Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100102

Alternative Method of Measuring Concentration

Zygmunt Jacek Zawistowski[1,∗], *Wiesław Szczesny*[1], *Bolesław Borkowski*[1], *Ryszard Kozera*[1] *and Yochanan Shachmurove*[2]

[1] Warsaw University of Life Sciences - SGGW, Faculty of Applied Informatics and Mathematics, Nowoursynowska Street 159, 02-776 Warsaw, Poland
[2] The City College of the City University of New York, Department of Economics and Business, 160 Convent Avenue at 138th Street, New York 10031, USA

Received: 5 Jun. 2015, Revised: 3 Aug. 2015, Accepted: 4 Aug. 2015
Published online: 1 Jan. 2016

Abstract: The most popular measure of concentration, Gini coefficient G, does not show uniformly all the changes taking place in a flow of goods between objects. This work presents alternative index of concentration IC that when used together with Gini index mitigate that inconvenience. Studies have shown that coefficient IC differs considerably from Gini index in sensitivity to flow of ”good”. The coefficient IC is more sensitive because it depends on value of change quadratically while Gini index only linearly. IC depends on coordinates more properly than G as it depends on difference in values of coordinates opposite to G which depends on difference in positions of coordinate.

Keywords: multidimensional comparative analysis, measures of concentration, Gini index, coefficient IC, Lorenz curves

1 Introduction

Measurement of concentration is one of key problems in economy. This measurement is done in research of inequality of income as well as studies of concentration in individual markets. The literature in this field is abundant. It is worth noting, that the Gini coefficient ([1], [2]) is the most well known and most widely used coefficient in the studies of inequality of income [3]. On the other hand, Herfindahl-Hirschman index, sometimes called Herfindahl index or simply abbreviated to HHI [4] and [5] is the most popular tool used in literature concerning studies of concentration in the market. For completeness see [6], [7], [8], [9], [10] and [11]. Gini coefficient is used as a main measurement tool within information systems of large commercial institutions with international reach as well as public ones which collect data on levels of differentiation in access to a specific good in individual subpopulations. Moreover, it also allows to plot the trend of the process, whether the differentiation does not grow too quickly during predefined periods within a specific group of objects (see [12]).

Our current research has shown that Gini coefficient does not always precisely show the changes taking place in the studied population. In order to present the problem, let us consider an example of changes of the structure of salaries in a large department of a fiscal company and see what the output of Gini coefficient is. The changes of salaries for full time employees (FTE) in individual decile groups during the studied period are presented in Tab. 1 and Tab. 2. These are values of a structure of an annual capital of salaries divided into decile groups. Under normal conditions, the number of employees and, as a result, the cardinality of a decile group is variable, which was included in the second period of Table 1. Common information systems in large corporations (when changes to salaries are considered) are limited mostly to juxtaposing (on every level of the corporate hierarchy) the number of employees and average salaries with Gini coefficient - for a period of a couple of years. Both Table 1 and Table 2, indicate that the structure of salaries in the given department is degenerating. Meaning, the topmost 10% of employees in terms of earnings is absorbing an increasing percentage of the capital of salaries (in the last period, this percentage was above 51%, see Table 2). At the same time, the value of Gini coefficient (see the penultimate row of Table 2) for all periods remains the same. In the last row of Table 2 we have presented values of a different coefficient G_R, based on a geometrical interpretation that uses radar charts (see [13]), proposed in [3], which, in this case, visibly shows the changes and indicates a substantial increase in concentration. This

∗ Corresponding author e-mail: zygmunt_zawistowski@sggw.pl

Table 1: Juxtaposition of average monthly salaries in decile groups within a large department of an international company in six one-year periods. Source: own research.

Decile	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
1	500.00	1365.00	2268.00	2545.20	3849.62	4554.46
2	1000.00	1365.00	2268.00	2545.20	3849.62	4554.46
3	1500.00	1365.00	2268.00	2545.20	3849.62	4554.46
4	2000.00	2094.75	2268.00	2545.20	3849.62	4554.46
5	4000.00	2625.00	2268.00	2545.20	3849.62	4554.46
6	5000.00	5074.13	3780.00	2545.20	3849.62	4554.46
7	7000.00	7350.00	8820.00	7635.60	3849.62	4554.46
8	8000.00	8400.00	10080.00	10180.80	7699.23	4554.46
9	9000.00	9450.00	11340.00	11453.40	9239.08	4554.46
10	12000.00	13411.13	17640.00	19089.00	33106.69	43699.87
AVG	5000.00	5250.00	6300.00	6363.00	7699.23	8469.15

Table 2: Annual structures of salaries in decile groups based on information from table 1. Source: own research.

Decile	S1 %	S2 %	S3 %	S4 %	S5 %	S6 %
1	1.0	2.60	3.6	4.0	5.0	5.3778
2	2.0	2.60	3.6	4.0	5.0	5.3778
3	3.0	2.60	3.6	4.0	5.0	5.3778
4	4.0	3.99	3.6	4.0	5.02	5.3778
5	8.0	5.0	3.6	4.0	5.0	5.3778
6	10.0	9.665	6.0	4.0	5.02	5.3778
7	14.0	14.0	14.0	12.0	5.0	5.3778
8	16.0	16.0	16.0	16.0	10.0	5.3778
9	18.0	18.0	18.0	18.0	12.0	5.3778
10	24.0	25.545	28.0	30.0	43.0	51.5998
Gini	0.416	0.416	0.416	0.416	0.416	0.416
G_R	0.51515	0.52791	0.54816	0.56188	0.61353	0.63768

