

# A Convenient Method of Decomposing the Gini Index by Population Subgroups

Tomson Ogwang<sup>1</sup>

We propose a convenient method of estimating the within-group, between-group, and interaction components of the overall traditional Gini index from the estimated parameters of underlying “trick regression models” involving known forms of heteroscedasticity related to income. Two illustrative examples involving both real and artificial data are provided. The issue of appropriate standard error of the subgroup decomposition is also discussed.

*Key words:* Subgroup decomposition; Stochastic approach; Gini index; pseudo-Gini.

## 1. Introduction

Subgroup decomposition of inequality measures entails determining the proportion of observed inequality that is accounted for by the within-group, between-group, and in some cases the interaction component. Analysis of the trends of overall inequality and its components aids policy makers in devising appropriate inequality-reduction strategies. Kanbur (2006) articulates the policy significance of such decompositions and Radaelli (2010) provides a comprehensive survey of the literature on Gini subgroup decomposition.

In the 1960s, Bhattacharya and Mahalanobis (1967) and Rao (1969) decomposed the traditional Gini index by population subgroups into within-group and between-group components. The two-component decomposition strategy is valid if the subgroup income ranges do not overlap, that is, the richest income-receiving unit (individual, household) in the subgroup with a lower mean income class is not richer than the poorest income-receiving unit in any subgroup with a higher mean income. Subsequently, Pyatt (1976), Silber (1989) and Sastry and Kelkar (1994) decomposed the Gini index by population subgroups into within-group, between-group, and interaction (overlapping) components. In the traditional three-component subgroup decomposition, the within-group component is zero when there is no income inequality within each of the subgroups; the between-group component is the value of the Gini index when the income-receiving units in each subgroup receive the subgroup mean income; and the interaction component indicates the degree of income overlap between the subgroups. The three-component approach to Gini subgroup decomposition is more appealing than its two-component counterpart since it also applies when some subgroup income ranges do overlap. Because of this appeal, the

<sup>1</sup> Department of Economics, Brock University, 500 Glenridge Avenue, St. Catharines, Ontario, Canada L2S 3A1. Email: [togwang@brocku.ca](mailto:togwang@brocku.ca)

**Acknowledgment:** The author wishes to thank L. Kwong, J.F. Lamarche, the Associate Editor and three anonymous referees for their valuable comments. However, the usual disclaimer applies.

development of convenient ways of conducting three-component decompositions of the Gini index and the search for alternative ways of viewing the three components continue.

Pyatt's, Silber's and Sastry and Kelkar's subgroup decompositions of the Gini index employ matrix formulations that are not easily amenable to empirical implementation using linear regression methods. Yao and Liu (1996) and Yao (1999) proposed convenient ways of conducting these decompositions using spreadsheets without invoking any regressions. The stochastic approach considered by Ogwang (2000, 2004, 2006, 2007) and Giles (2004) provides a simple way of computing the Gini index from the estimated parameters of an underlying regression model with a known form of heteroscedasticity related to income. The purpose of this article is to exploit the simplification provided by the stochastic approach for purposes of conducting three-component subgroup decomposition of the traditional Gini index.

The methodology considered in this article has three major advantages. First, it provides a new way of viewing the within-group, between-group and interaction components of the traditional Gini index. For example, the interpretation of the between-group and interaction components of the overall Gini as weighted averages of their respective pseudo-Ginis has not previously been featured in the literature on subgroup decomposition of the Gini index. The concept "pseudo-Gini" as used in the context of Gini subgroup decomposition is explained in Section 3, where its interpretation and its similarity to the same concept as used in the context of income source decompositions are also explained. Second, from a practical perspective, the proposed method provides a new and convenient way of computing the three aforementioned components using widely available regression software packages. Third, the calculations and decomposition processes involved are quite transparent, which facilitates understanding.

