The sensitivity of the Gini to changes in group sizes and mean incomes: an extension of ANOGI applied to Brazil

Pedro Mendes Loureiro
University of Cambridge, Cambridge, UK

Received 1 September 2021
Revised 22 April 2022
Accepted 19 May 2022

Abstract

Purpose – The purpose of this paper is to first develop indicators for how total inequality, measured through the ANalysis Of GIni (ANOGI) framework, is mapped onto each group – i.e. indicators for each group's share of total inequality. Second, to develop indicators for the sensitivity of total inequality and its structure to changes in the composition of the population. Specifically, to develop indicators for how the Gini index and its ANOGI components react to (1) changes in the population-share of each group, (2) migration between groups, (3) changes in group incomes and (4) income transfers between groups.

Design/methodology/approach – First, the expressions for these indicators are derived analytically. Following this, the indicators are applied to labour-market data from Brazil, contrasting the results to others available in the literature.

Findings – The indicators described above are presented and their characteristics discussed. Empirically, it is illustrated how labour formalisation in Brazil was an inequality-reducing process between 2002 and 2011, contrary to previous incorrect measurements of the phenomenon based on income-source decompositions for Latin American countries.

Originality/value – Indicators for how total inequality reacts to changes in group sizes and income were unavailable for the ANOGI framework, which this article provides. The empirical illustration shows how this leads to a reassessment of important inequality dynamics, using the example of labour formalisation in Brazil. Contrary to the existing literature, it is shown how this was a progressive development, with key implications for social and labour-market policy. This framework can be used to assess the impact of diverse processes in the ANOGI methodology.

Keywords Gini coefficient, Gini decompositions, ANOGI, Inequality, Brazil, Latin America

Paper type Research paper

1. Introduction

With a growing number of works advancing proposals for decompositions of the Gini index (e.g. Dagum, 1997; Mornet *et al.*, 2013; Mussard & Savard, 2012; Mussini, 2013c; Okamoto, 2009; Yitzhaki, 1994), it is now a long way since it could be said that the "interaction term" in group-wise decompositions "is impossible to interpret with any precision, except to say that it is the residual necessary to maintain the identity" (Mookherjee & Shorrocks, 1982, p. 889). It is reasonable to suppose that this research agenda has expanded for two reasons. First, the Gini index remains the most widespread measure of inequality, particularly of income and wealth, and so it is only natural that decompositions of inequality should focus on it. Second, as

The author would like to thank Abigail McKnight and Ben Fine for the comments and encouragement they have given to this work. The usual disclaimers apply.

Funding: The author would like to thank Capes for the funding (BEX 0840/14-9) that made the initial steps of this project possible.

EconomiA
Vol. 23 No. 1, 2022
pp. 123-142
Emerald Publishing Limited
e-ISSN: 2358-2820
p-ISSN: 1517-7580
DOI 10.1108/ECON-05-2022-0013

Lambert and Decoster (2005) aptly put it, "the Gini coefficient reveals more." That is, it offers a deeper breakdown of how inequality is structured across and within sub-populations than, say, the generalised entropy index (which is famously know to be the only class of additively decomposable inequality measures, Shorrocks, 1980). Essentially, this occurs because the relation between members of different groups is also taken into account, instead of representing groups only by their mean values, so it is possible to gain insight into their interaction and, hence, into issues such as stratification. This becomes particularly relevant when distributions are non-normal, as is usually the case with income and wealth.

Given these advantages, several decompositions of the Gini coefficient have been proposed and applied to a variety of cases. This article engages with a particular group-wise decomposition, ANalysis Of GIni (ANOGI) (Frick *et al.*, 2006; Yitzhaki & Schechtman, 2013), in order to offer two developments. First, it proposes a measure of each group's share of total inequality that includes their contribution to between-groups inequality. Second, and most important, it develops indicators of the sensitivity of inequality and its components to changes in group sizes and mean incomes, when longitudinal data are absent. This is a relatively unexplored topic, of practical relevance in several fields such as studying the (un)equalising effects of changes in employment status (such as formalisation of labour or the increase of self-employment), of geographical migration and of the growth of a particular demographic.

These two indicators – for changes in group sizes and mean incomes – discriminate between phenomena that similarly alter a group's income share and which are often confounded in the literature (see section 3). Take the example of labour income, classified by individuals' positions (e.g. fixed-term or permanent employees). If the share of permanent employees' income in total income rises, this can be due to a rise in either their average income or in the proportion of permanent contracts. Consider, furthermore, that those on permanent contracts have a higher mean income (and mean rank in the distribution), but lower intra-group concentration. In this scenario, raising their income is expected to increase inequality, whereas transitioning from a fixed-term to a permanent position is expected to decrease inequality – but both would increase the income-share of permanent employees. To distinguish between the two phenomena and draw appropriate conclusions, the measures presented in this article are necessary. The article illustrates the point by showing, as an example, how Amarante (2016) incorrectly considered a rising share of formal wages in five Latin American countries, between 2002 and 2011, to have been inequality-increasing. Using data for Brazil, it is shown how labour formalisation – the transition from informal to formal employment – took place and was progressive, whereas scaling up the income of formal workers, which would have been regressive, did not occur. Additionally, the article shows how most groups in the Brazilian labour market had opposite sensitivities to expanding their size and their mean income, further highlighting the relevance of the method.

After this introduction, the article is structured as follows. Section 2 briefly examines different decompositions of the Gini index, develops the new measure of each group's contribution to total inequality and the sensitivity measures. Section 3 illustrates these measures with data for Brazil, contrasting them to income-source decompositions. The fourth section concludes.

2. Measuring the sensitivity of inequality to changes in group sizes and mean incomes

2.1 Gini decompositions by sub-populations: a brief review

It is well known that the Gini index is not neatly decomposable into a between- and a within-groups component, with a third term – sometimes later merged into the others – necessarily arising (Pyatt, 1976). This overlapping, interaction or "transvariation" element is simultaneously at the heart of the Gini coefficient's advantages, offering further information about how the groups relate to each other (Lambert & Decoster, 2005), and the main reason why several authors prefer other, more easily decomposable indices (Mookherjee & Shorrocks, 1982) [1]. It is around how to (re-)define and interpret this element that most of the literature on decompositions of the Gini coefficient into subpopulations has revolved (Deutsch & Silber, 1999) [2].

The proposed decompositions can be classified into three broad groups. The first, as in Bhattacharya and Mahalanobis (1967), seeks to maintain only two terms in the decomposition, including the overlap element into only one of them. This will, however, distort the concepts and at least one of the components will not adequately measure its goal. The second group can be associated with Pyatt (1976), which offers initial propositions for both between- and within-groups inequality, but leaves the third term as a residual whose meaning is hard to pin down [3].

The third group of decompositions offers a set of analytically more interesting options, as it redefines the overlapping component to develop a socioeconomic interpretation of all the resulting terms, and this group usually offers adjusted versions of between- and within-groups components. Two of them have arguably become the most popular ones in recent years, Dagum's (1987, 1997) [4] and Yitzhaki and Lerman's (Yitzhaki, 1994; Yitzhaki & Lerman, 1991), later developed into ANOGI (Frick *et al.*, 2006; Yitzhaki & Schechtman, 2013) [5]. Whilst Dagum introduces the extended Gini between subpopulations, ANOGI is based on re-ranking individuals according to the other groups and to the whole population, leading to an overlapping index (the inverse of stratification) and a modified between-groups pseudo-Gini [6]. The result is a four-term equation, measuring within-groups inequality, the effect of overlapping on the latter, between-groups inequality and the effect of overlapping thereon. It can be reduced to a two-term format, which will then measure overlapping-adjusted within- and between-groups inequality.

