A method for minimizing the residual term of the decomposable Gini index.

**Keywords:** Decomposition of the Gini Index, between groups inequality, within groups inequality, small sample bias.

The Gini index of income inequality for grouped data, is decomposed into the between groups and the within groups inequality that arises between the subgroups. In case that the subgroup income ranges overlap, the decomposition of the Gini index obtains as the sum of three terms, the between groups index, the within group index and the residual term. In this study we propose a method for reducing the value of the Gini’s index residual term, in case of overlapping. The proposed method concerns the representation of Gini index as a matrix product. We study both the case of large subgroups and the case of small samples as well. Τhe accurate calculation of the Gini Index based on data grouped by categories, commonly arises with income data, that are usually grouped for confidentiality purposes.

# Introduction

A decomposable inequality measure is defined as a measure such that the total inequality of a population can be broken down into the inequality that existing within subgroups of the population and the inequality that existing between subgroups. Gini index of inequality can be expressed as a decomposable measure but in case that the subgroup income ranges overlap, the decomposition of the Gini index obtains as the sum of three terms, the between groups index, the within group index and the residual term. Our main goal in this study, is to propose a correction that reduces the value of the residual term. The correction concerns the inequality between groups and the inequality within groups. Τhe impact of the methodology is examined for both large and small subgroups. The second section describes the decomposition of the Gini index that obtains as a matrix product and the proposed methodology that leads to the correction of the Gini index while it reduces the value of the residual term and the small error bias.

The third section contains the simulation results of the proposed method. We calculate the residual term before and after the correction for small and large subgroups and we compare the results regarding the value of the residual term. We also calculate the standard error and the corresponding confidence intervals.

The last section describes the conclusions and the contribution of the proposed method.

# Methods

Gini index  for n individuals grouped into m groups obtains as the product of the vectors **e΄**, **G** and **S** :

(1)

where,  are the column vectors of nh elements which are equal to 1/n, while h=1, 2, …, m.are the nκ by nλ matrices while κ, λ= 1,2,…,m. If κ=λ then matrices have 0 on their diagonals (-1)’s in their upper right triangle and (+1)’s in their lower left triangle, if κ<λ the value of all elements of matrices are equal to -1 and if κ>λ the value of all elements of matrices are equal to 1.  is the column vector of the nh elements s(i,h), where s(i,h) is the share of the individual i belonging to class h in total income. The product  can be written as a sum two components, Iw and IB that corresponds to the index of within classes inequality and to the index of between classes inequality [1].



In case that the subgroup income ranges overlap, the overall value of the Gini index is smaller of the sum of value of the within index inequality and the between index inequality and obtains as the sum of three terms, the between groups index, the within group index and the residual term.

IG(overlapping) = IW +IB +Ip

Let the  be the adjusted Gini index for n individuals that belong to k classes, where np is the number of individuals in p class , while p=1,2,...k . Then the adjusted Gini index is obtained as the sum of the following three terms:



where and  are the adjusted Gini index within and between groups and  the residual term. [2], [3] The adjusted Gini index within and between groups are defined as follow,





# Results

This section presents same initial results of the simulations that apply the proposed method in random data samples. The results concern random data samples that grouped into small or big population subgroups. In each case the value of the adjusted Gini is calculated, then the value of the residual term before the adjustment compared with the value of the residual term after the adjustment. Figure 1 and figure 2 depicts how the term is minimized using the adjustment Gini Index. In addition to that we will provide the standard error and the confidence intervals that are calculated for each case and for more applications as well.

****

Figure 1. Residual term and adjusted residual term calculated for 3 subgroups



Figure 2. Residual term and adjusted residual term calculated for 4 subgroups

# Conclusions

A method for reducing the value of the Gini ’s index residual term, in case of overlapping is proposed. Τhe calculation of the Gini Index based on data grouped by categories, commonly arises with income data, that are usually grouped for confidentiality purposes. This method lead not only to the reduction of the residual term but also to the reduction of the small sample bias that occurs in small subgroups. We study both the case of large subgroups and the case of small subgroups as well.

# main References

[1] S, Jacques. Factor components, population subgroups and the computation of the Gini index of inequality. The Review of Economics and Statistics (1989), 107-115.

[2] G. Deltas. The small-sample bias of the Gini coefficient: results and implications for empirical research. Review of economics and statistics 85.1 (2003), 226-234.

[3] T.V Ourti, and P. Clarke. The Bias of the Gini Coefficient due to Grouping. No. 08-095/3. Tinbergen Institute Discussion Paper, 2008.