The format of the rest of the article is as follows. In Section 2 we provide a brief overview of the stochastic approach to the overall Gini index. In Section 3 we derive the salient results that are needed in order to extend the stochastic approach for purposes of subgroup decomposition of the Gini index. In Section 4 we describe the actual empirical implementation of the stochastic approach for Gini subgroup decomposition. The salient issues surrounding the estimation of appropriate standard errors of the Gini subgroup decompositions are discussed in Section 5. Section 6 provides two illustrative examples using both real and artificial data and Section 7 forms the conclusion.

## 2. The Stochastic Approach to the Gini Index

Let  $y_1, y_2, \dots, y_n$  be the individual incomes of  $n$  income-receiving units (individuals, households) which are arranged such that  $y_1 \leq y_2 \leq \dots \leq y_n$ , and hence the ranks of  $y_1$  and  $y_n$  are 1 and  $n$ , respectively. Tied incomes are assigned the average of the ranks they would get without ties.

For purposes of subgroup decomposition of the Gini index using the stochastic approach, it is convenient to utilize the following formula considered by Ogwang (2007)

$$\hat{G} = (1/n)\hat{\gamma} \quad (1)$$

where  $\hat{\gamma}$  is the weighted least squares (WLS) estimator of  $\gamma$  in the "trick regression model"  $i^* = \gamma + u_i$  where  $i^* = (2i - n - 1)$ ,  $i = 1, 2, \dots, n$ , assuming that the errors  $u_i$  are

heteroscedastic of the form  $E(u_i^2) = \sigma^2/y_i$  (equivalently,  $\hat{\gamma}$  is the ordinary least squares (OLS) estimator of  $\gamma$  in the transformed model  $i^*\sqrt{y_i} = \gamma\sqrt{y_i} + u_i\sqrt{y_i}$  where the transformed errors  $u_i\sqrt{y_i}$  are homoscedastic under the assumed heteroscedastic structure). Note that under the stipulated heteroscedastic structure, Equation (1) yields accurate point estimates of the Gini index regardless of the number of observations. This is because under this heteroscedastic structure, the WLS estimator of  $\gamma$  in Equation (1) divided by  $n$  is, in fact, the usual Gini statistic.

### 3. Extending the Stochastic Approach to Gini Subgroup Decomposition

Suppose that  $n$  income-receiving units are classified into  $k$  mutually exclusive and exhaustive subgroups by, for example, gender, age, race, education, occupation or region. The  $k$  subgroups are arranged in ascending order of their subgroup mean incomes but the incomes in each subgroup can be in any order. Subgroups with identical mean incomes are first merged into one subgroup.

To facilitate the exposition of subgroup decomposition of the Gini index, let  $n_j$ ,  $j = 1, 2, \dots, k$  denote the number of income-receiving units in the subgroup with the  $j$ th smallest mean income, in which case  $n = \sum_{j=1}^k n_j$ . Let  $y_{ij}$ ,  $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$  denote the income of the  $i$ th income-receiving unit in the subgroup with the  $j$ th smallest mean income and  $\bar{y}_j = (1/n_j)\sum_{i=1}^{n_j} y_{ij}$ , the mean income in the same subgroup.

With respect to the income ranks, let  $r_{ij}$ ,  $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$  denote the rank of  $y_{ij}$  in relation to the incomes of all the  $n = \sum_{j=1}^k n_j$  income-receiving units in all the  $k$  subgroups. Also, let  $r'_{ij}$ ,  $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$  denote the rank of  $y_{ij}$  in relation to the incomes of only the  $n_j$  income-receiving units in the subgroup with the  $j$ th smallest mean income. Finally, let  $\tilde{r}_{ij}$  ( $j = 1, 2, \dots, k$ ;  $i = 1, 2, \dots, n_j$ ) denote the rank of  $y_{ij}$  in relation to the incomes of all the  $n = \sum_{j=1}^k n_j$  income-receiving units, assuming that  $y_{ij} = \bar{y}_j$  (i.e., all the income-receiving units in each subgroup are assumed to receive the mean income for that subgroup).

