

**A NOTE ON THE MEASUREMENT OF STRATIFICATION AND
BETWEEN-GROUP INEQUALITY**

M. MONTI e A. SANTORO

JEL Classification: D31

Keywords : Between-group inequality, Gini index decomposition, Stratification, Transvariations

A note on the measurement of stratification and between-group inequality¹.

M. Monti and A. Santoro

Abstract

This note belongs to the recent literature that emphasizes that the residual, in the traditional Gini index decomposition, reveals more than what is commonly believed. We show that the residual reveals the so far unexplored contribution of stratification to between-group inequality and this suggests that the between inequality measure should be modified accordingly. We propose a measure of between-group inequality which is a function of stratification. This is our first result. Moreover, we show that this measure is numerically equivalent to the one suggested by Yitzhaki and Lerman (1991) yet, contrary to the latter, we are able to define precisely its range of variation. This is our second result.

JEL Classification Code: D31

Keywords: between-group inequality, Gini index decomposition, stratification, transvariations.

¹ We thank the participants to the Conference on Income Distribution and the Family held in Kiel, 1-3 September 2008 and particularly Ray Rees and Gary Burtless for their comments on a previous version of this paper. We also thank Conchita d'Ambrosio, Bruno Bosco, the Editors of this special issue and two anonymous referees for their useful suggestions. Usual disclaimers apply. Financial support from Italian Ministry of Education (PRIN) is gratefully acknowledged.

1. Introduction

For a long time, the literature has looked at the residual of the traditional Gini index decomposition as obscure and perversely responding to changes in subgroup characteristics (Mookherjee and Shorrocks, 1982, p. 891).

This paper belongs to the more recent literature which emphasizes that the residual reveals more than what is commonly believed (see Lambert and Decoster, 2005). We show that the residual reveals the so far unexplored contribution of stratification to between-group inequality and this suggests that the between inequality measure should be modified accordingly. We propose a measure of between-group inequality which is a function of stratification.

Moreover, we show that this measure is numerically equivalent to the one suggested by Yitzhaki and Lerman (1991) yet, as opposed to the latter, we are able to define its range of variation precisely.

Yitzhaki and Lerman (1991) define stratification as “a group’s isolation from members of other groups” (Yitzhaki and Lerman, 1991, p. 319) and, by changing the formulation of both the within-group and the between-group components, obtain a new expression for the residual which reflects the impact of stratification. Following this first contribution, Yitzhaki (1994) is able to proceed further by splitting the Gini index into two parts: the first is a measure of between-group inequality and the second, reflecting the impact of stratification, is the sum of the products of income shares, Gini indices and overlaps for all groups. This decomposition is used by Yitzhaki and Milanovic (2002) to analyse world income inequality. In this literature (Yitzhaki and Lerman, 1991; Yitzhaki, 1994; Wodon, 1999; Yitzhaki and Milanovic, 2002, among others), between-group inequality is always measured separately from stratification, so that these two phenomena look unrelated. This is an unsatisfactory point of view, since, as we will show in the paper, the Yitzhaki between-group inequality depends upon stratification.

By analysing and decomposing the residual of the Gini index decomposition, we propose a between-group inequality measure that is the product of three components, one of which is a measure of stratification, so that we are able to provide a simple and intuitive measure of the impact of stratification on between-group inequality. This result is obtained by means of an extensive use of transvariation (Gini, 1959), a statistical concept which has been almost neglected by the literature (main exceptions are Deutsch and Silber, 1997 and Dagum, 1997) notwithstanding its strict links with stratification (see Yitzhaki and Lerman, 1991, p. 315).