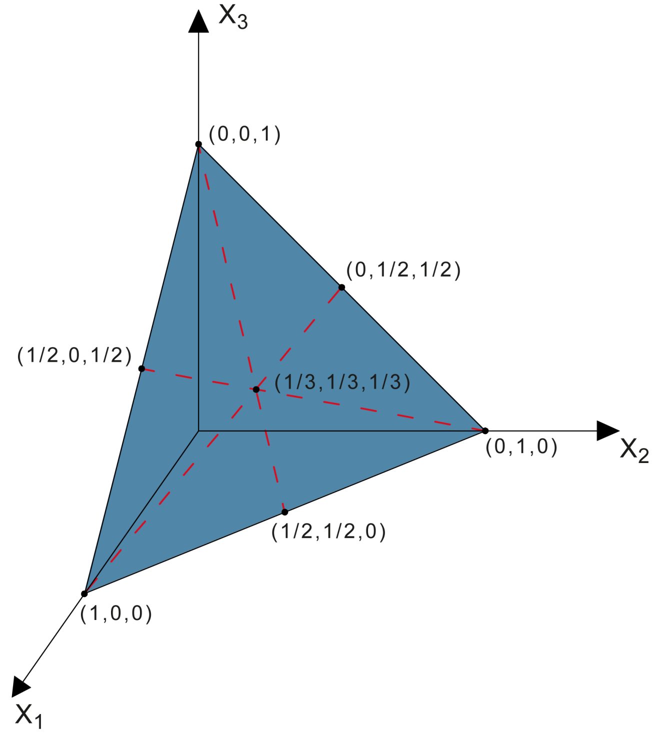


Fig. 1: Ω space in 3-dimensions.

example shows that in information systems that synthetically present phenomenon of this kind, one should not use a single indicator when compiling reports for senior management of large institutions. As such it appears that researching techniques to build alternative coefficients designed to support measurement of concentration is still an important task. The goal of this paper is to propose a new concentration index which is more sensitive to changes of concentration than Gini index.

2 Notations and definitions

For creating measure of value levels of concentration of a given good in a set of objects we choose the space of positive normalized vectors as a space of multidimensional data Ω:

$$\mathbb{R}_{+}^{n} \supset \Omega := \left\{ \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_{+}^{n} : \sum_{i=1}^{n} x_i = 1 \right\}. \quad (1)$$

Normalization follows from the fact that we can divide the value of a good possessed by a single object by the total value of that good possessed by the whole set of objects. The space Ω is the (n-1)-dimensional convex subspace of $\mathbb{R}_{+}^{n}$. The center of the set Ω :

$$\mathbf{e} := \left(\frac{1}{n}, \ldots, \frac{1}{n} \right) \quad (2)$$

is designated as an *egalitarian structure*, while the vertices of the space Ω, which are the vectors of the form

$$\mathbf{o} := (0, \ldots, 1, \ldots, 0) \quad (3)$$

represent an *extremely concentrated structures*. In the case of $n = 3$ the space Ω is shown in Fig. 1 with *characteristic directions* from center to vertices and midpoints of edges.

The construction of a measure of concentration a given structure is based on calculating how this structure differs from the egalitarian one. This means that using common tools of multidimensional comparative analysis one can construct many different coefficients by applying different measures of similarity and dissimilarity. Our approach is to apply a *utility function* U, that fulfils condition

$$\mathrm{U}(\mathbf{o}) \leq \mathrm{U}(\mathbf{x}) \leq \mathrm{U}(\mathbf{e}) \text{ for any } \mathbf{x} \in \Omega. \quad (4)$$

Then a coefficient of concentration IC can be defined by the following equation

$$\mathrm{I}C = 1 - \frac{\mathrm{U}(\mathbf{x})}{\mathrm{U}(\mathbf{e})}. \quad (5)$$

A natural choice of a utility function is a ”measure of discrete autocorrelation”

$$\mathrm{U}(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j. \quad (6)$$

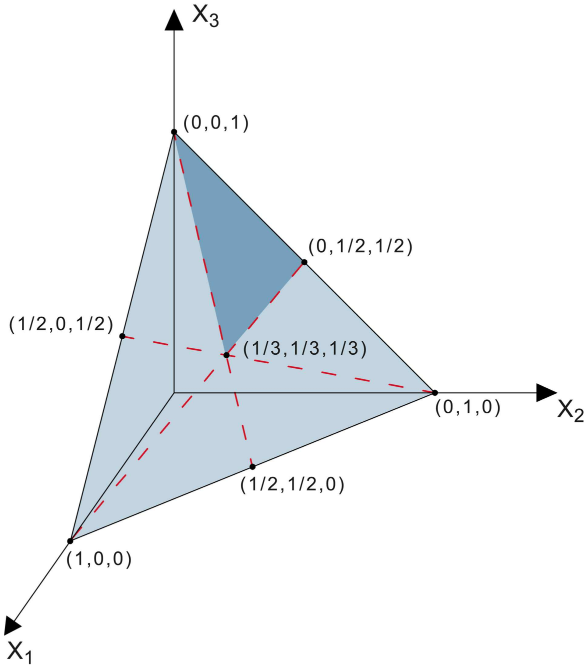


Fig. 2: Sector of ordered vectors in 3-dimensions.

Let us define a vector that had any pair of coordinates modified

$$\mathbf{x}'_{k,l} = (x_1,\dots,x_k-\varepsilon,\dots,x_l+\varepsilon,\dots,x_n), \quad (7)$$

where $1 \le k < l \le n,\ \ 0 < \varepsilon \le x_k$, and also *ordered vector* $\mathbf{x}^* = (x_1^*,\dots,x_n^*)$, which we derive from vector $\mathbf{x}$ by means of a permutation of its coordinates such that $x_1^* \le x_2^* \le \dots \le x_n^*$. The sector of ordered vectors in the case of $n = 3$ is shown in Fig. 2. An ordered vector with a modified pair of consequent coordinates is designated as follows

$$\mathbf{x}^*_{\varepsilon,i} = (x_1^*,\dots,x_i^*-\varepsilon,x_{i+1}^*+\varepsilon,\dots,x_n^*), \quad (8)$$

where $1 \le i \le n,\ \ 1 \le i \le n$, and ε is such that it does not spoil the ordering $x_{i-1}^* \le x_i^* - \varepsilon \le x_{i+1}^* + \varepsilon \le x_{i+2}^*$. In general, an ordered vector with a pair of any coordinates modified has the form

$$\mathbf{x}^*_{\varepsilon,i,j} = (x_1^*,\dots,x_i^*-\varepsilon,\dots,x_j^*+\varepsilon,\dots,x_n^*), \quad (9)$$

where $1 \le i < j \le n,\ \ 0 < \varepsilon \le x_i^*$ and where ε is such that it does not spoil the ordering $x_{i-1}^* \le x_i^* - \varepsilon \le x_{i+1}^*, x_{j-1}^* \le x_j^* + \varepsilon \le x_{j+1}^*$.