To facilitate the exposition of subgroup decomposition of the Gini index by exploiting Equation (1), it is also necessary to define the following three rank vectors:  $r_{ij}^* = 2r_{ij} - n - 1$  ( $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$ ),  $r_{ij}^{*'} = 2r'_{ij} - n_j - 1$  ( $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$ ) and  $\tilde{r}_{ij}^* = 2\tilde{r}_{ij} - n - 1$  ( $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$ ). It is easy to verify that

$$r_{ij}^* = \left[ 2r'_{ij} - n_j - 1 \right] + 2\left( r_{ij} - r'_{ij} \right) + (n_j - n), \quad i = 1, 2, \dots, n_j; \quad j = 1, 2, \dots, k. \quad (2)$$

Equation (2) can be conveniently rewritten as

$$r_{ij}^* = r_{ij}^{*'} + 2(r_{ij} - r'_{ij}) + (n_j - n), \quad i = 1, 2, \dots, n_j; \quad j = 1, 2, \dots, k. \quad (3)$$

Since the subgroups are arranged in ascending order of their mean incomes and in the absence of any subgroup mean income ties

$$\tilde{r}_{i1} = (n_1 + 1)/2, \quad i = 1, 2, \dots, n_1; \quad \tilde{r}_{ij} = \left[ \sum_{i=1}^{j-1} n_i \right] + (n_j + 1)/2, \quad (4)$$

$$i = 1, 2, \dots, n_j; \quad j = 2, \dots, k$$

(i.e.,  $\tilde{r}_{i1} = (n_1 + 1)/2$ ;  $\tilde{r}_{i2} = n_1 + (n_2 + 1)/2$ ;  $\tilde{r}_{i3} = n_1 + n_2 + (n_3 + 1)/2$ ; etc.).

Also,

$$\tilde{r}_{i1}^* = (n_1 - n), \quad i = 1, 2, \dots, n_1; \quad \tilde{r}_{ij}^* = \left[ \sum_{i=1}^{j-1} 2n_i \right] + (n_j - 1), \quad (5)$$

$$i = 1, 2, \dots, n_j; \quad j = 2, \dots, k$$

(i.e.,  $\tilde{r}_{i1}^* = (n_1 - n)$ ;  $\tilde{r}_{i2}^* = 2n_1 + (n_2 - n)$ ;  $\tilde{r}_{i3}^* = 2n_1 + 2n_2 + (n_3 - n)$ ; etc.).

Substituting  $(n_1 - n) = \tilde{r}_{i1}^*, i = 1, 2, \dots, n_1; (n_j - n) = \tilde{r}_{ij}^* - \sum_{i=1}^{j-1} 2n_i, i = 1, 2, \dots, n_j; j = 2, \dots, k$  from Equation (5) into Equation (3) and rearranging the terms yields

$$r_{ij}^* = r_{ij}^{*'} + \tilde{r}_{ij}^* + 2\tilde{r}_{ij}^*, \quad i = 1, 2, \dots, n_j; \quad j = 1, 2, \dots, k \quad (6)$$

where  $\tilde{r}_{i1}^* = (r_{i1} - r_{i1}')$ ,  $i = 1, 2, \dots, n_1$  and  $\tilde{r}_{ij}^* = (r_{ij} - r_{ij}' - \sum_{i=1}^{j-1} 2n_i)$ ,  $i = 1, 2, \dots, n_j; j = 2, \dots, k$ . The overall Gini index is given by

$$\hat{G} = \frac{1}{n} \frac{\sum_{i=1}^{n_j} \sum_{j=1}^k r_{ij}^* y_{ij}}{\sum_{i=1}^{n_j} \sum_{j=1}^k y_{ij}} \quad (7)$$

