

Decomposition of gini and multivariate gini indices

Masato Okamoto

Received: 6 March 2007 / Accepted: 9 October 2007 / Published online: 7 February 2008
© Springer Science + Business Media, LLC 2007

Abstract A new type of decomposition by population subgroup is proposed for the Gini inequality index. The decomposition satisfies the completely identical distribution (CID) condition, whereby the between-group inequality is null if and only if the distribution within each subgroup is identical to all the others. Thus, this decomposition contrasts strikingly with the subgroup decomposition of the generalized entropy measures, which satisfy the condition that the between-group inequality is null if the mean within each subgroup equals those of all the others. The new decomposition can be generalized to the distance-Gini index and the volume-Gini index, two multivariate Gini indices introduced by Koshevoy and Mosler, with some modification of the index definition and a somewhat loosened CID condition in the latter case. The source decomposition is also generalized to these multi-dimensional indices. Interaction terms appear among sources of different attributes in the decomposition for the modified volume-Gini index.

Keywords Brunn–Minkowski inequality · Completely Identical Distribution (CID) condition · Cramér test · Multi-dimensional Gini index · Multilevel decomposition · Source decomposition · Subgroup decomposition

1 Introduction

Decomposability of inequality measures into contributions of population subgroups and contributions of sources is a desirable property for studies of economic inequality status and trends in populations. In fact, several types of subgroup decomposition of the Gini inequality index have been proposed so far. However, these decompositions have disadvantages such as inconsistency and impracticality. In contrast, the new type of decomposition presented in this paper has good properties. It is notable that the new decomposition satisfies the completely identical distribution (CID) condition, whereby the between-group inequality is null if and only if the distribution within each subgroup is identical to all the others. This is in striking contrast to the well-known subgroup decomposition of the Theil index or the generalized

M. Okamoto (✉)
Statistical Research and Training Institute, Ministry of Internal Affairs and Communications, Shinjyuku,
Tokyo 162-8668, Japan
e-mail: mokamoto2@stat.go.jp

entropy measures, which satisfy the condition whereby the between-group inequality is null if the mean within each subgroup equals those of all the others. It should also be noted that the new decomposition can be generalized to multivariate Gini indices while essentially maintaining its properties, indicating its suitability for the Gini decomposition.

Generalization of the Gini index to multivariate settings is a relatively new research issue, although the Gini index has been the most popular inequality measure for many years. Koshevoy and Mosler [11] proposed two types of multivariate Gini index, the distance-Gini index and the volume-Gini index. The former was formulated using an approach involving generalization of the univariate relative mean difference. The latter is a modification of the multivariate index proposed by Oja [14], which can be formulated with an approach using generalization of the Lorenz curve. Koshevoy and Mosler [11] showed that their indices are decomposable into subgroups in a similar way as the two-term decomposition of the ordinary univariate Gini index [see (14) in the next section] proposed by several researchers, including Rao [16] and Dagum [4]. However, their decomposition has disadvantages in terms of practicality and consistency. The new decomposition in this paper can easily be generalized based on studies of the multivariate Cramér test [2] in the case of the distance-Gini index. The generalization can also be achieved in the case of the volume-Gini index based on the Brunn–Minkowski inequality or Minkowski's first inequality concerning mixed volume, with some modifications of the index definition. The CID condition needs to be loosened somewhat in the latter case. The source decomposition of Rao [16] can also be generalized to both multivariate indices. It is notable that interaction terms appear among sources of different attributes in the source decomposition of the modified volume-Gini index.

The paper is organized as follows. The next section is devoted to subgroup decomposition of the usual univariate Gini index. Section 2.1 introduces the new type of subgroup decomposition, which is extended to the multilevel decomposition in section 2.2, and compared with other types of subgroup decomposition previously proposed in section 2.3. In section 2.4, the new decomposition is applied to Japanese household income data. The results for age-group decomposition and regional decomposition are presented. Section 3 is devoted to subgroup and source decomposition of the multivariate Gini indices. The new subgroup decomposition is generalized to the distance-Gini index in section 3.1, and to the volume-Gini index in section 3.2, with modifications of the index definition. Source decomposition of both indices is introduced in section 3.3, followed by applications of the subgroup decomposition to Japanese household income and expenditure data in section 3.4. Section 4 concludes discussions, with some remarks concerning multivariate inequality measures.

2 Subgroup decomposition of the Gini index

2.1 New type of subgroup decomposition

Let $F(y)$ represent the cumulative distribution function of a nonnegative random variable Y such as income, with a finite positive expectation μ . The Gini mean difference $M(F)$ can be presented in several ways, as follows:

$$\begin{aligned} M(F) &= \frac{1}{2} \int \int |x - y| dF(x) dF(y) = \int F(y)(1 - F(y)) dy \\ &= \int (y - \mu) d((F - 1)F) = 2 \int (y - \mu) \left(F(y) - \frac{1}{2} \right) dF(y) \end{aligned} \quad (1)$$

In the literature, double $M(F)$ is often called the Gini mean difference; however, $M(F)$ is defined as the Gini mean difference in this paper. Among the four equivalent representations in Eq. 1, the first is the original expression of the Gini mean difference, the fourth express it as the covariance between the variable Y and its rank $F(Y)$ (see [13]). Strictly speaking, when taking non-continuous distributions into consideration, the fourth representation does not hold. The second representation expresses $M(F)$ as the co-moment between the rank function $F(y)$ and its reverse rank function $1-F(y)$. Since the integrand of the second representation $F(y)(1-F(y))$ equals the expected variance of the binary variable “whether the random variable Y takes a value less than or equal to y ”, the second representation is also interpretable as the total of the expected variance of the binary variable over various values of Y . The second representation can be proved using Lemma 2.1 of Baringhaus and Franz [2]. The third and fourth representations are derived from the second using integration by parts. As for the Gini relative mean difference, in other words, the Gini inequality index $R(F) = M(F)/\mu$, the corresponding representations are obtained by division by μ in Eq. 1.

Assume that the population consists of groups $1, 2, \dots, n$. Let $F_i(y)$, μ_i , and p_i represent the cumulative distribution function, the expected value and the share of group i in the overall population, respectively. Note that $F(y) = \sum p_i F_i(y)$. Then, using the second representations in Eq. 1, the Gini mean difference $M(F)$ and the Gini index $R(F)$ can be decomposed by subgroup, respectively, as follows:

$$M(F) = \sum p_i M(F_i) + \sum p_i \int (F_i(y) - F(y))^2 dy \tag{2}$$

$$R(F) = \sum p_i \frac{\mu_i}{\mu} R(F_i) + \sum p_i \frac{1}{\mu} \int (F_i(y) - F(y))^2 dy. \tag{3}$$

The proof is by direct calculation. The first term on the right-hand side of Eq. 3 corresponds to the within-group inequality, and the second term corresponds to the between-group inequality. The contribution of each group to the between-group inequality can be naturally defined as follows:

$$p_i \frac{1}{\mu} \int (F_i(y) - F(y))^2 dy = p_i \frac{1}{\mu} cv(F_i, F) (\geq 0). \tag{4}$$

$cv(G, F) := \int (G(y) - F(y))^2 dy$ satisfies the following equality [2], which is useful for generalizing the decompositions 2 and 3 to multivariate settings, as shown in the next section:

$$cv(G, F) = \int \int |x - y| dG(x) dF(y) - \frac{1}{2} \int \int |x - y| dG(x) dG(y) - \frac{1}{2} \int \int |x - y| dF(x) dF(y). \tag{5}$$

Equality 5 forms the basis of the Cramér two-sample test and its generalization to multivariate settings [2]. In this connection, I call the between-group inequality, the second term on the right-hand side of Eq. 3, the Cramér coefficient of variation among groups $1, \dots, n$.

It may be felt that the second term should not be regarded as the between-group inequality because the functional form appears different to that of the Gini index. However, the second term has essentially the same form as the Gini index, because the Gini index

holds the following equality, which can be regarded as a special case of Eq. 3 with the null within-group inequality:

$$R(F) = \int dF(x) \frac{1}{\mu} \int (I_{[x,\infty)}(y) - F(y))^2 dy = \frac{1}{\mu} \int cv(I_{[x,\infty)}, F) dF(x), \tag{6}$$

where $I_{[x,\infty)}(y)=1$ if $y \geq x$, or 0 if $y < x$. Equality 6 expresses the notion that the Gini index is identical to the Cramér coefficient of variation if each population unit forms an individual group. Note that the indicator function $I_{[x,\infty)}$ in Eq. 6 corresponds to the one-point distribution function for a random variable, which takes value x almost surely. For proof note that $\int (I_{[x,\infty)}(y) - F(y))^2 dF(y) = F(y)(1 - F(y))$.

The following decompositions are also true:

$$M(F) = \sum p_i M(F_i) + \sum_{i < j} p_i p_j cv(F_i, F_j) \tag{7}$$

$$R(F) = \sum p_i \frac{\mu_i}{\mu} R(F_i) + \sum_{i < j} p_i p_j \frac{1}{\mu} cv(F_i, F_j). \tag{8}$$

These decompositions can be proved by direct calculation. Decomposition 8 attributes the between-group inequality to the relative mean squared difference of distribution functions between each pair of groups. From Eqs. 5 and 1, the following equality is obtained for $cv(F_i, F)$:

$$cv(F_i, F) = 4 \left(M \left(\frac{1}{2} F_i + \frac{1}{2} F \right) - \frac{1}{2} M(F_i) - \frac{1}{2} M(F) \right). \tag{9}$$

Thus, $cv(F_i, F)$ is four-fold greater than the surplus of dispersion in terms of the Gini mean difference for the horizontal merger of group i and the overall population. Assuming the ε to $1-\varepsilon$ merger ratio, where ε is a small positive number, the following representation is also obtained for $cv(F_i, F)$ by substituting $\varepsilon, 1-\varepsilon, F_i, F$ and $\varepsilon F_i + (1 - \varepsilon)F$ for p_1, p_2, F_1, F_2 and F , respectively in Eq. 7:

$$cv(F_i, F) = \frac{M(\varepsilon F_i + (1 - \varepsilon)F) - \varepsilon M(F_i) - (1 - \varepsilon)M(F)}{\varepsilon} + o(\varepsilon). \tag{10}$$

Thus, $cv(F_i, F)$ equals the surplus of the dispersion relative to the merger ratio when merger with an infinitely small merger ratio of group i takes place.

Obviously, the between-group inequality in decompositions 3 and 8 is null if and only if the distribution within each group is identical to those of all other groups. I call this condition the completely identical distribution (CID) condition. For this reason, the decomposition is quite different from that of the generalized entropy measures of inequality, in which the between-group inequality is null if and only if the mean within each group is equal to all the others. Bhattacharya and Mahalanobis [3] mentioned that, intuitively, it is reasonable to lay down the between-group component should not change if the group distributions F_i are changed, keeping μ_i fixed. However, Dagum [4] was opposed to taking the income means of subpopulations as their representative values to estimate the between-subpopulation inequality because income distributions significantly depart from normality. I believe that the new decomposition 3 favors Dagum’s view, although he pursued a different approach that added an extra component besides the between-group component of Bhattacharya and Mahalanobis, as shown in section 2.3.

