the review of income and wealth

Review of Income and Wealth
Series 57, Number 3, September 2011
DOI: 10.1111/j.1475-4991.2010.00414.x

STRATIFICATION AND BETWEEN-GROUP INEQUALITY: A NEW INTERPRETATION

BY MARIA MONTI*

DSGE, University of Milan-Bicocca and DEAS, University of Milan

AND

ALESSANDRO SANTORO

DSGE, University of Milan-Bicocca and Econpubblica

Traditionally, the literature has seen stratification as linked closely to within-group inequality. More recently, some papers have focused on measuring the impact of stratification on between-group inequality. In this paper, we show that when two groups are involved, such an impact can be measured by a simple comparison of the two cumulative distribution functions. This approach allows an interpretation of stratification in terms of probabilities and paves the way for a neat and simple graphical illustration. We apply it to the analysis of between-continent inequality.

1. INTRODUCTION

Stratification means a group's isolation from members of other groups (Yitzhaki and Lerman, 1991, p. 319). A group is said to be *stratified* when it tends to form a perfect stratum in the overall distribution. Stratification has been used traditionally in sociological studies, but its rigorous definition and measurement are owed to Yitzhaki and Lerman (1991) and Yitzhaki (1994). To this end, they propose a decomposition of the Gini index into two parts: a component that is a weighted sum of groups' Ginis, and a between-group inequality measure. In the first component, the weight depends positively on the value of the overlapping index for each group.

In turn, the overlapping index for every group measures the extent to which the income ranges of the members of that group overlap with those of members of other groups. The less a group is stratified, the more it overlaps with other groups, and therefore Yitzhaki and Lerman (1991, p. 323) conclude that "inequality and stratification are inversely related."

Yitzhaki and Lerman argue that, though counterintuitive, this result is consistent with the relative deprivation theory. The idea is that groupings are leagues, so that each person "confines his aspirations to his assigned league" (Yitzhaki and Lerman, 1991, p. 323). To put it another way, an individual (i.e. a group member)

Note: We thank the Editor, two anonymous referees, Conchita d'Ambrosio and Bruno Bosco for their comments on a previous version of this paper. We also thank participants from seminars held at the State University of Milan and the University of Milan-Bicocca. The usual disclaimers apply. Financial support from the Italian Ministry of Education and Research (PRIN 2007HEWTBE) is gratefully acknowledged.

*Correspondence to: Maria Monti, DSGE, Università degli studi di Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy (maria.monti@guest.unimi.it).

cares more about the distribution in his/her own group than in other groups. Thus, "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" (Yitzhaki and Lerman, 1991, p. 323). Yitzhaki and Lerman (1991, p. 315) reinforce this interpretation, observing that "generally a rise in a subgroup's inequality will reduce the subgroup's stratification," so that, in general, if within-group inequality is low, overlapping is also low or, equivalently, stratification is high. Note that in this line of reasoning, the between-group inequality, which is the second component of the Gini decomposition in Yitzhaki (1994), is left completely aside: both Yitzhaki and Lerman (1991) and Yitzhaki (1994) treat stratification and between-group inequality as two completely separate objects. Nevertheless, Milanovic and Yitzhaki (2002), aware of the impact of stratification on between-group inequality, suggest evaluation by the ratio of the Yitzhaki and Lerman (1991) between-group Gini and the conventional between-group Gini coefficient, as obtained by Pyatt (1976).

In this paper, we take this ratio as our starting point. When two groups are considered, we show that such a ratio, i.e. the measure of the impact of stratification on between-group inequality, is a function of the probability that a random member of the poorer (on average) group is richer than a random member drawn from the (on average) richer group. Thus, this measure depends exclusively on the cumulative distribution functions of the two groups. The value of the ratio proposed by Milanovic and Yitzhaki (2002) can actually be expressed as one minus twice the area under the cumulative distribution function for the richer (on average) group, evaluated within a range of values of the cumulative distribution function of the poorer (on average) group. This expression naturally suggests a graphical interpretation. We provide the latter by applying our approach to the analysis of between-continent inequality using the data reported by Milanovic and Yitzhaki (2002). This is the first and main contribution of the paper.

The second contribution of the paper is the link we establish between the literature on stratification and the Gini concept of transvariation.[1] Building on Monti and Santoro (2009), it can be shown that the ratio proposed by Milanovic and Yitzhaki (2002) depends on the total number of transvariations occurring between the members of the two groups. Thus, such a ratio is ultimately a function of the probability of a transvariation, and this offers an additional interpretation of the results obtained here.

The paper is organized as follows. Section 2 summarizes the contribution of Yitzhaki and his colleagues to the measurement of stratification. Section 3 contains the theoretical results of the paper. We derive the measures of the impact of stratification on between-group inequality in terms of the cumulative distribution functions of the compared groups and provide an analysis of the range of variation of these measures. Section 4 applies these results to the analysis of between-continent inequality using a mainly graphical approach. Section 5 concludes.

[1]Given two groups with different average incomes, a transvariation occurs whenever a member of the poorer (on average) group has an income higher than a member of the richer (on average) group.

2. Overview of the Literature

The concept of stratification is frequently used in social science literature. Its definition can be traced back at least to Lasswell (1965, p. 10): "A stratum is a horizontal layer. Stratification is the process of forming observable layers . . . where the mass of society is constructed of layer upon layer of congealed population qualities."

