Income stratification and between-group inequality

Paul Allanson
University of Dundee
Income stratification and between-group inequality

Paul Allanson¹

Economics Studies, University of Dundee, Scotland.

Abstract

The paper demonstrates that the ratio of the Yitzhaki (1994) to the conventional measure of between-group inequality is in general equal to one minus twice the weighted average probability that a random member of a richer (on average) group is poorer than a random member of a poorer (on average) group, and may therefore be interpreted as an index of stratification in its own right.

JEL code: D31, D63

Key words: Gini coefficient; subgroup decomposability; income stratification

¹ Corresponding author. Economics Studies, University of Dundee, Perth Road, Dundee DD1 4HN, UK. Tel: +44 01382 384377 Fax: +44 01382 384691. Email: p.f.allanson@dundee.ac.uk.
Introduction

It is well known that the standard decomposition of the Gini index $G$ by groups does not yield an exact partition into between-group and within-group components, $G_B$ and $G_W$ respectively, unless the income ranges of the groups are non-overlapping (see, e.g., Mookherjee and Shorrocks, 1982). This has led to an extensive literature exploring the nature of the “residual” from the standard decomposition with the graphical interpretation in Lambert and Aronson (1993) showing clearly how it arises from the overlapping of incomes across groups and with Lambert and Decoster (2005) claiming to obtain, perhaps for the first time, a ‘transparent analytical expression’ for it in the two group case.

It has also led to a parallel search for alternative decomposition procedures that might prove more amenable to analysis and interpretation. Thus Yitzhaki and Lerman (1991) provides a partition of the Gini into between-group, within-group and overlapping components, $G_b$, $G_w$ and $G_o$ respectively, where overlapping is considered as the inverse of the sociological concept of ‘stratification’. Yitzhaki (1994) subsequently combines the latter two elements into a single within-group measure $G_{wo}$ that is explicitly written as a function of the degree of inequality within groups and the degree of overlapping between each pair of groups, but $G_b$ is also affected by overlapping and it remains unclear as to how this measure relates to the conventional between-group index $G_B$ (cf. Yitzhaki and Schechtman, 2013).

Monti and Santoro (2011) address this issue in the two group case by showing that the ratio $I = G_b/G_B$ is a function of the probability of transvariation (Gini, 1916), i.e. the probability that a random member of the richer (on average) group is poorer than a random member of the poorer (on average) group. The main contribution of this paper is to generalise their result to allow for more than two groups. Specifically we show that $I$ is in general equal to one minus twice the weighted average probability that a random member of a richer (on average) group is poorer than a random member of a poorer (on average) group, with the weights given by the share of each pair’s contribution to $G_B$. We thereby demonstrate fully how the residual from the conventional decomposition is absorbed into the between-group and within-group components proposed by Yitzhaki (1994). We illustrate our results through an elaboration of the empirical analysis of world inequality by regions presented in Milanovic and Yitzhaki (2002).²

---
² These regions are identified as ‘continents’ though the correspondence is not exact.
Group-wise decomposition of the Gini index

We consider a population divided into \( K (K \geq 2) \) mutually exclusive and exhaustive groups that are ordered by expected income from the poorest to the richest group. Let \( Y_k, F_k(Y_k), \mu_k, p_k \) and \( q_k \) represent the income (or some other relevant aspect of wellbeing) variable, the cumulative distribution function, the expected value, the population share and income share of group \( k \), respectively. The overall population \( Y_o = Y_1 \cup Y_2 \ldots \cup Y_K \) is the union of all groups with distribution function \( F_o(Y_o) = \sum_k p_k F_k(Y_k) \) and expected value \( \mu_o = \sum_k p_k \mu_k \).

The mean fractional rank of group \( k \) members within the income distributions of group \( l \) and the overall population are given as \( \bar{F}_{kl} \) and \( \bar{F}_{ko} \) respectively.