2.2 Extension to multilevel decomposition

Let $F_{ij}(y)$, μ_{ij} , and p_{ij} represent the cumulative distribution function, the expected value and the share of subgroup j in group i , respectively. Noting that $F_i(y) = \sum_j p_{ij}F_{ij}(y)$, and $\mu_i = \sum_j p_{ij}\mu_{ij}$, the two-level decomposition of the Gini index $R(F)$ is derived by further decomposing the Gini index within each group $R(F_i)$ in Eq. 3 as follows:

$$R(F) = \sum_{ij} p_i \frac{\mu_i}{\mu} p_{ij} \frac{\mu_{ij}}{\mu_i} R(F_{ij}) + \sum_{ij} p_i \frac{\mu_i}{\mu} p_{ij} \frac{1}{\mu_i} \int (F_{ij}(y) - F_i(y))^2 dy \quad (11)$$

$$+ \sum_i p_i \frac{1}{\mu} \int (F_i(y) - F(y))^2 dy.$$

The second term on the right-hand side of Eq. 11 corresponds to the sum of the between-subgroup inequalities within groups. One of the advantages of decomposition 3 is that the between-group inequality is consistently defined with hierarchical grouping systems¹, since the between-subgroup inequality in the overall population equals the sum of the between-subgroup inequalities within groups and the between-group inequality in the overall population, i.e. the following equality is true for each group:

$$\sum_j p_{ij} \frac{1}{\mu} \int (F_{ij}(y) - F(y))^2 dy = \sum_j \frac{\mu_i}{\mu} p_{ij} \frac{1}{\mu_i} \int (F_{ij}(y) - F_i(y))^2 dy + \frac{1}{\mu} \int (F_i(y) - F(y))^2 dy. \quad (12)$$

Obviously, decomposition 11 can be further extended to thicker-layered decompositions, retaining the consistency.

2.3 Comparison with other types of subgroup decomposition

Several researchers, such as Pyatt [15] and Dagum [4], proposed the following three-term decomposition:

$$R(F) = \sum_i p_i^2 \frac{\mu_i}{\mu} R(F_i) + \frac{2}{\mu} \sum_{i < j} p_i p_j D_{ij} + \frac{1}{\mu} \sum_{i < j} p_i p_j |\mu_i - \mu_j|, \quad (13)$$

where $D_{ij} = \int dF_j(x) \int_0^x (x - y) dF_i(y)$ if $\mu_i > \mu_j$, or $\int dF_i(x) \int_0^x (x - y) dF_j(y)$ if $\mu_i < \mu_j$. The first term can be regarded as the contribution of the within-group inequality since the term is a weighted sum of the within-group inequality values. The third term on the right-hand side of (13) equals the between-group inequality defined by Bhattacharya and Mahalanobis [3]. The second term is regarded as the contribution of the trans-variation intensity, which measures a degree of overlap between the within-group distributions. This three-term decomposition is less satisfactory because it is inconsistent with multilevel groupings, and the weights assigned to the subgroups in the first term do not sum up to one.

The sum of the second and third terms is called the gross between-group Gini index, which, with the first term, comprises the following two-term decomposition:

$$R(F) = \sum_i p_i^2 \frac{\mu_i}{\mu} R(F_i) + \frac{1}{\mu} \sum_{i < j} p_i p_j \int \int |x - y| dF_i(x) dF_j(y). \quad (14)$$

¹ The definition of ‘‘consistency with hierarchical grouping systems’’ in this paper is entirely different from that of ‘‘subgroup consistency’’ defined by Shorrocks [18].

Although the two-term decomposition 14 can be extended to a multilevel decomposition consistently in a sense, the gross between-group Gini index cannot be regarded as the between-group inequality because it does not take the minimum constant value (usually normalized to null) when the within-group distributions or means are identical to each other.

Making use of the fourth representation of the Gini mean difference in Eq. 1, Yitzhaki and Lerman [24] proposed a different type of three-term decomposition, which yields a between-group inequality that is much closer to the counterpart of the new decomposition 3, as shown in section 2.4.

$$R(F) = \sum p_i \frac{\mu_i}{\mu} R(F_i) + \sum p_i \frac{\mu_i}{\mu} R(F_i)(O_{0i} - 1) + \frac{2}{\mu} \sum p_i (\mu_i - \mu) \left(G_i - \frac{1}{2} \right), \quad (15)$$

where $O_{0i} = \frac{\int_{(y-\mu_i)F_i(y)dF_i(y)}^{\int_{(y-\mu)F(y)dF(y)}}{G_i = \int F(y)dF_i(y)}$. In Eq. 15, O_{0i} measures the degree to which the overall distribution is included in the range of the within-group distribution i , and G_i is the expected rank of observations belonging to group i if they are ranked according to the ranking of the overall population. The third term on the right-hand side of Eq. 15 is the covariance between the within-group means and the average ranks of the respective groups. Thus, the third term can be regarded as the between-group inequality, which vanishes if the within-group mean μ_i equals that of all the others, or the average rank of each group G_i equals that of all the others. Although their decomposition has disadvantages in that the between-group inequality may take a negative value and there is inconsistency with multilevel groupings, it notably takes a step towards the new decomposition presented in this paper, in that the between-group inequality is defined using more than a single type of aggregates.

The contribution of each group to the between-group inequality in decomposition 3 shows the following relation to the components in decomposition 15 if F_i is continuous for any group:

$$p_i \frac{1}{\mu} \int (F_i - F)^2 dy = p_i \frac{\mu_i}{\mu} R(F_i)(O_{0i} - 1) + 2p_i \frac{1}{\mu} (\mu_i - \mu) \left(G_i - \frac{1}{2} \right) + p_i R(F)(O_{i0} - 1), \quad (16)$$

where $O_{i0} = \frac{\int_{(y-\mu)F_i(y)dF_i(y)}^{\int_{(y-\mu)F(y)dF(y)}}{The proof is given in the Appendix. Note that the third term on the right-hand side of Eq. 16 vanishes on summation.$

2.4 Applications to Japanese family income data

2.4.1 Decomposition of income inequality into age groups for household heads

The most recent survey results drew attention to a sharp rise in income inequality among the young generation in Japan, although the overall inequality has not risen notably if age effects are excluded. Table 1 shows trends in income inequality within each age group for household heads measured by the Gini index and the squared coefficient of variance (SCV). These indices measure the annual income inequality among households with two or more members. The estimates of the indices are derived from the National Survey of Family Income and Expenditures, a large-scale family budget survey of approximately 50,000 households, conducted by the Statistics Bureau, Ministry of Internal Affairs and Communications every 5 years. The Gini indices are estimated from two-way tables of income class by age group of the household head using the composite Simpson’s rule for

Table 1 Income inequality by age group for household head

Age group of household head	Gini index by year				SCV by year			
	1989	1994	1999	2004	1989	1994	1999	2004
Overall	0.2942	0.2983	0.3014	0.3088	0.4651	0.4277	0.4017	0.4543
Overall adjusted*	0.2942	0.2928	0.2887	0.2892	0.4651	0.4144	0.3759	0.3894
< 30 years old	0.2121	0.2174	0.2214	0.2373	0.1764	0.3260	0.1819	0.1998
30–39	0.2275	0.2199	0.2212	0.2234	0.2735	0.2440	0.2166	0.2034
40–49	0.2411	0.2455	0.2436	0.2439	0.2809	0.3434	0.2544	0.2401
50–59	0.2914	0.2741	0.2760	0.2847	0.4173	0.3192	0.3466	0.3422
60–69	0.3515	0.3572	0.3376	0.3376	0.5791	0.6577	0.5170	0.6178
70 ≤	0.4167	0.3853	0.3483	0.3466	2.3195	0.8987	0.7489	1.0343

*Calculated after replacing the share of each age group with that of 1989.

approximations of Lorenz domains in a similar manner to the official Gini estimates. SCVs are picked up from the existing statistical tables, so the estimates are virtually calculated from the micro data. There are 10 income classes, ranging from <2 million yen to ≥15 million yen. Such Gini estimates are empirically considered to be good approximations to those estimated from the micro data.

As shown in Table 2, decomposition into age groups for household heads² reveals that the youngest group, with a household head <30 years old, did not contribute to the slight increase in overall inequality between 1999 and 2004, despite the sharp increase in the within-group inequality. The increase in relative income of the youngest group rather contributed to the decrease in the between-group inequality, which canceled out the positive contribution of the within-group inequality. Thus, the slight increase in overall inequality between 1999 and 2004 should be attributed to contributions of other age groups, which did not draw much attention.

The between-group inequality and contribution of each age group measured by the Gini decomposition of Yitzhaki and Lerman [24] are approximately two-fold greater than their counterparts in the new decomposition with the same sign.

2.4.2 Regional inequality in income distribution

It has recently been speculated that the between-region inequality, in particular the gap between the metropolitan areas including Greater Tokyo and other areas, is increasing, as well as the between-household inequality. However, the actual trend may differ somewhat from this speculation according to the results derived from the new decomposition. Table 3 shows the recent trends in regional income inequality in Japan, measured by three types of the Gini decomposition and decomposition of the generalized entropy measures. The single-parameter entropy family is defined as follows:

$$E_c(F) = \int \varphi_c(y/\mu) dF(y), \quad (17)$$

² As noted above, estimates for the new decomposition are made from aggregates using the composite Simpson's rule (with Eq. 9 for calculations of the between-group inequalities).

Table 2 Decomposition of income inequality into age groups for household heads*

Year	Inequality indices and related quantities		Age group of household head						
			Overall	< 30	30–39	40–49	50–59	60–69	70 ≤
2004	Population share	p_i	1.0000	0.0476	0.2295	0.2982	0.2303	0.1415	0.0528
	(in 1989)								
	Relative income	μ_i/μ	1.0000	0.6607	0.8407	1.0948	1.2336	0.8813	0.7622
	Gini Contribution		0.2892	0.0108	0.0467	0.0826	0.0878	0.0443	0.0169
	Within-group	$p_i\mu_i/\mu R_i$	0.2672	0.0075	0.0431	0.0796	0.0809	0.0421	0.0140
	Between-group	$p_i cv_i/\mu$	0.0221	0.0034	0.0036	0.0030	0.0069	0.0022	0.0030
	Within-group	R_i		0.2373	0.2234	0.2439	0.2847	0.3376	0.3466
	Gini								
	cvc**	cv_i/μ		0.0708	0.0158	0.0100	0.0299	0.0154	0.0565
	Yitzhaki–Lerman between-group Gini		0.0456	0.0081	0.0072	0.0047	0.0169	0.0037	0.0051
	SCV Contribution		0.3894	0.0094	0.0390	0.0933	0.1401	0.0721	0.0356
	Within-group	$p_i(\mu_i/\mu)^2 scv_i$	0.3562	0.0043	0.0343	0.0892	0.1247	0.0706	0.0330
	Between-group	$p_i(\mu_i/\mu-1)^2$	0.0332	0.0051	0.0047	0.0040	0.0153	0.0014	0.0026
	Within-group SCV	scv_i		0.1998	0.2034	0.2401	0.3422	0.6178	1.0343
Change from 1999	Population share	p_i	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(in 1989)								
	Relative income	μ_i/μ	0.0000	0.0366	-0.0021	0.0013	-0.0203	0.0054	0.0427
	Gini Contribution		0.0005	-0.0002	0.0000	0.0003	-0.0001	0.0002	0.0002
	Within-group	$p_i\mu_i/\mu R_i$	0.0035	0.0009	0.0003	0.0002	0.0012	0.0003	0.0007
	Between-group	$p_i cv_i/\mu$	-0.0030	-0.0010	-0.0003	0.0001	-0.0013	-0.0000	-0.0005
	Within-group	R_i		0.0160	0.0022	0.0003	0.0087	0.0000	-0.0017
	Gini								
	cvc**	cv_i/μ		-0.0220	-0.0013	0.0004	-0.0057	-0.0001	-0.0094
	Yitzhaki–Lerman between-group Gini		-0.0057	-0.0018	-0.0004	0.0001	-0.0026	0.0002	-0.0011
	SCV Contribution		0.0136	-0.0005	-0.0021	-0.0027	-0.0047	0.0129	0.0108
	Within-group	$p_i(\mu_i/\mu)^2 scv_i$	0.0175	0.0009	-0.0018	-0.0035	-0.0035	0.0133	0.0121
	Between-group	$p_i(\mu_i/\mu-1)^2$	-0.0039	-0.0014	-0.0004	0.0007	-0.0011	-0.0004	-0.0013
	Within-group SCV	scv_i		0.0179	-0.0132	-0.0143	-0.0044	0.1008	0.2854

*Calculated after replacing the share of each age group with that of 1989.

