically observed that "...within the most common degrees of concentration the $\zeta$ index is more sensitive than the Gini index" (Zenga, 1984). In particular, it has been recognized to be smaller than the Gini index $G$ whenever $G < 0.33$ and bigger (than the Gini index) whenever $G > 0.362$ (see Zenga, 1984). Just to provide a practical example, taking into account that the range of the Gini index for the 32 countries in European Union is the interval $[0.209, 0.396]$ (Eurostat, 2018), it follows that the range of the values assumed by $\zeta$ for the EU countries of 2018 is larger than 0.187, and therefore the "compression effect" due to the Gini index is dulled, allowing easier comparisons among the countries.

The $\zeta$ index is suitable, for its additive structure, to be easily decomposed by sources and by subpopulations. In this paper we describe a new joint decomposition that takes into account simultaneously the sources and the subpopulations. The joint decomposition presented allows to split the contribution related to each source among the subpopulations: in this way it permits deeper and more detailed analyses, since it is possible to evaluate the contribution to each source due to each subpopulation. This result is original and informative, since it allows to identify the contribution to the overall inequality of each source due to each subpopulation. To better explain the potentialities of our approach, we propose an application in a current and sensitive topic: we analyze the $CO_2$ emission distribution in different European countries (OECD and non-OECD) and referring to different sources (coal, oil, natural gas, and other). More in details, we are interested in evaluating the overall emission inequality and in jointly decomposing such inequality by sources and by subpopolations.

The paper is organized as follows. Section 2 describes some useful preliminaries, definitions, and notation. Section 3 is devoted to the explanation of the joint decomposition. Section 4 presents the results of the decomposition of inequality in $CO_2$ emissions. Finally, section 5 ends up with some final remarks.

## 2 Preliminaries and notation

First of all, it is useful to recall the definitions of the Zenga-84 inequality function $Z(p)$ and of the inequality index $\zeta$, introduced in Zenga (1984). For further details, see the literature review on them in Porro and Zenga (2020).

**Definition 1** Let $Y$ be a non-negative continuous random variable with probability density function $f$, distribution function $F$, and positive finite expectation $\mu$. Let $Q(y)$ be the first incomplete moment of $Y$, defined by $Q(y) = \frac{1}{\mu}\int_0^y tf(t)dt$. 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:

$$y_{(p)} = \begin{cases} \inf\{x : F(x) \geq p\} & \text{if } p \in (0,1] \\ \inf\{x : F(x) > p\} & \text{if } p = 0 \end{cases}$$

$$y^*_{(p)} = \begin{cases} \inf\{x : Q(x) \geq p\} & \text{if } p \in (0,1] \\ \inf\{x : Q(x) > p\} & \text{if } p = 0. \end{cases}$$

The inequality index $\zeta$ of $Y$ is the area below the $Z(p)$ function, and it is defined as

$$\zeta = \int_0^1 Z(p)dp.$$

The definition of the function $Z(p)$ and of the index $\zeta$ for a discrete random variable is straightforward. In the remainder of the paper we focus on such kind of variables, therefore the following definition deserves to be reported.

**Definition 2** 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:

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

In this case the function $Z(p)$ is a step function, and it is defined as in the continuous case, by formula (1). Its support is finite and the index $\zeta$ 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.

It is worth remarking here that $\zeta$ is bounded as it takes on values in the interval $(0,1)$. In particular, it tends to 0 as the distribution approaches the equalitarian case (where all the values are equal), and tends to 1 as the distribution approaches the case of maximum inequality (where only one value is not zero). The $Z(p)$ curve is constant as the distribution of the values follows a lognormal law (Zenga, 1984): this property is important since many papers in the literature prove that several economical phenomena, like for example the consumptions (Battistin et al., 2009), have a (quite) lognormal distribution. The issue of the distribution unicity for a given $Z(p)$ curve has been already investigated in the litearature: it has been found that the Zenga-84 curve of a random variable $X$ does not always determine the distribution of $X$ up to a scale factor (Arnold, 2015). This issue should be taken into account whenever the distribution of $X$ is obtained from the $Z(p)$ curve: see also Arnold and Sarabia (2018) for further details.

As detailed in Porro and Zenga (2020), the calculation of the index $\zeta$ can be simplified by using a procedure based on the cograduation table of $Y$, which makes explicit an association rule between two discrete variables. For further examples and details see also Zenga (1991) and Arcagni (2017).

The Appendix A provides an example about how data must be pre-processed, and how the inequality index $\zeta$ can be computed through the filling of the cograduation table.

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 $y_{\tau(t)}$ and $y^*_{\tau^*(t)}$, providing $Z_t = \frac{y^*_{\tau^*(t)} - y_{\tau(t)}}{y^*_{\tau^*(t)}}$, as defined in formula (1), which can be used to calculate the index $\zeta$ by formula (2). In this setting, the functions $\tau^*(t)$ and $\tau(t)$ map $t$ into the position of the values of $Y^*$ and $Y$ associated to the $t^{th}$ cell with non-zero weight, respectively.

In the following we briefly report the main results about two decomposition procedures of the index $\zeta$ already proposed in the literature: the former one, introduced in Arcagni (2017), is about sources, while the latter, described in Porro and Zenga (2020), deals with subpopulations. Both are presented here according to the described notation.

**Theorem 1 (Decomposition by sources)** *Let the statistical variable $Y$ be the sum of $c$ sources (with $c \geq 2$), meaning that $Y = \sum_{j=1}^{c} X_j$. Then the inequality index $\zeta$ can be decomposed as*

$$\zeta = \sum_{t=1}^{NC} \sum_{j=1}^{c} H_t(X_j) w_t = \sum_{j=1}^{c} H_{\cdot}(X_j) \tag{3}$$

*where:*

- $H_{\cdot}(X_j) = \sum_{t=1}^{NC} H_t(X_j) w_t$ *is the contribution to the total inequality of the source $X_j$ (with $j = 1, \ldots c$);*

**Table 1** Distribution of the variable $Y$, according to the $k$ subpopulations

| $Y$ | Subpopulations | | | | |
|---|---|---|---|---|---|
| | $S_1$ | $S_2$ | $\cdots$ | $S_k$ | |
| $y_1$ | $n_{11}$ | $n_{12}$ | $\cdots$ | $n_{1k}$ | $n_{1\cdot}$ |
| $y_2$ | $n_{21}$ | $n_{22}$ | $\cdots$ | $n_{2k}$ | $n_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $y_r$ | $n_{r1}$ | $n_{r2}$ | $\cdots$ | $n_{rk}$ | $n_{r\cdot}$ |
| | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $\cdots$ | $n_{\cdot k}$ | $N$ |

- $H_t(X_j) = \frac{\bar{x}^*_{j\tau^*(t)} - \bar{x}_{j\tau(t)}}{y^*_{\tau^*(t)}}$ *is the contribution to the pointwise inequality measure* $Z_t$ *of the source* $X_j$ *(with* $j = 1, \dots c$*);*
- $\bar{x}_{j\tau(t)}$ $(\bar{x}^*_{j\tau^*(t)})$ *is the arithmetic mean of all the values of* $X_j$ *which concur by the different combinations of the sources to the value* $y_{\tau(t)}$ $(y^*_{\tau^*(t)}$*, respectively).*

*Remark 1* We have to highlight the importance of $\bar{x}^*_{j\tau(t)}$ and $\bar{x}_{j\tau(t)}$, since real dataset may show repeated values: it can happen that different units have the same value of the total variable $Y$, but different values of the $c$ sources (for further details see Section 6 in Arcagni, 2017).