Until Yitzhaki and Lerman (1991), however, a rigorous approach to definition and measurement was lacking. Yitzhaki and Lerman's (1991) contribution is threefold. First, they obtain an index of relative stratification, index Q, which captures the extent of stratification of every group with respect to the entire population, taking into account the size of the group. Second, they derive an index of absolute overlapping, O, which is inversely related to Q. Third, they decompose the Gini index into three parts: a within-inequality component, a component that reflects the impact of stratification, and a measure of between-group inequality. When commenting upon the dynamics of these three parts, Yitzhaki and Lerman (1991, p. 323; our emphasis) note that "some changes in Q_i's may leave Ginis unchanged, *and influence only component two*." This implies that, in general, there is no relationship between stratification and between-group inequality and that these can be treated as separate concepts.

Yitzhaki (1994) further develops the index of overlapping O, focusing on overlapping between subpopulations. He obtains a decomposition of the Gini index into two components: the between-group inequality measure, defined by Yitzhaki and Lerman (1991), and a term that is the sum of the products of income shares, Ginis and overlaps for all groups. This decomposition is used by Milanovic and Yitzhaki (2002) to measure the world's income inequality. To summarize the main results obtained by Yitzhaki (1994) let us introduce some notation. We consider a population of n individuals and we confine our attention to the case of two population groups,[2] which stand for a given socioeconomic partition of the population based on the individuals' characteristics. We call affluent, denoted by a, the group with the higher average income and we call poor, denoted by p, the other group. The population size is $n_a + n_p = n$, with $n, n_a, n_p \in \mathbb{N}$, where n_a is the number of individuals belonging to group a, and n_p is the number of individuals belonging to group p. By y_{ih} we denote the income of individual h belonging to group i ($i = a, p$), and by μ, μ_p, and μ_a, the overall, the group p, and the group a average income, respectively.[3] Consistently with the notation above, we have $\mu_a > \mu_p$. Finally, $\{y\}$ is the set of all income units.

Following Yitzhaki (1994), the Gini index decomposes as

$$G(y) = (\omega_a G_a O_a + \omega_p G_p O_p) + G_b, \tag{1}$$

where

$$O_i = s_i + \sum_{j \neq i} s_j O_{ji}, \tag{2}$$

[2]As suggested by an anonymous referee, it is worth emphasizing that only when two groups are considered do stratification and between-group inequality relate so directly.

[3]We assume that these means are all positive.

(3) $$O_{ji} = \text{cov}_i\left(y, F_j(y)\right) / \text{cov}_i\left(y, F_i(y)\right),$$

and

(4) $$G_b = 2\,\text{cov}(\mu_i, \bar{F}_{Oi}) / \mu.$$

In (1–3), $F_i(y)$, μ_i, G_i, and s_i represent the cumulative distribution, the average income, the Gini index, and the share of group i in the overall distribution, respectively. Let $\omega_i = s_i\mu_i/\mu$ denote the share of total income owned by group i, and O_i denote the overlapping index of the same group i with the population's distribution. The index O_i is a function of the overlapping of group j by group i, O_{ji}, which in turn, is equal to the ratio between "the covariance between incomes of group i and their rank, had they been considered as belonging to the group j" (Yitzhaki, 1994, p. 149) and the covariance between incomes and own ranking in group i, the latter being a normalizing factor.[4] Finally, in expression (4), G_b is twice the covariance between each group's average income and group's average rank in the overall population ($\bar{F}_{Oi}$), divided by the overall mean income. The overlapping index O_i reflects the overlapping of group i with itself and with the other groups, and can be interpreted as a measure of stratification.

In (1), in the round brackets, the subgroup Gini indices, G_i, and the overlap indices, O_i, have symmetrical impacts on overall inequality, since inequality rises in both. Nevertheless, high stratification implies low overlapping so that if G_b is ignored, one concludes that "inequality and stratification are inversely related" (Yitzhaki and Lerman, 1991, p. 323). According to Milanovic and Yitzhaki (2002, p. 161), however, more overlapping (i.e. less stratification) leads to lower correlation between average income and average rank and this decreases the between-group component. To measure the impact of stratification on between-group inequality, Milanovic and Yitzhaki (2002) refer to the conventional decomposition of the Gini index as proposed by Pyatt (1976):

(5) $$G(y) = G_W + G_B + R,$$

where,

(6) $$G_W = \frac{1}{n^2\mu}\left(G_a n_a^2 \mu_a + G_p n_p^2 \mu_p\right) \quad \text{and} \quad G_B = \frac{n_a n_p}{n^2\mu}(\mu_a - \mu_p);\ \mu_a > \mu_p.$$

In (6), the terms G_a and G_p denote the group a and group p Gini index, respectively, so that G_W measures within-group inequality, G_B captures between-group inequality, and R is the residual, which depends on the overlapping between the two group income distributions. In the conventional Gini decomposition, G_B is different from G_b. In both G_B and G_b, each group is jointly represented by its mean income and rank. In G_B, however, the rank of a group is the rank of its mean whereas in G_b one "takes account of each observation's ranking in the overall

[4]To interpret these definitions recall that Yitzhaki and Lerman (1991, p. 321) estimate the cumulative distribution, $F(y)$, by the rank of y.

distribution by averaging these rankings within each group" (Yitzhaki and Lerman, 1991, p. 322). The two between-group inequality components are equal if there is no overlap between groups and it can be shown that $G_b < G_B$ when groups overlap (we return to this in Section 3). The ratio G_b/G_B is suggested by Milanovic and Yitzhaki (2002, p. 161) as an index representing the loss in between-group inequality owing to an increase (decrease) in overlapping (stratification).