ANOGI has seen several empirical applications with powerful results, particularly as regards stratification and the integration of different communities in unequal societies [7]. Ceccarelli *et al.* (2014) and D'Agostino *et al.* (2016) have looked at the integration of immigrants in Italy, pointing to greater inequality and stratification since 2008, whilst Frick and Goebel (2008) showed a persistent income divide between East and West Germany. Agrawal (2014) and Zacharias and Vakulabharanam (2011) studied rural-urban and caste inequality in India, whereas Castellano *et al.* (2016) analysed the occupational stratification of income in France and Italy. Staple topics in the field, such as global income inequality, have also benefited from the approach. For example, Liberati (2015) analyses world income between 1970 and 2009 to show how there has been a recent increase of within-country inequality, counterbalanced by a decrease of between-country inequality led by rising income in China.

One dimension that has not been explored, however, is a measurement of how total inequality and its components react to changes in the sizes of the subpopulations. If there is access to panel data, it might be possible to directly study the mobility of individuals between groups, such as in Yitzhaki and Wodon (2004) or in some multi-decompositions following Dagum's tradition (Mussard & Savard, 2012; Mussini, 2013b, c). If such data are not available – an all-too-common feature – or if mobility is impossible (e.g. how does labour income inequality react to greater female labour market participation), a different measure is called for. Will the increase of immigration increase or decrease inequality and stratification? How does self-employment impact inequality and the heterogeneity of workers? Which group's increase will have the strongest impact on inequality? These are the sorts of questions that such a measure can help to address.

2.2 Measuring each group's contribution to total inequality in the ANOGI framework

For a more detailed exposition of the method and proofs in what follows, please consult Frick *et al.* (2006) and Yitzhaki and Schechtman (2013). Consider a population comprising k

mutually exclusive groups with n_i members each, who receive non-negative income y [8]. The overall population $y_U = y_1 \cup y_2 \ldots \cup y_k$ is denoted by the subscript U. Let μ_i be the mean income of group i, so that $p_i = \frac{n_i}{n_U}$, $s_i = \frac{n_i \mu_i}{\sum_{j=1}^{k} n_j \mu_j}$ and $\eta_i = \frac{\mu_i}{\mu_U}$ are respectively the population share, the income share and the relative income of group i. Let $F_i = (y_i)$ represent the cumulative distribution of y in group i. F_i, with a single subscript, indicates the expected value of $F_i(y_i)$ – estimated in the sample by the rank of observations, normalised to be between 0 and 1 – and F_{ji}, with two subscripts, indicates the expected rank of individuals from group i had their income been ranked according to the distribution of group j (note the order of the notation).

The overall Gini index, G, can then be decomposed into four terms:

$$\begin{aligned} G &= G_{\mathrm{IG}} + G_{\mathrm{IGO}} + G_{\mathrm{BP}} + (G_{\mathrm{B}} - G_{\mathrm{BP}}) \\ &= \sum_{h=1}^{k} s_h G_h + \sum_{h=1}^{k} [s_h G_h (O_h - 1)] + G_{\mathrm{BP}} + (G_{\mathrm{B}} - G_{\mathrm{BP}}) \end{aligned} \quad (1)$$

where G_i is the intra-group Gini index of group i, O_i is the overlapping parameter of group i (see below), and the following indices are:

(1) G_{IG}: pure within-groups inequality, it is an income-share-weighted average of Gini coefficients calculated over members of each group, disregarding overlapping;

(2) G_{IGO}: the same as above, but multiplied by the overlapping indices minus 1, to assess the impact that overlapping (less-than-perfect stratification) has on within-groups inequality;

(3) G_{BP}: the pure between-groups Gini coefficient, which disregards overlapping. It is the Gini index that would obtain had all individuals received the mean income of their group;

(4) $(G_{\mathrm{B}} - G_{\mathrm{BP}})$: G_{B} is the overlapping-adjusted (pseudo-)Gini for between-groups inequality, calculated as though all individuals received the mean income of their group, but, differently from G_{BP}, the groups are represented by the mean rank of their members (and not their rank in the population of group-mean incomes). $(G_{\mathrm{B}} - G_{\mathrm{BP}})$ then measures the impact of overlapping on between-groups inequality.

O_i in turn is a measure of how much are the distributions of all groups contained in that of i. It is the population-share-weighted average of the group-by-group overlapping parameters, O_{ji}, which measure how much is the distribution of group j is contained in the range of i [9]:

$$O_i = \sum_{j=1}^{k} p_j O_{ji} \quad (2)$$

Equation (1) can also be simplified into two terms, by adjusting both the within-groups and the between-groups components for overlapping. This leads to the following formulation:

$$G = G_{\mathrm{WO}} + G_{\mathrm{B}} = \sum_{h=1}^{k} s_h G_h O_h + 2 \sum_{h=1}^{k} p_h \dot{\eta}_h \dot{F}_{Uh} \quad (3)$$

where the dot accent represents deviation from the expected values for the overall population, so that $\dot{\eta}_i$ stands for the i-th group's deviation from mean relative income $(\eta_i - 1)$ and $\dot{F}_{Ui}$ for its deviation from mean rank, assessed according to the overall population $(F_{Ui} - 0.5)$.

Before developing the group-sensitivity measures, a different presentation of (3) can shed light on each group's contribution to total inequality. By substituting the income-share for the product of each group's population share and relative income, we arrive at this formulation that puts the population share in evidence:

$$G = \sum_{h=1}^{k} p_h \left(\eta_h G_h O_h + 2\dot{\eta}_h \dot{F}_{Uh} \right) = \sum_{h=1}^{k} p_h G_h^C \tag{4}$$

where G_i^C is the i-th group's contribution to G. This shows how the Gini coefficient can be expressed as a population-share-weighted sum of each group's overlapping-adjusted intra-group Gini coefficient, multiplied by its relative mean income, and their contribution to between-groups inequality, in turn a function of deviations from mean relative income and mean rank.

Some elements of this expression are intuitive: the more representative is a group in terms of population (p_i), the more concentrated amongst its members is its income (G_i) and the higher its relative income (η_i), the higher will be its contribution to the overall Gini through the within-groups component. The multiplication by O_i, however, might be harder to grasp. It essentially indicates that, other things equal, subpopulations constituting a well-defined stratum contribute less to within-groups inequality [10].

The second term shows that even if a group's income is fairly de-concentrated amongst its members, if it is located at the top of the distribution, it might still add significantly to total inequality through the between-groups element. In fact, this is the "perfect" setting for a high contribution to the latter: being a cohesive group located at the top of the distribution, so that $\dot{\eta}_i$ and $\dot{F}_{Ui}$ are large. Likewise, groups at the bottom of the distribution also tend to increase the between-groups component, given that also rises.

The expression in (4) is the appropriate way of assessing each group's contribution to total inequality. This is different from the path Yitzhaki and Schechtman (2013, p. 317) and Milanovic and Yitzhaki (2002) take, for example, as they only consider the first term ($p_i \eta_i G_i O_i$). This disregards, however, the between-groups dimension of inequality, and hence incorrectly measures each group's contribution to overall inequality. The importance of this difference in measurement arises clearly when between-groups inequality is a large share of overall inequality, often reversing conclusions about the relative importance of the several groups in accounting for inequality.