**Theorem 2 (Decomposition by subpopulations)** *Let* $Y$ *be a statistical variable evaluated on a finite population of* $N$ *units, partitioned into* $k$ *subpopulations* $(k \geq 2)$*. Let* $n_{hl}$ *denote the frequency of the value* $y_h$ *in the* $l^{th}$ *subpopulation, with* $h = 1, \dots r$ *and* $l = 1, \dots, k$ *(see Table 1). Then the inequality index* $\zeta$ *can be decomposed as*

$$\zeta = \sum_{l=1}^{k} \sum_{g=1}^{k} \sum_{t=1}^{NC} Z_{tlg} w_t = \sum_{l=1}^{k} \sum_{t=1}^{NC} Z_{tl\cdot} w_t = \sum_{l=1}^{k} \sum_{g=1}^{k} Z_{\cdot lg} = \sum_{l=1}^{k} Z_{\cdot l \cdot} \qquad (4)$$

*where:*

- $Z_{\cdot l \cdot} = \sum_{g=1}^{k} \sum_{t=1}^{NC} Z_{tlg} w_t$ *is the contribution to the total inequality of the subpopulation* $S_l$ *(with* $l = 1, \dots k$*);*
- $Z_{\cdot lg} = \sum_{t=1}^{NC} Z_{tlg} w_t$ *is the contribution to the total inequality related to the couple of the subpopulations* $S_l$ *and* $S_g$ *(with* $l = 1, \dots k$*, and* $g = 1, \dots k$*);*
- $Z_{tl\cdot} = \sum_{g=1}^{k} Z_{tlg}$ *is the contribution to the pointwise measure* $Z_t$ *of the subpopulation* $S_l$ *(with* $l = 1, \dots k$*);*
- $Z_{tlg} = \frac{y^*_{\tau^*(t)} - y_{\tau(t)}}{y^*_{\tau^*(t)}} \cdot \frac{n_{\tau^*(t)g}}{n_{\tau^*(t)\cdot}} \cdot \frac{n_{\tau(t)l}}{n_{\tau(t)\cdot}}$ *is the contribution to the pointwise measure* $Z_t$ *of the couple of the subpopulations* $S_l$ *and* $S_g$ *(with* $l = 1, \dots k$*, and* $g = 1, \dots k$*).*

## 3 Joint decomposition

In this section we present a new multi-decomposition procedure, that can be considered a merge of the two decompositions described in the previous section. The aim is to provide a joint decomposition, which considers together sources and subpopulations. Like the other two, the proposed procedure follows a two-step approach: in particular, the first step is the decomposition of each inequality pointwise measure, taking into account $c$ sources and $k$ subpopulations. This means that each $Z_t$ is split in $c \times k$ values, one for each source in each subpopulation. At the second step, by averaging, the decomposition of the inequality index is obtained.

**Theorem 3 (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. Then the inequality index $\zeta$ can be decomposed as:*

$$\begin{aligned}\zeta &= \sum_{j=1}^{c}\sum_{l=1}^{k}\sum_{g=1}^{k}\sum_{t=1}^{NC} H_{tlg}(X_j)w_t = \sum_{j=1}^{c}\sum_{l=1}^{k}\sum_{t=1}^{NC} H_{tl\cdot}(X_j)w_t \\ &= \sum_{j=1}^{c}\sum_{l=1}^{k}\sum_{g=1}^{k} H_{\cdot lg}(X_j) = \sum_{j=1}^{c}\sum_{l=1}^{k} H_{\cdot l\cdot}(X_j). \qquad (5)\end{aligned}$$

*where*

- $H_{\cdot l\cdot}(X_j) = \sum_{g=1}^{k}\sum_{t=1}^{NC} H_{tlg}(X_j)w_t$ *is the contribution to the total inequality due to the source $X_j$ related to the subpopulation $S_l$;*
- $H_{\cdot lg}(X_j) = \sum_{t=1}^{NC} H_{tlg}(X_j)w_t$ *is the contribution to the total inequality due to the source $X_j$ related to the couple of the subpopulations $S_l$ and $S_g$;*
- $H_{tl\cdot}(X_j) = \sum_{g=1}^{k} H_{tlg}(X_j)$ *is the contribution to the pointwise measure $Z_t$ due to the source $X_j$ related to the subpopulation $S_l$;*
- $H_{tlg}(X_j) = \frac{\bar{x}^*_{j\tau^*(t)} - \bar{x}_{j\tau(t)}}{y^*_{\tau^*(t)}} \cdot \frac{n_{\tau^*(t)g}}{n_{\tau^*(t)\cdot}} \cdot \frac{n_{\tau(t)l}}{n_{\tau(t)\cdot}}$ *is contribution to the pointwise measure $Z_t$ due to the source $X_j$ related to the couple of the subpopulations $S_l$ and $S_g$.*

The detailed proof of Theorem 3 can be found in the Appendix B.

*Remark 2* It is worth mentioning that the decomposition described in formula (5) allows to split the value of $H_{\cdot}(X_j)$ and of $H_t(X_j)$, both defined in Theorem 1, in two $k \times k$ matrices, with entries $H_{\cdot lg}(X_j)$ and $H_{tlg}(X_j)$, respectively.

*Remark 3* From the joint decomposition, since all the sums are finite, their order can be exchanged, and the decomposition by subpopulation described in Theorem 2 can be easily achieved, by remarking that:

$$Z_{\cdot l\cdot} = \sum_{j=1}^{c}\sum_{g=1}^{k}\sum_{t=1}^{NC} H_{tlg}(X_j)w_t \qquad l = 1, \dots, k.$$

For the same reason, also the decomposition by sources described in Theorem 1 can be obtained, by noting that:

$$H_{\cdot}(X_j) = \sum_{l=1}^{k}\sum_{g=1}^{k}\sum_{t=1}^{NC} H_{tlg}(X_j)w_t \qquad j = 1, \dots, c.$$

## 4 Decompositions of $CO_2$ emissions

In 2008 the first commitment period of the Kyoto Protocol began. Thirty-seven industrialized countries and the European Union are committed to reduce their emissions by an average of 5% compared to 1990 by 2012. In this analysis we evaluate the situation of European countries in 2015.

Data are from the IEA, the International Energy Agency (2017). The datasets used in the applications and the R-code to replicate the analyses can be provided to the interested researchers upon request.