Substituting  $r_{ij}^*$ , given by Equation (6), into Equation (7) yields

$$\hat{G} = \frac{1}{n} \frac{\sum_{i=1}^{n_j} \sum_{j=1}^k r_{ij}^{*'} y_{ij}}{\sum_{i=1}^{n_j} \sum_{j=1}^k y_{ij}} + \frac{1}{n} \frac{\sum_{i=1}^{n_j} \sum_{j=1}^k \tilde{r}_{ij}^* y_{ij}}{\sum_{i=1}^{n_j} \sum_{j=1}^k y_{ij}} + \frac{2}{n} \frac{\sum_{i=1}^{n_j} \sum_{j=1}^k \tilde{r}_{ij}^* y_{ij}}{\sum_{i=1}^{n_j} \sum_{j=1}^k y_{ij}}. \quad (8)$$

The first, second, and third terms on the right-hand side of Equation (8) are the within-group component ( $\hat{G}_W$ ), the between-group component ( $\hat{G}_B$ ), and the interaction component ( $\hat{G}_I$ ), respectively, of the Gini index. It turns out that the within-group, between-group and interaction components are weighted averages of the within-group Ginis, the between-group pseudo-Ginis, and the interaction pseudo-Ginis, respectively.

The concept “pseudo-Gini” as used in this article refers to a numerical quantity which is computed using a Gini-like formula as an intermediate step in computing the between-group component of the overall Gini (in the case of the between-group pseudo-Gini) or the interaction component of the overall Gini (in the case of the interaction pseudo-Gini). [Fei et al. \(1978\)](#) and [Shorrocks \(1982\)](#) also use the concept “pseudo-Gini” in a similar manner, but in the context of income-source decompositions of the Gini index. As will be seen below, the subgroup decomposition pseudo-Ginis pertain to the various population

subgroups whereas the income-source decomposition pseudo-Ginis pertain to the various income sources.

For the within-group and between-group components in Equation (8), the weights are equal to the product of the subgroup population shares and the corresponding income shares; for the interaction component, the weights are equal to twice the product of the subgroup population shares and the corresponding income shares. To see these results, we note that the within-group component can be written as

$$\hat{G}_W = \sum_{j=1}^k p_j s_j \hat{G}_{Wj}, \quad j = 1, 2, \dots, k \quad (9)$$

where  $p_j = n_j/n$  is the population share of group  $j$ ;  $s_j = (\sum_{i=1}^{n_j} y_{ij}) / (\sum_{j=1}^k \sum_{i=1}^{n_j} y_{ij})$  is the income share of group  $j$ ; and

$$\hat{G}_{Wj} = \frac{1}{n_j} \frac{\sum_{i=1}^{n_j} r_{ij}^* y_{ij}}{\sum_{i=1}^{n_j} y_{ij}}, \quad j = 1, 2, \dots, k$$

is the within-group Gini for group  $j$ ; the between-group component can be written as

$$\hat{G}_B = \sum_{j=1}^k p_j s_j \hat{G}_{Bj}, \quad j = 1, 2, \dots, k \quad (10)$$

where  $p_j$  and  $s_j$  are defined in Equation (9) and

$$\hat{G}_{Bj} = \frac{1}{n_j} \frac{\sum_{i=1}^{n_j} \tilde{r}_{ij}^* y_{ij}}{\sum_{i=1}^{n_j} y_{ij}}, \quad j = 1, 2, \dots, k$$

is the between-group pseudo-Gini for subgroup  $j$ . Note that  $0 \leq \hat{G}_B \leq 1$  even though some between-group pseudo-Ginis may be negative as explained below; and the interaction component can be written as

$$\hat{G}_I = 2 \sum_{j=1}^k p_j s_j \hat{G}_{Ij}, \quad j = 1, 2, \dots, k \quad (11)$$

where  $p_j$  and  $s_j$  are defined in Equation (9) and

$$\hat{G}_{Ij} = \frac{1}{n_j} \frac{\sum_{i=1}^{n_j} \hat{r}_{ij}^* y_{ij}}{\sum_{i=1}^{n_j} y_{ij}}, \quad j = 1, 2, \dots, k$$

is the interaction pseudo-Gini for group  $j$ . Note also that  $0 \leq \hat{G}_I \leq 1$  even though some interaction pseudo-Ginis may be negative as explained below.