Take the example of Milanovic and Yitzhaki (2002), who study the world distribution of income. For the year of the study, between-groups inequality (with groups defined as regions of the world) accounted for about half of total inequality. In this case, qualitatively different conclusions arise with the adequate measurement of each region's contribution. The authors state that "Asia is the most important contributor to world inequality: it contributes some 20 Gini points which is almost one-third of total world inequality, and 57 percent of intra-continent inequality" (Milanovic & Yitzhaki, 2002, p. 164). Using the same data and assessing shares according to (4), however, Western Europe, North America and Oceania (WENAO) are seen to represent about 77% of between-groups inequality, which, summed to its 18% of intra-group inequality, leads to 46% of total world inequality – above Asia's share of 39%. Therefore, *WENAO, and not Asia, was the most important contributor to world inequality in 1993*, a conclusion which follows from the measurement presented above.

A note of caution in interpreting (4) is in order. This equation does not imply that the group whose contribution to inequality is being assessed *causes* the latter. This is a descriptive exercise, which maps overall inequality to the studied groups, and not an attribution of causality. A poor group will, for example, also enter with a positive sign, and this does not imply they are the causal agents behind this. What it indicates is that the existence of the group at hand, positioned as it is against other subpopulations (its contribution to the between-groups element) and with its internal distribution (its contribution to the within-groups component), accounts for a certain share of total observed inequality.

2.3 Group-sensitivity indicators

Expression (4) shows how the Gini index can be stated as a population-weighted average of group contributions, which will be above or below the resulting overall coefficient. Whilst an important result, analysing a group's disproportionate contribution to total inequality is a different question from investigating whether a marginal increase in its size will lead to a rise or fall in inequality. As Kimhi (2011) comments on income-source decompositions, the latter question is usually of greater relevance to ascertain the progressivity of a stream of income – or, in this case, of expanding a group. Therefore, in what follows this article develops group-size progressivity indices, and it adapts to the current framework the progressivity index for group incomes of Wodon (1999), which are similar to the income-source progressivity indices of Lerman and Yitzhaki (1994) and Hoffmann (2013).

Before presenting the indices themselves, a discussion of the hypotheses behind their validity is required. Whereas scaling up or down a revenue stream is a straightforward process, in some instances, even liable to policy intervention, population shares present a more nuanced picture. The substantive question, which must be assessed for each specific application, regards the causal processes attached to belonging to the groups and whether they extend to, and are affected by, new members. Throughout the discussion it is assumed that the individuals both leaving and entering a subpopulation are randomly drawn from their respective group distributions. There is thus no selection bias amongst "exiting" members, and "entrants" are subjected to the same forces that determine the income of the group to which they will belong. In this sense, this is not a means-preserving process for the overall population, and the individual group distributions are not affected by changes in their sizes. Whether this is an adequate set of hypotheses should be judged for each case. In the case of marginal changes to mean income, the key assumption is that there will be no re-ranking involved, which implies that no members of different groups have exactly equal incomes (or a negligible number does so).

Consider, then, a marginal absolute increase θ in the population-share of group i, offset by a corresponding decrease in the shares of the other groups. This is a random migration from the remainder of the population – alternatively, the independent growth of i – so the relative size of all h groups will remain constant. Therefore, adding and subtracting leads to:

$$\begin{aligned} &p_i + \theta \\ &p_h \quad \left(1 - \frac{\theta}{1-p_i}\right) \forall h \neq i \end{aligned} \tag{5}$$

The question, then, is how the Gini index and the components of its decomposition are affected by adding this arbitrarily small θ: $\frac{\partial G}{\partial \theta}, \frac{\partial G_{IG}}{\partial \theta}, \frac{\partial G_{IGO}}{\partial \theta}, \ldots, \theta \to 0$. As all individual y_i distributions are unchanged by this procedure, several parameters remain constant: the groups' mean income μ_i, their Gini indices G_i, their cumulative distribution when assessed according to other groups $F_j(y_i)$ and their mean rank in such groups F_{ji}, as well as the group-by-group overlapping indices O_{ji}. In general, though, the relative income η_i, the overall overlapping parameter O_i and the mean rank in the population F_{Ui} of all groups are affected, as they depend on a population-weighted summation over every group.

The setting is the same for a marginal multiplicative increase in a group's mean income, say $(1+\epsilon)\mu_i$. Given the hypothesis of there being no re-ranking, all individual group parameters except for their relative income are unchanged.

The derivation of expressions (6) and (7) is presented in Appendices 1 and 2 brings the impact on individual components of the ANOGI decomposition. Skipping the derivation and its details, the impacts on the overall Gini of an increase in the size and in the mean income of group i are respectively expressed by the following two expressions:

$$\begin{aligned} \frac{\partial G}{\partial \theta_i} &= \frac{1}{1-p_i}\left[G_i^C + G_i^O - (\eta_i + 1)G\right] \\ \text{where } G_i^O &= \sum_{h=1}^{k} p_h\left(\eta_h G_h O_{ih} + 2\dot{\eta}_h \dot{F}_{ih}\right) \end{aligned} \quad (6)$$

$$\frac{\partial G}{\partial \epsilon_i} = p_i\eta_i \sum_{h=1}^{k} p_h\eta_h[(G_iO_i - G_hO_h) + 2(F_{Ui} - F_{Uh})] \quad (7)$$

It follows that the increase of a group's size impacts the Gini coefficient through two channels. The first is a "direct" effect, based on its contribution to the overall Gini, G_i^C. The second is through its impact on overlapping (hence the superscript O), expressed in the second term G_i^O. This is the Gini that would obtain by substituting each group's overall overlapping and mean rank parameters, O_h and $\dot{F}_{Uh}$, by those referring to group i, O_{ih} and $\dot{F}_{ih}$ (see the definition in expression 6).

If the impact through these two channels is greater than the prevailing Gini (scaled by one plus the i-th group's relative income), an increase in the group's size will raise inequality. By looking at the impact on the individual terms of the ANOGI decomposition (see the expressions in Appendix 2), one can also assess how the increase in a group's size restructures inequality within and across subpopulations.

As for the impact of raising a group's mean income, in equation (7), it is an income-share-weighted sum of two differences, both between the i-th group and all the others: their overlapping-adjusted intra-group Gini and their mean rank. The interpretation is straightforward. The first difference is related to within-groups inequality – if group i has a higher than average overlapping-adjusted within-group Gini, it will increase this component. The second difference is in turn related to between-groups inequality: if it has a higher than average mean rank, it will raise this component.

Expression (7) takes into account the income-share of the groups, but it can easily be modified to express the sensitivity to equal transfers of resources. Notice that the first element in (7) is the income share of group i, reflecting that a proportional increase to a weighty group's average income will have a stronger effect on overall inequality, if compared to a proportional increase to a group that responds for a small share of the distribution. This measure thus answers questions along the lines of "if the income of a group is increased by 1%, which group will affect inequality the most?" If one wants, instead, to investigate by how much will inequality change if a given sum of resources is distributed proportionally to a certain group, (7) should be divided by the groups' respective income shares. This will answer questions along the lines of "if one unit of income is distributed across members of a group, maintaining the intra-group distribution, to which group should it be given to maximise the impact on inequality?"

Finally, we can construct "transfer matrices" to measure the impact of migration or income transfers from a given group to another, i.e. which only affect the size or income of the two groups involved. To construct two k-by-k skew-symmetric matrices $\boldsymbol{M}$ and $\boldsymbol{T}$, respectively, of the effect on inequality of migration and of income transfers, we can then define θ_{ji} as an absolute marginal decrease in group j and an increase in i, thus a migration from the former to the latter which leaves the other groups' population-shares unaffected, and

ϵ_{ji} as a marginal multiplicative increase in the income of group i entirely offset by a decrease in the income of group j. The elements of the matrices will all be differences between raising the size/income of group i and decreasing that of group j, taking into account their population and income [11].