Following the approach proposed by Chien and Hu (2007) we divided the EU countries "...into OECD (developed) economies and non-OECD (developing) economies. The OECD members are considered more developed than other economies in the world, and so we use the status of membership in OECD as a proxy variable for being a developed economy". In the non-OECD subpopulation also some countries of the former USSR and Yugoslavia are included. Mussini and Grossi (2015) for the first time analyzed the inequality for $CO_2$ emissions in such countries. This analysis can provide useful information in case of enlargement of the EU through the inclusion of new members, since the EU neighboring countries are usually the optimal candidates. Our attention is devoted to the inequality in $CO_2$ emissions, by simultaneously considering two different aspects: the first one is the membership of the country in OECD, the second one pertains the sources which generate the $CO_2$ emissions. The considered sources are: *Coal*, *Oil*, *Natural Gas* and *Other*; *Other* includes industrial and non-renewable municipal waste.

As already mentioned, we measure the inequality by the Zenga-84 index $\zeta$. We propose a new joint decomposition by sources and by subpopulations, to better analyze simultaneously the inequality in each subgroup of countries and for each emission source, by applying the procedure described in Section 3. From that, as showed in the Remark 3, we can obtain as particular cases the decomposition by sources, proposed in Arcagni (2017), and the decomposition by subpopulations, introduced in Porro and Zenga (2020).

We consider 51 countries: 26 of them are OECD countries, the other 25 non-OECD. See Appendix C for the complete lists of the members of the two subpopulations. Tables 2 and 3 show some descriptive statistics of the examined variables: all of them are expressed in tonnes $CO_2$ per capita. The last column of Table 2 reports the value of the Zenga-84 inequality index, computed for each source, singularly considered, by applying formula (2). The $Z(p)$ curves for the single four sources, and the one of the *Total Emissions* are showed in Figure 1. All these curves have been computed by following the formula (2). The amount of $CO_2$ emissions due to the source *Other* is not deeply analyzed in the following discussion, since it just represents a residual part of the *Total Emissions*.

For the whole population of the countries, the mean of the *Total Emissions* is equal to 5.999 tonnes per capita, and the standard deviation (SD) is 3.496; the median is lower than the mean, denoting a positive asymmetry. Indeed, the distribution of all the variables in both the subpopulations have positive

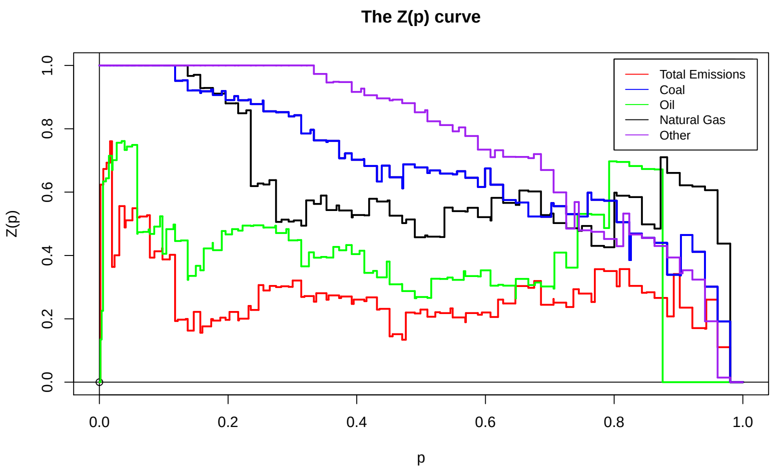


**Fig. 1** The $Z(p)$ curve of the *Total Emissions* (in red), and of the analyzed sources, singularly considered.

| *Variable* | *Min* | *Max* | *Median* | *Mean* | *SD* | $\zeta$ |
|---|---|---|---|---|---|---|
| *Coal* ($X_1$) | 0 | 8.548 | 1.181 | 1.742 | 1.930 | 0.6750 |
| *Oil* ($X_2$) | 0.230 | 17.390 | 2.357 | 2.735 | 2.790 | 0.4522 |
| *Natural Gas* ($X_3$) | 0 | 9.324 | 1.170 | 1.431 | 1.604 | 0.6315 |
| *Other* ($X_4$) | 0 | 0.435 | 0.035 | 0.091 | 0.119 | 0.7513 |
| *Total Emissions* ($Y$) | 0.510 | 17.390 | 5.450 | 5.999 | 3.496 | 0.2707 |

**Table 2** Descriptive statistics of the whole population of countries.

| *Variable* | *Subpopulation* | *Min* | *Max* | *Median* | *Mean* | *SD* |
|---|---|---|---|---|---|---|
| *Coal* ($X_1$) | OECD | 0.060 | 8.548 | 1.328 | 1.998 | 1.954 |
| | non-OECD | 0 | 6.908 | 0.323 | 1.476 | 1.907 |
| *Oil* ($X_2$) | OECD | 1.194 | 11.603 | 2.845 | 3.152 | 1.949 |
| | non-OECD | 0.230 | 17.390 | 1.212 | 2.303 | 3.446 |
| *Natural Gas* ($X_3$) | OECD | 0 | 3.672 | 1.269 | 1.434 | 0.874 |
| | non-OECD | 0 | 9.324 | 0.881 | 1.427 | 2.136 |
| *Other* ($X_4$) | OECD | 0 | 0.435 | 0.124 | 0.158 | 0.126 |
| | non-OECD | 0 | 0.260 | 0 | 0.021 | 0.056 |
| *Total Emissions* ($Y$) | OECD | 3.460 | 15.470 | 6.095 | 6.742 | 2.691 |
| | non-OECD | 0.510 | 17.390 | 3.800 | 5.227 | 4.087 |

**Table 3** Descriptive statistics of the two subpopulations (OECD and non-OECD) of countries.

asymmetry. The most important source of $CO_2$ emissions is *Oil* (which represents on average the 45.6%), followed by *Coal* (29%) and *Natural Gas* (23.8%). *Oil* is the source with the highest variability, meaning that it is the source with the most significant value differences: this can be noted also by remarking that *Oil* is the source with the biggest range.

From Table 3 it arises that for OECD countries the mean of per capita emissions is equal to 6.742 tonnes, with a range of 12.010 and a SD equal to 2.691; the median is close to the mean. Also in this subpopulation, *Oil* has the highest contribution to the *Total Emissions*. The lowest variability can be found for the source *Natural Gas*. The values of all the emissions are higher on average for OECD countries than for the whole population of countries, but the ranges and the variability are lower, denoting a more similar situation in OECD countries.

Conversely, the mean of emissions for non-OECD countries is equal to 5.227, lower than the overall mean. The median equal to 3.800 and the high value of SD (4.087) denote that there is a higher variability in the emissions in the non-OECD countries, with a very important positive asymmetry in the distribution. For the non-OECD countries, the more severe positive asymmetries correspond to the sources *Coal* and *Oil*.

The difference of the means for the *Total Emissions* in OECD and non-OECD countries is not surprising, taking into account the different levels of development in the two subpopulations.

The behavior of the source *Natural Gas* is particular: the mean is very similar for OECD countries (1.434) and non-OECD (1.427), but the median for non-OECD countries is remarkable lower than the one for OECD countries. Furthermore, the comparison of the SD for the two subpopulations allows to assert that the highest variability in the $CO_2$ emissions from *Natural Gas* is for non-OCED countries: also this can be related to more similar standards and restrictions adopted in OECD countries. An analogous interpretation holds for the source *Oil*, with a SD equal to 3.446 for non-OECD countries. The variability for *Coal* is similar for OECD and non-OECD. The comparison of the mean and the median suggests the presence of many non-OECD countries with very low levels of emissions due to *Coal*. Figure 2 displays the graphical representations of the source distributions for the two subpopulations.

