Kagawa University Economic Review
Vol. 67, No. 2, October 1994, 343-354

Decomposition of the Gini Coefficient*

Yoshimitsu Yokoyama

1 Introduction

The Gini coefficient has been extensively used to measure income inequality. Its many properties are made plain by use of graph and algebra. Especially the relation between the Lorenz curve and the Gini coefficient is popular and intuitively clear. In spite of such desirable characters, the Gini coefficient isn't always used to express subgroup decomposition of income inequality. There are some difficulties in decomposition. To avoid trouble, the alternative way is to use decomposable inequality measures. The entropy measure is useful in this case. But since the Gini coefficient is the most popular measure, it is important to consider the decomposition. This subject deserves more than a passing notice.

The Gini coefficient can be decomposed into three subparts : between-groups term, within-groups term and residual term. The implication of residual term is difficult to understand. Numerous works have been made to explain the meaning of the term. Mookherjee and Shorrocks [6] argued the question of decomposition of inequality measures. Their conclusion of residual term in the Gini coefficient is pessimistic. Silber [8] considered its term depended on income reranking. Lambert and Aronson [5] showed meaning of the residual term from a graphical point of view.

*The earlier version of this paper was written in my graduate school days at Kyushu University. I am deeply indebted to Prof. Ravindra R. Ranade, although of course I am solely responsible for any errors.

In this paper, we consider decomposition of the Gini coefficient and its implication. To begin with we consider decomposition of the Gini coefficient into two subgroups, in which we use matrix expression. Such an operation makes it possible to interpret the term. Then, we answer the question raised by Mookherjee and Shorrocks [6]. This analysis makes the residual term clear. Finally, we give decomposition in the general case. The form will give us one implication of the residual term.

2 Decomposition in the Two Cases

First, we consider that there are n members in a group, member i receives x_i where

$$x_1 \leq x_2 \leq \cdots \leq x_{n-1} \leq x_n$$

Total group consists of two subgroups. One is subgroup S_1 which has n_1 members and small mean income μ_1, and the other is subgroup S_2 which has n_2 members and large mean income μ_2, i.e. $\mu_1 \leq \mu_2$. In addition to this, an overlap exists. In other words, the richest member of subgroup 1 is wealthier than the poorest member of subgroup 2. Accordingly we define the overlapping set S,

$$S = \{x_l \mid x_l \in [\min_{x_i \in S_2}(x_i), \max_{x_j \in S_1}(x_j)]\}$$

and we take s_1 as the number of members in S_1 and S, similarly s_2 as in S_2 and S. Therefore we can divide the total members into four parts, $S_1 \cap S^c$, $S_1 \cap S$, $S_2 \cap S$, and $S_2 \cap S^c$. The numbers of them are $n_1 - s_1$, s_1, s_2, and $n_2 - s_2$, respectively. Clearly, the relation $n = n_1 + n_2$ holds.

Secondly, we define new vectors $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ to express members of subgroup in total group. The $n \times 1$ vector $\boldsymbol{t}_1$ has one in ith position, if ith income x_i belongs to the subgroup 1, and zeros elsewhere. It is the same with $\boldsymbol{t}_2$.

Next, we use a matrix to calculate the Gini coefficient decomposition.

The matrix $\boldsymbol{A}$ is defined as

$$\boldsymbol{A} = \begin{bmatrix} \min(x_1, x_1) & \min(x_1, x_2) & \cdots & \min(x_1, x_n) \\ \min(x_2, x_1) & \min(x_2, x_2) & \cdots & \min(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \min(x_n, x_1) & \min(x_n, x_2) & \cdots & \min(x_n, x_n) \end{bmatrix}$$

and vector $\boldsymbol{i}_n$ as

$$\boldsymbol{i}_n = [1, 1, 1, \cdots, 1, 1, 1]' : n \times 1$$

Now we can express the Gini coefficient by matrix and vectors.

$$n^2 \mu (1 - G) = \boldsymbol{i}_n' \boldsymbol{A} \boldsymbol{i}_n \tag{1}$$

because of

$$n^2 \mu (1 - G) = \sum_{i=1}^{n} \sum_{j=1}^{n} \min(x_i, x_j)$$

Next, we derive the between-group Gini coefficient G_b, in which each of the members in subgroup 1 has subgroup 1 mean μ_1 and equally each of the members in subgroup 2 gets subgroup 2 mean μ_2. That is to say n_1 members gets μ_1, and n_2 members take μ_2.

