ISSN 0819-2642
ISBN 978 0 7340 4018 3

THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 1054

October 2008

On Dagum's Decomposition of the Gini Coefficient

by

William Griffiths

Department of Economics
The University of Melbourne
Melbourne Victoria 3010
Australia.

On Dagum's Decomposition of the Gini Coefficient

William Griffiths*

University of Melbourne

Abstract

To measure the contributions to inequality from population subgroups, the Gini coefficient is often decomposed into inequality within groups, inequality between groups and a residual term arising from the overlapping of income distributions from different groups. In this paper we show that two decompositions presented separately in the literature, a traditional decomposition and a decomposition suggested by Dagum (1997), are identical, a fact not previously acknowledged in the literature.

Keywords: population subgroups; between inequality; within inequality

JEL classification: D31, D63

Author information

William Griffiths
Department of Economics
University of Melbourne 3010
Vic., Australia

Email: wegrif@unimelb.edu.au

Ph: +613 8344 3622

Fx: +613 8344 6899

* Thanks to Duangkamon Chotikapanich and Prasada Rao for comments on an earlier version of this paper. This research was supported by Australian Research Council grant DP0771334.

1. Introduction

When using the Gini coefficient to measure inequality, it is frequently of interest to assess the contributions to inequality from within and between designated population subgroups. For example, in countries like India and China who have experienced recent rapid growth, it is common to examine inequality within and between rural and urban population subgroups. When studying the global distribution of income (see, for example, Milanovic 2002), the contributions of inequality within countries and between countries to overall global inequality are of interest. To measure inequality between and within populations subgroups, a number of authors have suggested ways to decompose the Gini coefficient into within and between components, and these components have been given various interpretations. Examples of articles that do so are Dagum (1997, 2008), Deutsch and Silber (2008), Lambert and Aronson (1993), Sastry and Kelkar (1994), van de Ven et al (2001), Silber (1989), Yitzhaki (1994) and Yitzhaki and Lerman (1991). This list is not exhaustive; details of earlier contributions can be found in the reference lists of these papers.

In this paper we focus on two decompositions of the Gini coefficient, one that we call the traditional decomposition (e.g., Silber 1989, Lambert and Aronson 1993) and a second that we call the Dagum decomposition (Dagum 1997). We show that these two decompositions are identical, a fact that seems to have been overlooked in the literature. More specifically, the traditional decomposition can be written as

$$G = G_W + G_B + G_R \tag{1}$$

where G_W denotes within-group inequality, G_B denotes between-group inequality and G_R is a residual which is positive when some of the subpopulation income distributions overlap. More precise definitions are given below. Dagum writes his decomposition as

$$G = G_w + G_{nb} + G_t \tag{2}$$

where G_w denotes within-group inequality, G_{nb} measures the net contribution of between-group inequality, G_t measures the intensity of the "transvariation" between subpopulations, and $G_{gb} = G_{nb} + G_t$ measures the gross contribution of between-group inequality. Again, more precise definitions are given later in the paper. We observe that the definitions for G_W and G_w are the same. That is, $G_W = G_w$. What we establish that has not been previously acknowledged in the literature is that $G_B = G_{nb}$. It follows that $G_R = G_t$ and hence the decompositions are identical. Thus, when Dagum (1997) refers to his work as "A New Approach to the Decomposition of the Gini Income Inequality Ratio", it is true that he describes a new *approach*, but it is not true that he provides a new *decomposition*. Moreover, when Gertel et al (2008) describe their work as "…applying the Dagum decomposition analysis of the Gini ratio", they could equally have referred to it as the traditional decomposition.

In Section 2 we describe the traditional decomposition. The Dagum decomposition and its equivalence with the traditional decomposition are considered in Section 3.

2. The Traditional Decomposition

Suppose for the sake of exposition that the subpopulations are countries and that there are K of them which make up the total population that we call a region. Let μ_i denote mean income for the *i*-th country and λ_i denote the population share of the *i*-th country. Then mean income for the region is $\mu = \sum_{i=1}^{K} \lambda_i \mu_i$ and the income share for the *i*-th country is $s_i = \lambda_i \mu_i / \mu$. In the traditional decomposition $G = G_W + G_B + G_R$, the contribution of within-country inequality is given by a weighted average of the Gini coefficients for each of the countries, with weights given by the products of the population and income shares. That is,

$$G_W = \sum_{i=1}^{K} \lambda_i s_i G_i \tag{3}$$

where G_i is the Gini coefficient for the *i*-th country.