Before discussing the empirical implementation of the proposed method, it behooves us to clarify the interpretation of the between-group and interaction pseudo-Ginis as defined in Equations (10) and (11), respectively. Since the income shares and population shares in Equations (10) and (11) cannot be negative, and  $0 \leq \hat{G}_B, \hat{G}_I \leq 1$ , a positive/negative between-group pseudo-Gini for a particular subgroup indicates that the subgroup makes a positive/negative contribution to the between-group component. Likewise, a positive/

negative interaction pseudo-Gini for a particular subgroup indicates that the subgroup makes a positive/negative contribution to the interaction component. It should be noted that the pseudo-Ginis for the income source decompositions, as discussed by [Fei et al. \(1978\)](#) and [Shorrocks \(1982\)](#) among others, like those for the subgroup decompositions, can be positive or negative. With respect to the income source Gini decompositions, a positive/negative pseudo-Gini for a particular income source indicates that the source makes a positive/negative contribution to overall income inequality.

The issue of why the between-group pseudo-Gini or its interaction counterpart may be negative also deserves an explanation. A careful inspection of Equation (9) reveals that the transformed ranks,  $r_{ij}^{*'} , j = 1, 2, \dots, n_j$ , which are used in the computation of  $\hat{G}_{wj}$ , are identical to the transformed ranks which are used in the calculation of the Gini index for the  $j$ th subgroup. Hence,  $\hat{G}_{wj} , j = 1, 2, \dots, n_j$ , must lie between 0 and 1. Since the transformed ranks,  $\tilde{r}_{ij}^* , j = 1, 2, \dots, n_j$ , in Equation (10), which are used in the computation of  $\hat{G}_{Bj}$ , are different from the transformed ranks used in the calculation of the Gini index for the  $j$ th subgroup, there is no guarantee that  $\hat{G}_{Bj}$  will lie between 0 and 1 even though a Gini-like formula is used in its computation. Likewise, since the transformed ranks,  $\hat{r}_{ij}^*$  in Equation (11), which are used in the computation of  $\hat{G}_{ij}$ , are different from the transformed ranks used in the calculation of the Gini index for the  $j$ th subgroup, there is also no guarantee that  $\hat{G}_{ij}$  will lie between 0 and 1 even though a Gini-like formula is used in its computation. In the illustrative example provided below, negative between-group and interaction pseudo-Ginis are obtained.

#### 4. Empirical Implementation

First we create the seven rank vectors  $r_{ij} , r_{ij}' , \tilde{r}_{ij} , r_{ij}^* , r_{ij}^{*'} , \tilde{r}_{ij}^*$  and  $\hat{r}_{ij}^*$ .

##### 4.1. The Overall Gini Index

An inspection of Equation (7) reveals that the overall Gini index is given by

$$\hat{G} = (1/n)\hat{\gamma} \quad (12)$$

where  $\hat{\gamma}$  is the WLS estimator of  $\gamma$  in the model  $r_{ij}^* = \gamma + u_{ij}^* , i = 1, 2, \dots, n_j ; j = 1, 2, \dots, k$ , assuming that the errors,  $u_{ij}^*$ , are heteroscedastic of the form  $E(u_{ij}^{*2}) = \sigma^2/y_{ij}$ .