The expressions for the elements of matrices $\boldsymbol{M}$ and $\boldsymbol{T}$ are:

$$\boldsymbol{M}_{i,j} = \frac{\partial G}{\partial \theta_{ji}} = (1 - p_i)\frac{\partial G}{\partial \theta_i} - (1 - p_j)\frac{\partial G}{\partial \theta_j} \quad (8)$$

$$\boldsymbol{T}_{i,j} = \frac{1}{p_i \eta_i}\frac{\partial G}{\partial \epsilon_{ji}} = \frac{1}{p_i \eta_i}\frac{\partial G}{\partial \epsilon_i} - \frac{1}{p_j \eta_j}\frac{\partial G}{\partial \epsilon_j} \quad (9)$$

The use of the individual group indices or of the matrices depends on the phenomenon being studied. If individuals are migrating from one group to another, say self-employed workers are becoming employees, then $\boldsymbol{M}$ is the appropriate tool, whereas if individuals are being simply being added to a group, e.g. persons outside the labour market become self-employed, then the individual indices are in order. Likewise, if a group's income is rising independently of others, such as through wages having increased for a group of workers, the individual income-sensitivity indices should be used, whereas transferring income between groups requires $\boldsymbol{T}$ (with adequate attention paid to income shares). Together, these instruments can illuminate a wide range of analyses, as illustrated in the following section.

3. The equalising impacts of labour formalisation: an illustration with data from the Brazilian labour market

The method is now illustrated by examining income inequality in Brazil, with a focus on labour market income, particularly by highlighting how the measures (6) through (9) correct errors that arise from the interpretation of income-source decompositions. It is also shown how expression (4) is the one to adequately measure each group's contribution to total inequality. Whilst but an illustration of the method, without the aim of being a comprehensive analysis of income distribution in Brazil, it can nevertheless show the practical relevance of the measures proposed. Data from the 2002 and 2011 National Household Sample Survey (PNAD) are used, the objective variable for group-wise decompositions is income derived from the respondents' main occupation, and observations are grouped according to their position in employment. Four categories are considered: employers, self-employed workers, formal workers and informal workers. Respondents with null income are excluded.

Before proceeding to the ANOGI results, an income-source decomposition is performed, similar to that of Amarante (2016), to show an error the procedure proposed in the article corrects. As briefly mentioned above, the author analyses household per capita income inequality in five Latin American countries in 2002 and 2011, using a dynamic income-source decomposition to explain observed changes. She considers six income sources, amongst which are formal and informal wages, and concludes that "formal wages have contributed to inequality because of their increasing share in total income" (Amarante, 2016, p. S18) in all five countries, with the effect of such higher shares estimated to have been between 1.2 and 8.3 Gini points. However, given that the rate of labour informality decreased during this period in all countries, behind the higher share of formal wages potentially there were two distinct processes that were confounded: the transition from informal to formal occupations and higher relative wages for formal workers. This makes the effect Amarante measured incorrect, as shown below to be the case for Brazil [12].

The income-source decomposition of the Gini index (see Amarante, 2016; Lerman & Yitzhaki, 1985) can be expressed as an income-share-weighted average of concentration coefficients:

$$G = \sum_{g=1}^{f} s_g C_g \tag{10}$$

where there are f income sources, and C_g is the concentration coefficient of the g-th source (if it is higher than G, than this source is concentrated in relation to total income). The dynamic version uses a shift-share method, arriving at three effects: the share effect, which measures the impact of changing income shares; the concentration effect, which measures the impact of changes in the concentration of each source in relation to total income; and the dynamic effect, which measures simultaneous changes in shares and concentration:

$$\Delta G = G^{t=1} - G^{t=0} = \sum_{g=1}^{f} \Delta s_g C_g^{t=0} + \sum_{g=1}^{f} s_g^{t=0} \Delta C_g + \sum_{g=1}^{f} \Delta s_g \Delta C_g \tag{11}$$

The results of the income-source decomposition for Brazil, in 2002 and 2011, are presented in Table 1. Similar to the countries Amarante analysed, the share of formal wages in total income increased, in this case by 5.5 percentage points. Given that formal wages were concentrated relative to total income, this led to an estimated share effect of 3.3 Gini points, which would suggest that labour formalisation increased household per capita income inequality in Brazil.

However, as seen through the results of the ANOGI decomposition presented below (Tables 2 through 4, plus Tables A1, A2, A3 and A4 in Appendix 3), labour formalisation was actually an equalising process [13]. Moving on to the ANOGI results, as seen in Table 2, a

		Income share S_g	Concentration C_g	Contribution to G $S_g C_g$	Share effect $\Delta S_g C_g^{t=0}$	Concentration effect $S_g^{t=0} \Delta C_g$	Dynamic interaction $\Delta S_g \Delta C_g$
Employers' income	2002	0.119	0.857	0.102	−0.031	−0.001	0.000
	2011	0.083	0.852	0.07			
Self-employment income	2002	0.157	0.501	0.079	−0.002	0.000	0.000
	2011	0.154	0.501	0.077			
Formal wages	2002	0.387	0.603	0.233	0.033	−0.020	−0.003
	2011	0.442	0.551	0.243			
Informal wages	2002	0.110	0.338	0.037	−0.005	−0.005	0.001
	2011	0.094	0.293	0.028			
Other incomes	2002	0.226	0.601	0.136	0.000	−0.026	0.000
	2011	0.227	0.487	0.110			
Total	*2002*			*0.587*	*−0.005*	*−0.052*	*−0.002*
	2011			*0.528*			

Source(s): Prepared by the author based on data from PNAD, 2002/2011

Table 1. Income source decomposition of household per capita income in Brazil, 2002 and 2011

	2002	2011
Gini coefficient (G)	0.555	0.494
Overlapping-adjusted within-groups Gini (G_{wo})	0.424	0.404
Overlapping-adjusted between-groups Gini (G_b)	0.131	0.090
Intra-groups Gini (G_{IG})	0.503	0.458
Ef. of overlapping on within-groups Gini (G_{IGO})	−0.079	−0.054
Pyatt between-groups Gini (G_{BP})	0.239	0.174
Ef. of overlapping on between-groups Gini ($G_B - G_{BP}$)	−0.107	−0.084

Source(s): Prepared by the author based on data from PNAD, 2002/2011

Table 2. ANOGI for labour market income in Brazil, 2002 and 2011, decomposed by position in employment – main results

considerable portion of inequality (23.6% in 2002) was due to the between-groups dimension. This was true in spite of there also being substantial overlapping between the categories, which led to a 10-points adjustment in the "pure" between-groups coefficient (see the last row of Table 2).

Employers dominated the higher end of the distribution (see Table 3) in both years, with informal workers clearly in the bottom. Most categories were relatively well-formed strata (low overlapping), with the expected exceptions self-employed and informal workers. Whilst the income of these categories was mostly below average, given their usually precarious employment condition and lower coverage by labour laws, a few individuals nonetheless had higher income – hence their high internal Gini indices and overlapping parameters. Furthermore, formal workers were the less concentrated category, likely because of labour legislation protection which prevents payment below the minimum wage.

It can also be seen how there was a large difference between each group's contribution to total inequality (Table 3, column 7) and to overlapping-adjusted within-groups inequality (Table 3, column 6). This is the difference that expression (4) allows us to visualise and which has so far been overlooked in the literature (e.g. Milanovic & Yitzhaki, 2002).Take the example of employers: their contribution to total inequality was almost three times higher than that to within-groups inequality, leading to very different conclusions about how inequality is structured along sub-groups. This difference is larger when there are more extreme-income groups in the population. As argued above, then, expression (4) is the most appropriate one to study each group's role in the distribution of income.