The calculation of the cograduation matrix provides a number $NC = 101$ of cells with non-zero weight. The value of Zenga-84 inequality index for the *Total Emissions* related to all the countries is equal to $\zeta = 0.2707$. This value is not high and it denotes a moderate level of inequality in $CO_2$ emissions for the total analyzed population. For this reason, we do not expect to find in our analysis a subpopulation and/or a source with a high level of emission inequality.

By applying the decomposition proposed in Theorem 3, we obtain four (as the number of the considered sources) $2 \times 2$ matrices, each of them representing the contribution to the total inequality due to each source. They are all stored in Table 4.

In each matrix, the sum of the values in the central column represents the contribution due to the corresponding source to the total inequality related to subpopulation of OECD countries: the most relevant is due to the source *Coal*, with a value of 0.0692. The sum of the values in the last column is the contribution related to the non-OECD countries: the most significant one (that is equal to 0.0982) is due to the source *Oil*. It is worth remarking that

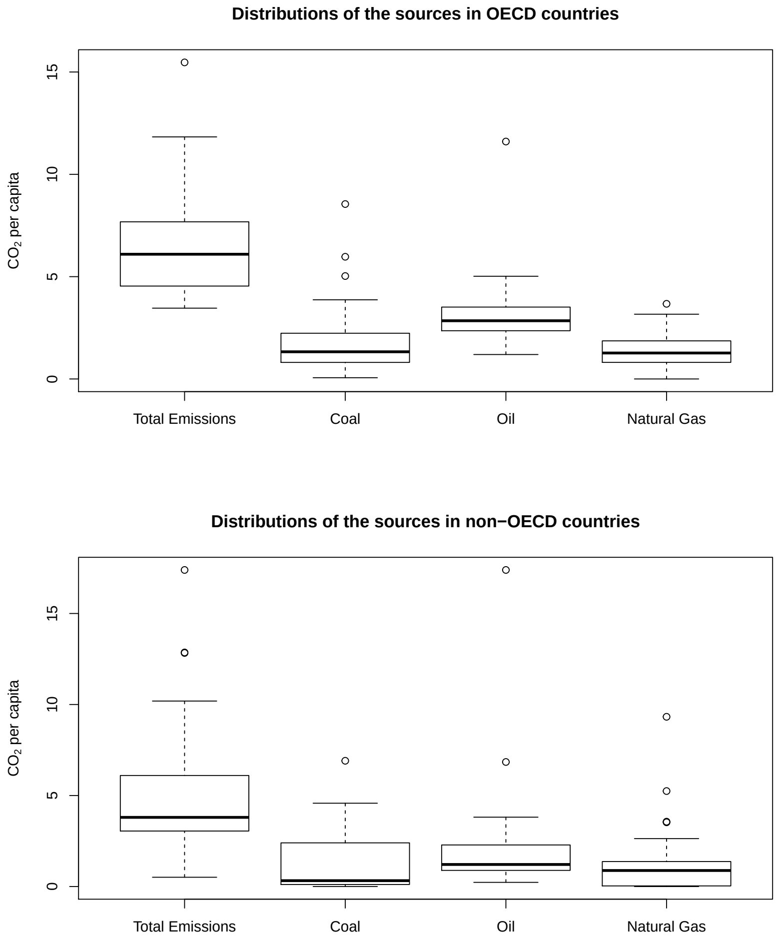


**Fig. 2** Boxplots of the variable *Total emissions* and of the sources *Coal*, *Oil*, and *Natural Gas* for OECD and non-OECD countries.

| *Coal* $(X_1)$ | *OECD* | *non-OECD* |
|---|---|---|
| *OECD* | 0.0470 | -0.0185 |
| *non-OECD* | 0.0222 | 0.0326 |
| *Total* | *0.0692* | *0.0141* |

| *Oil* $(X_2)$ | *OECD* | *non-OECD* |
|---|---|---|
| *OECD* | 0.0189 | 0.0608 |
| *non-OECD* | 0.0087 | 0.0374 |
| *Total* | *0.0276* | *0.0982* |

| *Natural Gas* $(X_3)$ | *OECD* | *non-OECD* |
|---|---|---|
| *OECD* | 0.0108 | 0.0211 |
| *non-OECD* | 0.0232 | 0.0017 |
| *Total* | *0.0340* | *0.0228* |

| *Other* $(X_4)$ | *OECD* | *non-OECD* |
|---|---|---|
| *OECD* | 0.0013 | 0.0064 |
| *non-OECD* | -0.0029 | $\approx 0$ |
| *Total* | *-0.0016* | *0.0064* |

**Table 4** The four $2 \times 2$ decomposition matrices for the $CO_2$ emissions sources.

| *Total Emissions* $(Y)$ | *OECD* | *non-OECD* |
|---|---|---|
| *OECD* | 0.0780 | 0.0698 |
| *non-OECD* | 0.0512 | 0.0717 |
| *Contribution to the total inequality:* $Z_{.l.}$ | 0.1292 (47.73%) | 0.1415 (52.27%) |

**Table 5** The *decomposition by subpopulations* matrix for the total inequality.

| *Source* | *Subpopulation* | | *Contribution to the total inequality:* $H_{\cdot}(X_j)$ |
|---|---|---|---|
| | *OECD* | *non-OECD* | |
| *Coal* $(X_1)$ | 0.0692 (83.07%) | 0.0141 (16.93%) | 0.0833 (100%) |
| *Oil* $(X_2)$ | 0.0276 (21.94%) | 0.0982 (78.06%) | 0.1258 (100%) |
| *Natural Gas* $(X_3)$ | 0.034 (59.86%) | 0.0228 (40.14%) | 0.0568 (100%) |

**Table 6** Partition among the subpopulations of the contributions of the sources.

three values:

$$H_{\cdot 21}(X_1) = -0.0185, \quad H_{\cdot 12}(X_4) = -0.0029, \quad H_{\cdot 1\cdot}(X_4) = -0.0016$$

turn out to be negative: this can happen in the joint decomposition, because for all $t$, $y^*_{\tau^*(t)}$ is always greater or equal to $y_{\tau(t)}$ (and therefore $Z_t \geq 0$), but it is not true that $\bar{x}^*_{j\tau^*(t)}$ is always greater or equal to $\bar{x}_{j\tau(t)}$. Here the value $H_{\cdot 1\cdot}(X_4) = -0.0016$ means that the source *Other* $(X_4)$ has a mitigating impact on the total inequality in the subpopulation OECD.

From the joint decomposition, the *decomposition by subpopulations* can be easily achieved. By summing the four matrices in Table 4 we obtain the decomposition by subpopulations matrix in Table 5, with the values of $Z_{\cdot 1\cdot} = 0.1292$ for OECD countries, and $Z_{\cdot 2\cdot} = 0.1415$ for non-OECD ones. They show that the contribution to the total inequality of $CO_2$ emissions is higher for non-OECD countries than for OECD ones, which adopt a more similar behavior, since they must follow shared and more restrictive guidelines.