Given two groups, 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. It follows that a group can be defined a perfect stratum when there are no transvariations involving the members of the groups (see Wodon, 1999, p. 23). Using transvariations we obtain an innovative decomposition of the Gini index, where both the residual (\bar{R}) and the between component (\hat{G}_b) are represented by new expressions, while the measure of within inequality is unchanged. Our between-group inequality component (\hat{G}_b) depends i) positively on the difference between the means of the groups and ii) negatively on the number of transvariations between the groups. It can be said that the between component “loses its innocence”² and its simplicity, changing its formulation. This is true, but it is also true that this simplicity comes at a cost. In the traditional between inequality component the mean values are taken as representative of the whole distributions. Then, the difference between the means is substituted for each difference between the values belonging to different distributions. Now, as recently observed (Giorgi, 2005), the theory of transvariation was introduced by Gini precisely for “measuring the typicality of the difference between the means of the groups”. To be more specific, Gini (1959, p. 47), argues that the comparison between the means “cannot summarize the comparison between the single values of the character (e.g. income) since such a comparison is much more complex”. He introduces the transvariation concept to face this problem and to take into account this complexity.

The use of transvariations in our between-group inequality measure needs further investigation. Considering the non-decreasing overall income ordering, the idea is that households of the poorer (on average) group experience *as a group* less inequality when any member of the group improves its position with respect to any member of the richer (on average) group. This may be associated with the idea of a group as a league, i.e. as an aggregation of individuals or households who have a strong group identity. When considering households, this league-feeling may be associated with cultural and sociological factors. For example, in Italy, and possibly in other Roman Catholic-based cultures, families with many children feel they share common religious and ethical values and they often act as a lobby in society (for example, calling for preferential tax treatment). Using transvariations, we are able to capture this league-feeling since in our measure of between inequality the inequality felt by the poorer group depends (negatively) on the average rank (in the overall distribution) of its members.

As said above, we obtain our between inequality measure by expressing the residual of the Gini decomposition in a new way. The meaning of \bar{R} can be interpreted by exploiting our second result,

² We thank an anonymous referee for this remark.

that is the numerical equivalence of our between-group measure and the one proposed by Yitzhaki and his coauthors (Yitzhaki, 1994; Yitzhaki and Milanovic, 2002). We can write $G(y) = G_W + \hat{G}_B + \hat{R}$, where $G(y)$ is the Gini index, G_W is the conventional measure of within-group inequality, \hat{G}_B is our measure of between-group inequality, numerically equivalent to the Yitzhaki between inequality measure, and \hat{R} is the residual of our decomposition. In Yitzhaki and Milanovic (2002, p.159), the sum $G_W + \hat{R}$ is defined as “the sum of income shares, Ginis and overlaps for all groups” and its value is interpreted accordingly. Actually, we show that \hat{R} has a meaning *per se*, since considering two groups, it is the sum of two terms representing the overlapping of the first group with the second and vice-versa. Given the relation linking overlapping and stratification, following Frick et al. (2006), this term represents the extent to which the different groups are stratified.

The paper is organized as follows. In Section 2, we introduce the concept of transvariation and we derive a decomposition of the Gini index where a new measure of between-group inequality is proposed. Section 3 is devoted to the interpretation and discussion of this measure. In Section 4 we show that the between-group inequality measure is numerically equivalent to the one proposed by Yitzhaki and Lerman (1991) and we decompose it in a way which makes explicit the contribution of stratification to between-group inequality. In the same section, we provide an interpretation of the residual \hat{R} . Finally, Section 5 draws some conclusions.

2. Using transvariations to decompose the Gini index

We consider a population of n households and, for the sake of simplicity, we confine our attention to the case of two population groups, which we will call “ a ” and “ c ” with $i = a, c$. The two groups stand for a given socioeconomic partition of the population on the basis of the households’ characteristics. The population size is $n_a + n_c = n$, with $n, n_a, n_c \in \mathbb{N}$, where n_a is the number of households belonging to group a , $j = 1, \dots, n_a$, and n_c is the number of households belonging to group c , $l = 1, \dots, n_c$. We denote by y_{ih} , $\{y_{ih}\} \in \mathbb{R}^+$, the equivalent income of the household h belonging to group i and by μ_p , μ_c and μ_a the overall, the group c and the group a average equivalent income, respectively.