$$n^2 \mu (1 - G_b) = (n_1^2 + 2 n_1 n_2) \mu_1 + n_2^2 \mu_2$$

Then we can derive G_b easily,

$$n^2 \mu G_b = n_1 n_2 (\mu_2 - \mu_1) \tag{2}$$

Next we consider partition of matrix and vectors.

$$\boldsymbol{t}_1 = \begin{bmatrix} \boldsymbol{i}_{n_1 - s_1} \\ \boldsymbol{g}_1 \\ \boldsymbol{0}_{n_2 - s_2} \end{bmatrix} \begin{matrix} n_1 - s_1 \text{ rows} \\ s_1 + s_2 \text{ rows} \\ n_2 - s_2 \text{ rows} \end{matrix} \qquad \boldsymbol{t}_2 = \begin{bmatrix} \boldsymbol{0}_{n1 - s1} \\ \boldsymbol{g}_2 \\ \boldsymbol{i}_{n_2 - s_2} \end{bmatrix} \begin{matrix} n_1 - s_1 \text{ rows} \\ s_1 + s_2 \text{ rows} \\ n_2 - s_2 \text{ rows} \end{matrix}$$

by definition the relations exists in the following

$$\boldsymbol{i}_n = \boldsymbol{t}_1 + \boldsymbol{t}_2$$

and

$$\boldsymbol{i}_{s1+s2} = \boldsymbol{g}_1 + \boldsymbol{g}_2$$

Likewise vectors, we consider 3×3 partition of matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \\ \boldsymbol{A}_{31} & \boldsymbol{A}_{32} & \boldsymbol{A}_{33} \end{bmatrix} \begin{matrix} \mathrm{n}_1 - \mathrm{s}_1 \text{ rows} \\ \mathrm{s}_1 + \mathrm{s}_2 \text{ rows} \\ \mathrm{n}_2 - \mathrm{s}_2 \text{ rows} \end{matrix}$$

$$\begin{matrix} \mathrm{n}_1 - \mathrm{s}_1 & \mathrm{s}_1 + \mathrm{s}_2 & \mathrm{n}_2 - \mathrm{s}_2 \\ \text{columns} & \text{columns} & \text{columns} \end{matrix}$$

By using this partition, the Gini coefficients of subgroup 1 and subgroup 2 are written as

$$n_1^2\mu_1(1 - G_1) = \boldsymbol{t}_1'\boldsymbol{A}\boldsymbol{t}_1 \tag{3}$$

$$n_2^2\mu_2(1 - G_2) = \boldsymbol{t}_2'\boldsymbol{A}\boldsymbol{t}_2 \tag{4}$$

Alternatively,

$$\boldsymbol{i}_{n1-s1}'\boldsymbol{A}_{12}\boldsymbol{g}_1 = s_2 \sum_{i=1}^{n_1-s_1} x_i = s_2(n_1\mu_1 - s_1\overline{\mu}_1) \tag{5}$$

$$\boldsymbol{i}_{n2-s2}'\boldsymbol{A}_{13}\boldsymbol{i}_{n1-s1} = (n_2 - s_2) \sum_{i=1}^{n_1-s_1} x_i = (n_2 - s_2)(n_1\mu_1 - s_1\overline{\mu}_1) \tag{6}$$

$$\boldsymbol{g}_1'\boldsymbol{A}_{23}\boldsymbol{i}_{n2-s2} = (n_2 - s_2) \sum_{x_1 \in S_1 \cap S} x_i = (n_2 - s_2)s_1\overline{\mu}_1 \tag{7}$$

where $\overline{\mu}_i$ is the mean in $S_i \cap S$, namely $\mu_i = \frac{1}{s_i}\sum_{x_i \in S_i \cap S} x_i$.

We consider the Gini coefficient in S as G_{in}. Next between-Gini coefficient G_c in S, namely each member in $S \cap S_1$, gets equally income $\overline{\mu}_1$, and each member in $S \cap S_2$ gets equal income $\overline{\mu}_2$. And the Gini coefficients $\overline{G}_1$ and $\overline{G}_2$ measure the income inequalities in $S \cap S_1$ and $S \cap S_2$, respectively.