Immediately G_b rewrites as

$$G_b = \frac{(\mu_a - \mu_p)}{\mu} \cdot \frac{n_a n_p}{n^2} \cdot I, \tag{7}$$

where[5]

$$I = \frac{G_b}{G_B}. \tag{8}$$

In the Appendix, we show that

$$I = 1 - \frac{2\sum_{h=1}^{n_p}(R_{ph} - r_{ph})}{n_a n_p}. \tag{9}$$

In expression (9), using the same notation of Yitzhaki and Lerman (1991), we denote by R_{ph} the rank of y_{ph} in the overall distribution and by r_{pk} the rank of y_{ph} in the distribution of group p. The difference $R_{ph} - r_{ph}$ evaluates the number of members of the affluent group a with an income lower than y_{ph} so that the sum $\sum_{h=1}^{n_p}(R_{ph} - r_{ph})$ is the number of instances where an income of group p is greater than an income of group a. Note that this number is equal to the number of instances where an income of group a is lower than an income of group p.

3. Index I

Here, we focus on I and on its expression (9). Since $n_a n_p$ is the total number of comparisons between members of the two groups, the ratio $\sum_{h=1}^{n_p}(R_{ph} - r_{ph})\big/ n_a n_p$ can be interpreted as the probability that a random member of the group which is on average poorer is richer than a random member drawn from the (on average) richer group. Then, recalling from the previous section that $\mu_a > \mu_p$, we represent the income sets of the two groups as two discrete random variables denoted by Y_p and Y_a, respectively, so that[6]

[5]This ratio is used by Frick *et al.* (2006), who also suggest statistical tests that are not covered in the present paper.

[6]We observe that in the definition of the Gini index and in its decomposition, there is an implicit assumption of independence between Y_a and Y_p. For all j and l, the probability of the difference $(y_{aj} - y_{pl})$ is the product of the probability to observe y_{aj} in the distribution Y_a and the probability to observe y_{pl} in distribution Y_p. That is, given $\Pr(Y_a = y_{aj}) = 1/n_a$ and $\Pr(Y_p = y_{pl}) = 1/n_p$, one has $\Pr(Y_a = y_{aj}, Y_p = y_{pl}) = \Pr(Y_a = y_{aj}) \cdot \Pr(Y_p = y_{pl}) = \frac{1}{n_a} \cdot \frac{1}{n_p}$.

(10) $$\text{Prob}[Y_p > Y_a] \equiv \frac{\sum_{h=1}^{n_p}(R_{ph} - r_{ph})}{n_a n_p}$$

and we can write

(11) $$I = 1 - 2\text{Prob}[Y_p > Y_a].$$

We can now derive a number of results concerning the range of I. First, we note that index I assumes its maximum, $I = 1$ ($G_b = G_B$), when stratification is perfect:

(12) $$I = 1 \Leftrightarrow \text{Prob}[Y_p > Y_a] = 0.$$

Second, we note that

(13) $$\begin{aligned} I = 0 &\Leftrightarrow \text{Prob}[Y_p > Y_a] = 1/2 \\ &\Leftrightarrow \sum_{h=1}^{n_p}(R_{ph} - r_{ph}) = n_a n_p/2. \end{aligned}$$

The index I is equal to zero if, and only if, the probability that a random member of the poorer group is richer than a random member drawn from the (on average) richer group is exactly equal to 50 percent. Third, we note that the index $I = G_b/G_B$ is at its minimum when the latter probability reaches its maximum, i.e.

(14) $$\min\left\{I = \frac{G_b}{G_B}\right\} = 1 - 2\max\{\text{Prob}[Y_p > Y_a]\},$$

where:

(15) $$\max\{\text{Prob}[Y_p > Y_a]\} = \frac{1}{n_a n_c}\max\sum_{h=1}^{n_p}(R_{ph} - r_{ph}).$$

By denoting q_a as the number of members of group a whose income is higher than μ_a, and ℓ_p as the number of members of group p whose income is (weakly) lower than μ_p, one obtains[7]

(16) $$\max\sum_{h=1}^{n_p}(R_{ph} - r_{ph}) = n_a n_p - q_a \ell_p,$$

and the minimum value of $I = G_b/G_B$ is

(17) $$\min\left\{I = \frac{G_b}{G_B}\right\} = -1 + \frac{2q_a \ell_p}{n_a n_p}.$$

Expression (17) suggests immediately writing both the index I and its minimum in a continuous form. In what follows, we assume that (see footnote 6)

[7]The proof of this result can be obtained from the authors upon request.

$$(18)\quad \text{Prob}(Y_a < y_a, Y_p < y_p) = \text{Prob}(Y_a < y_a)\text{Prob}(Y_p < y_p) = F(y_a)G(y_p),$$

where $F(y_a)$ and $G(y_p)$ are the cumulative distributions of Y_a and Y_p, respectively. Then, since

$$(19)\qquad \text{Prob}(Y_a > \mu_a, Y_p < \mu_p) = [1 - F(\mu_a)]G(\mu_p),$$

expression (15) rewrites as

$$(20)\qquad \max\{\text{Prob}[Y_p > Y_a]\} = 1 - [1 - F(\mu_a)]G(\mu_p),$$

and expression (17) becomes

$$(21)\qquad \min\left\{I = \frac{G_b}{G_B}\right\} = 2[1 - F(\mu_a)]G(\mu_p) - 1.$$

Using (21), we can see that min I can be negative and that its value depends on the skewness of the two distributions. If the two distributions are both symmetric with respect to their mean, the minimum value of index I is −1/2. On the other hand, one has min $I > -1/2$ if either the distribution of Y_a is symmetric with respect to μ_a and the distribution of Y_p has positive asymmetry ($\mu_p >$ median, right obliquity), or if the distribution of Y_a has negative asymmetry ($\mu_a <$ median, left obliquity) and the distribution of Y_p is either symmetric or asymmetric with positive asymmetry.