If we increase the depth of the analysis, we can easily obtain the Table 6 from the data in Table 4: it contains the repartitions among the subpopulations of the contribution due to each source. From the Figure 3 it can be seen that such repartitions are very different: for the source *Oil*, and for the *Total Emissions*, the most relevant part is due to the non-OECD countries, while for the other sources the opposite is true. An interpretation is that the sources *Coal* and *Natural Gas* contribute to the total inequality with a more relevant weight in the OECD countries than in the other countries; while the source *Oil* counts much more (almost four times more) in the non-OECD countries. We would like to stress that all these information cannot be obtained by the two decompositions by sources and by subpopulations: only a joint decomposition allows to reach such deep detail level. This analysis can be very useful, since it indicates where the policymakers should act to reduce the inequality of $CO_2$

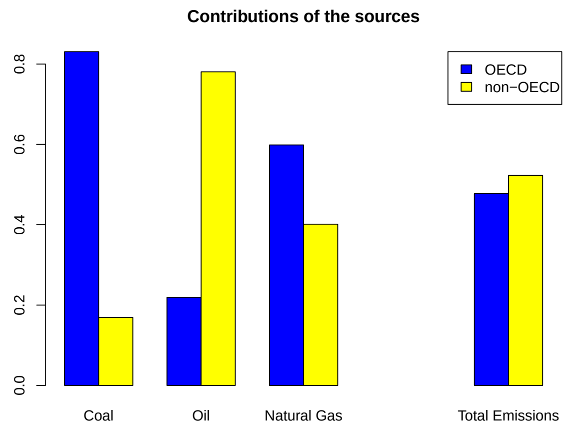


**Fig. 3** Representation of the partition among the subpopulations of the contributions of the sources.

emissions: it identifies exactly which source and which group of countries need an action.

An important decomposition is the classical one, in Between and Within components. Summing all the values in the four principal diagonals of the previous four matrices in Table 4 of the joint decomposition, we obtain the Within component; summing all the remaining values, we calculate the Between component. These quantities are respectively equal to:

$$\zeta_W = 0.1497 \qquad \text{and} \qquad \zeta_B = 0.121.$$

The Within component represents the 55.3% of the total inequality and the Between one the other 44.7%. In other words, the contribution to the total inequality due to comparisons of values in the same subpopulations is higher than the part related to the comparisons of countries in different subpopulations. This is a quite surprising result, since we could expect a more similar behavior within the two subpopulations. Probably, the sources have different performance into the two subpopulations and they contribute to obtain a Within component higher than the Between one. The joint decomposition performed can help to investigate this aspect, since it allows to assess the contribution to the Within and Between components due to each source.

From the matrices in Table 4, we can obtain the Table 7 and the chart in the Figure 4. It is easy to see that the repartitions in Within and Between components are very different for the four sources: the contribution of the source *Coal* is almost totally due to comparisons of values in the same subpopulations, while for the other sources, the disparity across the subpopulations, due to comparisons of values in different subpopulations, counts more. As expected, the repartition of the *Total Emissions Y* is a sort of average of those of all the sources.

| *Variable* | *Component* | | *Contribution to the total inequality:* $H.(X_j)$ |
|---|---|---|---|
| | *Within* | *Between* | |
| *Coal* ($X_1$) | 0.0796 (95.55%) | 0.0037 (4.45%) | 0.0833 (100%) |
| *Oil* ($X_2$) | 0.0563 (44.75%) | 0.0695 (55.25%) | 0.1258 (100%) |
| *Natural Gas* ($X_3$) | 0.0125 (22%) | 0.0443 (78%) | 0.0568 (100%) |
| *Other* ($X_4$) | 0.0013 (27.08%) | 0.0035 (72.92%) | 0.0048 (100%) |
| *Total Emissions* ($Y$) | 0.1497 (55.3%) | 0.121 (44.7%) | 0.2707 (100%) |

**Table 7** Within and Between components of the contribution related to each source.

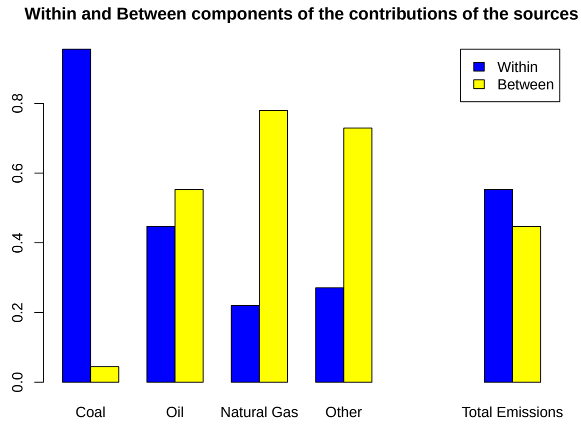


**Fig. 4** Representation of the partition among Within and Between components of the contributions of the sources.

From the joint decomposition, the *decomposition by sources* can also be achieved, identifying the contribution to the total inequality due to each source: the values obtained are reported in Table 8. From such data, it arises that the highest contribution to the total inequality of *Total Emissions* is due to the source *Oil* (for the 46.5%); *Coal* and *Natural Gas* follow with percentages equal to 30.8% and 20.9%, respectively. All these values regards the whole population: we can achieve more specific information, at subpopulation level, by using the results of the joint decomposition. From Table 4, the Table 9 can be filled. The advantage is that the contribution of each source is split among the subpopulations: in other words, we can assess the contribution of each source in each subpopulation. The data, illustrated in the Figure 5, show that the source *Oil* largely dominates the others in the contribution to the inequality in non-OECD countries (with 69.40%), while in OECD ones its relevance is much reduced (21.36%). The most important source in the OECD countries is

| *Variable* | *Contribution to the total inequality:* $H.(X_j)$ |
|---|---|
| *Coal* $(X_1)$ | 0.0833 (30.8%) |
| *Oil* $(X_2)$ | 0.1258 (46.5%) |
| *Natural Gas* $(X_3)$ | 0.0568 (20.9%) |
| *Other* $(X_4)$ | 0.0048 (1.8%) |
| *Total Emissions* $(Y)$ | 0.2707 (100%) |

**Table 8** The *decomposition by sources* matrix for the total inequality.

| *Variable* | *Subpopulation* | |
|---|---|---|
| | *OECD* | *non-OECD* |
| *Coal* $(X_1)$ | 0.0692 (53.56%) | 0.0141 (9.97%) |
| *Oil* $(X_2)$ | 0.0276 (21.36%) | 0.0982 (69.40%) |
| *Natural Gas* $(X_3)$ | 0.034 (26.32%) | 0.0228 (16.11%) |
| *Other* $(X_4)$ | -0.016 (-1.24%) | 0.0064 (4.52%) |
| *Total Emissions* $(Y)$ | 0.1292 (100%) | 0.1415 (100%) |

**Table 9** Partition among the sources of the contributions of the subpopulations.

*Coal* (53.56%): in non-OECD countries, the role of this source is very different, since it represents only the 9.97% of the contribution of that subpopulation. As expected, similarly as before, the repartition of the contributions of the whole population is in the middle between those of the two subpopulations.