We can express these as

$$s_1^2\overline{\mu}_1(1 - \overline{G}_1) = \boldsymbol{g}_1'\boldsymbol{A}_{22}\boldsymbol{g}_1 \tag{8}$$

$$s_2^2\overline{\mu}_2(1 - \overline{G}_2) = \boldsymbol{g}_2'\boldsymbol{A}_{22}\boldsymbol{g}_2 \tag{9}$$

Now consider the cross effect. This is the counterpart to the between-Gini coefficient in subgroups. Suppose s_1 members have each income $\overline{\mu}_1$, and s_2 members have each income $\overline{\mu}_2$. If $\overline{\mu}_1 \geq \overline{\mu}_2$,

$$(s_1 + s_2)(s_1\overline{\mu}_1 + s_2\overline{\mu}_2)(1 - G_c) = (s_1^2 + 2s_1s_2)\overline{\mu}_1 + s_2^2\overline{\mu}_2$$

else if $\overline{\mu}_1 < \overline{\mu}_2$,

$$(s_1 + s_2)(s_1\overline{\mu}_1 + s_2\overline{\mu}_2)(1 - G_c) = (s_2^2 + 2s_1s_2)\overline{\mu}_2 + s_1^2\overline{\mu}_1$$

by a simple calculation

$$(s_1 + s_2)(s_1\overline{\mu}_1 + s_2\overline{\mu}_2)G_c = \pm s_1s_2(\overline{\mu}_1 - \overline{\mu}_2) \tag{10}$$

where + part of the ± sigh is used if $\overline{\mu}_1 \geq \overline{\mu}_2$ and − part is used if $\overline{\mu}_1 < \overline{\mu}_2$

Next, the interaction effect is defined as

$$\begin{aligned}(s_1 + s_2)(s_1\overline{\mu}_1 + s_2\overline{\mu}_2)(1 - G_{in}) &= \boldsymbol{i}'_{s1+s2}\boldsymbol{A}_{22}\boldsymbol{i}_{s1+s2} \\ &= (\boldsymbol{g}_1 + \boldsymbol{g}_2)'\boldsymbol{A}_{22}(\boldsymbol{g}_1 + \boldsymbol{g}_2) \\ &= \boldsymbol{g}_1\boldsymbol{A}_{22}\boldsymbol{g}_1 + 2\boldsymbol{g}'_1\boldsymbol{A}_{22}\boldsymbol{g}_2 + \boldsymbol{g}_2\boldsymbol{A}_{22}\boldsymbol{g}_2\end{aligned}$$

and we solve for $2\boldsymbol{g}'_1\boldsymbol{A}_{22}\boldsymbol{g}_2$ by using (8) and (9)

$$\begin{aligned}2\boldsymbol{g}'_1\boldsymbol{A}_{22}\boldsymbol{g}_2 &= (s_1 + s_2)(s_1\overline{\mu}_1 + s_2\overline{\mu}_2)(1 - G_{in}) \\ &\quad - s_1^2\overline{\mu}_1(1 - \overline{G}_1) - s_2^2\overline{\mu}_2(1 - \overline{G}_2) \\ &= -(s_1 + s_2)(s_1\overline{\mu}_1 + s_2\overline{\mu}_2)G_{in} \\ &\quad + s_1s_2(\overline{\mu}_1 + \overline{\mu}_2) + s_1^2\overline{\mu}_1\overline{G}_1 + s_2^2\overline{\mu}_2\overline{G}_2\end{aligned} \tag{11}$$

using (5), (6), (7) and (11), so

$$\begin{aligned}2\boldsymbol{t}'_1\boldsymbol{A}\boldsymbol{t}_2 &= 2(\boldsymbol{i}'\boldsymbol{A}_{12}\boldsymbol{g}_1 + \boldsymbol{i}'\boldsymbol{A}_{13}\boldsymbol{i} + \boldsymbol{g}'_1\boldsymbol{A}_{22}\boldsymbol{g}_2 + \boldsymbol{g}'_1\boldsymbol{A}_{23}\boldsymbol{i}) \\ &= 2s_2(n_1\mu_1 - s_1\overline{\mu}_1) + 2(n_2 - s_2)(n_1\mu_1 - s_1\overline{\mu}_1) \\ &\quad + 2\boldsymbol{g}'_1\boldsymbol{A}_{22}\boldsymbol{g}_2 + 2(n_2 - s_2)s_1\overline{\mu}_1 \\ &= 2(n_1n_2\mu_1 - s_1s_2\overline{\mu}_1) + 2\boldsymbol{g}'_1\boldsymbol{A}_{22}\boldsymbol{g}_2\end{aligned}$$