Likewise, Brazilian data show the difference between having an above-average contribution to inequality and having a positive marginal impact on the latter (Table 4). Employers are the only group who had the same sign for their deviation from average contribution to inequality (column 1), their size's marginal impact (column 2) and their mean income's marginal impact (column 3). All the other groups had a below-average contribution to total inequality, but increasing the size of self-employed or informal workers would raise inequality, given their low mean income and their low mean rank (it would increase the size of the deprived population, in a sense). Formal workers, on the other hand, despite being a relatively income-rich group (i.e. slightly above the overall population's mean income and mean rank), had the lowest intra-group Gini index, so increasing their size would be expected to reduce inequality (it would increase the middle-class, so to say). For all three categories of workers, but not employers, the impact of raising their mean income had the opposite sign to that of raising their population share, highlighting the relevance of distinguishing between the two phenomena.

Finally, the estimated impact between 2002 and 2011 of changes in group sizes and mean income can be seen in columns 6 and 7 of Table 4. It can readily be seen that all impacts were negative (i.e. equalising), indicating that the movements in the size and in the relative income of all groups followed a progressive direction. In particular, it should be noticed that labour formalisation as a whole (the exit from self-employment and informal employment into formal occupations), measured by the sum of the group-size impacts of the three categories of workers, is estimated to have contributed with a *decrease of 1.6 Gini points – precisely the opposite conclusion that arises from the (incorrectly applied) income-source decomposition*.

In sum, the illustration with Brazilian data has shown three important results. First, the groups' contribution to overlapping-adjusted within-groups inequality and to total inequality can differ widely, with the latter being the appropriate measure of how inequality is structured along sub-groups. Second, that a group might have an above- or below-average contribution to total inequality is not indicative of the marginal impact its size has on inequality, particularly in the case of groups at the extremes of the distribution. Third, the marginal impact measures developed in this article are the appropriate ones to study phenomena that involve individuals entering or exiting positions, leading to adequate measurements that can correct common errors in income-source decompositions. In the case

Group		Population share (p_i)	Relative income (η_i)	Concentration (G_i)	Overlapping (O_i)	Rank (F_{Ui})	Contr. to G_{WO}: $p_i(\eta_i G_i O_i)$	Contr. to G: $p_i(\eta_i G_i O_i + 2\dot{\eta}_i \dot{F}_{Ui})$
1: Employers	2002	0.047	3.250	0.523	0.485	0.814	0.039	0.105
	2011	0.035	3.087	0.526	0.514	0.804	0.030	0.074
2: Self-employed	2002	0.249	0.818	0.572	1.104	0.432	0.129	0.135
	2011	0.225	0.885	0.531	1.120	0.441	0.118	0.121
3: Formal workers	2002	0.429	1.166	0.472	0.769	0.611	0.181	0.197
	2011	0.522	1.091	0.422	0.822	0.580	0.198	0.205
4: Informal workers	2002	0.275	0.524	0.491	1.060	0.335	0.075	0.118
	2011	0.217	0.560	0.443	1.069	0.318	0.058	0.092

Source(s): Prepared by the author based on data from PNAD, 2002/2011

Table 3.
ANOGI for labour market income in Brazil, 2002 and 2011, decomposed by position in employment – group parameters

of labour formalisation in Latin America, it was shown how opposite conclusions apply when using the two methods.

4. Final remarks

This article has discussed a neglected aspect of group-wise decompositions of the Gini coefficient, namely, how total inequality and its sub-components react to changes in group sizes. In this context, it has advanced two propositions and illustrated them with an analysis of changes to income inequality in Brazil.

First, this article defined a different expression for calculating each group's contribution to total inequality. It takes account not only of the share of overlapping-adjusted within-groups inequality but also of between-groups inequality. With data for the Brazilian labour market, it was shown how this can lead to considerably different conclusions regarding how inequality is structured along sub-populations, as groups with non-concentrated income can nevertheless contribute highly if they are well-formed strata located at extremes of the distribution. Likewise, calculations based on data from Milanovic and Yitzhaki (2002) reveal qualitatively different conclusions from those of the authors (namely, that Western Europe, North America and Oceania represented the highest share of world inequality, and not Asia as Milanovic and Yitzhaki had proposed).

Second, indicators for the sensitivity of inequality to the size of sub-populations were developed. A measure was proposed for the impact on inequality of an increase in a group's population-share, and income-sensitivity measures were developed for the ANOGI framework, all of which are decomposed into changes in the ANOGI components. Based on this, the article also developed a migration and an income transfer matrix between all groups, which measures the impact of population migration or income transfers between any two groups (leaving the population and income of other groups unchanged).

Illustrating the method with data from Brazil, it was shown how it is particularly useful when panel data are unavailable, for example, to study the impact of formalisation on labour market income inequality. Taking account of this can remedy inappropriate conclusions in the literature, such as Amarante's (2016) study that confounds the effects of labour formalisation (the transition between informal and formal work, which is equalising) and those of higher formal wages (scaling up the income of existing formal workers, which is unequalising) to arrive at the incorrect conclusion that the greater share of formal wages in total income was unequalising in Latin America. As shown for the case of Brazil, in this article, labour formalisation between 2002 and 2011 reduced overall income inequality. Important policy differences are forthcoming by using the appropriate measurements.

Table 4. Labour market income inequality in Brazil: sensitivity of inequality and impact of changes in group sizes and mean income

Group	Deviation from average contribution to G $(\eta_i G_i O_i + 2\dot{\eta}_i \dot{F}_{Ui}) - G$	Sens. to group size $\frac{\partial G}{\partial \theta_i}$	Sens. to group mean income $\frac{\partial G}{\partial \epsilon_i}$	Pop. change Δp_i	Mean income change[a] $\Delta_{\%}\mu_i - \Delta_{\%}\mu_U$	Group size impact $\Delta p_i \frac{\partial G}{\partial \theta_i}$	Group mean income impact $(\Delta_{\%}\mu_i - \Delta_{\%}\mu_U)\frac{\partial G}{\partial \epsilon_i}$
1: Employers	1.683	0.492	0.050	−0.011	−0.107	−0.006	−0.005
2: Self-employed	−0.014	0.029	−0.012	−0.024	0.176	−0.001	−0.002
3: Formal workers	−0.095	−0.137	0.015	0.093	−0.137	−0.013	−0.002
4: Informal workers	−0.125	0.032	−0.053	−0.058	0.148	−0.002	−0.008

Note(s): [a]The change in mean income is defined subtracted from the overall change in mean income, given that the Gini coefficient is scale-invariant
Source(s): Prepared by the author based on data from PNAD, 2002/2011