Moreover, one has min $I < -1/2$ if either the distribution of Y_a is symmetric with respect to μ_a and the distribution of Y_p has negative asymmetry, or if the distribution of Y_a has positive asymmetry ($\mu_a >$ median, right obliquity) and the distribution of Y_p is either symmetric or asymmetric with negative asymmetry. Nothing can be said about min I if the two distributions are asymmetric with the same asymmetry.

Let us now consider the continuous expression of index I (expressions (10) and (11)). Observe that

$$(22)\qquad \text{Prob}(Y_p > Y_a) = \text{Prob}(Y_a - Y_p < 0).$$

Then, if the difference variable ($Y_a - Y_p$) is denoted by Z, and its cumulative distribution is denoted by $H(z)$, one has

$$(23)\quad \text{Prob}(Z < z) = H(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z+y_p} dF(y_a)dG(y_p) = \int_{-\infty}^{\infty} F(z + y_p)dG(y_p),$$

and expression (10) becomes

$$(24)\qquad \text{Prob}(Z < 0) = H(0) = \int_{-\infty}^{\infty} F(y_p)dG(y_p).$$

Using (24) back in (11), we can write

$$I = 1 - 2\int_{-\infty}^{\infty} F(y_p)\, dG(y_p). \tag{25}$$

Expression (25) says that the measure of impact of overlapping on between-inequality can be expressed as a function of the cumulative distribution functions of the two groups. More precisely, expression (25) says that $I = G_b/G_B$ is equal to one minus twice the area under the cumulative distribution function of the richer group, evaluated as a function of the cumulative distribution function of the poorer group.

For readers familiar with Gini's concept of transvariation (Gini, 1959; Giorgi, 2005; Monti and Santoro, 2009), additional insights can be provided. In general, a transvariation occurs whenever a member of the poorer (on average) group is richer than a member of the richer (on average) group (Gini, 1959). It implies that the sign of a difference between the two incomes is opposite with respect to the sign of the difference between the means of the two groups. In our case, since $\mu_a > \mu_p$, a transvariation occurs whenever a member of group p is richer than a member of group a.

Now, the term $R_{ph} - r_{ph}$ represents the number of transvariations in which y_{ph} is involved so that the term $\sum_{h=1}^{n_p}(R_{ph} - r_{ph})$ is simply equal to the total number of transvariations. Therefore index I rewrites as

$$I = 1 - \frac{2N^{TR}}{n_a n_p}. \tag{26}$$

where N^{TR} is the total number of transvariations. Since $n_a n_p$ is the total number of comparisons between members of the two groups, the ratio $N^{TR}/n_a n_p$ can be interpreted as the probability that the sign of a difference between two incomes belonging to different groups is opposite with respect to the difference between the means of the two groups. In other words, this ratio corresponds to the probability of a transvariation.[8]

We now discuss the implications of expression (25) using a graphical approach where we analyze between-continent inequality in a way that immediately relates to the research of Milanovic and Yitzhaki (2002).

4. A Graphical Interpretation

Using the national income/expenditure distribution data from 111 countries, Milanovic and Yitzhaki (2002) decomposed total inequality between individuals in the world by continents and regions. In particular, they partitioned the world into five continents: Africa; Asia; Western Europe, North America, and Oceania (WENAO); Eastern Europe and the former Soviet Union (EUFSU); and Latin America and the Caribbean (LAC). Commenting on the results relating to between-continent inequality, they note that "between-continent inequality Gini is 0.309; had we used Pyatt's between-group component, we would have gotten a

[8]See Gini (1959, p. 8) on this point.

TABLE 1

Values of I for Continent-by-Continent Comparisons

Mean Income	Africa ($\mu = 1310$)	Asia ($\mu = 1594.6$)	EUFSU ($\mu = 2780.9$)	LAC ($\mu = 3639.8$)	WENAO ($\mu = 10012.4$)
Africa					
Asia	32.6%				
EUFSU	32.0%	–4.5%			
LAC	77.4%	30.0%	45.0%		
WENAO	99.0%	72.2%	96.4%	92.7%	

Notes: μ = mean income in \$PPP (1993), weighted for the size of the population.

EUFSU, Eastern Europe and former Soviet Union; LAC, Latin America and Caribbean; WENAO, Western Europe, North America, and Oceania.

Source: Authors' calculation from Milanovic and Yitzhaki (2002).

between-continent Gini of 0.398 which means that overlapping has decreased the between-continent component by about 9 Gini points" (Milanovic and Yitzhaki, 2002, p. 163).