## 5 Conclusions

The contribution of the article is twofold. First, we introduce a joint decomposition of Zenga-84 inequality index $\zeta$, splitting its value by sources and by subpopulations, simultaneously. This result generalizes the decompositions of such index, already introduced in the literature.

Second, we analyze the disparity of the $CO_2$ emissions for European countries in 2015, considering jointly the sources (*Coal*, *Oil*, *Natural Gas*, and *Other*) and the development level of the countries. For this task we use the $\zeta$ inequality index, since it allows to evaluate also little changes in inequality, observed in different situations. Applying the decomposition by subpopulations, we can argue that the contribution to the inequality of $CO_2$ emissions is higher for non-OECD countries ($Z_{\cdot 2 \cdot} = 0.1415$) than for OECD ($Z_{\cdot 1 \cdot} = 0.1292$): perhaps that is caused by more similar standards and restrictions that OECD

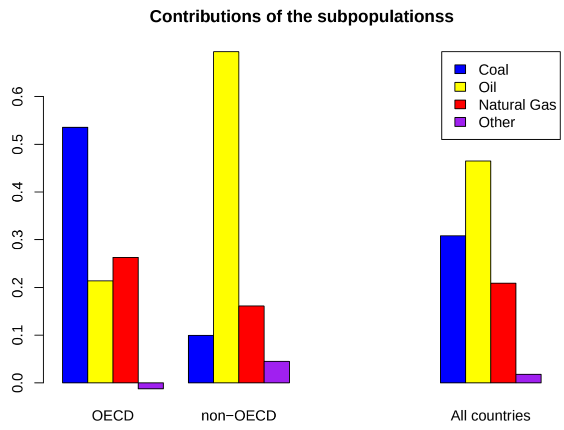


**Fig. 5** Representation of the partition among the sources of the contributions of the subpopulations.

countries agreed to respect. The decomposition by sources shows that the most relevant contribution to the inequality in $CO_2$ emissions is due to the source *Oil* (46.5%), followed by *Coal* (30.8%) and *Natural Gas* (20.9%). The proposed joint decomposition allows a deeper investigation, because it highlights that the behavior of the sources in each subpopulation is different, and specularly the partition of the subpopulations in each source is dissimilar. It shows indeed that for OECD countries the first source of the disparity in $CO_2$ emissions is *Coal*, while for non-OECD countries is *Oil*. This information provides to the policymakers the intervention areas to be considered for reducing the inequality of $CO_2$ emissions. In conclusion, supported by the information gain obtained by the multi-decomposition, we believe that the proposed procedure can be considered a valuable methodology that can help to improve the analysis of a large number of real and complex phenomena.

## A Example of pre-processing data procedure, computation of the cograduation table, and calculation of the index $\zeta$

The starting point is the original data matrix, where the values of the variable $Y$ and of the sources are stored, according to the splitting among the subpopulations. A numerical example is in the Table 10, containing $N = 20$ units, from $k = 3$ different subpopulations, and the values of the variable $Y$, obtained by the sum of $c = 2$ sources (labelled by $X_1$ and $X_2$). For each unit are reported in the last three columns the indicator variables ($Subpop_1, Subpop_2$ and $Subpop_3$), with value equal to 1 if the unit belongs to the corresponding subpopulation and zero otherwise.

| $i$ | $Y$ | $X_1$ | $X_2$ | $Subpop_1$ | $Subpop_2$ | $Subpop_3$ |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | 0 | 0 | 0 | 0 | 0 | 1 |
| 8 | 10 | 5 | 5 | 0 | 1 | 0 |
| 9 | 10 | 9 | 1 | 0 | 1 | 0 |
| 10 | 10 | 4 | 6 | 0 | 1 | 0 |
| 11 | 15 | 10 | 5 | 0 | 1 | 0 |
| 12 | 15 | 2 | 13 | 0 | 1 | 0 |
| 13 | 15 | 6 | 9 | 0 | 0 | 1 |
| 14 | 15 | 3 | 12 | 0 | 1 | 0 |
| 15 | 15 | 9 | 6 | 1 | 0 | 0 |
| 16 | 15 | 6 | 9 | 1 | 0 | 0 |
| 17 | 24 | 20 | 4 | 1 | 0 | 0 |
| 18 | 24 | 14 | 10 | 0 | 1 | 0 |
| 19 | 24 | 11 | 13 | 0 | 1 | 0 |
| 20 | 28 | 10 | 18 | 1 | 0 | 0 |
| *Total:* | | | | *6* | *10* | *4* |

**Table 10** Original Data Matrix.

| $y_h$ | $n_{h\cdot}$ | $\bar{x}_{1h}$ | $\bar{x}_{2h}$ | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|---|---|---|
| 0 | 7 | 0 | 0 | 2 | 2 | 3 |
| 10 | 3 | 6 | 4 | 0 | 3 | 0 |
| 15 | 6 | 6 | 9 | 2 | 3 | 1 |
| 24 | 3 | 15 | 9 | 1 | 2 | 0 |
| 28 | 1 | 10 | 18 | 1 | 0 | 0 |
| *Total:* | *20* | | | *6* | *10* | *4* |

**Table 11** Derived Data Matrix.

Now, the data can be easily synthetized in Table 11, which consists of the frequency distribution of $Y$ and the frequencies $n_{hl}$ regarding the subpopulations (as in Table 1). The two central columns (denoted by $\bar{x}_{1h}$ and $\bar{x}_{2h}$) contain the averages of the values of each source $X_1$ and $X_2$, corresponding to the same value of $y_h$ (cf. Section 6 in Arcagni, 2017).

By using the first two columns of Table 11, the cograduation table can be obtained. The first step is a table with in the first column the different positive values $y_i^*$ and in the last column the corresponding shares:

$$s_i = \frac{y_i n_{i\cdot}}{\sum_{h=1}^{r} y_h n_{h\cdot}}.$$

The bottom row is filled with the values $y_h$, and the first row with the corresponding relative frequencies $freq(y_h)$:

$$freq(y_h) = \frac{n_{h\cdot}}{N}.$$

The inner cells are filled using the following procedure. First of all, the cell in the lower left-hand corner of the table, corresponding to the couple ($y_1^* = 10, y_1 = 0$) is filled with $s_1 = 0.136$. After that, the second cell to be completed is the one corresponding to the couple ($y_2^* = 15, y_1 = 0$). There, it is allocated 0.214, which is the smallest of two values: $s_2 = 0.409$ and the "remaining frequency" of $y_1$ ($0.35 - 0.136 = 0.214$). At this point the first column is saturated, since all the first relative frequency $freq(y_1) = 0.35$ has been used. Then the procedure moves to the adjacent cell, corresponding to the couple ($y_2^* = 15, y_2 = 10$). Using the same rule as before, the cell is filled with 0.15, since 0.15 is the minimum between the "remaining share" of $y_2^*$ ($0.409-0.214 = 0.195$) and the frequency of $y_2$, which is 0.15. Then the procedure moves to the cell on the right ($y_2^* = 15, y_3 = 15$), and it is filled by the remaining share of $s_2$: $0.409 - 0.214 - 0.15 = 0.045$, and so on. The procedure stops when the upper right-hand cell is filled with the last frequency $freq(y_5) = 0.05$. The result is the following table in which a weight is associated to each couple $(y_i^*, y_h)$.