3 Properties of the coefficient IC

A properly defined coefficient of correlation IC should fulfill the following conditions:

1. Zero value for an egalitarian structure:
   $IC(\mathbf{e}) = 0$,
2. Maximum value for nodes (vertices) of the set Ω, normalized to 1:
   $IC(0,\dots,x_i,\dots,0) = 1 \quad$ for any $\ x_i = 1$,
3. Independence of the order of features:
   $IC(\mathbf{x}) = IC(\mathrm{perm}(\mathbf{x}))$,
   where $\mathrm{perm}(\mathbf{x})$ is given by any permutation of the set of indices $\{1,\dots,n\}$,
4. Increase in value when differentiation of any pair of coordinates increases:
   $|x_i' - x_j'| \equiv$
   $|(x_i-\varepsilon)-(x_j+\varepsilon)| \ge |x_i - x_j| \Rightarrow IC(\mathbf{x}') \ge IC(\mathbf{x})$,
5. Increase in value when differentiation of any pair of coordinates in an ordered vector increases
   $1 \le i \le j \le n \Rightarrow IC(\mathbf{x}^*_{\varepsilon,i,j}) \ge IC(\mathbf{x}^*)$.

Proofs of the properties 1 – 5 of our coefficient IC defined in (5) are given below.

Re. 1 Value of 0 for an egalitarian structure follows directly from the definition (5),

Re. 2 Nodes $(0,\dots,1,\dots0)$ are vectors with only one coordinate not equal to 0 (and equal to 1), so every product in definition (6) of utility function U contains an element equal to 0 . Thus the coefficient IC reaches 1 as per definition (5). It is interesting to calculate $U(\mathbf{e})$ as well. In his case all products from (6) are equal to $1/n$. The number of those products is equal to

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$$

and thus

$$U(\mathbf{e}) = \frac{n-1}{2n}. \quad (10)$$

Re. 3 This feature follows directly from definition (6). The value of a sum of elements is independent from the ordering of elements, and in (6) one has all different $(i \ne j)$ products of coordinates (each one appearing only once).

Re. 4 We calculate from (6) the value of function U for vector $\mathbf{x}'$ defined by (7) and group products such that:

– the first group does not contain $x_k' = (x_k-\varepsilon)$ nor $x_l' = (x_l+\varepsilon)$,
– the second group contains products of $x_k' = (x_k-\varepsilon)$ multiplied by $x_j' = x_j$ with $j \ne k,l$,
– the third group contains products of $x_l' = (x_l+\varepsilon)$ multiplied by $x_j' = x_j$ with $j \ne k,l$,
– the fourth group contains only the product $x_k'x_l' = (x_k-\varepsilon)(x_l+\varepsilon)$,

$$
\begin{aligned}
\mathrm{U}(\mathbf{x}') = &\sum_{\substack{i=1\\ i\neq k,l\\ j=1+1\\ j\neq k,l}}^{n} x_i x_j + (x_k-\varepsilon)\sum_{\substack{j=1+1\\ j\neq k,l}}^{n} x_j + (x_l+\varepsilon)\sum_{\substack{j=1+1\\ j\neq k,l}}^{n} x_j +\\
&+(x_k-\varepsilon)(x_l+\varepsilon) =\\
&= \sum_{\substack{i=1\\ i\neq k,l\\ j=1+1\\ j\neq k,l}}^{n} x_i x_j + x_k\sum_{\substack{j=1+1\\ j\neq k,l}}^{n} x_j + x_l\sum_{\substack{j=1+1\\ j\neq k,l}}^{n} x_j +\\
&+x_k x_l + \varepsilon(x_k-x_l) - \varepsilon^2 =\\
&= \sum_{\substack{i=1\\ j=1+1}}^{n} x_i x_j + \varepsilon(x_k - x_l - \varepsilon)
\end{aligned}
$$

which means

$$\mathrm{U}(\mathbf{x}') = \mathrm{U}(\mathbf{x}) + \varepsilon(x_k - x_l - \varepsilon). \tag{11}$$

Thus, the sign of the expression $(x_k - x_l - \varepsilon)$ decides which one of $\mathrm{U}(\mathbf{x}')$ or $\mathrm{U}(\mathbf{x})$ is greater. We encounter the following cases:

$$
\begin{aligned}
&x_k \le x_l,\ 0<\varepsilon\le x_k \ \Rightarrow\ \mathrm{U}(\mathbf{x}') < \mathrm{U}(\mathbf{x}),\\
&x_k > x_l,\ \varepsilon < x_k - x_l \ \Rightarrow\ \mathrm{U}(\mathbf{x}') > \mathrm{U}(\mathbf{x}),\\
&x_k > x_l,\ \varepsilon = x_k - x_l \ \Rightarrow\ \mathrm{U}(\mathbf{x}') = \mathrm{U}(\mathbf{x}),\\
&x_k > x_l,\ \varepsilon > x_k - x_l \ \Rightarrow\ \mathrm{U}(\mathbf{x}') < \mathrm{U}(\mathbf{x}).
\end{aligned}
$$

From that we have the following theorem for the coefficient IC.