**Contribution to the between-group inequality relative to its population share.

where $\varphi_c(y/\mu) = ((y/\mu)^c - 1)/c(c - 1)$ if $c \neq 0, 1$, $\varphi_1(y/\mu) = (y/\mu) \log(y/\mu)$, $\varphi_0(y/\mu) = \log(\mu/y)$. The inequality indices drawn from the entropy family 17 satisfy the following subgroup decomposition:

$$E_c(F) = \sum p_i \left(\frac{\mu_i}{\mu}\right)^c E_c(F_i) + \sum p_i \varphi_c\left(\frac{\mu_i}{\mu}\right). \tag{18}$$

The first and second terms on the right-hand side of Eq. 18 correspond to the within- and between-group inequality, respectively. In Table 3, the between-region inequality indices in the case of $c=0, 1, 2$ are denoted by E_0, E_1 and E_2 , respectively. E_2 is equivalent to SCV.

A two-level regional grouping is used for calculation of the regional inequality. The whole country consists of 47 prefectures. Each prefecture was subdivided into approximately 3,000 municipalities (cities, wards, towns and villages) before many

Table 3 Regional income inequality

Type of index/ decomposition	Between-prefecture inequality by year				Between-municipality-group inequality by year		
	1989	1994	1999	2004	1994	1999	2004
Gini							
New decomposition	0.0086	0.0067	0.0048	0.0045	0.0090	0.0074	0.0079
Yitzhaki–Lerman	0.0178	0.0141	0.0100	0.0101	0.0185	0.0145	0.0155
Bhattacharya–Mahalanobis	0.0716	0.0614	0.0521	0.0577	0.0697	0.0629	0.0672
E_2 (SCV)	0.0161	0.0123	0.0088	0.0099	0.0156	0.0127	0.0149
E_1 (Theil)	0.0083	0.0063	0.0045	0.0051	0.0081	0.0065	0.0075
E_0	0.0086	0.0066	0.0047	0.0052	0.0084	0.0066	0.0077

municipality mergers took place in 2005. The municipalities in each prefecture were grouped for the National Survey of Family Income and Expenditures, based on the more detailed municipality grouping used for the Establishment and Enterprise Census. Each municipality group consists of neighboring municipalities within the same prefecture. These groups were determined by taking into consideration spheres of residential life or economic relations among municipalities. There are 274 municipality groups in total. Similar to the application in section 2.4.1, Gini indices and their breakdowns by the new decomposition are estimated from two-way tables of income class by regions using the composite Simpson's rule for approximations of Lorenz domains.

The Gini decomposition of Bhattacharya and Mahalanobis [3] and decomposition of the generalized entropy measures indicate an increase in the between-prefecture inequality between 1999 and 2004 as well as an increase in the between-municipality-group inequality. In contrast, the new Gini decomposition indicates a continuance of the downtrend of the between-prefecture inequality, despite an upturn of the between-municipality-group inequality. The Gini decomposition of Yitzhaki and Lerman shows little change in the between-prefecture inequality in the same period. Thus, the new decomposition and the decomposition of Yitzhaki and Lerman imply that the regional inequality within prefectures should be an issue rather than the gaps among prefectures. If attaching greater importance to consistency with the measurement of between-household inequality usually made by the Gini index in Japan, the implication derived from the new Gini decomposition should be noted, and it deserves further investigation.

Shorrocks and Wan [19] pointed out that the Gini decomposition of Bhattacharya and Mahalanobis produces considerably greater shares for the between-group inequality in the overall inequality compared to the decompositions of other indices. However, the new Gini decomposition produces slightly smaller shares for the between-group inequality compared to the decompositions of other indices, as shown in Table 4. The Gini decomposition of Yitzhaki and Lerman produces slightly greater shares. Similar to the decomposition into age groups for household heads shown in section 2.4.1, the between-prefecture inequality derived from the Yitzhaki and Lerman decomposition is approximately two-fold greater than that of the new decomposition. However, for the between-municipality-group inequality, the relative difference is less than double. It seems intuitive to suppose that the more minutely the population is subdivided, the smaller is the relative difference becomes.

Table 4 Ratio of regional income inequality to the overall income inequality (1999)

Type of index/ decomposition	Between-prefecture inequality (%)	Between-municipality-group inequality within prefecture (%)	Between-municipality- group inequality (%)
Gini			
New decomposition	1.6	0.8	2.4
Yitzhaki–Lerman	3.3	1.5	4.4
Bhattacharya–Mahalanobis	17.2	10.0	20.8
E_2 (SCV)	2.2	1.0	3.2
E_1 (Theil)	2.9	1.2	4.1
E_0	2.9	1.2	4.2

3 Decomposition of the multivariate Gini index

3.1 Subgroup decomposition of the distance Gini index

In this section, the new type of subgroup decomposition for the Gini index is generalized to multivariate Gini indices. First, the corresponding decomposition of the distance-Gini index, a variation of the multivariate Gini index proposed by Koshevoy and Mosler [11], is introduced, applying the achievement of Baringhaus and Franz [2] concerning the multivariate Cramér test.

Let $\mathbf{x}=\{x_k\}$ and $\mathbf{y}=\{y_k\}$ be d -dimensional vectors, and let $F(\mathbf{y})$ represent the distribution function of a d -variate random variable \mathbf{Y} on the orthant R^d_+ with a finite positive expectation vector $\boldsymbol{\mu}=\{\mu_k\}$. Koshevoy and Mosler [11] defined the distance-Gini mean difference $M_D(F)$ as follows:

$$M_D(F) = \frac{1}{2d} \int \int \|\mathbf{x} - \mathbf{y}\| dF(\mathbf{x})dF(\mathbf{y}), \text{ where } \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{k=1}^d (x_k - y_k)^2}. \quad (19)$$

Let $\mathbf{x}/\boldsymbol{\mu}$ be a vector $\{x_k/\mu_k\}$, and $\tilde{F}(y)$ be the distribution function of a random variable $\mathbf{Y}/\boldsymbol{\mu}$. Then Koshevoy and Mosler [11] defined the distance-Gini relative mean difference (the distance-Gini index) $R_D(F)$ as $M_D(\tilde{F})$. In the univariate case ($d=1$), the distance-Gini index is identical to the ordinary Gini index.

The Euclidean norm can be represented as follows (e.g. [9]); the proof is also given in a more general form in the Appendix):

$$\|\mathbf{x}\| = C_d \int_{S^{d-1}} |\mathbf{a}' \cdot \mathbf{x}| dv(\mathbf{a}), \quad (20)$$

where v is the uniform distribution on the unit sphere $S^{d-1} = \{\mathbf{a} \in R^d \mid \|\mathbf{a}\| = 1\}$, and $C_d = \Gamma((d + 1)/2)/2\pi^{(d-1)/2}$. Equation 20 allows the distance-Gini mean difference to be represented as follows [12]:

$$M_D(F) = \frac{C_d}{2d} \int_{S^{d-1}} dv(\mathbf{a}) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |u - v| dF(u, \mathbf{a})dF(v, \mathbf{a}) = \frac{C_d}{2d} \int_{S^{d-1}} dv(\mathbf{a}) \int_{-\infty}^{\infty} F(u, \mathbf{a})(1 - F(u, \mathbf{a}))du, \quad (21)$$

where $F(\cdot, \mathbf{a})$ denotes the distribution function of $\mathbf{a}\cdot\mathbf{X}$, the projection of the random variable \mathbf{X} on the line spanned by vector \mathbf{a} . The corresponding representation for the distance-Gini index is obtained by substituting $\tilde{F}(\cdot, \mathbf{a})$ for $F(\cdot, \mathbf{a})$, where $\tilde{F}(\cdot, \mathbf{a})$ denotes the distribution function

of $\mathbf{a} \cdot \mathbf{X} / \mu$. Similar to the second representation of the Gini mean difference in Eq. 1, the second representation of $M_D(F)$ in Eq. 21 can be interpreted as the total expected variance for whether or not $\mathbf{a} \cdot \mathbf{X} \leq u$ occurs over the relative level u and the projection direction \mathbf{a} .

Koshevoy and Mosler [12] mentioned that the distance-Gini index is decomposable into subgroup in a similar manner to the two-term decomposition 14. However, such decomposition has the same disadvantages as Eq. 14.

An extension of Eq. 5 to multivariate settings allows the (new) subgroup decomposition for the distance-Gini index. Let $G(\mathbf{y})$ represents another d -variate distribution function. Baringhaus and Franz [2, Theorem 2.1) proved the following inequality, where the equality holds if and only if $F=G$:

$$cv_D(G, F) := \frac{1}{d} \left(\int \int \| \mathbf{x} - \mathbf{y} \| dG(\mathbf{x})dF(\mathbf{y}) - \frac{1}{2} \int \int \| \mathbf{x} - \mathbf{y} \| dG(\mathbf{x})dG(\mathbf{y}) - \frac{1}{2} \int \int \| \mathbf{x} - \mathbf{y} \| dF(\mathbf{x})dF(\mathbf{y}) \right) \geq 0. \tag{22}$$

Using Eqs. 4, 5 and 20, $cv_D(G, F)$ can be represented as follows:

$$cv_D(G, F) = \frac{C_d}{d} \int_{S^{d-1}} d\nu(\mathbf{a}) \int_{-\infty}^{\infty} (G(u, \mathbf{a}) - F(u, \mathbf{a}))^2 du. \tag{23}$$

Assume the population consists of groups $1, 2, \dots, n$. Let $F_i(\mathbf{y})$, μ_i , and p_i represent the d -variate distribution function, the expectation vector and the share of group i in the overall population, respectively. Let $\tilde{F}_i(\mathbf{y})$ and $\bar{F}_i(\mathbf{y})$ be the distribution functions of \mathbf{Y} / μ_i and \mathbf{Y} / μ within group i , respectively. Then, the distance-Gini mean difference $M_D(F)$ and the distance-Gini index $R_D(F)$ can be decomposed as follows:

$$M_D(F) = \sum p_i M_D(F_i) + \sum p_i cv_D(F_i, F) \tag{24}$$

$$R_D(F) = M_D(\tilde{F}) = \sum p_i r_D(F_i) R_D(F_i) + \sum p_i cv_D(\tilde{F}_i, \tilde{F}), \tag{25}$$

where $r_D(F_i)=0$ if $R_D(F_i)=0$ or $M_D(\tilde{F}_i) / M_D(\tilde{F})$ otherwise. The proof is given in the Appendix. $r_D(F_i)$ in Eq. 25 corresponds to the average relative level of group i . If $\mu_i = \mu$, then $r_D(F_i)=1$. However, if $\mu_i \neq \mu$, $r_D(F_i)$ depends on the distribution F_i unlike the univariate case. The second term on the right-hand side of Eq. 25 corresponds to the between-group inequality. It is null if and only if $F_i=F$ for any group. That is, subgroup decomposition 25 satisfies the CID condition.