from (10)

$$
\begin{aligned}
&= 2(n_1 n_2 \mu_1 - s_1 s_2 \overline{\mu}_1) - (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2) G_{in} \\
&\quad + s_1 s_2 (\overline{\mu}_1 + \overline{\mu}_2) + s_1^2 \overline{\mu}_1 \overline{G}_1 + s_2^2 \overline{\mu}_2 \overline{G}_2 \\
&= 2 n_1 n_2 \mu_1 - (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2) G_{in} \\
&\quad - s_1 s_2 (\overline{\mu}_1 - \overline{\mu}_2) + s_1^2 \overline{\mu}_1 \overline{G}_1 + s_2^2 \overline{\mu}_2 \overline{G}_2 \\
&= 2 n_1 n_2 \mu_1 - (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2) G_{in} \\
&\quad \mp (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2) G_c + s_1^2 \overline{\mu}_1 \overline{G}_1 + s_2^2 \overline{\mu}_2 \overline{G}_2 \\
&= 2 n_1 n_2 \mu_1 - (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2)(G_{in} \pm G_c) \\
&\quad + s_1^2 \overline{\mu}_1 \overline{G}_1 + s_2^2 \overline{\mu}_2 \overline{G}_2 \qquad (12)
\end{aligned}
$$

Put (3) and (4), we can get

$$
\begin{aligned}
n^2 \mu (1 - G) &= \boldsymbol{i}_n' \boldsymbol{A} \boldsymbol{i}_n \\
&= (\boldsymbol{t}_1 + \boldsymbol{t}_2)' \boldsymbol{A} (\boldsymbol{t}_1 + \boldsymbol{t}_2) \\
&= \boldsymbol{t}_1' \boldsymbol{A} \boldsymbol{t}_1 + 2 \boldsymbol{t}_1' \boldsymbol{A} \boldsymbol{t}_2 + \boldsymbol{t}_2' \boldsymbol{A} \boldsymbol{t}_2 \\
&= n_1^2 \mu_1 (1 - G_1) + n_2^2 \mu_2 (1 - G_2) + 2 n_1 n_2 \mu_1 \\
&\quad - (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2)(G_{in} \pm G_c) + s_1^2 \overline{\mu}_1 \overline{G}_1 + s_2^2 \overline{\mu}_2 \overline{G}_2
\end{aligned}
$$

Then the transformation is

$$
\begin{aligned}
n^2 \mu G &= n_1^2 \mu_1 G_1 + n_2^2 \mu_2 G_2 + n_1 n_2 (\mu_1 + \mu_2) - 2 n_1 n_2 \mu_1 \\
&\quad + (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2)(G_{in} \pm G_c) - s_1^2 \overline{\mu}_1 \overline{G}_1 - s_2^2 \overline{\mu}_2 \overline{G}_2 \\
&= n_1^2 \mu_1 G_1 + n_2^2 \mu_2 G_2 + n_1 n_2 (\mu_2 - \mu_1) \\
&\quad + (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2)(G_{in} \pm G_c) - s_1^2 \overline{\mu}_1 \overline{G}_1 - s_2^2 \overline{\mu}_2 \overline{G}_2
\end{aligned}
$$

substituting (2) in these

$$
\begin{aligned}
&= n_1^2 \mu_1 G_1 + n_2^2 \mu_2 G_2 + n^2 \mu G_b + (s_1 + s_2)(s_1 \overline{\mu}_1 + s_2 \overline{\mu}_2)(\pm G_c + G_{in}) \\
&\quad - s_2^2 \overline{\mu}_2 \overline{G}_2 - s_1^2 \overline{\mu}_1 \overline{G}_1 \qquad (13)
\end{aligned}
$$