Theorem 1. *The following implications hold*

$$
\begin{aligned}
&x_k \le x_l,\ 0<\varepsilon\le x_k \ \Rightarrow\ \mathrm{I}C(\mathbf{x}') > \mathrm{U}(\mathbf{x}),\\
&x_k > x_l,\ \varepsilon < x_k - x_l \ \Rightarrow\ \mathrm{I}C(\mathbf{x}') < \mathrm{U}(\mathbf{x}),\\
&x_k > x_l,\ \varepsilon = x_k - x_l \ \Rightarrow\ \mathrm{I}C(\mathbf{x}') = \mathrm{U}(\mathbf{x}),\\
&x_k > x_l,\ \varepsilon > x_k - x_l \ \Rightarrow\ \mathrm{I}C(\mathbf{x}') > \mathrm{U}(\mathbf{x}).
\end{aligned}
$$

The behavior of the coefficient IC as described by this theorem is expected and desired.

An increase in differentiation of any pair (k,l) of coordinates means that

$$
\begin{aligned}
|x_k' - x_l'| = \max(x_k', x_l') - \min(x_k', x_l') \ > \ |x_k - x_l| =\\
= \max(x_k, x_l) - \min(x_k, x_l)
\end{aligned}
$$

so using Theorem 1 (first and last cases of the theorem can be realized for all permissible ε: $0 < \varepsilon \le x_k$) we obtain the following

Corollary 1.
$|x_k' - x_l'| = |x_k - x_l - 2\varepsilon| > |x_k - x_l| \Rightarrow \mathrm{I}C(\mathbf{x}') > \mathrm{U}(\mathbf{x})$

which proves the feature 4.
Re. 5 This features follows from Theorem 1, as it is a special case of feature 4. Further, for an ordered vector $\mathbf{x}^*$ only the first implication (first line) of Theorem 1 is realized, regardless of changes of (9) when $\mp\varepsilon$ is applied.

In special practical problems, additional features which follow from actual applications are often added. In particular, a special requirement in today's globalized economy should be increased sensitivity of the coefficient to changes in allocation of the good that are interesting for the researcher. In order to illustrate this problem we will compare the popular Gini index and the proposed coefficient IC. To this end we investigate the sensitivity of these coefficients to value of coordinates change. For IC we have from (11) and (10)

$$
\begin{aligned}
\delta\mathrm{IC}_{\varepsilon,k,l}(\mathbf{x}) \equiv \mathrm{IC}(\mathbf{x}'_{\varepsilon,k,l}) - \mathrm{IC}(\mathbf{x}) =\\
= \frac{2n}{n-1}\varepsilon((x_l - x_k) + \varepsilon).
\end{aligned} \tag{12}
$$

In the case of ordered vectors $\mathbf{x}^*$ we always have $l > k \Rightarrow x_l^* > x_k^*$ so we obtain

Theorem 2. *Increase of* IC *under* $\mp\varepsilon$ *operation for ordered vectors* $\mathbf{x}^*$ *is linearly dependent on* $(x_l^* - x_k^*)$ *and* quadratically *on* ε.

It is interesting to consider relative increase in value of IC ("speed" of changes of IC) according to various arguments

$$\frac{\delta\mathrm{IC}_{\varepsilon,k,l}}{\delta(x_l - x_k)} = \frac{2n}{n-1}\left(\varepsilon + \frac{\varepsilon^2}{x_l - x_k}\right), \tag{13}$$

$$\frac{\delta\mathrm{IC}_{k,l,\varepsilon}}{\delta\varepsilon} = \frac{2n}{n-1}((x_l - x_k) + \varepsilon)$$

or better yet

$$\frac{\delta\mathrm{IC}_{\varepsilon,k,l}}{\delta(2\varepsilon)} = \frac{n}{n-1}((x_l - x_k) + \varepsilon) \tag{14}$$

because the total increase of a good is tied to the change

$$
\begin{aligned}
|x_k' - x_l'| - |x_k - x_l| = (x_l' - x_k') - (x_l - x_k) =\\
= (x_k - \varepsilon) - (x_l + \varepsilon) - x_l + x_k = 2\varepsilon
\end{aligned}
$$

as for ordered vectors $x_l \ge x_k \ge 0,\ \ x_l' \ge x_k' \ge 0,\ \ \varepsilon > 0$.

Let us now consider the behavior of the coefficient IC alongside chosen characteristic directions (see the Fig.1).
1. Alongside a straight line passing through the center of the set and any one of its vertices $(0,\ldots,x_i = 1,\ldots,0)$.

We are always interested in the knowledge of dependance of IC on *coordinates*, on x_i this time which measures the distance on projection of the above mentioned line onto this coordinate. When we move alongside the coordinate $x_i \equiv x \in [0,1]$ then all other coordinates are equal to

$$x_j = \frac{1-x}{n-1} \quad \text{for} \quad j \neq i \tag{15}$$

and then the utility function (6) depends on x as follows

$$\mathrm{U}(x) = \mathrm{U}(x_i) = \sum_{\substack{j=1\\ j\neq i}}^{n} x_i x_j + \sum_{\substack{j=1\\ k=j+1\\ j,k\neq i}}^{n} x_j x_k =$$
$$= x\frac{1-x}{n-1}(n-1) + \frac{(1-x)^2}{(n-1)^2}\left[\binom{n}{2} - (n-1)\right] =$$
$$= x(1-x) + \frac{(1-x)^2}{(n-1)^2}\frac{(n-1)(n-2)}{2} =$$
$$= -\frac{n}{2(n-1)}x^2 + \frac{1}{n-1}x + \frac{n-2}{2(n-1)}.$$