It is interesting to verify how continent-by-continent comparisons have contributed to this result. We treat countries as units of observation and continents as groups.[9] Ordering the continents by their per capita average income in international dollars, in Table 1 we report the values of I, i.e. the values of the ratio G_b/G_B, for each pair of continents.

Recalling the discussion in Section 3, we note that Table 1 covers the whole range of possible values of I. When WENAO is compared with other continents, I reaches very high values. In particular, I is very close to unity when WENAO is contrasted with Africa (99 percent), EUFSU (96.4 percent), and LAC (92.7 percent). According to expression (12) above, in all these cases, Prob ($Y_p > Y_a$) is close to zero, so stratification dominates and there is virtually no overlapping between WENAO countries and countries belonging to other continents. Thus, in these cases, using Pyatt's (1976) decomposition and Yitzhaki's (1994) decomposition, we would obtain almost equivalent measures of between-continent inequality. The exceptions involving WENAO arise in the comparison with Asian countries. In this case, I equals 72.2 percent, so that we know from (11) that the probability of an Asian country having a mean income higher than a WENAO country is equal to 13.9 percent.

At the other extreme, I reaches a (small) negative value (i.e. –4.5 percent) when EUFSU and Asia are compared. According to expression (13), this means that, although the mean income of EUFSU is 75 percent higher than the mean income of Asia, it is more likely that an Asian country has a mean income higher than an EUFSU country than the reverse. More precisely, using (11), this probability amounts to 52.2 percent. This result is associated clearly with a high polarization within both these continents, which generates a *negative* value of I. This signals low stratification and high overlapping.

[9]Note, however, that using this level of aggregation we cannot fully explain findings by Milanovic and Yitzhaki (2002), since, in this paper, countries are presented by deciles and even smaller aggregated observations, and stratification is affected by the level of aggregation.

Finally, remaining comparisons are somewhat in between these two extremes. For example, when LAC and EUFSU are compared, the value of I is close to 50 percent, which indicates, again from (11), that the probability of an EUFSU country having a mean income higher than a LAC country is around 25 percent. According to Table 1 and expression (11), the probability of an African country having a mean income higher than either an Asian or an EUFSU country, and the probability of an Asian country having a mean income higher than a LAC country, ranges between 25 and 50 percent.

These, and more results, can be illustrated representing, for each pair of continents, the cumulative distribution of the richer continent as a function of the cumulative distribution of the poorer one, using unweighted mean income for each country involved in the comparison. In Figure 1, we present four comparisons that represent the graphical counterpart of expression (25).

In each of the four diagrams, the kinked line represents the cumulative distribution of the richer continent (function $F(y)$ using the notation of expression (25)) plotted against the cumulative distribution of the poorer continent (function $G(y)$). It follows that the coordinates of each point on the kinked line are given by: (i) the proportion of countries, which, in the poorer continent, have an average income smaller than $\bar{y}$ on the horizontal axis; and (ii) the percentage of countries, which, in the richer continent, have an average income smaller than $\bar{y}$ on the vertical axis. In each diagram, areas under the $F(y)$ curves are equal to Prob ($Y_p > Y_a$) between countries belonging to the corresponding continents, as can be verified by simple numerical computation.

The 45° line (when present) indicates the values that the cumulative distribution of the richer continent should possess to be exactly equal to the cumulative distribution of the poorer continent [$F(y) = G(y)$] *at the same average income level*. Where the slope of $F(y)$ is higher than one (the slope of the 45° line) in a given range of $G(y)$, it means that, for *each point* belonging to that region, the percentage of countries having an average income smaller than $\bar{y}$ in the richer continent is higher than the percentage of countries having an average income smaller than the same value $\bar{y}$ in the poorer continent.

We choose to present two comparisons involving EUFSU countries and two comparisons involving WENAO countries; these continents are both compared with Asian and African countries. The polarization among EUFSU countries is visible in the shape of its cumulative distribution when plotted against Asia and, to some extent, against Africa. In both these cases, at low income levels, $F(y)$ is above the 45° line since many of the absolute poorest countries belong to EUFSU (Georgia, Uzbekistan, Armenia; see Milanovic and Yitzhaki, 2002, p. 176). This means that the minimum average income level, i.e. the lower boundary of the region R, where the integral in (25) is evaluated, belongs to the richer (on average) continent, EUFSU in both cases. As higher income levels are considered, the cumulative distribution for EUFSU falls behind the 45° degree line with respect to both Africa and Asia. This signals that the probability of an EUFSU country having an average income below a given level is lower than the probability of finding an African or an Asian country with an average income below the same level. In the comparison with Asia, however, there is another region, at middle-high income levels, where $F(y)$ lies above the 45° line at the top. The latter reflects