Between country inequality G_B is the Gini coefficient that would be obtained if everybody in a given country was given the mean income for that country. To obtain what will be a convenient representation for G_B, some more notation is needed. Let y_{ih} be the income of the *h*-th income unit in country *i*, let n_i be the number of income units in country *i*, and let $n = \sum_{i=1}^{K} n_i$ be the number of income units in the region. Then, the Gini coefficient for the complete region can be written as

$$G = \frac{1}{2n^2\mu} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{h=1}^{n_i} \sum_{k=1}^{n_j} \left| y_{ih} - y_{jk} \right| \tag{4}$$

Replacing y_{ih} and y_{jk} by their country means μ_i and μ_j yields

$$\begin{aligned} G_B &= \frac{1}{2n^2\mu} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{h=1}^{n_i} \sum_{k=1}^{n_j} \left| \mu_i - \mu_j \right| \\ &= \frac{1}{2n^2\mu} \sum_{i=1}^{K} \sum_{j=1}^{K} n_i n_j \left| \mu_i - \mu_j \right| \\ &= \frac{1}{2\mu} \sum_{i=1}^{K} \sum_{j=1}^{K} \lambda_i \lambda_j \left| \mu_i - \mu_j \right| \\ &= \frac{1}{\mu} \sum_{i=2}^{K} \sum_{j=1}^{i-1} \lambda_i \lambda_j \left| \mu_i - \mu_j \right| \end{aligned} \tag{5}$$

Interpretations of the residual term $G_R = G - G_W - G_B$ have been given by Silber (1989) and by Lambert and Aronson (1993). If none of the country income distributions overlap, then a ranking of income units in the whole region, ranked according to increasing income, is the same as the ranking that is obtained when countries are first ranked according to their mean incomes, and then income units within each country are ranked. Under these circumstances $G_R = 0$. If the two rankings are different, then some country distributions overlap and $G_R > 0$. Then, G_R can be viewed as the contribution to inequality that occurs when income units are re-ranked from the country-based ranking to the ranking for the whole region. Lambert and Aronson (1993) show that G_R is given by the area between the concentration curve from the country-based ranking and the Lorenz curve for the regional ranking.

3. The Dagum Decomposition

As mentioned in the introduction, in the Dagum decomposition within inequality is given by

$$G_w = \sum_{i=1}^{K} \lambda_i s_i G_i = G_W \tag{6}$$

To describe the other two components, Dagum assumes countries have been ordered according to increasing mean income. That is, $\mu_1 < \mu_2 < \cdots < \mu_K$. Then,

$$G_{nb} = \sum_{i=2}^{K} \sum_{j=1}^{i-1} \left(\lambda_j s_i + \lambda_i s_j \right) G_{ij} D_{ij} \tag{7}$$

$$G_t = \sum_{i=2}^{K} \sum_{j=1}^{i-1} \left(\lambda_j s_i + \lambda_i s_j \right) G_{ij} \left(1 - D_{ij} \right) \tag{8}$$

We will examine the components G_{ij} and D_{ij} and then prove that $G_{nb} = G_B$.

The term G_{ij} is what Dagum calls the extended Gini ratio between the *i*-th and *j*-th subpopulations

$$G_{ij} = \frac{\Delta_{ij}}{\mu_i + \mu_j} = \frac{1}{\left(\mu_i + \mu_j \right) n_i n_j} \sum_{h=1}^{n_i} \sum_{k=1}^{n_j} \left| y_{ih} - y_{jk} \right| \tag{9}$$

where Δ_{ij} is the mean difference between the *i*-th and *j*-th subpopulations.

Dagum calls D_{ij} the *relative* economic affluence between subpopulations *i* and *j*, when $\mu_i > \mu_j$. It is given by

$$D_{ij} = \frac{d_{ij} - p_{ij}}{\Delta_{ij}} \tag{10}$$

where $d_{ij} - p_{ij}$ is the *net* economic affluence between subpopulations *i* and *j*, when $\mu_i > \mu_j$, and Δ_{ij} is the maximum possible value of $d_{ij} - p_{ij}$. The components d_{ij} and p_{ij} are called, respectively, the *gross* economic affluence, and the first-order moment of transvariation, between subpopulations *i* and *j*, when $\mu_i > \mu_j$.

The definition of gross economic affluence d_{ij} for two countries such that $\mu_i > \mu_j$ is the weighted average of the income difference $y_{ih} - y_{jk}$ over all pairs such that $y_{ih} > y_{jk}$. In terms of continuous density functions $f_i(y)$ and $f_j(y)$ for income distributions *i* and *j*

$$\begin{aligned} d_{ij} &= \int_0^\infty \int_0^y (y-x) f_j(x)dx\, f_i(y)dy \\ &= m_{ij} + m_{ji} - \mu_j \end{aligned} \tag{11}$$

where

$$m_{ij} = \int_0^\infty yF_j(y) f_i(y)dy = E_i\left[yF_j(y)\right]$$

A proof of the second line in (11) is given by Dagum (1997).

The definition of the first-order moment of transvariation p_{ij} for two countries such that $\mu_i > \mu_j$ is the weighted average of the income difference $y_{jk} - y_{ih}$ over all pairs such that $y_{jk} > y_{ih}$. The number of such pairs will only be nonzero if the distributions overlap. It is given by

$$\begin{aligned} p_{ij} &= \int_0^\infty \int_0^y (y-x) f_i(x)dx\, f_j(y)dy \\ &= m_{ij} + m_{ji} - \mu_i \end{aligned} \tag{12}$$

It can also be shown that $\Delta_{ij} = d_{ij} + p_{ij}$ from which it follows that Δ_{ij} is the maximum possible value of $d_{ij} - p_{ij}$, the maximum occurring when $p_{ij} = 0$.