Following Mookherjee and Shorrocks (1982), the conventional group-wise decomposition of the Gini index of the overall population may be written as

\[
G = G_B + G_W + R ,
\]

where:

\[
G_B = \frac{1}{2} \sum_k \sum_{l > k} p_k p_l \left| \mu_l - \mu_k \right| / \mu_o = \sum_k \sum_{l > k} p_k p_l (\mu_l - \mu_k) / \mu_o = \sum_k \sum_{l > k} (p_k + p_l)(q_k + q_l) G_{B}^{kl} \quad (1)
\]

\[
G_W = \sum_k p_k q_k G_k \quad (2)
\]

with \( G_{B}^{kl} \) denoting the between-group Gini in the sub-population consisting of groups \( k \) and \( l \) only; \( G_k \) is the Gini index within group \( k \); and the residual \( R \) is interpreted as an ‘interaction effect’. Pyatt (1976) shows that \( R \geq 0 \) implying that the overall effect of interaction due to the overlapping of group income ranges is to increase inequality \textit{ceteris paribus}.

The alternative approach of Yitzhaki (1994) yields the exact decomposition

\[
G = G_b + G_{wo} ,
\]

where:

\[
G_b = 2 \sum_k p_k (\mu_k - \mu_o) \left( \bar{F}_{ko} - 0.5 \right) / \mu_o \quad (3)
\]

\[
G_{wo} = \sum_k q_k G_k O_k = \sum_k q_k G_k \left( \sum_l p_l O_{lk} \right) = \sum_k q_k G_k \left( \sum_l p_l \frac{\text{cov}(Y_k, F_l(Y_k))}{\text{cov}(Y_k, F_k(Y_k))} \right) \quad (4)
\]

with \( F_l(Y_k) \) denoting the (fractional) ranking of group \( k \) incomes in the group \( l \) income distribution. \( O_{lk} \) is interpreted as a pairwise overlapping index that measures of the degree to
which incomes in group $l$ are included in the income range of group $k$. In particular, $O_{lk}$ will take a value of zero when there is no overlap, which will be the case if there is perfect stratification in the sense of Lasswell (1965), since the ranks of members of the $k$'th group within the income distribution of group $l$ will all be identical. More generally, $O_{lk}$ is an increasing function of the fraction of group $l$ that is located in the income range of group $k$, taking a value of one if the income distributions of the two groups are identical, i.e. $F_l(Y_k) = F_k(Y_k)$.

Thus $G_{wo} = \sum_k q_k G_k p_k = G_w$ if there is perfect stratification, since $O_{kk} = 1$ by definition, whereas $G_{wo} > G_w$ if the income ranges of the various groups overlap to any extent with the difference $R_w = G_{wo} - G_w$ given as:

$$R_w = \sum_k q_k G_k \left( \sum_{l \neq k} p_l O_{lk} \right) = 2 \sum_k p_k \left( \sum_{l \neq k} p_l \text{cov}(Y_k, F_l(Y_k)) \right) \mu_o \geq 0 \quad (5)$$

Yitzhaki and Lerman (1991, p.323) conclude that “inequality and stratification are inversely related”, arguing that this relationship is consistent with relative deprivation theory in that “stratified societies can tolerate higher inequality than unstratified societies” since “As people become more (less) engaged with each other, they have less (more) tolerance for a given level of inequality”. However, as Monti and Santori (2011) observe, this conclusion ignores the effect of overlapping on the between-group component $G_b$, which will also affect the overall level of inequality perceived by the society.

Yitzhaki and Lerman (1991, p.322) note that $G_b = G_B$ if there is no overlapping and $G_b < G_B$ otherwise. Monti and Santori (2011) further demonstrate in the two group case that the ratio of $G_b$ to $G_B$ is equal to:

$$I = \frac{G_b}{G_B} = 1 - 2 \text{Prob}(Y_1 > Y_2) \quad (6)$$

where $\text{Prob}(Y_1 > Y_2)$ is the probability that the income of a random member of the richer (on average) group is less than that of a random member of the poorer (on average) group. To extend this result to the general case of $K \geq 2$ groups, we note that $G_b$ may be re-written from (3) as:
where the first line follows since \( F_{kl} = \sum_l p_l F_{kl} \) and \( F_{kk} = 0.5 \), while \( G_{kl}^{kl} \) denotes the Yitzhaki (1994) between-group index in the sub-population consisting only of groups \( k \) and \( l \).