Supposing $\mu_a > \mu_c$, the standard decomposition of the Gini index is the following

$$G(y) = G_W + G_B + R, \quad (1)$$

where $\{y\}$ is the set of all income units,

$$G_W = \frac{1}{n^2 \mu_p} (G_a n_a^2 \mu_a + G_c n_c^2 \mu_c) \text{ and } G_B = \frac{1}{n^2 \mu_p} (n_a n_c (\mu_a - \mu_c)). \quad (2)$$

In (2), the terms G_a and G_c denote the group a and group c 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 distributions (Lambert and Aronson, 1993).

We now introduce the concept of transvariation. 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). In our case, a transvariation occurs whenever a household of group c is richer than a household of group a . It can immediately be noted that, when no transvariation occurs, the two groups do not overlap at all, i.e. in our context, the first n_a richest households in the overall distribution are the n_a households belonging to group a . Equivalently, one can say that, when there is no transvariation between the two groups, the groups are perfect strata.

Monti (2007, p. 8) shows that, when only two groups are considered, the residual term R may be written as

$$R = 2 \frac{\sum_{\{j,l: y_{cl} > y_{aj}\}} (y_{cl} - y_{aj})}{n^2 \mu_p}. \quad (3)$$

In (3), the term $\sum_{\{j,l: y_{cl} > y_{aj}\}} (y_{cl} - y_{aj})$ is the sum of the transvariation values (intensity of transvariation), that is the sum of the income differences $y_{cl} - y_{aj}$ for all pairs of household incomes such that the income of the household of group c is greater than the income of the household of group a , as the average income of households of group c is lower than the average income of households of group a .

Let us now introduce the following additional notation. N^{TR} is the total number of transvariations, n_{cl}^{TR} is the number of transvariations involving y_{cl} and n_{aj}^{TR} is the number of transvariations involving y_{aj} . Observing that $N^{TR} = \sum_l n_{cl}^{TR} = \sum_j n_{aj}^{TR}$, it can be shown³ that

$$\sum_{\{j,l: y_{cl} > y_{aj}\}} (y_{cl} - y_{aj}) = -N^{TR} (\mu_a - \mu_c) + \sum_l n_{cl}^{TR} (y_{cl} - \mu_c) + \sum_j n_{aj}^{TR} (\mu_a - y_{aj}), \quad (4)$$

so that we can state our first result as follows

$$G(y) = G_W + \hat{G}_B + \bar{R}, \quad (5)$$

³ For this and following results of the paper proofs are available upon request.

$$\hat{G}_B = \frac{1}{n^2 \mu_p} \left[(n_a n_c - 2N^{TR}) (\mu_a - \mu_c) \right], \quad (6)$$

$$\hat{R} = \frac{2}{n^2 \mu_p} \left[\sum_l n_{cl}^{TR} (y_{cl} - \mu_c) + \sum_j n_{aj}^{TR} (\mu_a - y_{aj}) \right]. \quad (7)$$

3. Discussion

There are two main differences between the standard Gini decomposition, i.e. equation (1), and our decomposition, i.e. equation (5): the between inequality measure (\hat{G}_B) and the residual measure (\hat{R}).

In (5), the measure of between-group inequality \hat{G}_B is the product of G_B and of a factor which takes into account transvariations. To give a sound interpretation of the term \hat{G}_B we briefly recall some of Gini's remarks (Gini, 1959). It is well-known that, in his inequality index, Gini measures inequality by the mean difference, that is he considers the difference between two income values as the measure of the inequality between those two incomes. Then, taking two groups as a whole and measuring their inequality by the difference between their means, implies that we are assuming the group mean values as representative of the whole group distributions. Now, up to which point is this assumption acceptable? Gini answers that the reality is so complex that "neither the difference between the means, nor the variability of the two distributions and not even their form may answer this question which depends or may depend on all these elements" (Gini, 1959, p. 47, authors' translation). He introduces the transvariation concept to evaluate the attitude of the difference between the means to represent the comparison between the distributions (see Giorgi, 2005 and Gini, 1959).