Notes

1. See also Fine and Loureiro (2020, 2021) for recent approaches to group-wise decompositions of inequality from first principles whilst relaxing the assumption of group symmetry (i.e. allowing income to be valued differently across groups), drawing on the foundational approaches of Atkinson (1970), Shorrocks (1980) and Fine (1975, 1985, 1996).
2. The most wide-ranging and systematic coverage of the different decomposition, including comparison with other inequality measures, is in Deutsch and Silber (1999), even if it does not cover recent forays into multi-decompositions and the like. Giorgi (1990, 1993) presents the history of debates around the coefficient, and other authors discuss various proposals for decompositions as they present their own (e.g. Lambert & Decoster, 2005; Yitzhaki, 1994; Yitzhaki & Schechtman, 2013).
3. Although Lambert and Decoster (2005) have provided an extensive discussion of the behaviour of the interaction term, determining under which conditions it will rise or fall, a decomposition that attaches clear meaning to it is nevertheless much easier to interpret.
4. For recent developments of this approach, see the multi-decompositions in Mussard and Savard (2012) and Mussini (2013a), as well as Mussini (2013b, c), who offers a version specific to longitudinal data.
5. For a discussion of the differences between Dagum's and Yitzhaki and Lerman's proposals, see Yitzhaki (1994). The arguments can be carried over to ANOGI without significant losses.
6. A somewhat similar, though independent, approach can be found in Sastry and Kelkar (1994).
7. Other uses include Frick *et al.* (2006) and Ceccarelli and Giorgi (2009), who studied attrition in longitudinal surveys, focussing on the overlapping parameter to see if different time-cohorts constitute separate strata according to several variables (income, satisfaction etc.), and whether there are learning-effects due to repeated sampling.
8. The presentation uses income for shorthand.
9. The mean rank parameter can also be expressed as the population-weighted average of group-by-group mean rank parameters.
10. Naturally, although G_i might vary without impacting O_i, overlapping is in general not independent of the spread of observations, so the precise relationship between stratification and within-groups inequality is not easily spelled out in a general fashion.
11. Notice, as such, that matrix $\boldsymbol{T}$ analyses which marginal transfers impact inequality the most, investigating between which groups one can have the "greatest bang for a transferred buck".
12. As a further illustration of how income-source decompositions are often employed incorrectly, and for which the current method can correct, see Judzik *et al.* (2017). The authors study income inequality in Argentina from 1996 to 2014, and correctly conclude that labour formalisation was an equalising process from the 2000s onwards, but the methods employed do not allow them to discriminate between transition from informality to formality (equalising) and rising relative income of formal workers (unequalising). The present method does.
13. The ANOGI procedure analyses labour market income inequality for individuals' main occupations, whereas the income-source decomposition analyses household per capita income to follow Amarante (2016). Qualitatively similar results arise for an ANOGI decomposition of household per capita income classified according to the head of household's main occupation, but the analysis of labour market income inequality is more intuitive.

References

Agrawal, T. (2014). Educational inequality in rural and urban India. *International Journal of Educational Development*, *34*, 11–19.

Amarante, V. (2016). Income inequality in Latin America: A factor component analysis. *Review of Income and Wealth*, *62*(Supplement 1), S4–S21.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263.

Bhattacharya, N., & Mahalanobis, B. (1967). Regional disparities in household consumption in India. *Journal of the American Statistical Association*, *62*(317), 143–161.

Castellano, R., Manna, R., & Punzo, G. (2016). Income inequality between overlapping and stratification: A longitudinal analysis of personal earnings in France and Italy. *International Review of Applied Economics*, *30*(5), 567–590.

Ceccarelli, C., & Giorgi, G. M. (2009). Analysis of Gini for evaluating attrition in Italian survey on income and living condition. *RIEDS – Rivista Italiana di Economia, Demografia e Statistica – Italian Review of Economics, Demography and Statistics*, (1-2), 49.

Ceccarelli, C., Giorgi, G.M., & Guandalini, A. (2014). Is Italy a melting pot? *RIEDS – Rivista Italiana di Economia. Demografia e Statistica – Italian Review of Economics, Demography and Statistics*, (3-4), 23.

Dagum, C. (1987). Measuring the economic affluence between populations of income receivers. *Journal of Business and Economic Statistics*, *5*(1), 5–12.

Dagum, C. (1997). A new approach to the decomposition of the Gini income inequality ratio. *Empirical Economics*, *22*(4), 515–531.

Deutsch, J., & Silber, J. (1999). Inequality decomposition by population subgroups and the analysis of interdistributional inequality. In Silber, J. (Ed.), *Handbook of income inequality measurement*, (pp. 363–404). Boston: Springer Science.

D'Agostino, A., Regoli, A., Cornelio, G., & Berti, F. (2016). Studying income inequality of immigrant communities in Italy. *Social Indicators Research*, *127*(1), 83–100.

Fine, B. (1975). A note on "interpersonal aggregation and partial comparability". *Econometrica*, *43*(1), 169–172.

Fine, B. (1985). A note on the measurement of inequality and interpersonal comparability. *Social Choice and Welfare*, *1*(4), 273–277.

Fine, B. (1996). Reconciling interpersonal comparability and the intensity of preference for the utility sum rule. *Social Choice and Welfare*, *13*(3), 319–325.

Fine, B., & Loureiro, P. M. (2020). A note on the relationship between additive separability and decomposability in measuring income inequality. *Review of Social Economy*, 1. doi: 10.1080/00346764.2020.1802055.

Fine, B., & Loureiro, P. M. (2021). From social choice to inequality-decomposition: In the spirit of Arrow and Atkinson by way of Sen and Shorrocks. *International Review of Applied Economics*, *35*(5), 765–791. doi:10.1080/02692171.2021.1892039.

Frick, J. R., & Goebel, J. (2008). Regional income stratification in unified Germany using a Gini decomposition approach. *Regional Studies*, *42*(4), 555–577.

Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G., & Yitzhaki, S. (2006). 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*(4), 427–468.

Giorgi, G. M. (1990). Bibliographic portrait of the Gini concentration ratio. *METRON – International Journal of Statistics*, *XLVIII*(1-4), 183–221.

Giorgi, G. M. (1993). A fresh look at the topical interest of the Gini concentration ratio. *METRON – International Journal of Statistics*, *LI*(1-2), 83–98.

Hoffmann, R. (2013). How to measure the progressivity of an income component. *Applied Economics Letters*, *20*(4), 328–331.

Judzik, D., Trujillo, L., & Villafañe, S. (2017). A tale of two decades: Income inequality and public policy in Argentina (1996–2014). *Cuadernos de Economía*, *36*(72), 233–264.

Kimhi, A. (2011). Comment: On the interpretation (and misinterpretation) of inequality decompositions by income sources. *World Development*, *39*(10), 1888–1890.

Lambert, P. J., & Decoster, A. (2005). The Gini coefficient reveals more. *Center for Economic Studies DPS*, *2005*(8), 1–26.

Lerman, R. I., & Yitzhaki, S. (1985). Income inequality effects by income source: A new approach and applications to the United States. *The Review of Economics and Statistics*, *67*(1), 151–156.

Lerman, R. I., & Yitzhaki, S. (1994). Effect of marginal changes in income sources on U.S. income inequality. *Public Finance Quarterly*, *22*(4), 403–417.

Liberati, P. (2015). The world distribution of income and its inequality, 1970–2009. *Review of Income and Wealth*, *61*(2), 248–273.

Milanovic, B., & Yitzhaki, S. (2002). Decomposing world income distribution: Does the world have a middle class? *Review of Income and Wealth*, *48*(2), 155–178.

Mookherjee, D., & Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal*, *92*(368), 886–902.

Mornet, P., Zoli, C., Mussard, S., Sadefo-Kamdem, J., Seyte, F., & Terraza, M. (2013). The (α, β)-multi-level α-Gini decomposition with an illustration to income inequality in France in 2005. *Economic Modelling*, *35*, 944–963.

Mussard, S., & Savard, L. (2012). The gini multi-decomposition and the role of Gini's transvariation: Application to partial trade liberalization in the Philippines. *Applied Economics*, *44*(10), 1235–1249.

Mussini, M. (2013a). A matrix approach to the Gini index decomposition by subgroup and by income source. *Applied Economics*, *45*(17), 2457–2468.

Mussini, M. (2013b). On decomposing inequality and poverty changes over time: A multi-dimensional decomposition. *Economic Modelling*, *33*, 8–18.

Mussini, M. (2013c). A subgroup decomposition of the inequality change over time. *Applied Economics Letters*, *20*(4), 386–390.