The distance-Gini mean difference and the distance-Gini index also have decompositions that correspond to Eqs. 7 and 8, respectively. The decomposition 8 is extended as follows:

$$R_D(F) = \sum p_i r_D(F_i) R_D(F_i) + \sum_{i < j} p_i p_j cv_D(\tilde{F}_i, \tilde{F}_j). \tag{26}$$

The decomposition 26 can be proved in a similar manner to the proof of Eq. 25. $cv_D(F_i, F)$ holds Eq. 27, which correspond to Eq. 10.

$$cv_D(F_i, F) = \frac{M_D(\varepsilon F_i + (1 - \varepsilon)F) - \varepsilon M_D(F_i) - (1 - \varepsilon)M_D(F)}{\varepsilon} + o(\varepsilon). \tag{27}$$

3.2 Subgroup decomposition of the modified volume-Gini index

3.2.1 Modified Torgersen index

Several types of multivariate Gini index proposed in the past can be defined based on the generalized Lorenz domain. For the introduction of such indices, let $\boldsymbol{\gamma}=\{\gamma_i\}$ be a non-negative $d(\geq 2)$ -dimensional constant vector, and define the quantity $M_T(F|\boldsymbol{\gamma})$ as follows:

$$M_T(F|\boldsymbol{\gamma}) = \frac{1}{d!(1 + \sum \gamma_i)} \int \cdots \int |\det(\mathbf{y}_1 - \boldsymbol{\gamma}, \dots, \mathbf{y}_d - \boldsymbol{\gamma})| dF(\mathbf{y}_1) \cdots dF(\mathbf{y}_d). \tag{28}$$

Thus, a type of multivariate Gini index $R_T(F|\boldsymbol{\gamma})$ is defined as $M_T(\tilde{F}|\boldsymbol{\gamma})$. $R_T(F|\boldsymbol{\gamma})$ ranges from zero to unity, as proved in the Appendix. Since $R_T(F|\boldsymbol{\gamma})$ equals the multivariate Gini index of Torgersen [21] if $\boldsymbol{\gamma}=\mathbf{0}$, $R_T(F|\boldsymbol{\gamma})$ is called the modified Torgersen index. $M_T(F|\boldsymbol{\gamma})$, which corresponds to the Gini mean difference, is hereafter called the modified Torgersen mean difference or mean volume. As explained later, $R_T(F|\mathbf{1})$, where $\mathbf{1}=\{1, \dots, 1\}$, is preferable to the original Torgersen index $R_T(F|\mathbf{0})$. Similarly, $M_T(F|\boldsymbol{\mu})$ is preferable to $M_T(F|\mathbf{0})$. Oja [14] proposed a different generalization, as follows:

$$\begin{aligned} M_O(F) = \sigma_1(F) &= \frac{1}{(d+1)!} \int \cdots \int \left| \det \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \mathbf{y}_{d+1} \end{pmatrix} \right| dF(\mathbf{y}_1) \cdots dF(\mathbf{y}_{d+1}) \\ &= \frac{1}{d+1} \int \left(1 + \sum \gamma_i \right) M_T(F|\boldsymbol{\gamma}) dF(\boldsymbol{\gamma}). \end{aligned} \tag{29}$$

$R_O(F) = M_O(\tilde{F})$. If $d=1$, $M_O(F)$ is equivalent to the univariate Gini mean difference, and $R_O(F)$ is equivalent to the ordinary Gini index. $M_O(F)$ was introduced as a variation of the generalized variance of Wilks [23], which can be presented as follows:

$$\begin{aligned} \sigma_2^2(F) &= \frac{1}{(d+1)!} \int \cdots \int \left| \det \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \mathbf{y}_{d+1} \end{pmatrix} \right|^2 dF(\mathbf{y}_1) \cdots dF(\mathbf{y}_{d+1}) \\ &= \frac{1}{d!} \int \cdots \int |\det(\mathbf{y}_1 - \boldsymbol{\mu}, \dots, \mathbf{y}_d - \boldsymbol{\mu})|^2 dF(\mathbf{y}_1) \cdots dF(\mathbf{y}_d) \\ &= \det \left(\int (\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T dF(\mathbf{y}) \right). \end{aligned} \tag{30}$$

The third representation (rightmost side) in Eq. 30 expresses the notion that the generalized variance equals the determinant of the variance–covariance matrix of distribution F . $M_O(F)$ and $M_T(F|\boldsymbol{\mu})$ can be regarded as counterparts of the first and second representations of the generalized variance, respectively.

3.2.2 Multidimensional Lorenz domain

The modified Torgersen index and the Oja index can be defined based on the volume of the Lorenz zonoids, which is a generalization of the Lorenz domain, introduced by Koshevoy and Mosler [10]. The relation between the indices and the Lorenz zonoids was studied by Koshevoy and Mosler [11]. This relation is utilized to derive the subgroup decomposition

of the indices in this paper. For a measurable function: $\phi : R_+^d \rightarrow [0, 1]$, consider a d -dimensional vector $\mathbf{z}(\phi, F|\gamma)$, where

$$\mathbf{z}(\phi, F|\gamma) = \int (\mathbf{y} - \gamma)\phi(\mathbf{y})dF(\mathbf{y}). \tag{31}$$

The following set $Z_T(F|\gamma)$, consisting of all $\mathbf{z}(\phi, F|\gamma)$, is called the γ -zonoid of the distribution F .

$$Z_T(F|\gamma) = \{\mathbf{z}(\phi, F|\gamma) | \phi : R_+^d \rightarrow [0, 1]\}. \tag{32}$$

$Z_T(\tilde{F}|\gamma)$, the γ -zonoid of \tilde{F} that corresponds to the distribution function of \mathbf{Y}/μ , is called the γ -Lorenz zonoid of F . $Z_O(F)$, the lift zonoid of F , is defined in the $d+1$ dimensional space $[0, 1] \times R_+^d$ as follows:

$$Z_O(F) = \{(p(\phi, F), \mathbf{z}(\phi, F|\mathbf{0})) | \phi : R_+^d \rightarrow [0, 1]\}, \text{ where } p(\phi, F) = \int \phi(\mathbf{y})dF(\mathbf{y}). \tag{33}$$

$Z_O(\tilde{F})$ is called the Lorenz zonoid of F . The γ -zonoids and lift zonoids belong to the family of the convex bodies – i.e. nonempty, compact and convex subsets of R^d (regardless of whether they contain interior points or not). $Z_O(F)$ and $Z_O(\tilde{F})$ are projected onto $Z_T(F|\gamma)$ and $Z_T(\tilde{F}|\gamma)$, respectively, by the following linear transformation from $[0, 1] \times R_+^d$ onto R_+^d :

$$(p, \mathbf{z}) \rightarrow \mathbf{z} - p\gamma. \tag{34}$$

In the univariate case ($d=1$), the Lorenz zonoid has the shape shown in Fig. 1. The boundary of the univariate Lorenz zonoid consists of the Lorenz curve and the inverse Lorenz curve, which is equivalent to the Lorenz curve if rotated on the center point $(1/2, 1/2)$ at an angle of 180° .

The volumes of $Z_T(F|\gamma)$ and $Z_T(\tilde{F}|\gamma)$ multiplied by the reciprocal of $1 + \sum \gamma_i$ equal the modified Torgersen mean difference $M_T(F|\gamma)$ and the modified Torgersen index $R_T(F|\gamma)$, respectively. Similarly, the volumes of $Z_O(F)$ and $Z_O(\tilde{F})$ equal the Oja mean difference $M_O(F)$ and the Oja index $R_O(F)$, respectively. The relation between the Lorenz zonoid and the Oja index was proved by Koshevoy and Mosler [11, Theorem 5.1]. The relation between the γ -Lorenz zonoid and the modified Torgersen index can also be proved along the same lines. In the case of finite-point distributions, it is essentially the relation between the Minkowski sum of line segments and its volume (e.g. [20]). Koshevoy and Mosler generalized this using the existence of a sequence of finite-point distributions, which converges weakly to any distribution.

3.2.3 Subgroup decomposition of the modified Torgersen index

To introduce the subgroup decomposition of the modified Torgersen index, we define the mixed volume of $Z_T(F|\gamma)$ and $Z_T(G|\gamma)$ with $d-1$ repetitions of $Z_T(F|\gamma)$ as follows:

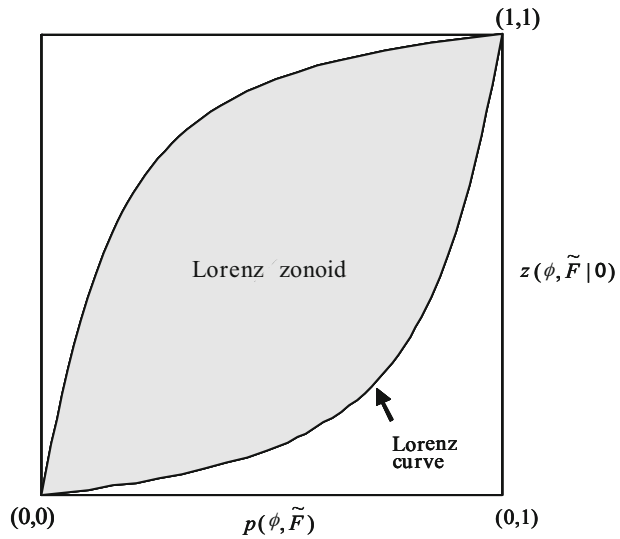
$$MV_{d-1}(F, G|\gamma) = \frac{1}{d!} \int \dots \int |\det(\mathbf{y}_1 - \gamma, \dots, \mathbf{y}_d - \gamma)| dF(\mathbf{y}_1) \dots dF(\mathbf{y}_{d-1}) dG(\mathbf{y}_d). \tag{35}$$

Definition 35 is equivalent to the following ordinary definition (e.g. [8]):

$$MV_{d-1}(F, G|\gamma) = \lim_{\epsilon \rightarrow +0} \frac{\text{vol}(Z_T(F|\gamma) + \epsilon Z_T(G|\gamma)) - \text{vol}(Z_T(F|\gamma))}{\epsilon d}, \tag{36}$$

where $Z_T(F|\gamma) + \epsilon Z_T(G|\gamma) = \{x + \epsilon y | x \in Z_T(F|\gamma), y \in Z_T(G|\gamma)\}$ is the Minkowski sum of $Z_T(F|\gamma)$ and $\epsilon Z_T(G|\gamma)$, and $\text{vol}(\bullet)$ denotes the volume of the γ -zonoid. Note that

Fig. 1 Illustration of the Lorenz zonoid in the univariate case



$\text{vol}(Z_T(F|\gamma)) = (1 + \sum \gamma_i)M_T(F|\gamma)$. According to Minkowski’s first inequality concerning the mixed volume (e.g. [8]), the following inequality is true if $M_T(F|\gamma) > 0$.

$$cv_T(G, F|\gamma) := \frac{MV_{d-1}(F, G|\gamma)}{(1 + \sum \gamma_i)M_T(F|\gamma)^{(d-1)/d}} - M_T(G|\gamma)^{1/d} \geq 0. \tag{37}$$

The equality holds if and only if $Z_T(F|\gamma)$ and $Z_T(G|\gamma)$ are homothetic – i.e. $Z_T(F|\gamma) = \alpha Z_T(G|\gamma)$, where α is a positive constant.