Finally we get new decomposition of the Gini coefficient

$$
G = \frac{n_1^2 \mu_1}{n^2 \mu} G_1 + \frac{n_2^2 \mu_2}{n^2 \mu} G_2 + G_b
$$

$$+\frac{(s_1+s_2)(s_1\overline{\mu}_1+s_2\overline{\mu}_2)}{n^2\mu}(\pm G_c+G_{in})$$
$$-\frac{s_1^2\overline{\mu}_2}{n^2\mu}\overline{G}_1-\frac{s_2^2\overline{\mu}_2}{n^2\mu}\overline{G}_2 \qquad (14)$$

more simply we can write

$$G=a_1G_1+a_2G_2+G_b+R$$

where

$$a_i=\frac{n_i^2\mu_i}{n^2\mu}$$

and

$$R=\pm bG_c+bG_{in}-c_1\overline{G}_1-c_2\overline{G}_2$$
$$b=\frac{(s_1+s_2)(s_1\overline{\mu}_1+s_2\overline{\mu}_2)}{n^2\mu}$$
$$c_i=\frac{s_i^2\overline{\mu}_i}{n^2\mu}$$

Although the 'difficult to interpretation term' R is always regarded as interaction effects, the implication of such effects is often obscure. In this paper, we get more useful decomposition, which consists of four parts. Two of them are the Gini coefficients of interaction area of subgroup 1 and subgroup 2. The term G_{in} expresses 'pure' diverse interaction. The term G_c explains the mean difference effects. It gives positive effect if $\overline{\mu}_1 \geq \overline{\mu}_2$ or negative effect if $\overline{\mu}_1 < \overline{\mu}_2$. And $\overline{G}_1$ and $\overline{G}_2$ show the income inequalities in overlapping interval. They give the negative effects, if the income inequalities increase in each overlapping area.

We can give one implication to the 'troublesome example' proposed by Mookherjee and Shorrocks [6]. They regarded the term R as vague one, and cite the example to illustrate it. It has two subgroups, subgroup 1 : $\{3,4,14\}$ and subgroup 2 : $\{8,11\}$, then $G_1=0.349$, $G_2=0.079$, and the overall Gini coefficient is 0.29. If a redistribution of subgroup 1 incomes results in the new set of incomes $\{1,7,13\}$, G_1 increases to 0.381, and G_2 is unchanged. But the overall Gini coefficient declines to 0.27. Although this example is difficult to understand intuitively, we can disaggrigate the effect to some factors. In case 1, $S = \{8, 11, 14\}$ and in case 2, $S = \{8, 11, 13\}$, we show the values of terms in table.

Coefficient	G	G_1	G_2	G_b	G_c	G_{in}	$\overline{G}_1$	$\overline{G}_2$
Case 1	0.29	0.349	0.079	0.075	0.091	0.121	0.000	0.079
Case 2	0.27	0.381	0.079	0.075	0.073	0.104	0.000	0.079

and their weights are

Coefficient	a_1	a_2	b	c_1	c_2
Case 1	0.270	0.190	0.495	0.070	0.190
Case 2	0.270	0.190	0.480	0.065	0.190

Then, we can express the overall Gini coefficients by factors.

The case 1

$$\begin{aligned} G_1 &= 0.270 \times 0.349 + 0.190 \times 0.079 + 0.075 \\ &\quad +0.495 \times 0.091 + 0.495 \times 0.121 - 0.070 \times 0.000 - 0.190 \times 0.079 \\ &= 0.09423 + 0.01501 + 0.075 \\ &\quad +0.045045 + 0.059895 - 0.00000 - 0.01501 \end{aligned}$$

The case 2

$$\begin{aligned} G_2 &= 0.270 \times 0.381 + 0.190 \times 0.079 + 0.075 \\ &\quad +0.480 \times 0.073 + 0.480 \times 0.104 - 0.070 \times 0.000 - 0.190 \times 0.079 \\ &= 0.10287 + 0.01501 + 0.075 \\ &\quad +0.03504 + 0.04992 - 0.00000 - 0.01501 \end{aligned}$$

This calculation helps to account for the result. The example cited by them shows the overlapping ratio is fairly large. Therefore the interaction Gini coefficients and the cross Gini coefficients in the area have a great deal of influence on the total Gini coefficient. In this case, the redistribution gives a gain in G_1, but it gives losses in G_c and G_{in}. This change can be explained by the decrease in mean difference in overlapping area and by the decrease of inequality in overlapping area.