Let us consider an example to illustrate why the difference between the means may be misleading, Consider two groups of households, say a group of households with children and a group of only-adult households and suppose that the incomes of only-adult households are {4, 8, 18} while the incomes of the households with children are {2, 6, 13}. The traditional between component would be based on the difference between the means, which is equal to 3 (10-7) in this case. Now, compare this situation with a different one where the incomes of the households with children are unchanged, while the incomes of only-adult households are {4, 5, 21}. The inequality within only-adult households is surely increased, but what happens to between-group inequality? The traditional measure G_B is unchanged, while, on the contrary, our measure \hat{G}_B is reduced since the number of

transvariations increases (from 3 to 4).

The number of transvariations is the number of pairs of incomes belonging to different groups for which inequality behaves differently with respect to the inequality represented by the difference between the means of their respective groups. Then, following Gini's suggestion, we weigh the inequality between two groups (expressed by the differences of their means) by a factor (the number of transvariations) representing the attitude of the comparison between the means of the groups to represent the comparisons between the members of the groups. Given the difference between the two groups' mean incomes, between-group inequality as measured in (6) is smaller than the standard one when there are transvariations between the two groups.⁴ Considering the non-decreasing overall income ordering, the idea is that households of group *c as a group* experience less inequality when any of its members improves her position with respect to any member of the other group. Referring to the example, households with children *as a group* experience less inequality in second case since the second poorest household with children is richer than the second poorest only-adult household.

4. Relating between-group inequality to stratification

The stratification literature (Yitzhaki and Lerman, 1991; Yitzhaki, 1994; Yitzhaki and Milanovic, 2002) decomposes the Gini index differently from (1). Following Yitzhaki (1994) and Yitzhaki and Milanovic (2002), when only two groups are considered, the Gini index decomposes as

$$G(y) = s_c G_c O_c + s_a G_a O_a + G_b, \quad (8)$$

where

$$O_i = p_i + \sum_{i \neq j} p_j O_{ji}, \quad (9)$$

$$O_{ji} = \text{cov}_i(y, F_j(y)) / \text{cov}_i(y, F_i(y)). \quad (10)$$

In (8), s_i is the share of total income owned by group i ($s_i = p_i \mu_i / \mu_p$, $p_i = n_i / n$), O_i is the total overlapping of group i and G_b is the between inequality as defined in Yitzhaki and Lerman (1991). In (9), O_{ji} is the overlapping of group j by group i . O_{ji} is defined in (10) as 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.

Yitzhaki and Lerman (1991) define the between inequality as

⁴ Note that \hat{G}_B is negative if $2N^{TR} > n_a n_c$. We consider the sign of \hat{G}_B in the next section.

$G_b = 2 \text{cov}(\mu_i, \bar{F}_{O_i}^G) / \mu_p$. Thus, G_b is twice the covariance between each groups' average income and groups' average rank in the overall population divided by the overall mean income. That is, each group is represented by its mean income and by the mean rank of its members in the overall distribution. 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.⁵ This implies that, using (8), there is no easy way to relate stratification to between-group inequality, since the latter is apparently separate.

However, it can be shown that

$$\hat{G}_B = G_b, \quad (11)$$

i.e. the between inequality measure that we have derived above, see expression (6), is numerically equivalent to the between inequality measure proposed by Yitzhaki and Lerman (1991). Expression (11) justifies our claim that \hat{G}_B is a measure of between-group inequality, since it is equivalent to a measure which is widely used in the literature.

Rewriting \hat{G}_B as in (12), we show that our measure is a function of the stratification impact on between-group inequality measurement

$$\hat{G}_B = \frac{(\mu_a - \mu_c)}{\mu_p} \cdot \frac{n_a n_c}{n^2} \cdot I, \quad (12)$$

where

$$I = 1 - \frac{2N^{TR}}{n_a n_c}. \quad (13)$$

In expression (12) we recognize three parts of \hat{G}_B . The first is the amount of income that should be redistributed to achieve complete between-group equality and it is equal to $(\mu_a - \mu_c) / \mu_p$. The second is the ratio $n_a n_c / n^2$ which is a weight depending on the composition of the population.⁶ The third is the index I that we propose as the measure of the stratification impact on between-group inequality. This index is a decreasing function of the number of transvariations and it assumes its maximum, $I=1$, when stratification is absolute, i.e. when there are no transvariations ($N^{TR} = 0$). More precisely, the index I varies in the interval $[\beta, 1]$, where

$$\beta = \left(1 - \frac{2MaxN^{TR}}{n_a n_c} \right). \quad (14)$$

In (14), $MaxN^{TR}$ is the maximum number of transvariations between the two groups under the