Thus

$$\begin{aligned} \mathrm{IC}(x) &= 1 - \frac{2n}{n-1}\mathrm{U}(x) = \\ &= \frac{n^2}{(n-1)^2}x^2 - \frac{2n}{(n-1)^2}x + \frac{1}{(n-1)^2}. \end{aligned} \tag{16}$$

Formula (16) describes a parabola opening upwards. Minimum equal to zero is achieved for an egalitarian structure $\mathbf{e}$ (center of the set Ω) which corresponds to $x = 1/n$. Naturally, for $x = 1$ coefficient IC vanishes: $\mathrm{IC}(1) \equiv \mathrm{IC}(\mathbf{o}) = 0$. For $x = 0$ which corresponds to the center of the $(n-2)$-dimensional wall we have

$$\mathrm{IC}(0) = \frac{1}{(n-1)^2}.$$

Asymptotically

$$n \to \infty \;\Rightarrow\; \mathrm{IC}(x) \sim x^2 \tag{17}$$

with $\mathrm{IC}(0) \to 0$, while for a typical value $n = 10$ (division into deciles) $\mathrm{IC}(0) = 1/81 \approx 0.012$.

2. Alongside a straight line passing through $\mathbf{e} = (1/n, \ldots, 1/n)$ and the center of an edge connecting vertices i and $j = i+1$. We look for dependance on $x' = \sqrt{2}x$, where $x = x_i = x_j \in [0, 0.5]$ that is on distance along diameter between axes x_i and x_j (projection of the above mentioned line on the plane x_i, x_j). In this case all other coordinates are equal to

$$x_k = \frac{1-2x}{n-2} \quad \text{for} \quad k \neq i, j \tag{18}$$

and

$$\mathrm{U}(x) = x_i x_j + x_i \sum_{\substack{k=1\\ k\neq j}}^{n} x_k + x_j \sum_{\substack{k=1\\ k\neq i}}^{n} x_k + \sum_{\substack{k=1\\ k\neq i,j\\ l=k+1\\ l\neq i,j}}^{n} x_k x_l =$$
$$= x^2 + x\frac{1-2x}{n-2}(n-2) + x\frac{1-2x}{n-2}(n-2) +$$
$$+ \frac{(1-2x)^2}{(n-2)^2}\binom{n-2}{2} =$$
$$= x^2 + 2x(1-2x) + \frac{(1-2x)^2}{(n-2)^2}\frac{(n-2)(n-3)}{2} =$$
$$= -\frac{n}{n-2}x^2 + \frac{2}{n-2}x + \frac{n-3}{2(n-2)}$$

so

$$\begin{aligned} \mathrm{IC}(x) &= \frac{2n^2}{(n-1)(n-2)}x^2 - \frac{4n}{(n-1)(n-2)}x + \\ &+ \frac{2}{(n-1)(n-2)}. \end{aligned} \tag{19}$$

$\mathrm{IC}(0) = 2/(n-1)(n-2)$, in the center $\mathrm{IC}(\mathbf{e}) = 0$, and in the center of the edge $\mathrm{IC}(1/2) = (n-2)/2(n-1)$. Asymptotically $n \to \infty \;\Rightarrow\; \mathrm{IC}(x) \sim 2x^2$, and then $\mathrm{IC}(0) = 0$ and $\mathrm{IC}(1/2) = 1/2$.

To compare dependance (19) of IC with (16) we have to change variable in (19): $x \to x' = \sqrt{2}x$. In this way we compare values of IC in the same distances for each direction in the plane x_i, x_j). For typical value $n = 10$ (division into deciles) the result of this comparison is depicted in Fig. 3.

4 Comparison of the concentration coefficient IC and Gini index

Features of coefficient IC (especially 4 and 5), that were discussed in section 2, will be compared with corresponding features of Gini index which is defined as follows

$$\mathrm{G}(\mathbf{x}) = \frac{1}{n-1}\sum_{\substack{i=1\\ j=i+1}}^{n} |x_i - x_j|. \tag{20}$$

For $\mp\varepsilon$ operations (7) we have

$$\begin{aligned} \mathrm{G}(\mathbf{x}'_{k,l}) = \frac{1}{n-1}\Bigg(&\sum_{\substack{i=1\\ i\neq k,l\\ =i+1\\ j\neq k,l}}^{n} |x_i - x_j| + \sum_{\substack{j=k+1\\ j\neq l}}^{n} |x'_k - x_j| + \\ &+ \sum_{j=l+1}^{n} |x'_l - x_j| + |x'_k - x'_l| \Bigg). \end{aligned} \tag{21}$$

The last term is simple

$$|x'_k - x'_l| = |x_k - \varepsilon - x_l + \varepsilon| = |x_k - x_l| \tag{22}$$

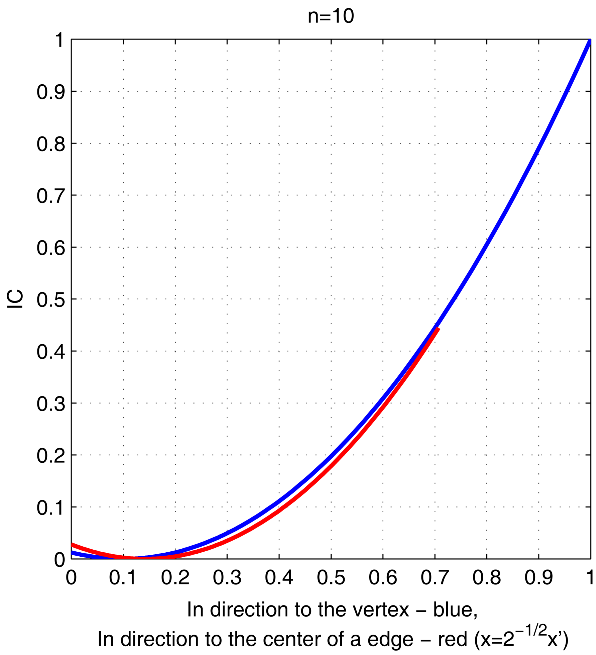


Fig. 3: Changes of coefficient IC along two exemplary characteristic directions: blue line - in the direction of a vertex, red line - in the direction of a center of an edge.