3 Decomposition in the General Case

In the previous section, we derived new decomposition of one-overlapping case. We consider more general multi-overlapping case in this section. To expand the decomposition, we use a different notation. There are $k(< n)$ subgroups, we denote ith subgroups S_i and consider overlapping set of subgroup i and subgroup j. We express this as S_{ij}, that is to say

$$S_{ij} = \{x_l \mid x_l \in [\min_{x_l \in S_j}(x_l), \max_{x_l \in S_i}(x_l)]\}$$

when $\mu_j \geq \mu_i$. The number in the $S_i \cap S_{ij}$ is expressed by the s_{ij}, and the mean in this area is denoted by $\overline{\mu}_{ij}$. $\overline{G}_{ij}$ is the Gini coefficient in $S_i \cap S_{ij}$. $G_{c[ij]}$ is the Gini coefficient where s_{ij} members get each $\overline{\mu}_{ij}$ income and s_{ji} members get each $\overline{\mu}_{ji}$ income. $G_{in[ij]}$ is the Gini coefficient in S_{ij}. [1]

Similarly (12), $2\boldsymbol{t}_i'\boldsymbol{A}\boldsymbol{t}_j$ can be expressed as

$$\begin{aligned} 2\boldsymbol{t}_i'\boldsymbol{A}\boldsymbol{t}_j &= 2n_in_j\min(\mu_i,\mu_j) \\ &\quad -(s_{ij}+s_{ji})(s_{ij}\overline{\mu}_{ij}+s_{ji}\overline{\mu}_{ji})(\pm G_{c[ij]}+G_{in[ij]}) \\ &\quad +s_{ij}^2\overline{\mu}_{ij}\overline{G}_{ij}+s_{ji}^2\overline{\mu}_{ji}\overline{G}_{ji} \end{aligned} \tag{15}$$

when $i < j$, and

$$\boldsymbol{t}_i'\boldsymbol{A}\boldsymbol{t}_i = n_i^2\mu_i(1-G_i) \tag{16}$$

By using (15) and (16)

$$\begin{aligned} n^2\mu(1-G) &= \boldsymbol{i}'\boldsymbol{A}\boldsymbol{i} \\ &= (\boldsymbol{t}_1+\boldsymbol{t}_2+\cdots+\boldsymbol{t}_k)'\boldsymbol{A}(\boldsymbol{t}_1+\boldsymbol{t}_2+\cdots+\boldsymbol{t}_k) \\ &= \sum_{i=1}^{k}\boldsymbol{t}_i'\boldsymbol{A}\boldsymbol{t}_i+\sum_{i<j}2\boldsymbol{t}_i'\boldsymbol{A}\boldsymbol{t}_j \end{aligned}$$

[1] Here are $s_{ij} \neq s_{ji}$, $\overline{\mu}_{ij} \neq \overline{\mu}_{ji}$ and $\overline{G}_{ij} \neq \overline{G}_{ji}$. The subscript ij expresses the set $S_i \cap S_{ij}$ and the ji $S_j \cap S_{ij}$. Alternatively, $G_{c[ij]}$ and $G_{in[ij]}$ are defined only if $\mu_j \geq \mu_i$.

$$
\begin{aligned}
&= \sum_{i=1}^{k} n_i^2\mu_i(1-G_i) + \sum_{i<j} 2n_in_j\min(\mu_i,\mu_j) \\
&\quad - \sum_{i<j}(s_{ij}+s_{ji})(s_{ij}\overline{\mu}_{ij}+s_{ji}\overline{\mu}_{ji})(\pm G_{c[ij]}+G_{in[ij]}) \\
&\quad + \sum_{i<j} s_{ij}^2\overline{\mu}_{ij}\overline{G}_{ij} + \sum_{i<j} s_{ji}^2\overline{\mu}_{ji}\overline{G}_{ji} \\
&= -\sum_{i=1}^{k} n_i^2\mu_iG_i + \sum_{i,j} n_in_j\min(\mu_i,\mu_j) \\
&\quad - \sum_{i<j}(s_{ij}+s_{ji})(s_{ij}\overline{\mu}_{ij}+s_{ji}\overline{\mu}_{ji})(\pm G_{c[ij]}+G_{in[ij]}) \\
&\quad + \sum_{i<j} s_{ij}^2\overline{\mu}_{ij}\overline{G}_{ij} + \sum_{i<j} s_{ji}^2\overline{\mu}_{ji}\overline{G}_{ji} \qquad (17)
\end{aligned}
$$