It follows immediately from (6) and (7) that:

\[
G_b = \sum_{k} \sum_{l>k} p_k p_l G_{kl}^{kl} / \mu_0
\]

where \( G_{kl}^{kl} = \sum_{l>k} p_k p_l (\mu_l - \mu_k) \{(1 - 2 \text{Prob}(Y_k > Y_l)) / \mu_0 \}
\] (8)

Hence \( I \) will be equal to:

\[
I = \frac{G_b}{G_B} = \sum_{k} \sum_{l>k} w_{kl} (1 - 2 \text{Prob}(Y_k > Y_l)) = 1 - 2 \sum_{k} \sum_{l>k} w_{kl} \text{Prob}(Y_k > Y_l)
\]

\[
= \sum_{k} \left\{ \sum_{l<k} w_{kl} (0.5 - (1 - \text{Prob}(Y_k < Y_l))) + \sum_{l>k} w_{kl} (0.5 - (\text{Prob}(Y_k > Y_l))) \right\}
\]

(9)

where \( w_{kl} = p_k p_l (\mu_l - \mu_k) / \left( \sum_k \sum_{l>k} p_k p_l (\mu_l - \mu_k) \right) \), with \( \sum_k \sum_{l>k} w_{kl} = 1 \) by definition, and the final line holds since \( \text{Prob}(Y_k > Y_l) = (1 - \text{Prob}(Y_k < Y_l)) \).

Hence \( I \) is in general equal to one less twice the weighted average probability of transvariation between the various pairs of groups in the population. Gastwirth (1975) in his study of earnings differentials proposes \( TPROB = 2 \text{Prob}(Y_1 > Y_2) \) as an index of overlapping between two groups, taking an “ideal” value of one when the two distributions are identical since \( \text{Prob}(Y_k > Y_k) = 0.5 \) by definition. Thus \( I \) in (6) may be interpreted as the complementary index of non-overlapping or stratification, with (9) providing a generalisation to two or more groups. \( I \) is a unit-free index that will take a maximum value of one when there is no overlap between any of the groups such that \( \text{Prob}(Y_k > Y_l) = 0 \ \forall k, l \); and will equal zero when the income distributions of all the groups are identical.\(^3\) The extent to which non-overlapping between any pair of groups contributes to the overall level of stratification is

\(^3\) Negative values of \( I \) are also possible when mean incomes by group are negatively correlated with mean ranks.
an increasing function of their population shares and the difference in average incomes between them. \( I \) is invariant to both the scaling and translation of incomes. It is also invariant to the replication both of the populations within existing groups and of groups.

\( I \) has previously been identified by Milanovic and Yitzhaki (2002, p.161) “as an index indicating the loss of between group inequality due to overlapping”. The difference \( R_B = G_\beta - G_B \) can be written from (9) as:

\[
R_B = -2G_B \sum \sum w_{kl} \operatorname{Prob}(Y_k > Y_l) = -2 \sum \sum p_k p_l (\mu_l - \mu_k) \operatorname{Prob}(Y_k > Y_l) / \mu_o \leq 0
\] (10)

on which basis it may be argued, in contrast to Yitzhaki and Lerman (1991), that unstratified societies can tolerate more between-group inequality than stratified societies because individuals’ positions within society are less narrowly determined by group membership. Nevertheless, with \( G_\theta \) and \( G_\beta \) held constant, overlapping per se must increase overall inequality since \( R \geq 0 \) by definition, where from (5) and (10) we obtain a novel expression for \( R = R_\theta + R_B \) as:

\[
R = 2 \sum p_k \left( \sum_{l > k} \left( p_l \left( \operatorname{cov}(Y_k, F_l(Y_k)) + \operatorname{cov}(Y_l, F_k(Y_l)) - (\mu_l - \mu_k) \operatorname{Prob}(Y_k > Y_l) \right) \right) \right) / \mu_o
\] (11)

where the final line makes use of the expression for \( R \) presented in Lambert and Decoster (2005) for the two group case.