⁵ In Yitzhaki (1994), differently from Yitzhaki and Lerman (1991), O_i includes overlapping of group i with itself.

⁶ If $n_a = n_c$ the weight attains its maximum value, i.e. $1/4$.

assumption $\mu_a > \mu_c$. If we denote as q_a the number of the member of the group a whose income is higher than μ_a and as p_c the number of the group c whose income is (weakly) lower than μ_c , it can be shown that $Max N^{TR} = n_a n_c - q_a p_c$. Therefore we have

$$\beta = \left(\frac{2q_a p_c}{n_c n_a} - 1 \right), \quad (15)$$

where the ratio $q_a p_c / n_a n_c$ depends on the skewness of the two distributions and it holds that $\beta = -1/2$ when both distributions are symmetric with respect to their mean.

Expression (12) sheds some light on the range of \hat{G}_B . As noted by Yitzhaki and Lerman (1991, p. 322) with reference to their measure, the maximum of \hat{G}_B is G_B and $\hat{G}_B = G_B$ when there is no stratification, i.e. in our context when $N^{TR} = 0$. On the other hand, Yitzhaki and Lerman (1991) note that \hat{G}_B can be negative in some extreme cases. More generally, we have shown that the possibility of a negative \hat{G}_B depends on the number of transvariations and, consequently, it depends on the importance of stratification as measured by I . More precisely, $\hat{G}_B \leq 0$ if and only if

$$N^{TR} \geq n_c n_a / 2 \text{ and it is easy to see that } \min \hat{G}_B = -G_B \left(1 - 2 \frac{q_a p_c}{n_a n_c} \right).$$

Expression (11) allows us to interpret the term \hat{R} better.

Considering the overlapping of the two groups with themselves one obtains the conventional expression for within-group inequality

$$G_W = s_c p_c G_c + s_a p_a G_a. \quad (16)$$

Then, substituting expression (11) in expression (8) and, taking into account (16), we have

$$G(y) = G_W + (s_c p_a G_c O_{ac} + s_a p_c G_a O_{ca}) + G_b. \quad (17)$$

Let us compare the Gini decomposition (17) with our decomposition (5). In the two decompositions the within components are the same, the between components have the same value, then it holds

$$\hat{R} = s_c p_a G_c O_{ac} + s_a p_c G_a O_{ca}. \quad (18)$$

Thus, the term \hat{R} represents the amount of overlapping between the two groups as measured by Yitzhaki (1994) and Yitzhaki and Milanovic (2002). Given that overlapping can be interpreted as the inverse of stratification (Yitzhaki and Milanovic, 2002, p. 160), \hat{R} represents the extent to which the different groups are stratified.

approach, is now less than 9% (against 16% in Table 2).

5. Concluding remarks

Between-group inequality measurement has recently received renewed attention (Elbers et al., 2008). The reason probably lies in the fact that many policies, rather than being purported to decrease overall inequality, are restrained to reduce between-group inequality while ensuring that overall inequality does not increase. This appears to be particularly true when policies aiming at reducing inequality among households are considered.

