ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY
Nr 16(22)

A NOTE ON THE ZENGA INDEX WITH RESPECT TO DIFFERENT RESULTS DEPENDING ON GROUPING OR NOT GROUPING THE DATA

Katarzyna Ostasiewicz

Wrocław University of Economics

e-mail: katarzyna.ostasiewicz@ue.wroc.pl

ORCID: 0000-0002-0115-3696

ISSN 1644-6739
e-ISSN 2449-9765

DOI: 10.15611/sps.2018.16.10
JEL Classification: I31, I32

Summary: The Zenga index of inequality is a new proposal of measuring this phenomenon. The properties and empirical applications of Zenga index (and the underlying Zenga curve) have been recently widely investigated. Its advantages (and differences, as compared to other existing measures) have been pointed out. However, one of the possible problems associated with the use of the Zenga index has not yet been appropriately addressed.Namely, the Zenga index assumes different values depending on whether it is applied to grouped or ungrouped data. As it may seem that due to contemporary computers power it is not necessary to group data, the problem still exists as a conceptual one. Moreover, in some situations – such as applying equivalence scales – avoiding grouping of the data is not possible even in principle. The problem is stated, illustrated by simple numerical examples and briefly discussed in this paper.

Keywords: inequality, Zenga index, grouped data.

M. Zenga proposed a new inequality index (see [Zenga 2007]). Suppose that some good is distributed among$N$ individuals, and the allotments (observations) are organized in increasing order with their frequencies given:

$$\{\{x_1, n_1\}, \dots, \{x_k, n_k\}\},$$

where $0 \le x_1 < \dots < x_k$ and $\sum_{i=1}^{k} n_k = N$.

Let us define U_i as:

$$U_i = \frac{M_i^-}{M_i^+} \text{ for } i = 1, \dots, k,$$

with

$$M_i^- = \frac{\sum_{j=1}^{i} x_i n_i}{\sum_{j=1}^{i} n_i}, \text{for } i = 1, \dots, k$$

and

$$M_i^+ = \frac{\sum_{j=i+1}^{k} x_i n_i}{\sum_{j=i+1}^{k} n_i}, \text{ for } i = 1, \dots, k-1 \text{ and } M_k^+ = x_k.$$

ŚLĄSKI PRZEGLĄD STATYSTYCZNY
Nr 16(22)

The inequality point index, for any$i = 1, \dots, k$ is defined as:

$$I_i = 1 - U_i,$$

and the pairs of coordinates $\left(\frac{\sum_{j=1}^{i} n_i}{N}, I_i\right)$ determines the Zenga inequality curve.

The synthetic inequality measure, the Zenga index, is the weighted average of all point indexes:

$$I = \sum_{i=1}^{k} \frac{n_i}{N} I_i.$$

Apart from calculating the Zenga curve and the Zenga index for grouped data, as presented above, bothconcepts may be also applied for ungrouped data (see [Zenga, 2012]). That is, one deals in this case with a non-decreasing sequence: $0 \le x_1 \le x_2 \le \dots \le x_N$ and all above formulae are validwith $n_i \equiv 1$.

The properties of the index have been intensively studied and its usefulness in empirical cases as well (see [Pollastri 1987; Greselin et al. 2010; Radaelli 2010; Ostasiewicz, Mazurek 2013; Jedrzejczak 2015; Greselin et al. 2017]).

However there is a problem that has been not yet satisfactory solved, namely the differences in values of this indexdepending on whether the data is grouped or not grouped.

For an illustrative example, suppose that in N allotments the first $N - 1$ are all equal to 1, and the Nth is equal to $x > 1$. This situation can be equivalently expressed either in the form of ungrouped data or grouped:

A) $\{1,1, \dots ,1, x\}$ (ungrouped) or

B) $\{\{1, N-1\}, \{x, 1\}\}$ (grouped).

The lower meanM^- will be the same no matter whether we group observations or not for all elements apart from the last one. However, the upper mean will successively increase in case A, while in case B it will be x for both elements (note, in case A there are N elements, while in case B only two elements, due to grouping).

Thus in case B, for grouped data, we have $U_1 = \frac{1}{x}$, $U_2 = \frac{N+x-1}{Nx}$, and the weighted average, U_B:

$$U_B = \frac{N-1}{N} \cdot \frac{1}{x} + \frac{1}{N} \cdot \frac{N+x-1}{Nx} = \frac{N^2-N+N+x-1}{N^2x} = \frac{N^2+x-1}{N^2x}.$$

On the other hand, in case A one has: $M_1^+ = \frac{x+N-2}{N-1}$, $M_2^+ = \frac{x+N-3}{N-2}$ (in general: $M_i^+ = \frac{x+N-(i+1)}{N-i}$ for $i = 1, .., N-1$ and $M_N^+ = x$), and thus:

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

$$U_i = \frac{N-i}{x+N-(i+1)} \text{ for } i = 1,..,N-1 \text{ and } U_N = \frac{N-1+x}{Nx}.$$

Thus:

$$U_A = \frac{1}{N}\sum_{i=1}^{N-1}\frac{N-i}{x+N-(i+1)} + \frac{N+x-1}{N^2x}.$$

The difference is the more striking the more elements are grouped (or not grouped). For example, the figure below presents plots of the Zenga index for $x = 100$ and $N = 1,2,\dots,100$ for both cases.