Assume that the population consists of groups $1, 2, \dots, n$. Let $F_i(\mathbf{y})$, μ_i , and p_i represent the d -variate distribution function, the expectation vector and the share of group i in the overall population, respectively. Let $\tilde{F}_i(\mathbf{y})$ be the distribution functions of \mathbf{Y}/μ within group i . Then inequality 37 allows the following decompositions by subgroup:

$$M_T(F|\gamma)^{1/d} = \sum p_i M_T(F_i|\gamma)^{1/d} + \sum p_i cv_T(F_i, F|\gamma) \tag{38}$$

$$R_T(F|\gamma)^{1/d} = \sum p_i r_T(F_i|\gamma) R_T(F|\gamma)^{1/d} + \sum p_i cv_T(\tilde{F}_i, \tilde{F}|\gamma), \tag{39}$$

where $r_T(F_i|\gamma) = 0$ if $R_T(F|\gamma) = 0$, or $M_T(\tilde{F}_i|\gamma)^{1/d} / R_T(F_i|\gamma)^{1/d}$ otherwise, and $cv_T(G, F|\gamma) = 0$ if $M_T(F|\gamma) = 0$. Note the following equality for the derivation of decompositions 38:

$$M_T(F|\gamma) = \sum p_i \frac{1}{1 + \sum \gamma_i} MV_{d-1}(F, F_i|\gamma) \tag{40}$$

The corresponding equality for the derivation of Eq. 39 is obtained by substituting \tilde{F} and \tilde{F}_i for F and F_i , respectively. The second term in decomposition 39 corresponds to the between-group inequality. According to Minkowski’s first inequality concerning the mixed volume, the second terms on the right-hand side of Eqs. 38 and 39 vanishes if and only if $Z_T(F_i|\gamma)$ is homothetic to $Z_T(F|\gamma)$ for any group. This is true if $M_T(F|\gamma) > 0$ – i.e. $Z_T(F|\gamma)$

has interior points. If $M_T(F|\gamma) = 0$ – i.e. $Z_T(F|\gamma)$ is on some hyperplane, the Brunn-Minkowski inequality (e.g. [8]) asserts that $Z_T(F_i|\gamma)$ is on the same hyperplane as $Z_T(F|\gamma)$; however, $Z_T(F_i|\gamma)$ does not need to be homothetic to $Z_T(F|\gamma)$ in this case.

On the assumption that $R_T(F|\gamma) > 0$, if the between-group inequality is null, the mean within each group μ_i equals $\alpha_i\mu + (1 - \alpha_i)\gamma\mu$ with some homothetic ratio $\alpha_i > 0$. Thus, if $\gamma = \mathbf{1}$, μ_i equals μ if the between-group inequality is null, while μ_i equals $\alpha_i\mu$ if $\gamma = \mathbf{0}$. On this basis, $R_T(F|\mathbf{1})$ is preferable to the original Torgersen index $R_T(F|\mathbf{0})$, whereas $r_T(F_i|\gamma)$ can be expressed as the simpler form $(\prod \mu_i/\mu)^{1/d}$ in the latter case.

It may be reasonable to assert that $R_T(F|\gamma)^{1/d}$ should be used as a multivariate inequality index instead of $R_T(F|\gamma)$, taking the decomposability into consideration. This question is left open in this paper, since the subsequent discussions do not require any specific decision on this, although $R_T(F|\gamma)^{1/d}$ is used for definition of the modified volume-Gini index.

$cv_T(F_i, F|\gamma)$ can be regarded as the contribution of each group to the between-group mean difference relative to its population share. It has the following representation, which corresponds to Eq. 10 for the univariate Gini mean difference and Eq. 27 for the distance-Gini mean difference:

$$cv_T(F_i, F|\gamma) = \lim_{\varepsilon \rightarrow +0} \frac{1}{\varepsilon} \left(M_T((1 - \varepsilon)F + \varepsilon F_i|\gamma)^{1/d} - (1 - \varepsilon)M_T(F|\gamma)^{1/3} - \varepsilon M_T(F_i|\gamma)^{1/d} \right). \tag{41}$$

The proof is given in the Appendix. The representation 41 expresses the notion that $cv_T(F_i, F|\gamma)$ equals the surplus of the dispersion relative to the merger ratio when a merger with an infinitely small ratio of group i takes place.

3.2.4 Modified volume-Gini index and its subgroup decomposition

The Oja and the modified Torgersen indices vanish not only when the distribution is egalitarian – i.e. for a one-point distribution – but also when the distribution is on a hyperplane. In this extreme case in which only one population unit monopolizes all income and property, the inequality measures zero. To avoid this drawback, Koshevoy and Mosler [11] proposed the volume-Gini mean difference, as follows:

$$M_{KM}(F) = \frac{1}{2^d - 1} \sum_{s=1}^d \sum_{1 \leq j_1 < \dots < j_s \leq d} M_O(F^{j_1 \dots j_s}), \tag{42}$$

where $F^{j_1 \dots j_s}$ is the marginal distribution in the space of sub-coordinate axes $\{j_1, \dots, j_s\}$. the volume-Gini index $R_{KM}(F)$ is defined as $M_{KM}(\tilde{F})$, namely, the average of the Oja sub-indices for the distribution F and all its marginal distributions in the spaces of the sub-coordinate axes. Since the Oja sub-index for any univariate marginal distribution (identical to the ordinary Gini index) vanishes if $R_{KM}(F)$ equals zero, $R_{KM}(F)$ vanishes if and only if the distribution is egalitarian. Thus, the drawback is surely overcome. However, further modification seems to be desirable, taking it into consideration that the Oja sub-indices for marginal distributions vary in homothetic degree to the following enlargement with dilation factor $\lambda (> 0)$ and center at the mean μ :

$$T_{\lambda, \mu} : \mathbf{y} \rightarrow \lambda(\mathbf{y} - \mu) + \mu. \tag{43}$$

Note that the d -variate Oja index is of homothetic degree d – i.e. $R_O(T_{\lambda, \mu}(F)) = \lambda^d R_O(F)$. For this reason, the greater the dilation, the higher is the relative contribution of higher sub-dimensional marginal distributions, although the shape of the distribution and the mean

remains the same. Furthermore, the decomposability of the volume-Gini index is also questionable. Koshevoy and Mosler [12] showed a type of generalization of the two-term decomposition 14. However, their decomposition has a very complex form, in addition to the same disadvantages as Eq. 14.

Thus, the volume-Gini mean difference should be modified by replacing the Oja mean sub-volumes, except for the univariate cases, with the modified Torgersen mean sub-volumes, power-transformed by the reciprocal of the dimensions, as follows:

$$M_V(F|\gamma) = \frac{1}{2^d - 1} \left(\sum_{s=1}^d M(F^s) + \sum_{s=2}^d \sum_{1 \leq j_1 < \dots < j_s \leq d} M_T(F^{j_1 \dots j_s}|\gamma)^{1/s} \right). \tag{44}$$

Strictly speaking, $M_T(F^{j_1 \dots j_s}|\gamma)^{1/s}$ in Eq. 44 should be denoted as $M_T(F^{j_1 \dots j_s}|\gamma^{j_1 \dots j_s})^{1/s}$, where $\gamma^{j_1 \dots j_s} = \{\gamma_{j_1}, \dots, \gamma_{j_s}\}$; However, the above notation is used for simplicity. The modified volume-Gini index $R_V(F|\gamma)$ is defined as $M_V(\tilde{F}|\gamma)$. $R_V(F|\gamma)$ as well as the original volume-Gini index $R_{KM}(F)$ equals zero if and only if F is egalitarian. In addition, if $\gamma = \mathbf{1}$, $R_V(F)$ is of homothetic degree one to the enlargement $T_{\lambda, \mu}$, and the relative contribution of any sub-index is invariant to $T_{\lambda, \mu}$. Furthermore, $M_V(F|\gamma)$ and $R_V(F|\gamma)$ are decomposable into subgroups, as shown below. If considering only the homotheticity, the Oja sub-indices need not be replaced and the power transformations by reciprocal dimensions are sufficient. However, the decomposability is an open question in this case.

Assume $M_T(F|\gamma) > 0$. Then, according to inequalities 4 and 37, the following inequality is true:

$$cv_V(G, F|\gamma) := \frac{1}{2^d - 1} \left(\sum_{s=1}^d cv(G^s, F^s) + \sum_{s=2}^d \sum_{1 \leq j_1 < \dots < j_s \leq d} cv_T(G^{j_1 \dots j_s}, F^{j_1 \dots j_s}|\gamma) \right) \geq 0. \tag{45}$$

The equality holds if and only if $G^s = F^s$ for any coordinate axis and $Z_T(F|\gamma) = Z_T(G|\gamma)$. Proof of the condition for the equality is given in the Appendix.

Assume that the population consists of groups 1, 2, ..., n . Let $F_i(\mathbf{y})$, μ_i , and p_i represent the d -variate distribution function, the expectation vector and the share of group i in the overall population, respectively. Let $\tilde{F}_i(\mathbf{y})$ be the distribution functions of \mathbf{Y}/μ within group i . Then inequality 45 allows the following subgroup decomposition of the modified volume-Gini mean difference and the modified volume-Gini index:

$$M_V(F|\gamma) = \sum p_i M_V(F_i|\gamma) + \sum p_i cv_V(F_i, F|\gamma) \tag{46}$$

and

$$R_V(F|\gamma) = \sum p_i r_V(F_i|\gamma) R_V(F_i|\gamma) + \sum p_i cv_V(\tilde{F}_i, \tilde{F}|\gamma), \tag{47}$$

where $r_V(F_i|\gamma) = 0$ if $R_V(F|\gamma) = 0$, or $M_V(\tilde{F}_i|\gamma) / R_V(F_i|\gamma)$ otherwise. The second term on the right-hand side of Eq. 47, which corresponds to the between-group inequality, equals zero if and only if $F_i^s = F^s$ for any group and any coordinate axis, and $Z_T(F_i|\gamma) = Z_T(F|\gamma)$ for any group. The proof is given in the Appendix. Thus, $M_V(F_i|\gamma) = M_V(F|\gamma)$, and $R_V(F_i|\gamma) = R_V(F|\gamma)$ for any group if the between-group inequality is null. Unfortunately, equality of the γ -zonoids plus equality of the univariate marginal distributions is not equivalent to equality of the multivariate distributions. This is dissimilar to the equality of the lift zonoids [11]. A counterexample is given below.

Example Let F_1 be a bivariate distribution evenly distributed at the six points $\{1,1\}$, $\{4,4\}$, $\{0,2\}$, $\{3,2\}$, $\{2,0\}$ and $\{2,3\}$, and let F_2 be another bivariate distribution evenly distributed at the six points $\{0,0\}$, $\{3,3\}$, $\{1,2\}$, $\{4,2\}$, $\{2,1\}$ and $\{2,4\}$. Let λ and $1-\lambda$ be the population share of F_1 and F_2 , respectively – i.e. $F = \lambda F_1 + (1 - \lambda)F_2$. Then, F_1 and F_2 have the identical marginal distributions to F . Their means equal $\mathbf{2}=\{2,2\}$. Their γ -zonoids $Z_T(F_1|\mathbf{2})$ and $Z_T(F_2|\mathbf{2})$ are also identical to $Z_T(F|\mathbf{2})$ (Fig. 2).

A pair of bivariate distributions evenly distributed within triangles with vertices at $\{0,2\}$, $\{2,0\}$, $\{4,4\}$ and $\{0,0\}$, $\{4,2\}$, $\{2,4\}$, respectively, is also a counterexample, yet decomposition 47 can be considered to nearly satisfy the CID condition, since the condition of the null between-group inequality ensures some equivalence among within-group distributions in terms of dilation ordering defined by the γ -zonoid 32 in addition to the equivalence of the univariate marginal distributions. At least, the mean and dispersion of any group measured by the modified volume-Gini index must agree with each other if the between-group inequality vanishes.