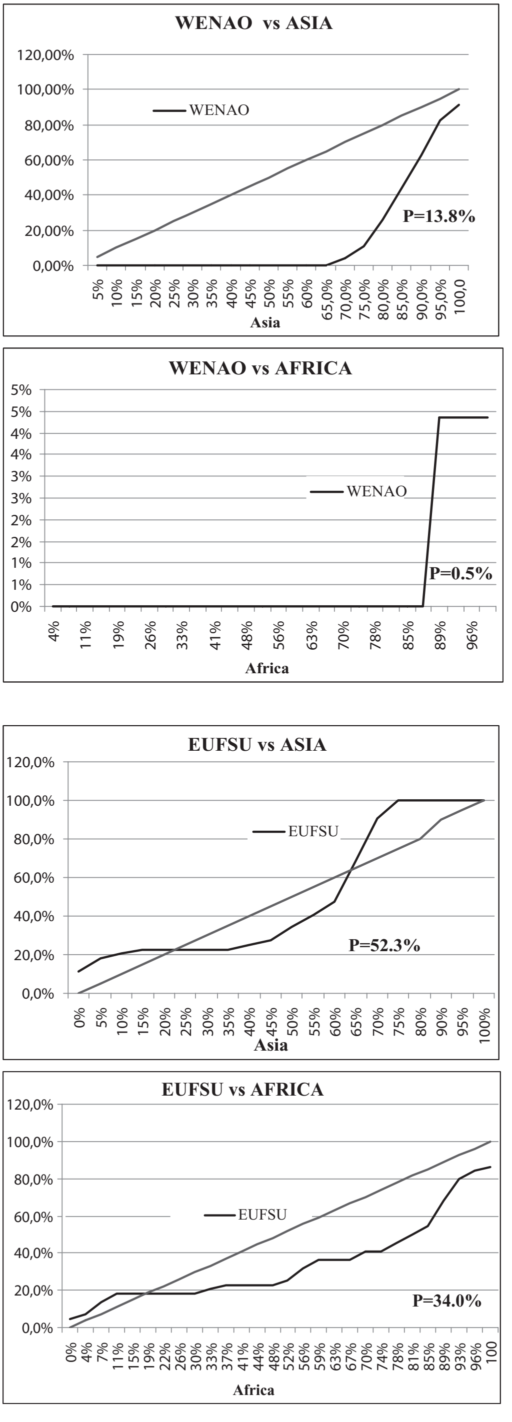

Figure 1. Continent-by-Continent Comparisons

polarization of income across Asian countries, namely the presence of high-income countries, such as Singapore, Taiwan, Korea, Japan, and Hong Kong.

Diagrams involving WENAO are much more conventional. Since the poorest WENAO country (Turkey) is always far richer than the poorest African or Asian country, $F(y)$ (in these cases) lies on the horizontal axes for a large interval of values of $G(y)$. More precisely, there is no WENAO country with a mean income lower than an Asian country until the seventh decile of the Asian distribution. At higher mean income levels, some Asian countries are richer than WENAO countries, but the kinked line never crosses the 45° line. Overall the probability that an Asian country is richer than a WENAO country amounts to 13.8 percent, again, a value that can be approximated by calculating the area under the kinked curve. The comparison between WENAO and African countries is much more dramatic, since the stratification, as indicated by the value of I at 99 percent in Table 1, is almost absolute. The probability of an African country being richer than a WENAO country is only marginally different from zero; thus the 45° degree line cannot be represented and $F(y)$ lies almost everywhere on the horizontal axis.

The entire discussion above can be reformulated in terms of transvariations. Indeed, when a WENAO country is considered, the probability of a transvariation is generally negligible or very low, the exception being the possibility that an Asian country is richer. On the other hand, when EUFSU and Asia are compared, the probability of a transvariation is high, again because of the existence of very rich Asian countries. Finally, in the intermediate cases such as LAC vs EUFSU, the value of I is close to 50 percent since the probability of a transvariation is about ¼.

Also, in the figures above, the slope of $F(y)$ becoming higher than one signals that in that region transvariations are being originated, because, at the same levels of average income, there are countries belonging to the richer continent whose incomes are lower than at least one of the countries of the poorer continent.

5. Concluding Remarks

The traditional literature on stratification measurement (Yitzhaki and Lerman, 1991; Yitzhaki, 1994) tends to see stratification as inversely related to inequality. This view derives from the fact that higher stratification, i.e. lower overlapping, is usually associated with lower within-group inequality. The most recent literature, however, focuses on the impact of stratification on between-group inequality and proposes a measure to evaluate it (Milanovic and Yitzhaki, 2002; Monti and Santoro, 2009). This measure is such that, *ceteris paribus*, a higher stratification is associated with higher between-group inequality.

In this paper, we interpret this measure as a function of the probability that a random member of the poorer (on average) group is richer than a random member drawn from the (on average) richer group. We show that when two groups are considered, this approach leads to rewriting such a measure as one minus twice the area under the cumulative distribution function of the richer group, expressed as a function of the cumulative distribution function of the poorer group. This formula is, to some extent, similar to the expression of the Gini index and naturally suggests the graphical illustration that we provide to analyze between-continent inequality. The major advantage of our approach is that a lot of information about the impact of stratification on between-group inequality can be obtained by a simple graphical inspection of the plot of the cumulative distribution function of the group with a higher mean income against the cumulative distribution function of the group with a lower mean income. Moreover, such an inspection immediately suggests where the probability of a transvariation as defined by Gini in 1916 (Gini, 1959) is increasing.

What rationale can be provided for this interpretation of the impact of stratification on between-group inequality? We think an answer to this question can be provided by the concept of *group deprivation*; it applies when group members share a strong body of common moral, social, and cultural values. By *group deprivation* we mean the feeling of deprivation that a group has whenever the income of any of its members is lower than the income of any member of the other group. In such a case, every member of a group feels empathy for any other member of their own group, the group as a whole being affected by the probability that any of its members is richer than any of the member of the other group. Group deprivation, thus increases in this probability and this drives the impact of stratification on between-group inequality.