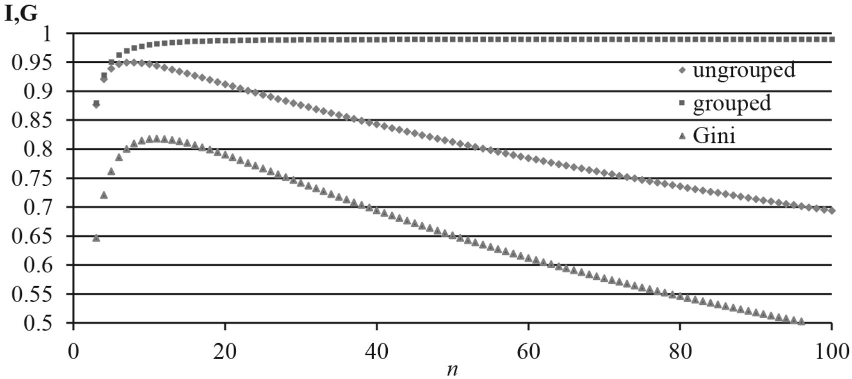

Fig. 1. The Zenga index for grouped and ungrouped data, compared with Gini index

Source: own construction.

As can be seen in Figure 1, the qualitativebehavior of the Zenga index for the example examinedabove is quite different from Gini index (and from other popular indexes, like the Atkinson measure, not illustrated here) while calculated for grouped data.

There is a relationship between the Lorenz curve $L(p)$ (the basis for the Gini index) and the Zenga curve $I(p)$, which holds both for ungrouped and grouped data:

$$I(p) = \frac{p-L(p)}{p[1-L(p)]}.$$

For discrete data this relationship holds only for some particular values: for $p = \frac{i}{N}, i = 1,\dots,N-1$ (for ungrouped data) and for $p = \frac{n_i^c}{N}, i = 1,\dots,s-1$ (where n_i^c are cumulative frequencies for subsequent classes, for grouped data), while between these points both curves are to be interpolated. For the Lorenz curve the linear interpolation is in strict accordance with the Lorenz curve for ungrouped data. On the other hand, for the Zenga curve the linear interpolation does not correspond strictly to the curve for ungrouped data. However, while calculating the Zenga index for grouped data the linear interpolation isapplied. Roughly (and not rigorously) speaking, the discrepancy arises because grouping the same values of data is linear, while the Zenga index is nonlinear with respect to equal values. This discrepancy is illustrated in Figures 2 and 3, for the following example of ungrouped data: {0,1,1,1,1,1,2}. TheLorenz curves for both grouped and ungrouped data are exactly the same (Figure 2).

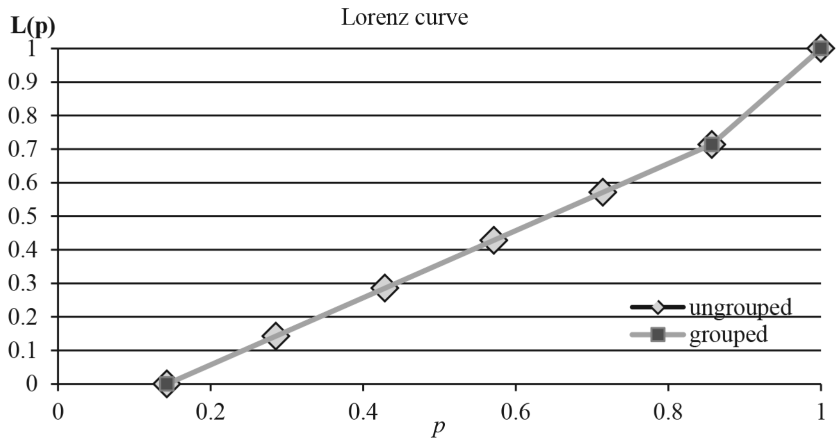

Fig. 2. The Lorenz curves for ungrouped and grouped data

Source: own construction.

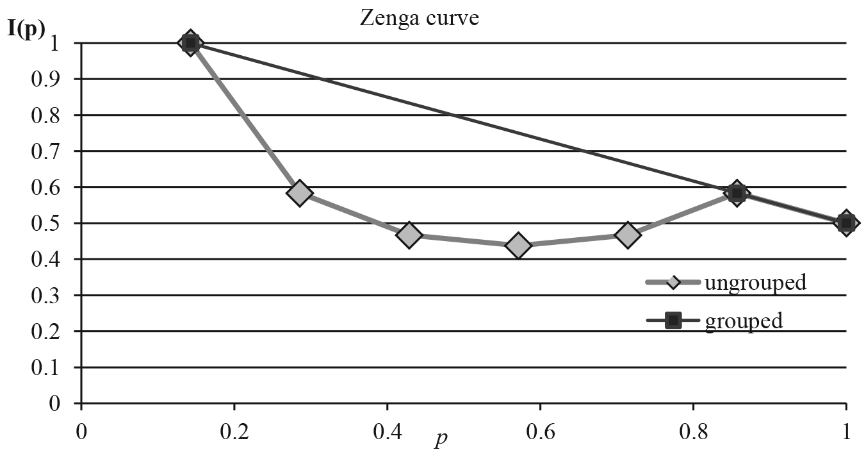

Fig. 3. The Zenga curves for ungrouped and grouped data

Source: own construction.