Before closing this subsection, the representation of $cv_V(F_i, F|\gamma)$, which corresponds to that of $cv_T(F_i, F|\gamma)$ in Eq. 41, is given below:

$$cv_V(F_i, F|\gamma) = \lim_{\varepsilon \rightarrow +0} \frac{1}{\varepsilon} (M_V((1 - \varepsilon)F + \varepsilon F_i|\gamma) - (1 - \varepsilon)M_V(F|\gamma) - \varepsilon M_V(F_i|\gamma)). \quad (48)$$

Similar to Eq. 41, representation 48 expresses the notion that $cv_V(F_i, F|\gamma)$ equals the surplus of the dispersion relative to the merger ratio for a merger with an infinitely small ratio of group i .

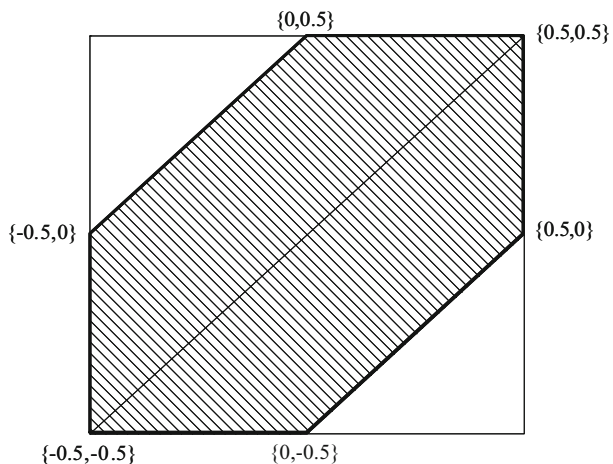
3.3 Source decomposition of the multivariate Gini indices

In this subsection, source decomposition of the distance-Gini index and the modified volume-Gini index is introduced.

3.3.1 Source decomposition of the distance-Gini index

Assume that attribute i consists of contributions from m_i types of sources ($i=1, \dots, d$). Let $x_i^{(k_i)}$ and $y_i^{(k_i)}$ be the expected contributions from source k_i to attribute i for the conditions $\mathbf{x} =$

Fig. 2 γ -zonoid of distributions F_1 , F_2 and F in Example 1



$\{x_i\}$ and $\mathbf{y}=\{y_i\}$, respectively, and $\mu_i^{(k_i)}$ be the unconditional mean of the contribution. Taking it into consideration that the distance-Gini mean difference is proportional to the average of the Gini mean differences for univariate marginal distributions on lines in all directions (see Eq. 21), it is intuitive to define the contribution of each source as the average of the quasi-Gini mean differences for univariate marginal distributions on lines in all directions with the same multiplier as Eq. 21, as follows:

$$\begin{aligned}
 M_D(F) &= \sum_{i=1}^d \sum_{k_i=1}^{m_i} \frac{C_d}{2d} \int \int \left(x_i^{(k_i)} - y_i^{(k_i)}\right) dF(\mathbf{x})dF(\mathbf{y}) \int_{S^{d-1}} \operatorname{sgn}(\mathbf{a} \cdot (\mathbf{x} - \mathbf{y})) a_i d\nu(\mathbf{a}) \\
 &= \sum_{i=1}^d \sum_{k_i=1}^{m_i} M_D^{(k_i)}(F),
 \end{aligned}
 \tag{49}$$

where ν is the uniform distribution on the unit sphere $S^{d-1} = \{\mathbf{a} \in R^d \mid \|\mathbf{a}\| = 1\}$, and $\operatorname{sgn}(x)=1$ if $x \geq 0$, or -1 otherwise. $M_D^{(k_i)}(F)$, the contribution of each source to $M_D(F)$, should be called the quasi-distance-Gini mean difference. The corresponding decomposition for the distance-Gini index is obtained by substituting \tilde{F} for F . Taking the definition of the univariate quasi-Gini index into consideration, the quasi-distance-Gini index $R_D^{(k_i)}(F)$ should be defined as the contribution of each source to $R_D(F)$ relative to the amount share, as follows: $M_D^{(k_i)}(\tilde{F}) = \frac{\mu_i^{(k_i)}}{\mu_i} R_D^{(k_i)}(F)$. Since the following equality is true (the proof is given in the Appendix),

$$C_d \int_{S^{d-1}} \operatorname{sgn}(\mathbf{a} \cdot \mathbf{x}) a_i d\nu(\mathbf{a}) = \frac{x_i}{\|\mathbf{x}\|}, \text{ where } \mathbf{x} \neq \mathbf{0},
 \tag{50}$$

the quasi-distance-Gini mean difference can be expressed as follows:

$$M_D^{(k_i)}(F) = \frac{1}{2d} \int \int \frac{\left(x_i^{(k_i)} - y_i^{(k_i)}\right)(x_i - y_i)}{\|\mathbf{x} - \mathbf{y}\|} dF(\mathbf{x})dF(\mathbf{y})
 \tag{51}$$

Note that the integrand in Eq. 51 is assumed to be zero if $\mathbf{x}=\mathbf{y}$.

The quasi-distance-Gini mean difference of each source is also derived in the following manner. Assume that the contributions from source k_i increase by an infinitely small rate ε uniformly – i.e. $x_i^{(k_i)}$ increases to $(1 + \varepsilon)x_i^{(k_i)}$ on any population unit; then the increase in the distance-Gini mean difference relative to rate ε equals $M_D^{(k_i)}(F)$. This derivation is also applicable to the univariate Gini source decomposition of Rao [16].

3.3.2 Source decomposition of the modified volume-Gini index

The modified volume-Gini mean difference and the modified volume-Gini index seemingly do not have such intuitive derivation of source decomposition as the distance-Gini mean difference and the distance-Gini index. However, if adhering to the reasoning that the contribution of each source to the mean difference is derived from differentiation by rate of a uniform increase in amount for each source as mentioned in section 3.3.1, then a type of source decomposition is obtained for the modified volume-Gini mean difference and the modified volume-Gini index if $M_T(F^{i_1 \dots i_s} \mid \gamma) > 0$ for any marginal distribution in space of

any sub-coordinate axes. Due to space limitations, only the result of the derivation for the modified volume-Gini index is presented here.

For introduction of the source decomposition, the following $s \times s$ matrices are first defined:

$$M(\mathbf{y}_1, \dots, \mathbf{y}_s | i_1 \dots i_s) = [\mathbf{y}_1 - \boldsymbol{\gamma}, \dots, \mathbf{y}_s - \boldsymbol{\gamma}], \text{ and } N^{(k_{i_1} \dots k_{i_s})}(\mathbf{y}_1, \dots, \mathbf{y}_s | i_1 \dots i_s) = \left[\frac{\mathbf{y}_1^{(k_{i_1} \dots k_{i_s})}}{\boldsymbol{\mu}^{(k_{i_1} \dots k_{i_s})}} - \boldsymbol{\gamma}, \dots, \frac{\mathbf{y}_s^{(k_{i_1} \dots k_{i_s})}}{\boldsymbol{\mu}^{(k_{i_1} \dots k_{i_s})}} - \boldsymbol{\gamma} \right], \tag{52}$$

where $\mathbf{y}_l^{(k_{i_1} \dots k_{i_s})} = \{y_{li}^{(k_{it})}\}_{l=1, \dots, s}$, and $\boldsymbol{\mu}^{(k_{i_1} \dots k_{i_s})} = \{\mu_{it}^{(k_{it})}\}_{l=1, \dots, s}$. Note that abbreviated notations are used in (52) for simplicity. $[\mathbf{y}_1 - \boldsymbol{\gamma}, \dots, \mathbf{y}_s - \boldsymbol{\gamma}]$ is an abbreviation of $[\mathbf{y}_1 - \boldsymbol{\gamma}^{i_1 \dots i_s}, \dots, \mathbf{y}_s - \boldsymbol{\gamma}^{i_1 \dots i_s}]$, for instance. In the t -th row of $N^{(k_{i_1} \dots k_{i_s})}(\mathbf{y}_1, \dots, \mathbf{y}_s | i_1 \dots i_s)$, each element is equal to the conditional expected contribution of source k_{it} to attribute i_t relative to the unconditional expected contribution $\mu_{it}^{(k_{it})}$ minus γ_{it} .

By performing some manipulation after the above-mentioned derivation, the source decomposition of the modified volume-Gini index can be expressed as follows:

$$R_V(F|\boldsymbol{\gamma}) = \frac{1}{2^d - 1} \left(\sum_{i=1}^d \sum_{k_{it}=1}^{m_i} R^{(k_{it})}(F^i) + \sum_{s=2}^d \sum_{1 \leq i_1 < \dots < i_s \leq d} \sum_{k_{i_1}=1}^{m_{i_1}} \dots \sum_{k_{i_s}=1}^{m_{i_s}} \prod_{l=1}^s \frac{\mu_{i_l}^{(k_{i_l})}}{\mu_{i_l}} R_{\Gamma}^{(k_{i_1} \dots k_{i_s})}(F^{i_1 \dots i_s} | \boldsymbol{\gamma}) / R_T(F^{i_1 \dots i_s} | \boldsymbol{\gamma})^{(s-1)/s} \right), \tag{53}$$

where $R^{(k_i)}(F^i) = \int \int \text{sgn}(x_i - y_i) \left(\frac{x_i^{(k_i)}}{\mu_i^{(k_i)}} - \frac{y_i^{(k_i)}}{\mu_i^{(k_i)}} \right) dF^i(x_i) dF^i(y_i)$ and $R_{\Gamma}^{(k_{i_1} \dots k_{i_s})}(F^{i_1 \dots i_s} | \boldsymbol{\gamma}) = \frac{1}{s! \left(1 + \sum_{t=1}^s \gamma_{it}\right)} \int \dots \int \text{sgn}(\det(M(\mathbf{y}_1, \dots, \mathbf{y}_s | i_1 \dots i_s))) \det(N^{(k_{i_1} \dots k_{i_s})}(\mathbf{y}_1, \dots, \mathbf{y}_s | i_1 \dots i_s)) dF(\mathbf{y}_1) \dots dF(\mathbf{y}_s)$.

$R^{(k_i)}(F^i)$ in the first term is the quasi-Gini index of the univariate marginal distribution in subspace of attribute i . The second term can be regarded as an interaction term among sources of different attributes, which makes a notable difference with the decomposition of the distance-Gini index.