To prove that $G_{nb} = G_B$, first note that $d_{ij} - p_{ij} = \mu_i - \mu_j$ and, from (9) and (10),

$$G_{ij}D_{ij} = \frac{\Delta_{ij}}{\mu_i + \mu_j} \times \frac{d_{ij} - p_{ij}}{\Delta_{ij}} = \frac{\mu_i - \mu_j}{\mu_i + \mu_j} \qquad \text{for } \mu_i > \mu_j \tag{13}$$

Substituting (13) into (7) yields

$$G_{nb} = \sum_{i=2}^{K} \sum_{j=1}^{i-1} \left(\lambda_j s_i + \lambda_i s_j \right) \frac{\mu_i - \mu_j}{\mu_i + \mu_j} \qquad \text{for } \mu_i > \mu_j \tag{14}$$

While the ordering $\mu_1 < \mu_2 < \cdots < \mu_K$ is needed for Dagum's interpretations of the components of G_{nb}, it is not needed for computing (14) providing we replace $\left(\mu_i - \mu_j\right)$ by $\left|\mu_i - \mu_j\right|$. Doing so yields

$$\begin{aligned} G_{nb} &= \sum_{i=2}^{K} \sum_{j=1}^{i-1} \left(\lambda_j s_i + \lambda_i s_j \right) \frac{\left|\mu_i - \mu_j\right|}{\mu_i + \mu_j} \\ &= \sum_{i=2}^{K} \sum_{j=1}^{i-1} \left(\frac{\lambda_j \lambda_i \mu_i}{\mu} + \frac{\lambda_i \lambda_j \mu_j}{\mu} \right) \frac{\left|\mu_i - \mu_j\right|}{\mu_i + \mu_j} \\ &= \frac{1}{\mu} \sum_{i=2}^{K} \sum_{j=1}^{i-1} \lambda_j \lambda_i \left(\mu_i + \mu_j \right) \frac{\left|\mu_i - \mu_j\right|}{\mu_i + \mu_j} \\ &= \frac{1}{\mu} \sum_{i=2}^{K} \sum_{j=1}^{i-1} \lambda_j \lambda_i \left|\mu_i - \mu_j\right| \end{aligned} \tag{15}$$

This expression is identical to that in (5) from the traditional decomposition, giving the desired result. Thus, it follows that $G_w = G_W$, $G_{nb} = G_B$ and $G_t = G_R$.

4. Conclusion

The new approach to decomposing the Gini coefficient proposed by Dagum (1997) for measuring inequality contributions from between and within subpopulations yields a decomposition that is identical to the traditional decomposition which is commonly applied in the literature. Thus, while the "Dagum decomposition" provides a new interpretation, and new insights into the measurement of inequality, it does not lead to new expressions for the different components of the decomposition.

References

Dagum, C., 1997. A new approach to the decomposition of the Gini income inequality ratio. Empirical Economics 22, 515-531.

Dagum, C., 2008. Inequality decomposition, directional economic distance, metric distance, and Gini dissimilarity between income distributions. In: Betti, G., Lemmi, A. (Eds.), Advances on Income Inequality and Concentration Measures. Routledge, London, pp. 131-160.

Deutsch, J., Silber, J., 2008. On the Shapley value and the decomposition of inequality by population subgroups with special emphasis on the Gini index. In: Betti, G., Lemmi, A. (Eds.), Advances on Income Inequality and Concentration Measures. Routledge, London, pp. 161-178.

Gertel, H.R., Giuliodori, R.F., Rodríguez, A., 2008. Analysis of the short term impact of the Argentine social assistance program '*Plan Jefes y Jefas*' on income inequality applying the Dagum decomposition analysis of the Gini ratio. In: Betti, G., Lemmi, A. (Eds.), Advances on Income Inequality and Concentration Measures. Routledge, London, pp. 179-200.

Lambert, P.J., Aronson, J.R., 1993. Inequality Decomposition and the Gini coefficient revisited. Economic Journal 103, 1221-1227.

Milanovic, B., 2002. True world income distribution, 1988 and 1993: First calculation based on household surveys alone. Economic Journal 112, 51-92.

Sastry, D.V.S., Kelkar, U.R., 1994. Note on the decomposition of the Gini inequality. Review of Economics and Statistics 76, 584-586.

Silber, J., 1989. Factor components, population subgroups and the computation of the Gini index of inequality. Review of Economics and Statistics 71, 107-115.

van de Ven, J., Creedy, J., Lambert, P.J., 2001. Close equals and calculation of the vertical, horizontal and reranking effects of taxation. Oxford Bulletin of Economics and Statistics 63, 381-394.

Yitzaki, S., 1994. Economic distance and overlapping of distributions. Journal of Econometrics 61, 147-159.

Yitzaki, S., Lerman, R.I., 1991. Income stratification and income inequality. Review of Income and Wealth 37, 313-329.