but values of modules $|x'_k - x_j| = |x_k - \varepsilon - x_j|$ and $|x'_l - x_j| = |x_l - x_j + \varepsilon|$ depend on the sign of inner expressions, which means that they depend on relations of x_k to x_j and of x_l to x_j. So, generally, for unordered vectors it is impossible to define the behavior of coefficient IC as everything depends on actual values of vector's $\mathbf{x}$ coordinates. However for ordered vectors $\mathbf{x}^*$ we have $x^*_k - x^*_j - \varepsilon < 0$ for $j > k$ because $x^*_j \geq x^*_k$, therefore

$$\sum_{\substack{j=k+1\\ j\neq l}}^{n} |x^*_k - x^*_j - \varepsilon| = \sum_{\substack{j=k+1\\ j\neq l}}^{n} (x^*_j - x^*_k + \varepsilon) = = \sum_{\substack{j=k+1\\ j\neq l}}^{n} |x^*_j - x^*_k| + \varepsilon(n-k-1). \tag{23}$$

Similarly $x^*_l - x^*_j + \varepsilon < 0$ for $j > l$ because $x^*_l + \varepsilon \leq x^*_{l+1} \leq x^*_j$ for $j > l$ so

$$\sum_{j=l+1}^{n} |x^*_l - x^*_j + \varepsilon| = \sum_{j=l+1}^{n} (x^*_j - x^*_l - \varepsilon) = = \sum_{j=l+1}^{n} |x^*_j - x^*_l| - \varepsilon(n-l). \tag{24}$$

From (21)–(24) it follows that for *ordered* vectors the increase of Gini index as a result of $\mp\varepsilon$ operation is equal to

$$\delta \mathrm{IC}_{\varepsilon,k,l}(\mathbf{x}^*) \equiv \mathrm{IC}(\mathbf{x}^*_{\varepsilon,k,l}) - \mathrm{IC}(\mathbf{x}^*) = \varepsilon(l-k-1). \tag{25}$$

This result can be formulated as the following theorem

Theorem 3. *Increase of Gini index* G *under* $\mp\varepsilon$ *operation for ordered vectors* $\mathbf{x}^*$ *is linearly dependent on* ε *and on the difference* $(l-k)$ *of positions of coordinates.*

One can also compare speed of change for index IC and Gini index as defined by equations analogous to equations (13)–(14)

$$\frac{\delta \mathrm{G}_{\varepsilon,k,l}(\mathbf{x}^*)}{\delta(\varepsilon)} = (l-k-1), \tag{26}$$

$$\frac{\delta \mathrm{G}_{\varepsilon,k,l}(\mathbf{x}^*)}{\delta(l-k)} = \varepsilon. \tag{27}$$

As a result of the limitation caused by the constraints on the length of this article, we will present the behavior of the coefficient IC compared to the changes in value of Gini coefficient only for a single family of models. We will use this family to model the changes of income divided into decile groups. It is important to normalize the values of the coefficient in the scope of changes concerning concentration. In essence, a user wants to operate on ranges of values that define small, medium or large level of concentration. Normally, this is done via Lorenz curves. In order to do so, let us define a family of curves as

$$L(u;a) = u^a, \quad u \in [0,1],\ a \in [1,\infty[\tag{28}$$

presented in Fig. 4 with the assumption that we possess data for decile groups in a similar fashion as income is presented by GUS, Eurostat, etc. (compare [14]). The value of Gini index is equal to (a-1)(a+1). In Fig. 5 the volatility of Gini index in comparison to the volatility of coefficient IC for the 14 distribution displayed in Fig. 4. is shown. Fig. 5 presents significant differences in sensitivities of the two coefficients for deviations from an egalitarian structure. Gini index already takes a value of 0.5 for the fifth curve, while IC not until the eleventh. Which means that Gini index is more sensitive to changes in concentration when the distribution differs only slightly from an egalitarian one. This follows from the fact that the change of value of Gini index depends on the position on the list of changes of a good, not the real change in the value of that good (equation (25) and Theorem 3). This causes that even for slight changes of a good we see an artificially increased change of the value of Gini index. However, for large deviations from an egalitarian distribution, Gini index is less sensitive than the coefficient IC. This is because Gini index does not take into account the real change of value of a good and then a change of position is less significant.

Using the same formula for utility functions, we can construct, by means of simple modifications, coefficients of different sensitivities. One of the simplest being

$$\mathrm{IC}^*(\mathbf{x}) = (\mathrm{IC}(\mathbf{x}))^p, \quad p \in]0,1[\tag{29}$$

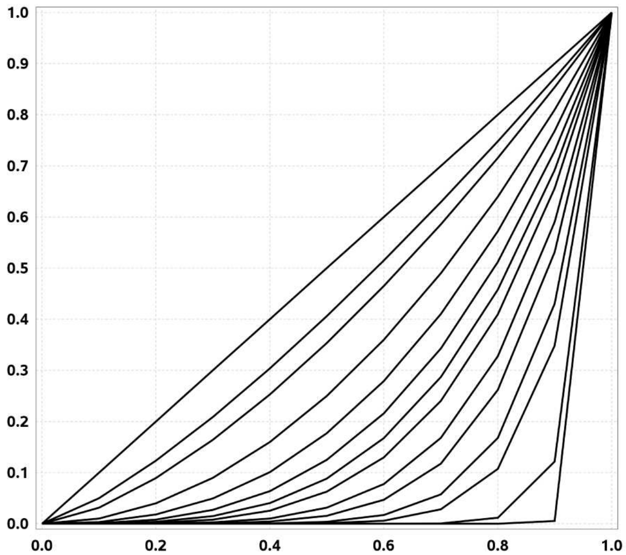


Fig. 4: Lorenz curves given by equation (28) for a=1,1.3, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 8,10, 20, 50 to the nearest decile group. Source: own research.