3.4 Application to Japanese family budget data

Several types of multivariate Gini indices are estimated for annual income and consumption of Japanese households with two or more members using tabulated data from the National Survey of Family Income and Expenditures. Three-way tables of consumption class by income class by age class of household head are available for the household distribution. However, owing to the unavailability of the three-way table for average income and consumption, average income in two-way tables of income class by age class of household head are used for the estimation instead, irrespective of consumption class, and the intermediate value between the lower and upper limits of each consumption class is used as average consumption, irrespective of income class and age class of the household head. The estimates after adjustment by excluding the age effects are presented in Table 5. These estimates should be treated carefully because of the above-mentioned approximation; nevertheless, it is notable that the multivariate Gini indices

Table 5 Multivariate Gini indices for distribution of income and consumption*

Year	Univariate Gini		R_D	R_O	$R_O^{1/2}$	R_T $\gamma=1$	$R_T^{1/2}$ $\gamma=1$	R_{KM}	R_V $\gamma=1$
	Income	Consumption							
1989	0.2892	0.2564	0.2132	0.05591	0.2364	0.03205	0.1790	0.2005	0.2416
1994	0.2879	0.2457	0.2084	0.05377	0.2319	0.03128	0.1768	0.1958	0.2368
1999	0.2837	0.2486	0.2078	0.05389	0.2321	0.03135	0.1771	0.1954	0.2365
2004	0.2844	0.2516	0.2097	0.05532	0.2352	0.03196	0.1788	0.1971	0.2383
	(1989=100)								
1989	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1994	99.5	95.8	97.8	96.2	98.1	97.6	98.8	97.6	97.9
1999	98.1	97.0	97.5	96.4	98.2	97.8	98.9	97.4	97.7
2004	98.3	98.1	98.4	99.0	99.5	99.7	99.9	98.3	98.6

*Calculated after replacing the share of each age group with that of 1989.

for 2004 relative to 1989 are higher than the Gini indices of both univariate marginal distributions for income and consumption. For example, the distance-Gini index and the modified volume-Gini index for 2004 relative to 1989 are 98.4 and 98.6, respectively, while the Gini indices for annual income and consumption are 98.3 and 98.1, respectively. This indicates that consumption tends to vary more widely than before within the same income class, although the whole consumption distribution does not disperse as before.

It is also notable that the modified volume-Gini index is relatively close to the distance-Gini index in comparison with the Oja index, the modified Torgersen index or their 1/2-th power transformations.

The contributions of age groups for household heads to changes in these indices are also estimated using the subgroup decomposition technique (Table 6). The multivariate Gini indices show similar tendencies to the univariate Gini indices, although the magnitudes of the contributions vary somewhat.

4 Concluding remarks

In this paper, a new type of subgroup decomposition for the Gini index is proposed. The new decomposition is consistent with multilevel sub-groupings³, and is characterized by the CID condition – i.e. the between-group inequality vanishes if and only if distributions within groups are identical to all the others.

The new decomposition is then generalized to two types of multivariate Gini indices introduced by Koshevoy and Mosler [11]. In the case of the distance-Gini index, the decomposition strictly satisfies the CID condition, while for the volume-Gini index, the decomposition satisfies the condition not strictly, but nearly after the index definition is modified to be of homothetic degree one to enlargement with the center at the mean.

Source decompositions of the two types of multivariate Gini indices are also introduced as a generalization of the Gini decomposition of Rao [16].

I hope this new decomposition will advance studies of economic inequality. The following remarks concerning the definition or concept of multivariate Gini indices may be helpful for further research.

³ See footnote 1.

Table 6 Contributions of age groups for household heads to changes in multivariate Gini indices for distribution of income and consumption*

Age	Univariate Gini for income			Univariate Gini for consumption		
	1994	1999	2004	1994	1999	2004
Total	-0.0014	-0.0042	0.0007	-0.0107	0.0030	0.0030
Within-Group	-0.0065	-0.0036	0.0038	-0.0129	0.0037	0.0056
< 30	0.0003	-0.0001	0.0009	-0.0004	-0.0001	0.0007
30-39	-0.0019	0.0004	0.0004	-0.0010	0.0009	0.0002
40-49	0.0023	0.0003	0.0002	-0.0022	0.0019	0.0013
50-59	-0.0037	0.0005	0.0014	-0.0058	0.0010	0.0027
60-69	-0.0003	-0.0037	0.0002	-0.0023	0.0002	0.0004
70 ≤	-0.0031	-0.0010	0.0007	-0.0013	-0.0001	0.0003
Between-Group	0.0051	-0.0006	-0.0031	0.0022	-0.0007	-0.0027
< 30	-0.0001	0.0005	-0.0011	0.0003	0.0000	-0.0009
30-39	0.0006	-0.0003	-0.0003	0.0006	0.0002	0.0000
40-49	0.0003	0.0003	0.0002	0.0002	-0.0002	-0.0007
50-59	0.0022	-0.0006	-0.0014	0.0009	0.0001	-0.0003
60-69	0.0007	0.0002	0.0000	-0.0001	-0.0002	-0.0004
70 ≤	0.0014	-0.0005	-0.0005	0.0002	-0.0006	-0.0004

Age	R_D			$R_T^{1/d}, \gamma=1$			$R_V, \gamma=1$		
	1994	1999	2004	1994	1999	2004	1994	1999	2004
Total	-0.0048	-0.0006	0.0019	-0.0022	0.0002	0.0017	-0.0048	-0.0003	0.0018
Within-Group	-0.0074	-0.0003	0.0039	-0.0023	0.0003	0.0019	-0.0072	0.0001	0.0038
< 30	-0.0001	-0.0001	0.0006	-0.0001	-0.0001	0.0001	-0.0001	-0.0001	0.0006
30-39	-0.0012	0.0004	0.0002	-0.0003	0.0002	0.0000	-0.0011	0.0005	0.0002
40-49	0.0000	0.0007	0.0006	0.0007	0.0006	0.0003	0.0003	0.0009	0.0006
50-59	-0.0036	0.0007	0.0016	-0.0014	0.0007	0.0009	-0.0037	0.0008	0.0016
60-69	-0.0008	-0.0014	0.0004	-0.0003	-0.0007	0.0003	-0.0009	-0.0014	0.0003
70 ≤	-0.0017	-0.0004	0.0005	-0.0009	-0.0004	0.0003	-0.0017	-0.0005	0.0004
Between-Group	0.0026	-0.0004	-0.0020	0.0001	-0.0001	-0.0002	0.0025	-0.0005	-0.0020
< 30	0.0001	0.0002	-0.0007	0.0000	0.0001	-0.0001	0.0001	0.0002	-0.0007
30-39	0.0004	0.0000	-0.0001	0.0000	0.0000	0.0000	0.0004	0.0000	-0.0001
40-49	0.0002	0.0000	-0.0002	-0.0001	-0.0001	0.0000	0.0001	0.0000	-0.0002
50-59	0.0012	-0.0002	-0.0006	0.0000	0.0000	0.0000	0.0011	-0.0002	-0.0006
60-69	0.0002	0.0000	-0.0001	0.0001	0.0000	0.0000	0.0002	0.0000	-0.0001
70 ≤	0.0006	-0.0004	-0.0003	0.0001	-0.0001	0.0000	0.0005	-0.0004	-0.0003

*Calculated after replacing the share of each age group with that of 1989.

Anderson [1] used the following multivariate Gini index, which is similar to the distance-Gini index:

$$GINIMCW = \frac{1}{2\sqrt{d}} \int \int \left\| \frac{\mathbf{x}}{\boldsymbol{\mu}} - \frac{\mathbf{y}}{\boldsymbol{\mu}} \right\|_W dF(\mathbf{x})dF(\mathbf{y}), \tag{54}$$

where $\left\| \frac{\mathbf{x}}{\boldsymbol{\mu}} - \frac{\mathbf{y}}{\boldsymbol{\mu}} \right\|_W = \sqrt{\sum_{i=1}^d w_i \left(\frac{x_i}{\mu_i} - \frac{y_i}{\mu_i} \right)^2}$, $w_i > 0$ and $\sum_{i=1}^d w_i = d$.

The differences between Eq. 54 and the distance-Gini index highlight two issues. First, the Anderson index (Eq. 54) almost equals unity if the amounts for all attributes are

monopolized by only one population unit and all other units have no contributions. The index exceeds unity if all the amounts for each attribute belong to only one population unit, but the monopolist for each attribute differs. The index almost equals $\frac{1}{\sqrt{d}} \sum_{i=1}^d \sqrt{w_i} (> 1)$ in this case. The index can be limited to less than unity by modifying d for \sqrt{d} in Eq. 54 and w_i^2 for w_i in definition of the between-unit distance. Note that the weights constraint $\sum_{i=1}^d w_i = d$ needs not be replaced. Such an index can be regarded as the weighted distance-Gini index. The subgroup and source decomposition for the distance-Gini index presented in this paper can easily be extended to it. However, the following naïve question still remains open:

Which situations should be judged to be higher in inequality, only one population unit makes all contributions to all attributes (the absolute monopolistic situation) or different monopolists exist for each attribute?

Since the Lorenz zonoid in the latter situation is wider than that in the former, the Oja index, the (modified) Torgersen index and the (modified) volume-Gini index are higher in the latter case. So is the distance-Gini index because of its consistency with the dilation ordering of the Lorenz zonoid [11]. This characteristic is seemingly not necessarily a disadvantage, at least in terms of the weighting problem described in the next paragraph. Nevertheless, several researchers have pursued multivariate inequality measures that are higher in the absolute monopolistic situation. Tsui [22] studied the multidimensional generalized entropy measures satisfying a condition of consistency with the correlation increasing majorization (CIM) as well as some other conditions. If a multidimensional inequality measure satisfies the CIM condition, the absolute monopolistic situation is judged to be higher in inequality. However, imposition of this condition is seemingly too restrictive (see also [7]). For example, Tsui's multidimensional extension of the Theil measure is outside the restriction. The above question may rarely arise if wealthy population units contributing to one attribute tend to contribute to other attributes in practice. It thus seems likely that one of the approaches to determine the appropriateness of the CIM condition is to verify whether a set of attributes or the subject to be studied can be accounted by uni- or multi-dimensional factors after excluding measurement errors. If multiple factors are identified, then the way to extract each factor with mutual relations should be explored for measurement of inequality. In usual cases, attributes seem to be determined by one major common factor and some additional factors peculiar to individual attributes. In this context, it is notable that Easterlin [5, 6] pointed out that correlation between income and self-reported well-being is weak at least over the life cycle, implying that (subjective) well-being is affected by multiple factors.

Another issue raised by the Anderson index is the weight assignment to attributes concerned. If economic inequality is measured in terms of income, consumption and education level achieved, it seems to be reasonable to assign smaller weights to income and consumption relative to educational attainment because of their similarity. However, as the multivariate Gini indices automatically have lower outcomes if attributes are correlated with each other – in other words, the indices run counter to the CIM condition – the weighting problem is not considered to be serious relative to multivariate inequality measures that satisfy the CIM condition. The way to determine appropriate weights is not a trivial matter.

Thus, further research is still needed for applications of the multivariate Gini indices or other multivariate inequality measures.

Acknowledgement I wish to thank the editors and anonymous reviewers at this journal whose comments and suggestions have been very helpful to improve the paper.

Appendix

Proofs

Proof of Eq. 16

On the assumption that F_i is continuous, by applying integration by parts, $cv(F_i, F)$ can be expanded as follows:

$$\begin{aligned}
 cv(F_i, F) &= \int (F_i - F)^2 dy = -2 \int y(F_i - F) dF_i + 2 \int y(F_i - F) dF \\
 &= 2 \left(\int (y - \mu_i) F dF_i - \int (y - \mu_i) F_i dF_i \right) + 2\mu_i \left(\int F dF_i - \frac{1}{2} \right) \\
 &+ 2 \left(\int (y - \mu) F_i dF - \int (y - \mu) F dF \right) + 2\mu \left(\int F_i dF - \frac{1}{2} \right).
 \end{aligned}
 \tag{55}$$

Note that $\int F_i dF_i = \int F dF = 1/2$. Then, using the equality $\int F dF_i - 1/2 = -\int F_i dF + 1/2$ and the notation in section 2.3, Eq. 16 is obtained.