| | | $freq(y_h)$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.35 | 0.15 | 0.30 | 0.15 | 0.05 | 1 | |
| | 28 | - | - | - | 0.078 | 0.05 | 0.128 | |
| | 24 | - | - | 0.255 | 0.072 | - | 0.327 | |
| $y_i^*$ | 15 | 0.214 | 0.15 | 0.045 | - | - | 0.409 | $s_i$ |
| | 10 | 0.136 | - | - | - | - | 0.136 | |
| | | 0 | 10 | 15 | 24 | 28 | | |
| | | | | $y_h$ | | | | |

In this cograduation table, there are $NC = 8$ cells with non-zero weight. The inequality index $\zeta$ can be calculated, using the information provided by Table 12, where:

- $t$ counts the cells with non-zero weight;
- $w_t$ is the weight associated to the $t$-th cell;
- $p_t = \sum_{m=1}^{t} w_m$ is the cumulative weight;
- $y^*_{\tau^*(t)}$ and $y_{\tau(t)}$ are the values of $Y^*$ and $Y$ associated to the $t$-th cell with non-zero weight, since the two functions $\tau^*$ and $\tau$ are defined by:

$$\begin{aligned} \tau : \{1, 2, \ldots, 8\} &\to \{1, 2, \ldots, 5\} \\ t &\mapsto \tau(t) \\ \tau^* : \{1, 2, \ldots, 8\} &\to \{1, 2, \ldots, 5\} \\ t &\mapsto \tau^*(t) \end{aligned}$$

  such that $y_{\tau(t)}$ and $y^*_{\tau^*(t)}$ are the values of $Y$ and $Y^*$, respectively, corresponding to the $t$-th cell with non-zero weight of the cograduation table;
- $Z_t$ is the value assumed by the function $Z$ at the point $p_t$, calculated through formula (1).

Finally, the sum of the values in the last column of Table 12 provides the value of the index $\zeta$:

$$\zeta = \sum_{t=1}^{8} Z_t w_t = 0.5317634. \tag{6}$$

| $t$ | $w_t$ | $p_t$ | $y^*_{\tau^*(t)}$ | $y_{\tau(t)}$ | $Z_t$ | $Z_t w_t$ |
|---|---|---|---|---|---|---|
| 1 | 0.136 | 0.136 | 10 | 0 | 1 | 0.136 |
| 2 | 0.214 | 0.35 | 15 | 0 | 1 | 0.214 |
| 3 | 0.15 | 0.5 | 15 | 10 | 0.333 | 0.075 |
| 4 | 0.045 | 0.545 | 15 | 15 | 0 | 0 |
| 5 | 0.255 | 0.8 | 24 | 15 | 0.375 | 0,095625 |
| 6 | 0.072 | 0.872 | 24 | 24 | 0 | 0 |
| 7 | 0.078 | 0.95 | 28 | 24 | 0.1428 | 0,0111384 |
| 8 | 0.05 | 1 | 28 | 28 | 0 | 0 |
| | 1 | | | | | 0.5317634 |

**Table 12** Calculation of the values of the function $Z$ and the index $\zeta$.

## B Proof of Theorem 3.

As stated in Theorem 1, we can decompose the pointwise inequality measure $Z_t$ by sources as follows:

$$Z_t = \frac{y^*_{\tau^*(t)} - y_{\tau(t)}}{y^*_{\tau^*(t)}} = \sum_{j=1}^{c} \frac{\bar{x}^*_{j\tau^*(t)} - \bar{x}_{j\tau(t)}}{y^*_{\tau^*(t)}} = \sum_{j=1}^{c} H_t(X_j) \tag{7}$$

where $H_t(X_j)$ is the contribution to $Z_t$ of $X_j$, $\bar{x}_{j\tau(t)}$ and $\bar{x}^*_{j\tau^*(t)}$ are the arithmetic means of all the values of $X_j$ which concur by the different combinations to the values $y_{\tau(t)}$, and $y^*_{\tau^*(t)}$, respectively.

Consider now the $k$ subpopulations: by applying the decomposition procedure described in Theorem 2 at the contributions $H_t(X_j)$, we can obtain a $k \times k$ matrix for each value $H_t(X_j)$. In particular, for each $t \in \{1, \dots, NC\}$, and for each $l \in \{1, \dots, k\}$, let

$$p(l|t) = \frac{n_{\tau(t)l}}{n_{\tau(t).}}$$

be the relative frequency (corresponding to the value $y_{\tau(t)}$) of the subpopulation $l$, and for each $g \in \{1, 2, \dots, k\}$ let

$$a(g|t) = \frac{n_{\tau^*(t)g}}{n_{\tau^*(t).}}$$

be the relative frequency (corresponding to the value $y^*_{\tau^*(t)}$) of the subpopulation $g$. By setting these weights as marginal distributions, the inner cells of the following table can be fulfilled by the products of the marginal cells:

| | | $y$: $S_1$ | $S_2$ | $\cdots$ | $S_k$ | |
|---|---|---|---|---|---|---|
| $y^*$ | $S_1$ | $a(1\|t)p(1\|t)$ | $a(1\|t)p(2\|t)$ | $\cdots$ | $a(1\|t)p(k\|t)$ | $a(1\|t)$ |
| | $S_2$ | $a(2\|t)p(1\|t)$ | $a(2\|t)p(2\|t)$ | $\cdots$ | $a(2\|t)p(k\|t)$ | $a(2\|t)$ |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\cdots$ | $\vdots$ | $\vdots$ |
| | $S_k$ | $a(k\|t)p(1\|t)$ | $a(k\|t)p(2\|t)$ | $\cdots$ | $a(k\|t)p(k\|t)$ | $a(k\|t)$ |
| | | $p(1\|t)$ | $p(2\|t)$ | $\cdots$ | $p(k\|t)$ | 1 |

It is worth noting that for a fixed $t \in \{1, \ldots NC\}$ the aforementioned table is the same for every $j \in \{1, \ldots, c\}$, since the couple of the values $(y^*_{\tau^*(t)}, y_{\tau(t)})$ is the same for all the sources.
By replicating this procedure, for each $t \in \{1, \cdots, NC\}$, we can spread the value of each $H_t(X_j)$ among the subpopulations. We can then assign a value to each ordered pair of subpopulations (say $S_l$ and $S_g$), denoted by $H_{tlg}(X_j)$, which is the portion of the contribution $H_t(X_j)$ related to such ordered couple:

$$H_{tlg}(X_j) = H_t(X_j)a(g|t)p(l|t). \tag{8}$$

Now, we can perform the second step of the procedure: it consists of the aggregation of all the split pointwise inequality measures defined in formula 8. Since it holds that

$$\sum_{l=1}^{k} \sum_{g=1}^{k} H_{tlg}(X_j) = H_t(X_j)$$

the joint decomposition is achieved:

$$\zeta = \sum_{j=1}^{c} H_{\cdot}(X_j) = \sum_{j=1}^{c} \sum_{t=1}^{NC} H_t(X_j) w_t = \sum_{j=1}^{c} \sum_{l=1}^{k} \sum_{g=1}^{k} \sum_{t=1}^{NC} H_{tlg}(X_j) w_t.$$

## C The analyzed subpopulations

*Europe OECD countries*: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom.