Okamoto, M. (2009). Decomposition of Gini and multivariate Gini indices. *The Journal of Economic Inequality*, *7*(2), 153–177.

Pyatt, G. (1976). On the interpretation and disaggregation of gini coefficients. *The Economic Journal*, *86*(342), 243–255.

Sastry, D. V. S., & Kelkar, U. R. (1994). Note on the decomposition of Gini inequality. *The Review of Economics and Statistics*, *76*(3), 584–586.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, *48*(3), 613–625.

Wodon, Q. T. (1999). Between group inequality and targeted transfers. *Review of Income and Wealth*, *45*(1), 21–39.

Yitzhaki, S. (1994). Economic distance and overlapping of distributions. *Journal of Econometrics*, *61*, 147–159.

Yitzhaki, S., & Lerman, R. I. (1991). Income stratification and income inequality. *Review of Income and Wealth*, *37*(3), 313–329.

Yitzhaki, S., & Schechtman, E. (2013). *The Gini methodology: A primer on a statistical methodology*. New York: Springer.

Yitzhaki, S., & Wodon, Q. (2004). Inequality, mobility, and horizontal inequity. In Amiel, Y., Bishop, J. A., & Formby, J. P. (Eds.), *Studies on economic well-being: Essays in the honor of John P. Formby*, (pp. 177–198). Amsterdam: Elsevier.

Zacharias, A., & Vakulabharanam, V. (2011). Caste stratification and wealth inequality in India. *World Development*, *39*, 1820–1833.

Appendix 1
Derivation of sensitivity expressions

This appendix presents the derivation of expression (6). Expression (7) follows an analogous logic and is thus omitted.

Starting from expressions (4) and introducing the marginal change, it is a matter of finding the four partial derivatives:

$$\frac{\partial G}{\partial \theta_i}=\sum_{h=1}^{k}\left[\begin{array}{c}\frac{\partial p_h}{\partial \theta_i}\left(\eta_h G_h O_h+2\dot{\eta}_h \dot{F}_{Uh}\right)+\\ p_h\left(\frac{\partial \eta_h}{\partial \theta_i}G_h O_h+\eta_h G_h\frac{\partial O_h}{\partial \theta_i}+2\frac{\partial \eta_h}{\partial \theta_i}\dot{F}_{Uh}+2\dot{\eta}_h\frac{\partial F_{Uh}}{\partial \theta_i}\right)\end{array}\right] \tag{A.1}$$

The population-shares are trivial:

$$p=\begin{array}{ll} p_i: & p_i+\theta_i \\ p_{h,h\neq i}: & p_h\left(1-\frac{\theta_i}{1-p_i}\right)\forall h\neq i\end{array}$$

$$\frac{\partial p_h}{\partial \theta_i}=\begin{array}{ll} p_i: & 1 \\ p_{h,h\neq i}: & -\frac{p_h}{1-p_i}\end{array} \tag{A.2}$$

where h is used to index the k groups, given that the derivatives are in general different for the i-th and the other groups. Below, summations within the expression are indexed by j, to avoid confusion with the i-th group (the one under investigation) and the general index h, used outside summations.

Since this is not a means-preserving process, the relative income of all groups is liable to change as the overall distribution is altered:

$$\eta_h=\frac{\mu_h}{(p_i+\theta_i)\mu_i+\sum_{j=1,j\neq i}^{k}\left[p_j\left(1-\frac{\theta_i}{1-p_i}\right)\mu_j\right]}$$

and the derivative then is:

$$\frac{\partial \eta_h}{\partial \theta_i}=\frac{-\mu_h\left(\mu_i-\sum_{j=1,j\neq i}^{k}\frac{p_j}{1-p_i}\mu_j\right)}{\mu_U^2}$$

$$\frac{\partial \eta_h}{\partial \theta_i}=-\eta_h\left(\eta_i-\frac{1}{1-p_i}\sum_{j=1,j\neq i}^{k}p_j\eta_j\right)$$

$$\frac{\partial \eta_h}{\partial \theta_i}=-\frac{1}{1-p_i}\eta_h\left((1-p_i)\eta_i-\sum_{j=1,j\neq i}^{k}p_j\eta_j\right)$$

$$\frac{\partial \eta_h}{\partial \theta_i}=-\frac{1}{1-p_i}\eta_h\left(\eta_i-\sum_{j=1}^{k}p_j\eta_j\right)$$

since $\sum_{h=1}^{k}p_h\eta_h=1$ and using the dot accent to represent deviation from relative income equal to unity,

$$\frac{\partial \eta_h}{\partial \theta_i}=-\frac{1}{1-p_i}\eta_h\dot{\eta}_i \tag{A.3}$$

This indicates that if the i-th group's income is above average, then the relative income of all groups, including its own, will fall.

As for the overlapping index,

$$
\begin{aligned}
O_h &= (p_i+\theta_i)O_{ih} + \sum_{j=1, j\neq i}^{k} p_j\left(1-\frac{\theta_i}{1-p_i}\right)O_{jh} \\
\frac{\partial O_h}{\partial \theta_i} &= O_{ih} - \sum_{j=1, j\neq i}^{k} \frac{p_j}{1-p_i}O_{jh} \\
\frac{\partial O_h}{\partial \theta_i} &= \frac{1}{1-p_i}\left((1-p_i)O_{ih} - \sum_{j=1, j\neq i}^{k} p_j O_{jh}\right) \\
\frac{\partial O_h}{\partial \theta_i} &= -\frac{1}{1-p_i}\left(\sum_{j=1}^{k} p_j O_{jh} - O_{ih}\right)
\end{aligned}
$$

given that $\sum_{j=1}^{k} p_j O_{jh} = O_h$,

$$
\frac{\partial O_h}{\partial \theta_i} = -\frac{1}{1-p_i}(O_h - O_{ih}) \tag{A.4}
$$

This indicates that, as the pairwise overlapping indexes do not change, it is a matter of assessing how overlapping is affected by the different structure of the population. If the h-th group overlaps more with the i-th than with the overall population, then its final overlapping index will rise (it will be less of a stratum). The i-th group in turn will only reduce its overlapping if $O_i > 1$, meaning it is more spread than the population (a malformed stratum).

As the mean rank parameter is very similar to the overlapping one, some steps can be omitted:

$$
\begin{aligned}
F_{Uh} &= (p_i+\theta_i)F_{ih} + \sum_{j=1, j\neq i}^{k} p_j\left(1-\frac{\theta_i}{1-p_i}\right)F_{jh} \\
\frac{\partial F_{Uh}}{\partial \theta_i} &= F_{ih} - \sum_{j=1, j\neq i}^{k} \frac{p_j}{1-p_i}F_{jh} \\
&\dots \\
\frac{\partial F_{Uh}}{\partial \theta_i} &= -\frac{1}{1-p_i}(F_{Uh} - F_{ih})
\end{aligned} \tag{A.5}
$$

It shows that if a group is higher ranked in the overall population than according to the distribution of the i-th group, it will lose positions. The i-th group will in turn get close to the average.