Proof of Eqs. 24 and 25

By applying the definition of $cv_D(F_i, F)$ in Eq. 22, the distance-Gini mean difference $M_D(F)$ can be expanded as follows:

$$\begin{aligned}
 M_D(F) &= \frac{1}{2} \int \int \|x - y\| dF(x) dF(y) = \frac{1}{2} \sum p_i \int \int \|x - y\| dF_i(x) dF(y) \\
 &= \frac{1}{2} \sum p_i M_D(F_i) + \frac{1}{2} M_D(F) + \frac{1}{2} \sum p_i cv_D(F_i, F).
 \end{aligned}
 \tag{56}$$

Note that $F = \sum p_i F_i$. Subtracting $M_D(F)$ from both sides after doubling, decomposition 24 is obtained. Equation 25 is easily derived by replacing F with \tilde{F} .

Range of the modified Torgersen index

Since

$$z(\phi, \tilde{F} | \gamma) = \int \left(\frac{\mathbf{y}}{\mu} - \gamma \right) \phi(\mathbf{y}) dF(\mathbf{y}) = \int \frac{\mathbf{y}}{\mu} \phi(\mathbf{y}) dF(\mathbf{y}) - \gamma \int \phi(\mathbf{y}) dF(\mathbf{y}) \in \sum_i [0, \mathbf{e}_i] + [0, -\gamma],
 \tag{57}$$

where $\sum_i [0, \mathbf{e}_i] + [0, -\gamma]$ is the Minkowski sum of the line segments $[0, \mathbf{e}_i]$ ($i=1, \dots, d$) and $[0, -\gamma]$, and $\mathbf{e}_i = \{e_{ij}\}$ is a unit vector consisting of elements $e_{ij}=0$ if $j \neq i$, or $=1$ if $j=i$, then, the volume of $Z_T(\tilde{F} | \gamma)$ is less than or equal to the volume of $\sum [0, \mathbf{e}_i] + [0, -\gamma]$, which is calculated at $1 + \sum \gamma_i$ by using the following formula $vol \left(\sum_{i=1}^n [0, \mathbf{a}_i] \right) = \sum_{1 \leq i_1 < \dots < i_d < n} |\det(\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_d})|$

(e.g. [20]). Thus, the modified Torgersen index $R_T(F | \gamma)$ is less than or equal to unity.

Assume that F is an n -point d -variate distribution in which there is a different monopolist for each variable (attribute). Then, $R_T(F|\gamma) \rightarrow 1$ if $n \rightarrow \infty$.

Proof of Eq. 41

By applying Eq. 38, Eq. 58 is derived.

$$cv_T(F_i, G(\varepsilon)|\gamma) = \frac{M_T(G(\varepsilon)|\gamma)^{1/d} - \varepsilon M_T(F_i|\gamma)^{1/d} - (1 - \varepsilon)M_T(F|\gamma)^{1/d}}{\varepsilon} - (1 - \varepsilon) \frac{cv_T(F, G(\varepsilon)|\gamma)}{\varepsilon}, \tag{58}$$

where $G(\varepsilon) = \varepsilon F_i + (1 - \varepsilon)F$. The modified Torgersen mean difference $M_T(G(\varepsilon)|\gamma)$ and the mixed volume $MV_{d-1}(F, G(\varepsilon)|\gamma)$ have the following derivatives at $\varepsilon=0$:

$$\frac{d}{d\varepsilon} \frac{MV_{d-1}(F, G(\varepsilon)|\gamma)}{1 + \sum \gamma_i} = \frac{MV_{d-1}(F_i, F|\gamma)}{1 + \sum \gamma_i} - M_T(F|\gamma) = M_T(F|\gamma)^{(d-1)/d} \frac{d}{d\varepsilon} M_T(G(\varepsilon)|\gamma)^{1/d}. \tag{59}$$

The derivative of $cv_T(F, G(\varepsilon)|\gamma)$ at $\varepsilon=0$ equals zero from its definition and Eq. 59. Thus, the second term on the right-hand side of Eq. 58 $\rightarrow 0$ if $\varepsilon \rightarrow 0$. Since the left-hand side $\rightarrow cv_T(F_i, F|\gamma)$ if $\varepsilon \rightarrow 0$, Eq. 41 is derived.

Proof of the equality condition in Eq. 45 and the condition for null between-inequality in Eq. 47

It is trivial that $cv_V(G, F|\gamma) = 0$ if $G^s = F^s$ for any coordinate axis, and $Z_T(G|\gamma) = Z_T(F|\gamma)$. If $cv_V(G, F|\gamma) = 0$, all terms constituting $cv_V(G, F|\gamma)$ in Eq. 45 should be zero, including $cv_T(G, F|\gamma)$. On the assumption that $M_T(F|\gamma) > 0$, this means that $Z_T(G|\gamma) = \alpha Z_T(F|\gamma)$, where α is a positive constant, according to Minkowski’s first inequality concerning the mixed volume. Since $cv(G^s, F^s)$ should also be zero at the same time, $G^s = F^s$ for any coordinate axis. This forces α to be unity. Thus, the equality condition in Eq. 45 is proved.

If $M_T(F|\gamma) > 0$, the equality condition in Eq. 45 can be applied to the proof of the condition for the null between-inequality in Eq. 47. If $M_T(F|\gamma) = 0$, the Brunn-Minkowski inequality (e.g. [17]) asserts that $Z_T(F_i|\gamma)$ is on the same hyperplane as $Z_T(F|\gamma)$. Let k_1, \dots, k_i be the highest-dimensional sub-coordinate axes under the condition that $M_T(F^{j_1 \dots j_s}|\gamma) > 0$. The equality condition in Eq. 45 asserts that $Z_T(F_i^{k_1 \dots k_i}|\gamma) = Z_T(F^{k_1 \dots k_i}|\gamma)$ (Note that $cv_V(F_i^{k_1 \dots k_i}, F^{k_1 \dots k_i}|\gamma) = 0$ if $cv_V(F_i, F|\gamma) = 0$). From this, $Z_T(F_i|\gamma)$ should be identical to $Z_T(F|\gamma)$. $G^s = F^s$ is true for any coordinate axis, irrespective of the value of $M_T(F|\gamma)$. Thus, the condition for null between-inequality in Eq. 47 is proved.

Proof of Eqs. 20 and 50

First, the proof of Eq. 50 is given inductively. Without loss of generality, we can assume that $i=1$ and $\mathbf{x} = \{x_1, x_2, 0, \dots, 0\}$ because of invariance for rotation on axis i and axis permutation. If $d=2$, Eq. 50 can be proved as follows:

$$\begin{aligned} I_2(\mathbf{x}) &= \int_{S^1} \text{sgn}(\mathbf{a} \cdot \mathbf{x}) a_1 d\nu(\mathbf{a}) = \int_{-\pi}^{\pi} \text{sgn}(\mathbf{a} \cdot \mathbf{x}) \cos \theta d\theta \\ &= \int_{\varphi-\pi/2}^{\varphi+\pi/2} \cos \theta d\theta - \int_{\varphi+\pi/2}^{\varphi+3\pi/2} \cos \theta d\theta = 4 \cos \varphi = \frac{1}{C_2} \frac{x_1}{\|\mathbf{x}\|}. \end{aligned} \tag{60}$$

where $\mathbf{a} = \{\cos\theta, \sin\theta\}$, and $\mathbf{x} = \{x_1, x_2\} = \{\|\mathbf{x}\| \cos \varphi, \|\mathbf{x}\| \sin \varphi\}$. Assume that Eq. 50 is proved if the dimension is $d-1$. Then,

$$I_d(\mathbf{x}) = \int_{S^{d-1}} \text{sgn}(\mathbf{a} \cdot \mathbf{x}) a_1 d\nu(\mathbf{a}) = I_{d-1}(\mathbf{x}) \int_{-1}^1 \sqrt{1 - a_d^2}^{d-2} da_d = \frac{1}{C_d} \frac{x_1}{\|\mathbf{x}\|}. \quad (61)$$

Equation 50 is derived by summing Eq. 50 over i after multiplying by x_i .

References

- Anderson, G.: Indices and tests for multidimensional inequality: multivariate generalizations of the Gini coefficient and Kolmogorov–Smirnov two sample test. Paper prepare for the 28th general conference of the IARIW, Cork, Ireland, 22–28 Aug. 2004
- Baringhaus, L., Franz, C.: On a new multivariate two-sample test. *J. Multivar. Anal.* **88**, 190–206 (2004)
- Bhattacharya, N., Mahalanobis, B.: Regional disparities in household consumption in India. *J. Am. Stat. Assoc.* **62**, 143–161 (1967)
- Dagum, C.: A new approach to the decomposition of the Gini income inequality ratio. *Emp. Econ.* **22**, 515–531 (1997)
- Easterlin, R.: Does economic growth improve the human lot? Some empirical evidence. In: David P., Reder M. (eds.) *Nations and households in economic growth: essays in honour of Moses Abramobitz*, pp. 89–125. Academic, New York (1974)
- Easterlin, R.: Building a better theory of well-being. In: Bruni, L., Porta, P.L. (eds.) *Economics and happiness: framing the analysis*, pp. 29–64. Oxford University Press, New York (2006)
- Gajdos, T., Weymark, J.A.: Multidimensional generalized Gini indices. *Vanderbilt University Working Paper 3-W11R*, Department of Economics, Vanderbilt University, Nashville, Tennessee (2003)
- Gardner, R.J.: The Brunn–Minkowski inequality. *Bull. Am. Math. Soc.* **39**, 355–405 (2002)
- Helgason, S.: *The Radon Transform*. Birkhauser, Boston (1980)
- Koshevoy, G. A., Mosler, K.: The Lorenz zonoid of a multivariate distribution. *J. Am. Stat. Assoc.* **91**, 873–882 (1996)
- Koshevoy, G.A., Mosler, K.: Multivariate Gini indices. *J. Multivar. Anal.* **60**, 252–276 (1997)
- Koshevoy, G.A., Mosler, K.: Multivariate Lorenz dominance based on zonoids. *Advance in Statistical Analysis* **91**, 57–76 (2007)
- Lerman, R.I., Yitzhaki, S.: A note on the calculation and interpretation of the Gini index. *Econ. Lett.* **15**, 363–368 (1984)
- Oja, H.: Descriptive statistics for multivariate distributions. *Stat. Probab. Lett.* **1**, 327–332 (1983)
- Pyatt, G.: On the interpretation and disaggregation of Gini coefficients. *Econ. J.* **86**, 243–255 (1976)
- Rao, V.M.: Two decompositions of concentration ratio. *J. R. Stat. Soc. A* **132**, 418–425 (1969)
- Schneider, R.: *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge University Press, New York (1993)
- Shorrocks, A.: Aggregation issues in inequality measurement. In: W. Eichhorn (ed.) *Measurement in Economics*, pp. 429–451. Physica-Verlag, Heidelberg (1988)
- Shorrocks, A., Wan, G.: Spatial decomposition of inequality. *J. Economic Geography* **5**, 59–81 (2005)
- Shephard, G.C.: Combinatorial properties of associated zonotopes. *Can. J. Math.* **26**, 302–321 (1974)
- Torgersen, E.: *Comparison of Statistical Experiments*. Cambridge University Press, Cambridge (1991)
- Tsui, K.-Y.: Multidimensional inequality and multidimensional generalized entropy measures: an axiomatic derivation. *Soc Choice Welf* **16**, 145–157 (1999)
- Wilks, S.S.: Multidimensional statistical scatter. In: Olkin, et al. (eds.) *Contributions to Probability and Statistics*, pp. 486–503. Stanford University Press, Stanford, (1960)
- Yitzhaki, S., Lerman, R.I.: Income stratification and income inequality. *Rev. Income Wealth* **37**, 313–329 (1991)