On the other hand, the Zenga curves for grouped and ungrouped data are significantly different (Figure 3).

The discussion above could be an argument for treating ungrouped individual data. In general, nowadays there is no problem with the access to the individual data and due to computers power one can easily deal even with huge sets of individual, ungrouped data.

However for many surveys we get data grouped in a natural way, e.g. for individual households rather than individual persons. To calculate inequalities on an individual level one usually applies so-called equivalence scales, which are in general non-integer numbers. For example, if a household consistsof two adults it is treated (e.g. according to the OECD equivalence scale) as if consisting of1.7 equivalent standard "individuals". Now, there are different kinds of approach (for discussion see [Cowell 1984]). If the total income of this household is equal to *h,* some researchers treat this situation as if two individuals have$h$/1.7 (keeping the number of individuals as in reality), however others (e.g. [Kot 2012]) argue that the situation should be viewed as if 1.7 individuals have$h$/1.7 income (keeping the total income as in reality).

If adopting this second approach with non-integer in general numbers of individuals it is not possible to deal with ungrouped data.

Still another question is the continuous case, in which the difference between grouped and ungrouped data is irrelevant – however it might, by analogy and some demand of continuity of properties, throw some light on the discrete case.

The problem seems to be far from being simple and far from being solved, and requires further investigation.

Bibliography

Cowell F.A., 1984, *The structure of American income inequality*, Review of Income and Wealth, 30(3), pp. 351-375.

Greselin F., Pasquazzi L., Zitikis R., 2010, *Zenga's new index of economic inequality, its estimation, and an analysis of incomes in Italy*, Journal of Probability and Statistics, Article ID 718905.

Greselin F., Pellegrino S., Vernizzi A., 2017, *Lorenz versus Zenga Inequality Curves: a New Approach to Measuring Tax Redistribution and Progressivity*, Working papers 046, Department of Economics and Statistics (Dipartimento di Scienze Economico-Sociali e Matematico-Statistiche), University of Torino.

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

Kot S.M., 2012, *Ku stochastycznemu paradygmatowi ekonomii dobrobytu*, Oficyna Wydawnicza Impuls.

Ostasiewicz K., Mazurek E., 2013, *Comparison of the Gini and Zenga indexes using some theoretical income distributions abstract*, Operations Research and Decisions, 23, pp. 37-62.

Pollastri A., 1987, *Characteristics of Zenga's Concentration Index Z2*, [in:] M. Zenga (ed.), La Distribuzione Personale del Reddito: Problemi di Formazione, di Ripartizione e Misurazione, Vita e Pensiero, Milano, pp. 214-229.

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

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

Zenga M., Radaelli P., 2012, *Decomposition of Zenga's inequality index by sources*, Statistica & Applicazioni, 9(1), pp. 3-34.

UWAGI DOTYCZĄCE INDEKSU ZENGI W ODNIESIENIU DO OTRZYMYWANIA ODMIENNYCH WYNIKÓW DLA ZGRUPOWANYCH I NIEZGRUPOWANYCH DANYCH

Streszczenie: Indeks Zengi jest stosunkowo nową propozycją w kontekście pomiaru nierówności. Własności i praktyczne zastosowania zarówno indeksu Zengi, jak i związanej z nim krzywej Zengi są obecnie intensywnie badane. Wskazywano na korzyści płynące ze stosowania tej właśnie miary oraz różnice pomiędzy nią a innymi istniejącymi miernikami nierówności. Jednakże istnieje potencjalny problem, związany ze stosowaniem indeksu Zengi, który jak dotąd nie został dostatecznie przedyskutowany. Otóż okazuje się, iż indeks ten daje odmienne wyniki w zależności od tego, czy stosowany jest do danych szczegółowych czy też zgrupowanych. Nawet jeśli wydawać by się mogło, iż jest to problem nieistotny ze względów praktycznych – ze względu na moce obecnych komputerów, niewymagające grupowania danych – zagadnienie to wciąż pozostaje problemem konceptualnym. Ponadto w niektórych sytuacjach, na przykład w sytuacji stosowania skal ekwiwalentności, kwestii grupowania danych nie da się uniknąć. W artykule problem ten został sformułowany, zilustrowany na prostych przykładach oraz krótko przedyskutowany.

Słowa kluczowe: nierówności, indeks Zengi, grupowanie danych.