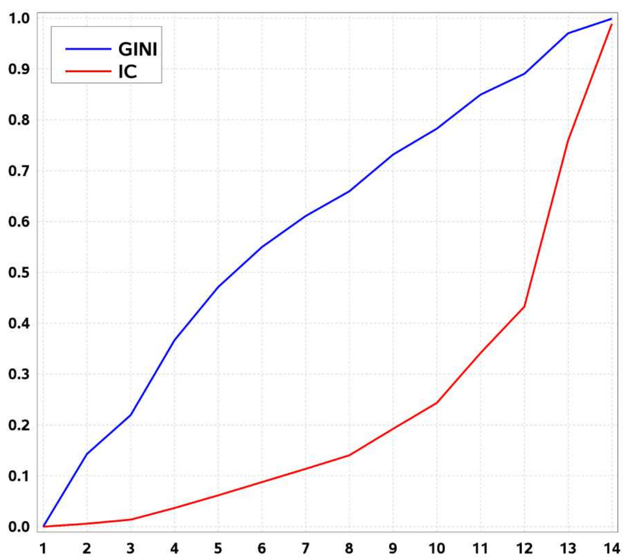


Fig. 5: Gini index and coefficient IC corresponding to the 14 distributions of a good defined by the depicted in Fig.4 Lorenz curves. Source: own research.

or

$$IC^{**}(\mathbf{x}) = 1 - \left(\frac{U(\mathbf{x})}{U(\mathbf{e})}\right)^k, \quad k \in \{1,2,...\}. \qquad (30)$$

Another possibility appears for cumulated structures $\mathbf{x}_{cum} = (x_1, x_1 + x_2, ..., x_1 + x_2 + ... + x_{n-1}, 1)$

$$IC_{cum}(\mathbf{x}) = IC(\mathbf{x}_{cum}). \qquad (31)$$

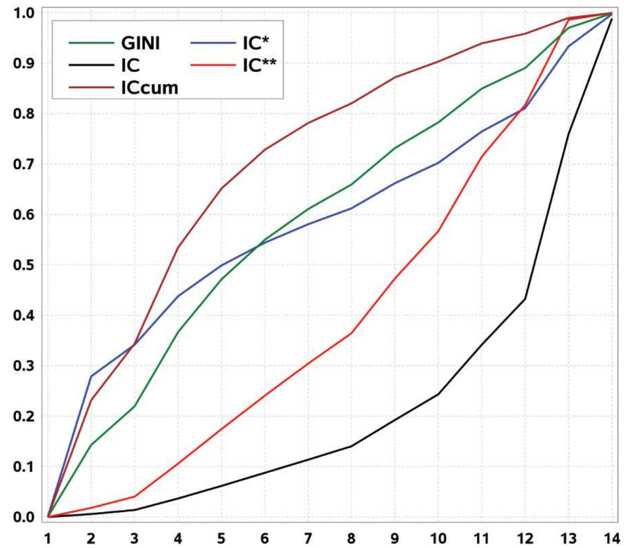


Fig. 6: Comparison of Gini index and coefficient IC and coefficients defined by formulae (29), (30) and (31) for decile structures of Fig. 4 for k=3 and p=0.25. Source: own research.

Sample charts of (29)–(31) for $p = 0.25$, $k = 3$ and decile structure of Fig. 4 are compared with Gini index and the coefficient IC in Fig. 6.

5 Conclusions

We proposed an alternative method of measuring concentration based on the new coefficient IC which fulfils all basic properties of a proper coefficient of concentration, as mentioned in section 3. Properties of IC were thoroughly examined and compared with properties of Gini index. Studies have shown that coefficient IC differs considerably from Gini index G in sensitivity to flow of "good" as it is concluded in Theorems 2 and 3. The coefficient IC is more sensitive because it depends on value of change *quadratically* while Gini index only linearly. IC depends on coordinates more properly than G as it depends on difference in *values* of coordinates opposite to G which depends on difference in *positions* of coordinate. This leads to an artificial amplifying of change of G in the case of small and medium deviations from an egalitarian distribution when the difference in positions is large and to weak reaction on big value of good flow. The coefficient IC is better tuned to structure of good as it accounts for the real difference of "richness" in different groups.

Further studies will concentrate on the applicability of coefficient IC in analysis of various empirical data and on its usability in information systems. Based on coefficient IC, one can construct new coefficients of concentration as functions of them. This allows to tune the sensitivity of a coefficient to various kinds of distributions.

In a situation when one needs to trace changes of concentration of a good possessed by multiple groups of objects, an analyst needs to ponder deeply *what* coefficient (better a few different coefficients) should be used. One should not unconditionally use popular coefficients in their classical forms (e.g. a curious modification of Gini coefficient has been presented in [2]), as they may prove to be insensitive to changes that should be carefully studied. In a specific situation, it is paramount for a coefficient to be especially sensitive to changes in precisely defined range of handpicked scenarios, while it may be less sensitive in other cases. In particular, we might be interested in its sensitivity to change in a situation that starts with high concentration of good or, conversely, we want to capture even minuscule changes in structures close to an egalitarian one. Coefficient IC seems like a good candidate to use in the former scenario.

References

[1] L.D. Hoffmann and G.L.Bradley, Calculus for Business, Economics, and the Social and Life Sciences, 9th ed., McGraw Hill, New York, 2007.

[2] T.H. Hannan, Review of Industrial Organization **12**, 23–35 (1997).

[3] R.A. Barnett, M.R Ziegler, K.E. Byleen, College Mathematics for Business, Economics, Life Sciences, and Social Sciences, 10th ed., Prentice-Hall, Upper Saddle River, 2005.