To provide an example, we refer back to the analysis of between-continent inequality and consider the viewpoint of a representative individual of an African country. By representative individual, we mean an individual whose income is exactly equal to the mean income of their country. Suppose this individual feels he/she belongs to the African continent, not only to his/her own country. When comparing Africa with any other continent, this individual would therefore care about the possibility that any representative African is richer than the representative individual of a Western or Asian country. This possibility corresponds to the probability that any representative African is richer than a representative individual of another (richer) continent. The higher this probability, i.e. the probability of a transvariation, the lower the feeling of group deprivation and between-group inequality.

Appendix

In this Appendix we show how expression (9) is obtained. We start with notation.

We define Y-ordering as the ordering of income units by their income level, where:

R_{ih} is the rank of the income y_{ih} in the overall population, $R_{ih} = 1 \ldots n$;

$F_O(y)$ is the overall cumulative distribution of Y, empirically estimated by R_{ih}/n;

$\bar{R}_i$ is the average rank of the group i in the overall population, or $\bar{R}_i = 1/n_i \sum_{k=1}^{n_i} R_{ik}$;

$\bar{F}_{Oi}$ is the mean of the cumulative function values for the group i, estimated by $\bar{R}_i/n$.

Also, we define g-ordering (group-ordering) as the order in which the groups are lined up following the non-decreasing ordering of their means and, within the groups, incomes are ordered by their non-decreasing order. In this ordering:

r_{ih} is the rank of the income y_{ih} in its own group;

$\bar{r}_i$ is the average rank of group i, or $\bar{r}_i = 1/n_i \sum_{h=1}^{n_i} r_{ih}$;

$F_O^g(y)$ is the overall cumulative distribution, empirically estimated by r_{ih}/n $[i = a, p;\ h = 1 \ldots n_p, \ldots, (n_p + 1), \ldots (n_p + j) \ldots n]$;

$\bar{F}_{Oi}^g$ is the mean of the cumulative distribution values for the group i, estimated by $\bar{r}_i/n$.

The proof of (9) is equivalent to the proof of the following equivalence:

$$G_B = G_b + 2(\mu_a - \mu_p)\sum_{h=1}^{n_p}(R_{ph} - r_{ph})/n^2\mu \tag{A1}$$

where μ_a, μ_p, μ, are the mean income of the group a, of the group p and of the overall population respectively. $G_B = 2\operatorname{cov}(\mu_i, \bar{F}_{Oi}^g)/\mu$ is the between component of the conventional Gini decomposition, $G_b = 2\operatorname{cov}(\mu_i, \bar{F}_{Oi})/\mu$ is the between component of the Yitzhaki (1994) Gini index decomposition, and $\sum_{k=1}^{n_p}(R_{ph} - r_{ph})$ is the number of instances where a member of group p has an income higher than a member of group a.

From the definition of covariance one has

$$\operatorname{cov}(\mu_i, \bar{F}_{Oi}^g) = \frac{n_p}{n}\mu_p\left(\bar{F}_{Op}^g - \frac{n+1}{2n}\right) + \frac{n_a}{n}\mu_a\left(\bar{F}_{Oa}^g - \frac{n+1}{2n}\right).$$

Using ranks to estimate the cumulative distribution function we write

$$\begin{aligned}
\operatorname{cov}(\mu_i, \bar{F}_{Oi}^g) &= \frac{n_p}{n}\mu_p\left(\sum_{h=1}^{n_p}\frac{r_{ph}}{n}\frac{1}{n_p} - \frac{n+1}{2n}\right) + \frac{n_a}{n}\mu_a\left(\sum_{k=n_p+1}^{n}\frac{r_{ak}}{n}\frac{1}{n_a} - \frac{n+1}{2n}\right) \\
&= \frac{1}{n}\left\{\mu_p\left(\frac{n_p(n_p+1)}{2n} - \frac{n_p}{n}\frac{n+1}{2}\right)\right. \\
&\quad \left. + \mu_a\left[\left(\frac{n(n+1)}{2n} - \frac{n_p(n_p+1)}{2n}\right) - \frac{n_a}{n}\frac{n+1}{2}\right]\right\} \\
&= \frac{1}{n}\frac{n_a n_p}{2n}(\mu_a - \mu_p).
\end{aligned}$$

Thus we can write

$$G_B = \frac{n_p n_a(\mu_a - \mu_p)}{n^2\mu} = 2\operatorname{cov}(\mu_i, \bar{F}_{Oi}^g)/\mu. \tag{A2}$$

That is, the between component in the conventional Gini decomposition, G_B, is twice the covariance between the group's average income and the average rank of the group divided by the overall mean. From the definition of covariance one has

$$\operatorname{cov}(\mu_i, \bar{F}_{Oi}^g) = \frac{n_p}{n}\mu_p\left(\bar{F}_{Op}^g - \frac{n+1}{2n}\right) + \frac{n_a}{n}\mu_a\left(\bar{F}_{Oa}^g - \frac{n+1}{2n}\right). \tag{A3}$$

Then using the rank to estimate the cumulative distribution, we obtain

$$\text{(A4)}\quad \operatorname{cov}\left(\mu_i, \bar{F}_{Oi}^{g}\right) = \frac{1}{n}\left[\frac{n_p}{n}\mu_p\left(\sum\nolimits_{h=1}^{n_p}\frac{r_{ph}}{n_p} - \frac{n+1}{2}\right) + \frac{n_a}{n}\mu_a\left(\sum\nolimits_{k=n_p+1}^{n}\frac{r_{ak}}{n_a} - \frac{n+1}{2}\right)\right].$$