*Europe non-OECD countries*: Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, FYR of Macedonia, Georgia, Gibraltar, Kazakhstan, Kosovo, Kyrgyzstan, Lithuania, Malta, Republic of Moldova, Montenegro, Romania, Russian Federation, Serbia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan.

## References

Aly, E. and Hervas, M. (1999), 'Nonparametric inference for Zenga's measure of income inequality', *Metron* **52**, 69–84.

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

Arnold, B. (2015), 'On Zenga and Bonferroni curves', *Metron* **73**(1), 25–30.

Arnold, B. C. and Sarabia, J. (2018), *A Majorization and the Lorenz order with applications in applied mathematics and economics*, Cham, Switzerland, Springer.

Battistin, E., Blundell, R. and Lewbel, A. (2009), 'Why is consumption more log normal than income? Gibrat's law revisited', *Journal of Political Economy* **117**(6), 1140–1154.

Belzunce, F., Candel, J. and Ruiz, J. M. (1995), 'Ordering of truncated distributions through concentration curves', *Sankhya - The Indian Journal of Statistics - Series A* **57**(3), 375–383.

Berti, P. and Rigo, P. (1995), 'A note on Zenga concentration index', *Journal of the Italian Statistical Society* **4**(3), 397–404.

Berti, P. and Rigo, P. (2006), Concentration curve and index, Zenga's, *in* N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. Teugels, eds, 'Wiley StatsRef: Statistics Reference Online', John Wiley & Sons, Inc.

Bertoli-Barsotti, L. (2001), 'Some remarks on Lorenz ordering-preserving functionals', *Statistical Methods & Applications* **10**, 99–112.

Bourguignon, F. (1979), 'Decomposable income inequality measures', *Econometrica* **47**, 901–920.

Chien, T. and Hu, J. L. (2007), 'Renewable energy and macroeconomic efficiency of OECD and non-OECD economies', *Energy Policy* **35**, 3606–3615.

Dagum, C. (1997*a*), 'A new approach to the decomposition of the Gini income inequality ratio', *Empirical Economics* **22**, 515–531.

Dagum, C. (1997*b*), 'Scomposizione ed interpretazione della misura di disuguaglianza di Gini e di entropia generalizzata', *Statistica* **3**, 295–307.

Ebert, U. (2010), 'The decomposition of inequality reconsidered: weakly decomposable measures', *Mathematical Social Sciences* **60**, 94–103.

Eurostat (2018), 'Gini coefficient of equivalised disposable income - EU-SILC survey'.
**URL:** *https://ec.europa.eu/eurostat/web/income-and-living-conditions*

Fernandez Morales, A., De Haro Garcia, J. and Martin Reyes, G. (1996), 'Medicion de la desigualdad y el bienestar social. analisis de la distribucion de la renta en España. (1981-1991)', *Revista de Estudios Regionales* **45**, 15–42.

Fiori, A. M. and Porro, F. (2020), 'A method to decompose the systemic risk in geographic areas', *Soft Computing* **24**, 8599–8606.

Frosini, B. V. (2012), 'Approximation and decomposition of Gini, Pietra-Ricci and Theil inequality measures', *Empirical Economics* **43**, 175–197.

International Energy Agency (2017), '$CO_2$ emissions from fuel combustion', *Highlights IEA Publishing* .

Jedrzejczak, A. (2013), Asymptotic properties of some estimators for income inequality measures - a simulation study, *in* 'Proceedings 59th ISI World Statistics Congress, CPS203 (2013)', International Statistical Institute.

Jedrzejczak, A. (2015), 'Asymptotic properties of some estimators for Gini and Zenga inequality measures: a simulation study', *Statistica & Applicazioni* **13**(2), 143–162.

Jedrzejczak, A. and Kubacki, J. (2013), 'Estimation of income inequality and the poverty rate in Poland, by region and family type', *Statistics in Transition - New series* **14**(3), 359–378.

Kleiber, C. and Kotz, S. (2003), *Statistical size distributions in economics and actuarial sciences*, John Wiley & Sons, Hoboken, NJ.

Lerman, R. and Yitzhaki, S. (1984), 'A note on the calculation and the interpretation of the Gini index', *Economics Letters* **15**(3-4), 363–368.

Lerman, R. and 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**(1), 151–156.

Mehran, F. (1975), A statistical analysis of income inequality based on a decomposition of the Gini index, *in* 'Proceedings of the 40-th Session of the I.S.I. (Warsaw)', Vol. 4, pp. 146–150.

Mookherjee, D. and Shorrocks, A. (1982), 'A decomposition analysis of the trend in UK income inequality', *The Economic Journal* **92**(368), 886–902.

Mussard, S. (2004), 'The bidimensional decomposition of the Gini ratio. A case study: Italy', *Applied Economics Letters* **11**, 503–505.

Mussini, M. and Grossi, L. (2015), 'Decomposing changes in $CO_2$ emission in inequality over time: the roles of re-ranking and changes in per capita $CO_2$ emission disparities', *Energy Economics* **49**, 274–281.

Porro, F. and Zenga, M. (2020), 'Decomposition by subpopulations of the Zenga-84 inequality curve and the related index $\zeta$: an application to 2014 Bank of Italy survey', *Statistical Methods & Applications* **29**, 187–207.

Porro, F. and Zenga, M. (2021), 'Decompositions by sources and by subpopulations of the Pietra index: two applications to professional football teams in Italy', *AStA Advances in Statistical Analysis - Online first article* .
**URL:** *https://link.springer.com/article/10.1007/s10182-021-00397-6*

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.

Radaelli, P. and Zenga, M. (2005), 'On the decomposition of the Gini's mean difference and concentration ratio', *Statistica & Applicazioni* **3**(2), 5–24.

Rao, V. (1969), 'Two decompositions of concentration ratio', *Journal of Royal Statistical Society - Serie A (General)* **132**, 418–425.

Shorrocks, A. F. (1980), 'The class of additively decomposable inequality measures', *Econometrica* **48**(3), 613–625.

Shorrocks, A. F. (1982), 'Decomposition by factor components', *Econometrica* **50**(1), 193–211.

Shorrocks, A. F. (1984), 'Inequality decomposition by population subgroups', *Econometrica* **52**(6), 1369–1385.

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. (1990), Concentration curves and concentration indexes derived from them, *in* C. Dagum and M. Zenga, eds, 'Income and Wealth Distribution, Inequality and Poverty', Springer-Verlag, Berlin Heidelberg.

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. (1993), 'Il principio dei trasferimenti', *Statistica* **53**(4), 647–660.

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

Zenga, M. and Valli, I. (2017), 'Joint decomposition by subpopulations and sources of the point and synthetic Bonferroni inequality measures', *Statistica & Applicazioni* **25**(2), 83–120.

Zenga, M. and Valli, I. (2018), 'Joint decomposition by subpopulations and sources of the point and synthetic Gini indexes', *Statistica & Applicazioni* **26**(1), 31–79.