The between-Gini coefficient is expressed as

$$
n^2\mu(1-G_b) = \sum_{i,j} n_in_j\min(\mu_i,\mu_j) \qquad (18)
$$

From (17) and (18)

$$
\begin{aligned}
n^2\mu G &= \sum_{i=1}^{k} n_i^2\mu_iG_i + n^2\mu G_b \\
&\quad + \sum_{i<j}(s_{ij}+s_{ji})(s_{ij}\overline{\mu}_{ij}+s_{ji}\overline{\mu}_{ji})(\pm G_{c[ij]}+G_{in[ij]}) \\
&\quad - \sum_{i<j} s_{ij}^2\overline{\mu}_{ij}\overline{G}_{ij} - \sum_{i<j} s_{ji}^2\overline{\mu}_{ji}\overline{G}_{ji} \qquad (19)
\end{aligned}
$$

The general form is

$$
G = \sum_{i=1}^{k} a_iG_i + G_b + \sum_{i<j} b_{ij}(\pm G_{c[ij]}+G_{in[ij]}) - \sum_{i\neq j} c_{ij}\overline{G}_{ij} \qquad (20)
$$

where

$$
a_i = \frac{n_i^2\mu_i}{n^2\mu}
$$

$$
b_{ij} = \frac{(s_{ij}+s_{ji})(s_{ij}\overline{\mu}_{ij}+s_{ji}\overline{\mu}_{ji})}{n^2\mu}
$$

$$
c_{ij} = \frac{s_{ij}^2\overline{\mu}_{ij}}{n^2\mu}
$$

Now we get the decomposition of the Gini coefficient. The residual term is largely effected by the condition in overlapping area. One of them is the product of population share and income share in overlapping area. The others are the cross effect, the interaction effects and diversity in overlapping area. In the k deviation case, the number of G_c, G_{in} and $\overline{G}$ are $k(k-1)/2$, $k(k-1)/2$ and $k(k-1)$, respectively. Therefore there are $2k(k-1)$ effects in residual term.

4 Conclusion

We have derived a new decomposition of the Gini coefficient. Although a large number of studies have been made on the interpretation of the residual term, most of studies give weight to income reranking. While there is a fairly general agreement that interaction effect increases the value of the residual term, no study has ever tried to express the interaction effect concretely. In this paper we derived new decomposition from another point of view, and got the form of the residual term.

The residual term is influenced by the overlapping ratio and the degree of the diversity of income. Such a shape is composed of two overlapping combinations. However the residual terms have a fine form, it is too complicated to state the economic implication directly. The overlapping is more limited in practice, but we cannot avoid the complexities when the number of groups increases. Because if the deviation increases, the number of interaction effects increases geometrically. The residual term has an original evaluation to the overlapping area. Its evaluation is based on the combination of two subsets. Although the Gini coefficient is not so strange to consider income inequality, the residual term has a complication to consider effects intuitively. In particular, when there are many subsets and overlapping parts, we cannot neglect that the Gini coefficient has difficulties to explain the income inequality.

References

[1] Atkinson A. B., (1970) "On the measurement of inequality", *Journal of Economic Theory* 2, pp. 244–263.

[2] Bhattacharya N. and Mahalanobis B., (1967) "Regional disparities in household consumption in India", *Journal of American Statistical Association* 62, pp. 143–161.

[3] Kakwani N. C., (1980) *Income, Inequality and Poverty*, Oxford University Press, Oxford

[4] Lambert P. J., (1989) *The Distribution and Redistribution of Income*. Blackwell, Oxford

[5] Lambert P. J. and Aronson J. R., (1993) "Inequality Decomposition Analysis and the Gini Coefficient Revisited", *The Economic Journal* 103, pp. 1221–1227.

[6] Mookherjee D. and Shorrocks A. F., (1982) "A Decomposition analysis of trend in U.K. Income Inequality", *The Economic Journal* 92, pp. 886–902.

[7] Pyatt G., (1976) "The interpretation and disaggregation of Gini Coefficients", *The Economic Journal* 86, pp. 243–255.

[8] Silber J., (1989) "Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality" *The Review of Economics and Statistics* 71, pp. 107–115.