Adding and subtracting R_{ik} in the sums of (A4), expression (A3) rewrites:

$$\text{(A5)}\quad \operatorname{cov}\left(\mu_i, \bar{F}_{Oi}^{g}\right) = \frac{1}{n}\left[\frac{n_p}{n}\mu_p\left(\sum\nolimits_{h=1}^{n_p}\frac{r_{ph} - R_{ph} + R_{ph}}{n_p} - \frac{n+1}{2}\right) + \frac{n_a}{n}\mu_a\left(\sum\nolimits_{k=n_c+1}^{n}\frac{r_{ak} - R_{ak} + R_{ak}}{n_a} - \frac{n+1}{2}\right)\right],$$

$$\text{(A6)}\quad \operatorname{cov}\left(\mu_i, \bar{F}_{Oi}^{g}\right) = \frac{1}{n}\left\{\left[\frac{n_p}{n}\mu_p\left(\bar{R}_p - \frac{n+1}{2}\right) + \frac{n_a}{n}\mu_a\left(\bar{R}_a - \frac{n+1}{2}\right)\right] - \frac{\mu_p}{n}\sum\nolimits_{h=1}^{n_p}\left(R_{ph} - r_{ph}\right) + \frac{\mu_a}{n}\sum\nolimits_{k=n_p+1}^{n}\left(r_{ak} - R_{ak}\right)\right\}.$$

We observe that the difference between the rank of y_{pk} in Y-ordering (R_{pk}) and the rank of y_{pk} in g-ordering (r_{pk}) represents the number of incomes belonging to group *a* less than y_{pk}. Analogously, looking at the comparisons between incomes (belonging to different groups) from the perspective of group *a*, the difference $r_{ak} - R_{ak}$ represents the number of elements of group *p* greater than y_{ak}. The sum $\sum\nolimits_{h=1}^{n_p}\left(R_{ph} - r_{ph}\right)$ is the number of times that an income of group *p* is greater than an income of group *a*, and the sum $\sum\nolimits_{k=n_c+1}^{n}\left(r_{ak} - R_{ak}\right)$ is the number of times that an income of group *a* is lesser than an income of group *p*, so that

$$\sum\nolimits_{k=n_p+1}^{n}\left(r_{ak} - R_{ak}\right) = \sum\nolimits_{h=1}^{n_p}\left(R_{ph} - r_{ph}\right).$$

Moreover, from Yitzhaki and Lerman (1991, p. 321) we have

$$\frac{1}{n}\left[\mu_p\frac{n_p}{n}\left(\bar{R}_p - \frac{n+1}{2}\right) + \mu_a\frac{n_a}{n}\left(\bar{R}_a - \frac{n+1}{2}\right)\right] = \operatorname{cov}\left(\mu_i, \bar{F}_{Oi}\right).$$

Then we can write:

$$2\operatorname{cov}\left(\mu_i, \bar{F}_{Oi}^{g}\right)/\mu = 2\operatorname{cov}\left(\mu_i, \bar{F}_{Oi}\right)/\mu + 2\left(\mu_a - \mu_p\right)\sum\nolimits_{h=1}^{n_p}\left(R_{ph} - r_{ph}\right)\Big/n^2\mu$$

or equivalently

$$2\operatorname{cov}\left(\mu_i, \bar{F}_{Oi}^{g}\right)/\mu = 2\operatorname{cov}\left(\mu_i, \bar{F}_{Oi}^{G}\right)/\mu + 2\left(\mu_a - \mu_p\right)\sum\nolimits_{k=n_p+1}^{n}\left(r_{ak} - R_{ak}\right)\Big/n^2\mu$$

which proves (A1) and thus (9).

References

Frick, J. R., J. Goebel, E. Schechtman, G. G. Wagner, and S. Yitzhaki, "Using Analysis of Gini (ANOGI) for Detecting Whether Two Subsamples Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience," *Sociological Methods and Research*, 34, 427–68, 2006.

Gini, C., "Il Concetto di Transvariazione e le Sue Prime Applicazioni," 1916, reproduced in *Corrado Gini Memorie di Metodologia Statistica, Volume II: Transvariazione*, Università degli Studi di Roma, 1–55, 1959.

Giorgi, G. M., "Gini's Scientific Work: An Evergreen," *Metron—International Journal of Statistics*, 63, 299–315, 2005.

Lasswell, T. E., *Class and Stratum*, Houghton Mifflin, Boston, MA, 1965.

Milanovic, B. and S. Yitzhaki, "Decomposing World Income Distribution: Does the World have a Middle Class?" *Review of Income and Wealth*, 48, 155–78, 2002.

Monti, M. and A. Santoro, "A Note on Between-Group Inequality with an Application to Households," *Journal of Income Distribution*, 18(3–4), 34–48, 2009.

Pyatt, G., "On the Interpretation and Disaggregation of Gini Coefficient," *Economic Journal*, 86, 243–55, 1976.

Yitzhaki, S., "Economic Distance and Overlapping of Distributions," *Journal of Econometrics*, 61, 147–59, 1994.

Yitzhaki, S. and R. Lerman, "Income Stratification and Income Inequality," *Review of Income and Wealth*, 37, 313–29, 1991.