[4] O.C. Herfindahl, Comment on Rosenbluth's measures of concentration, In: Stigler, G. (Ed.), Business Concentration and Price Policy, Princeton University Press, Princeton, USA, 1955.

[5] A.O. Hirschman, American Economic Review **54**(5), 761–762 (1964).

[6] S. Calkins, California Law Review **71**(2), 402–429 (1983).

[7] B. Czarny, R. Rapacki, Podstawy Ekonomii (In Polish), Polskie Wydawnictwo Ekonomiczne, Warszawa, 1992.

[8] M.G. Lijesen, Transportation Research, Part E **40**, 123–134 (2004).

[9] A. Matsumoto, U. Merlone, F. Szidarovszky, Applied Economics Letters, Taylor and Francis Journals **19**(2), 181-184 (2012).

[10] S.A. Rhoades, Federal Reserve Bulletin, March, 188–189 (1993).

[11] R. Jedut, Annales Universitatis Mariae Curie Skodowska, Series B, **16**(1), 119-156 (1961).

[12] P. Łysoń, R. Antczak, M. Barlik, A. Brochwicz-Lewiska, B. Lewandowska , J. Mirosaw, K. Siwiak and U. Wgrowska, Budety gospodarstw domowych w 2012 roku , GUS Informacje i Opracowania (In Polish), Warsaw 2012.

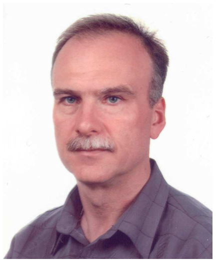
[13] Z. Binderman, B. Borkowski, W. Szczesny, Quantitative Methods in Economics **13**(2), 7–21 (2012).

[14] K. Jackowicz, O. Kowalewski, Koncentracja dziaalnoci sektora bankowego w Polsce w latach 1994 - 2000, Materiay i studia Narodowy Bank Polski, Departament Analiz i Badań (In Polish), 143-143 (2002).

Zygmunt Zawistowski born in Warsaw, Poland in 1946. He received his Master Degree in physics in 1970 from Warsaw University and Ph.D. in 1997 and D.Sc. (habilitation) in 2007 in technical sciences from Institute of Fundamental Research of Polish Academy of Sciences. In 1970-1977 he was with Faculty of Physics of Warsaw University. Next, up to 2009 with Institute of Fundamental Research of Polish Academy of Sciences. Now he is Associated Professor at Faculty of Applied Informatics and Mathematics of Warsaw University of Life Sciences in Warsaw, Poland. The main field of his scientific activity is mathematical modeling of complex structures and processes and mathematical methods of physics and technical sciences.

Wiesław Szczesny born in 1951 in Pułtusk, Poland. He graduated from the Warsaw University in 1974 with a Master's degree in mathematics and continued his academic career at the Institute of Informatics at the Warsaw University till 1981 while he was awarded a doctorate in mathematics. Since 1982 he has been working at Warsaw University of Life Sciences in Warsaw. Szczesny has attained habilitation in informatics from Institute of Computer Science, Polish Academy of Sciences in 2004. He is currently working as Associated Professor at Faculty of Applied Informatics and Mathematics of Warsaw University of Life Sciences in Warsaw, Poland. His main research activity are grade models and methods for data analysis. He has numerous publications on statistics and informatics issues in Poland and abroad.

Bolesław Borkowski is a full professor of economy at the Faculty of Applied of Informatics and Mathematics Sciences at Warsaw University of Life Sciences-SGGW. He was first dean of this faculty. He is also a visiting Professor at the Department of Operations Research in Management at Warsaw University. Main fields of his scientific interests are mathematical methods in economy, especially in economics agricultural. He is an author of 110 scientific publications of which 11 books. Conjuration of econometrical models and methods of their parameters estimation are the main points of interest of the author. Prognosis of financial instruments prices and modeling of risk management aspects at farm level are of great interest to prof. Borkowski.



Ryszard Kozera is a Associate Professor at The Faculty of Applied Informatics and Mathematics at Warsaw University of Life Sciences - SGGW. He is also the Adjunct Associate Professor at The School of Information Science and Software Engine- ering at The University of Western Australia in Perth. He received his M.Sc. degree in Pure Mathematics at Warsaw University, Poland (1985), a Ph.D. degree in Computer Science at Flinders University of South Australia, Adelaide, Australia (1991) and his D.Sc. degree in Computer Science (a Habilitation) at The Silesian University of Technology, Gliwice, Poland (2006). Professor Kozera was awarded three times Alexander von Humboldt Fellowships between 1995-2004. He has delivered about 60 seminars at various prestigious international research centers, published 80 refereed journal and conference papers, organized international conferences and participated in the collaborative international grants. He has 28 postgraduate supervisions including Ph.D. and M.Sc. students. Professor Kozera's research interests include computer vision, numerical analysis, optimization, interpolation and approximation, partial differential equations and artificial intelligence. More information can be found under Professor Kozera's homepage (see http://www.wzim.sggw.pl/ryszard_kozera or http://www.csse.uwa.edu.au/~ryszard).

Yochanan Shachmurove is Professor of Economics and Business at the City College and the Graduate Center of the City University of New York, and a Visiting Professor at the Department of Economics at The University of Pennsylvania. He has worked with Harvard Institute for International Development advising the government of Ukraine with regards to its macroeconomic policies, including energy and natural resources issues. He has served as a Guest Editor for three special issues of the International Journal of Business, two of them dealing with Financial Markets of the Middle East, and one co-edited with Nobel Laureate Professor Lawrence Klein about the BRICs (Brazilian, Russian, Indian and Chinese economies). Professor Shachmuroves main research interests are in applied econometrics (including vector auto regressions, Bayesian vector auto regressions, emulative neural networks, non-linear deterministic systems, and chaos), international economics, industrial organization, entrepreneurial finance, advanced futures and options.