It is now possible to expand (A.1). The only derivative that varies per group is the population-share, so group i is dealt with separately in the term that involves the former. Substituting expressions (A.2) through (A.5) in (A.1), putting $-\frac{1}{1-p_i}$ in evidence in the third term, then gives us:

$$
\frac{\partial G}{\partial \theta_i} = \begin{array}{c} \left(\eta_i G_i O_i + 2\dot{\eta}_i \dot{F}_{Ui}\right) - \sum_{j=1, j\neq i}^{k}\left[\frac{p_j}{1-p_i}\left(\eta_j G_j O_j + 2\dot{\eta}_j \dot{F}_{Uj}\right)\right] \\ -\sum_{j=1}^{k}\frac{p_j}{1-p_i}\left(\eta_j \dot{\eta}_i G_j O_j + \eta_j G_j (O_j - O_{ij}) + 2\eta_j \dot{\eta}_i \dot{F}_{Uj} + 2\dot{\eta}_j (F_{Uj} - F_{ij})\right) \end{array} \tag{A.6}
$$

Multiplying and dividing the first term by $(1-p_i)$ allows a full summation over all groups in the second term, so, putting $(1-p_i)$ in evidence for the whole expression:

$$\frac{\partial G}{\partial \theta_i} = \frac{1}{1-p_i}\left[\begin{array}{c} \left(\eta_i G_i O_i + 2\dot{\eta}_i \dot{F}_{Ui}\right) - \sum_{j=1}^{k}\left[p_j\left(\eta_j G_j O_j + 2\dot{\eta}_j \dot{F}_{Uj}\right)\right] \\ -\sum_{j=1}^{k} p_j\left(\eta_j \dot{\eta}_i G_j O_j + \eta_j G_j (O_j - O_{ij}) + 2\eta_j \dot{\eta}_i \dot{F}_{Uj} + 2\dot{\eta}_j (F_{Uj} - F_{ij})\right) \end{array}\right]$$

Since $\sum_{j=1}^{k}\left[p_j\left(\eta_j G_j O_j + 2\dot{\eta}_j \dot{F}_{Uj}\right)\right] = G$:

$$\frac{\partial G}{\partial \theta_i} = \frac{1}{1-p_i}\left[\begin{array}{c} \left(\eta_i G_i O_i + 2\dot{\eta}_i \dot{F}_{Ui}\right) - G \\ -\sum_{j=1}^{k} p_j\left(\eta_j \dot{\eta}_i G_j O_j + \eta_j G_j (O_j - O_{ij}) + 2\eta_j \dot{\eta}_i \dot{F}_{Uj} + 2\dot{\eta}_j (F_{Uj} - F_{ij})\right) \end{array}\right]$$

Define then the i-th group's overlapping Gini as $G_i^O = \sum_{h=1}^{k}\left[p_h\left(\eta_h G_h O_{ih} + 2\dot{\eta}_h \dot{F}_{ih}\right)\right]$, which is the Gini that would obtain by substituting each group's overall overlapping and mean rank parameters, O_h and $\dot{F}_{Uh}$, by those referring to group i, O_{ih} and $\dot{F}_{ih}$. Given that the i-th group's contribution to G is $G_i^C = \left(\eta_i G_i O_i + 2\dot{\eta}_i \dot{F}_{Ui}\right)$ (see expression 4), this finally leads to expression (6):

$$\frac{\partial G}{\partial \theta_i} = \frac{1}{1-p_i}\left[G_i^C + G_i^O - (\eta_i + 1)G\right]$$

Appendix 2
Sensitivity indicators for ANOGI components

Expressions for the impact of changes to group sizes on individual terms of the ANOGI decomposition:

$$\begin{aligned}
\frac{\partial G_{IG}}{\partial \theta_i} &= \frac{1}{1-p_i}\left[\eta_i (G_i - G_{IG})\right] \\
\frac{\partial G_{IGO}}{\partial \theta_i} &= \frac{1}{1-p_i}\left[G_{IGO_i}^C + G_{IGO_i}^O - (\eta_i + 1)G_{IGO}\right] \\
&\quad \text{where } G_{IGO_i}^C = \left[\eta_i G_i (O_i - 1)\right], G_{IGO_i}^O = \sum_{j=1}^{k} p_j\left[\eta_j G_j (O_{ij} - 1)\right] \\
\frac{\partial G_B}{\partial \theta_i} &= \frac{1}{1-p_i}\left[G_{B_i}^C + G_{B_i}^O - (\eta_i + 1)G_B\right] \\
&\quad \text{where } G_{B_i}^C = 2\dot{\eta}_i \dot{F}_{Ui}, G_{B_i}^O = 2\sum_{j=1}^{k} p_j\left[\dot{\eta}_j \dot{F}_{ij}\right] \\
\frac{\partial G_{WO}}{\partial \theta_i} &= \frac{1}{1-p_i}\left[G_{WO_i}^C + G_{WO_i}^O - (\eta_i + 1)G_{WO}\right] \\
&\quad \text{where } G_{WO_i}^C = \left[\eta_i G_i O_i\right], G_{WO_i}^O = \sum_{j=1}^{k} p_j\left[\eta_j G_j O_{ij}\right]
\end{aligned} \quad (A.7)$$

Expressions for the impact of changes to group mean incomes on individual terms of the ANOGI decomposition:

$$\frac{\partial G_{IG}}{\partial \epsilon_i} = p_i\eta_i \sum_{j=1}^{k} p_j\eta_j[G_i - G_j]$$

$$\frac{\partial G_{IGO}}{\partial \epsilon_i} = p_i\eta_i \sum_{j=1}^{k} p_j\eta_j[(G_iO_i - G_jO_j) - (G_i - G_j)]$$

$$\frac{\partial G_{B}}{\partial \epsilon_i} = p_i\eta_i \sum_{j=1}^{k} p_j\eta_j 2[F_{Ui} - F_{Uj}] \tag{A.8}$$

$$\frac{\partial G_{WO}}{\partial \epsilon_i} = p_i\eta_i \sum_{j=1}^{k} p_j\eta_j[G_iO_i - G_jO_j]$$

Appendix 3
Additional information for the ANOGI decomposition

Table A1. ANOGI for labour market income in Brazil, 2002, decomposed by position in employment – group-by-group overlapping matrix

	1	2	3	4
1	1.000	0.418	0.610	0.262
2	0.979	1.000	1.254	0.986
3	1.011	0.656	1.000	0.472
4	0.820	0.954	1.187	1.000

Note(s): Each cell indicates the extent to which the distribution of the column category is included in the range of the row one (each group's overall overlapping index is a population-weighted summation of its row)
Source(s): Prepared by the author based on data from PNAD, 2002

Table A2. ANOGI for labour market income in Brazil, 2002, decomposed by position in employment – group-by-group mean rank matrix

	1	2	3	4
1	0.500	0.843	0.769	0.911
2	0.157	0.500	0.329	0.579
3	0.231	0.671	0.500	0.795
4	0.089	0.421	0.205	0.500

Note(s): Each cell indicates the mean rank of the row group when assessed according to the distribution of the column category (each group's mean rank in the overall population is a population-weighted summation of its row)
Source(s): Prepared by the author based on data from PNAD, 2002

Table A3. ANOGI for labour market income in Brazil, 2002, decomposed by position in employment – i to j migration matrix of the impact on total inequality

	1	2	3	4
1	0	−0.447	−0.548	−0.446
2	0.447	0	−0.100	0.001
3	0.548	0.100	0	0.102
4	0.446	−0.001	−0.102	0

Note(s): Each cell indicates the impact on inequality of a transition from the row to the column category
Source(s): Prepared by the author based on data from PNAD, 2002

Table A4. ANOGI for labour market income in Brazil, 2002, decomposed by position in employment – i to j income transfer matrix of the impact on total inequality

	1	2	3	4
1	0	−0.385	−0.297	−0.692
2	0.385	0	0.088	−0.307
3	0.297	−0.088	0	−0.395
4	0.692	0.307	0.395	0

Note(s): Each cell indicates the sensitivity of inequality to an income transfer from the row to the column category
Source(s): Prepared by the author based on data from PNAD, 2002

Corresponding author
Pedro Mendes Loureiro can be contacted at: PML47@cam.ac.uk