By way of illustration, we elaborate the empirical analysis presented in Milanovic and Yitzhaki (2002) of world inequality in 1993 by regions. The top panel in Table 1 presents estimates from their Tables 4 and 7 of population shares, \( p_k \), mean incomes, \( \mu_k \), and mean rankings in the income distributions of each region, \( F_{kl} \), and the world \( F_{k0} \). The lower panel reports the pairwise components of \( I \) identified in the final line of (12) where the calculation of these estimates makes use of the identity \( F_{kl} = \operatorname{Prob}(Y_k > Y_l) \). The components sum to give the value of the stratification index \( I = 0.776 \), which is equal to the ratio of the reported

---

\[ \text{Lambert and Decoster (2005) state that attention is confined to the case of two population subgroups “for ease of presentation, but the results can clearly be extended.”} \]
estimates of $G_b = 0.309$ and $G_B = 0.398$.\(^5\) Examination of the individual entries shows that the main contribution to stratification, accounting for as much as two thirds of the total, is due to the Asia/WENAO pair as a result of a combination of the low degree of income overlap, the populousness of the two regions and the large difference in mean incomes between them. In contrast, the Africa/Asia pair contributes negatively to stratification, although the magnitude of this effect is negligible, because an African chosen at random is likely to be better off than a randomly chosen Asian despite the fact that average incomes are lower in Africa. Given that the value of $I$ implies a weighted average probability of transvariation of 11.2%, only the Africa/WENAO and Asia/WENAO pairs contribute more to $R_b$ than to $G_b$.

Table 1. Income stratification between regions

<table>
<thead>
<tr>
<th></th>
<th>Popn share (%)</th>
<th>Mean income ($PPP$)</th>
<th>Mean rank in income distribution of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Africa</td>
<td>Asia</td>
</tr>
<tr>
<td>Africa</td>
<td>10.0</td>
<td>1310.0</td>
<td>0.500</td>
</tr>
<tr>
<td>Asia</td>
<td>59.5</td>
<td>1594.6</td>
<td>0.485</td>
</tr>
<tr>
<td>EFSU</td>
<td>7.8</td>
<td>2780.9</td>
<td>0.725</td>
</tr>
<tr>
<td>LAC</td>
<td>8.4</td>
<td>3639.8</td>
<td>0.739</td>
</tr>
<tr>
<td>WENAO</td>
<td>14.3</td>
<td>10012.4</td>
<td>0.951</td>
</tr>
<tr>
<td>World</td>
<td>100.0</td>
<td>3031.8</td>
<td>0.915</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pairwise contribution to $I$</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>-0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>Asia</td>
<td>-0.000</td>
<td>0.011</td>
</tr>
<tr>
<td>EFSU</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>LAC</td>
<td>0.004</td>
<td>0.021</td>
</tr>
<tr>
<td>WENAO</td>
<td>0.047</td>
<td>0.258</td>
</tr>
<tr>
<td>World</td>
<td></td>
<td>0.776</td>
</tr>
</tbody>
</table>

Notes: Top panel. Source: Milanovic and Yitzhaki (2002) Tables 4 and 7 - see also Table 1 for country groupings (EFSU – Eastern Europe and Former Soviet Union; LAC – Latin America and Caribbean; WENAO – Western Europe, North America and Oceania). Bottom panel. Author’s own calculations.

\(^5\) Note that this is not the case with the results presented in Monti and Santori (2011) who base their analysis on country-level mean income data.
Conclusion

The paper demonstrates fully how the residual from the conventional decomposition of the Gini index is absorbed into the between-group and within-group components proposed by Yitzhaki (1994). In particular, we demonstrate that $I = I / G_b$ is in general equal to one minus twice the weighted average probability of transvariation and may therefore be interpreted as an index of stratification in its own right. We are thereby able to show that the main source of stratification between regions in 1993 was the limited overlap between the income distributions of Asia and WENAO given the relative populousness of the two regions and the difference in mean incomes between them. High per capita growth rates in some poorer Asian countries, most notably China and India, may be expected to have reduced levels of both stratification and inequality between regions in more recent years.⁶

⁶ See Milanovic (2012) for further discussion and evidence on trends in between-country inequality.
References