In this paper, we propose to measure between-group inequality using \hat{G}_B which depends i) positively on the difference between the means of the groups and ii) negatively on the number of transvariations between the groups. Given two groups, 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. Transvariations are useful for two reasons. First, they provide an easy way to take into account the complexity of the comparison between the single values of the character (e.g. income), an issue raised by Gini (Gini, 1959; Giorgi, 2005). Second, they allow to obtain an index measuring the impact of stratification on between-group inequality. This impact is relevant whenever a group is conceived as a league as suggested by Yitzhaki and Lerman (1991) who propose a measure of between-group inequality to which \hat{G}_B is numerically equivalent. The league-feeling is such that the households of the poorer (on average) group experience *as a group* less inequality when any member of the group improves its position with respect to any member of the richer (on average) group.

Maria Monti, DSGE, Università degli studi di Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milano (Italy) and DEAS, Università degli studi di Milano, maria.monti@unimi.it

Alessandro Santoro, DSGE, Università degli studi di Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milano (Italy) and Econpubblica (Università Bocconi), Alessandro.santoro@unimib.it (corresponding author).

References

- Dagum, C., 1997, A New Approach to the Decomposition of the Gini Income Inequality Ratio, *Empirical Economics*, 22, 551-531.
- Del Boca, D., and S. Pasqua, 2003, Employment Patterns of Husbands and Wives and Family Income Distribution in Italy, *Review of Income and Wealth*, 49, 2, 221-245.
- Deutsch, J. and J. Silber, 1997, Gini's "Transvariazione" and the Measurement of Distance Between Distributions, *Empirical Economics*, 22, 547-554.
- Elbers, C., Lanjouw, P., Mistiaen, J.A. and B. Ölzer, 2008, Reinterpreting between-group

- inequality, *Journal of Economic Inequality*, 6, 3, 231-245.
- Frick, J., Goebel, J., Schechtman, E., Wagner, G. and S Yitzhaki, 2006, Using analysis of Gini for detecting whether two sub-samples represent the same universe: the German socio-economic panel study experience, *Sociological Methods & Research*, 34, 427–468.
- Gini, C., 1959, Il concetto di transvariazione e le sue prime applicazioni. (1916). Reproduced by Corrado Gini Memorie di metodologia statistica. Volume secondo: Transvariazione, pp 1-55, Università degli Studi di Roma (in Italian).
- Giorgi, G.M., 2005, Gini's scientific work: an evergreen, *Metron -International Journal of Statistics*, 63, 3, 299-315.
- Lambert, P. and J.R. Aronson, 1993, Inequality decomposition Analysis and the Gini Coefficient Revisited, *The Economic Journal*, 103, 1221-1227.
- Lambert, P. and A. Decoster, 2005, The Gini Coefficient reveals more, *Metron*, 63, 3, 373-400.
- Martin, Molly A., 2006, Family Structure and Income Inequality in Families with Children , 1976 to 2000, *Demography*, 43, 3, 421-445.
- Michellini, C., 2002, Equivalence Scales and Consumption Inequality: A Study of Household Consumption Patterns in Italy, *DEAS, Università degli studi di Milano, Working Paper*. 2002-99.
- Monti, M., 2007, On the Dagum decomposition of the Gini inequality index, *DEAS, Università degli studi di Milano, Working Paper*. 2007-16.
- Monti, M and A. Santoro, 2007, The Gini Decomposition: an Alternative Formulation with an Application to Tax Reform, paper presented at the Second Conference of the Society for the Study of Economic Inequality (ECINEQ), available at http://www.diw.de/documents/dokumentenarchiv/17/60106/papermonti_santoro.pdf.
- Mookherjee, D. and A. Shorrocks. 1982. "A Decomposition Analysis of the Trend in UK Income Inequality", *The Economic Journal* 92: 886-902.
- Yitzhaki, S. and R. Lerman, 1991, Income stratification and income inequality, *Review of income and wealth*, 37, 313-329.
- Yitzhaki, S., 1994, Economic distance and overlapping of distributions, *Journal of Econometrics*, 61, 147-159.
- Yitzhaki, S. and B. Milanovic, 2002, Decomposing World Income Distribution: Does the World Have a Middle Class?, *Review of Income and Wealth*, 48, 2, 155-178.
- Wodon, Q., 1999, Between-group inequality and targeted transfers, *Review of Income and Wealth*, 45, 1, 21-39.