##### 4.2. The Within-Group Component

An inspection of the first expression on the right-hand side of Equation (8) reveals that the within-group component is given by

$$\hat{G}_w = (1/n)\hat{\gamma}_w \quad (13)$$

where  $\hat{\gamma}_w$  is the WLS estimator of  $\gamma_w$  in the model  $r_{ij}^{*'} = \gamma_w + u_{ij}^{*'} , i = 1, 2, \dots, n_j ; j = 1, 2, \dots, k$ , assuming that the errors,  $u_{ij}^{*'}$ , are heteroscedastic of the form  $E(u_{ij}^{*'}{}^2) = \sigma^2/y_{ij}$ .

#### 4.3. The Between-Group Component

An inspection of the second expression on the right-hand side of Equation (8) reveals that the between-group component is given by

$$\hat{G}_B = (1/n)\hat{\gamma}_B \quad (14)$$

where  $\hat{\gamma}_B$  is the WLS estimator of  $\gamma_B$  in the model  $\hat{r}_{ij}^* = \gamma_B + \tilde{u}_{ij}^*$ ,  $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$ , assuming that the errors,  $\tilde{u}_{ij}^*$ , are heteroscedastic of the form  $E(\tilde{u}_{ij}^{*2}) = \sigma^2/y_{ij}$ .

#### 4.4. The Interaction Component

An inspection of the third expression on the right-hand side of Equation (8) reveals that the interaction component is given by

$$\hat{G}_I = (2/n)\hat{\gamma}_I \quad (15)$$

where  $\hat{\gamma}_I$  is the WLS estimator of  $\gamma_I$  in the model  $\hat{r}_{ij}^* = \gamma_I + \hat{u}_{ij}^*$ ,  $i = 1, 2, \dots, n_j$ ;  $j = 1, 2, \dots, k$ , assuming that the errors,  $\hat{u}_{ij}^*$ , are heteroscedastic of the form  $E(\hat{u}_{ij}^{*2}) = \sigma^2/y_{ij}$ .

Two points about these subgroup decompositions are noteworthy. First,  $\hat{\gamma} = \hat{\gamma}_w + \hat{\gamma}_B + 2\hat{\gamma}_I$  and  $\hat{G} = \hat{G}_w + \hat{G}_B + \hat{G}_I$ . Second, the heteroscedastic structures for the regressions underlying  $\hat{G}$ ,  $\hat{G}_w$ ,  $\hat{G}_B$  and  $\hat{G}_I$  are similar.

### 5. Standard Errors of the Gini Subgroup Decompositions

When reporting estimates of the inequality measures and their decompositions, it is also important to report estimates of their standard errors or confidence intervals to facilitate hypotheses tests about their significance. The case for reporting standard errors of inequality measures as well as their decompositions seems strong given that large standard errors may arise even though the number of income-receiving units is large, as pointed out by [Maasoumi \(1994\)](#).

In the case of the overall Gini, four types of standard errors have been employed in the literature, namely the asymptotic standard errors (e.g., [Cowell 1989](#); [Davidson 2009](#)), the bootstrap (e.g., [Dixon et al. 1987](#); [Mills and Zandvakili 1997](#); [Davidson 2009](#)), the jackknife (e.g., [Yitzhaki 1991](#); [Karoly 1992](#); [Ogwang 2000](#)), and WLS/OLS (e.g., [Giles 2004](#)). However, most Gini subgroup decomposition proposals so far do not consider the issue of appropriate standard errors of the relevant components, which is surprising given that the standard errors or confidence intervals of the decompositions of other inequality measures, such as the Generalized Entropy class of measures, are widely reported in the literature. For example, [Mills and Zandvakili \(1997\)](#), [Biewen \(2002\)](#), and [Gray et al. \(2003\)](#) report bootstrap standard errors of the subgroup decompositions of several inequality measures, but they do not report the standard errors of the decompositions of the Gini index. [Mussard and Richard \(2012\)](#) is a rare paper that reports confidence intervals of the Gini decompositions.

Although OLS/WLS standard errors of the decompositions could also be obtained from the estimated standard errors of the parameters of the underlying regressions following

Giles (2004), we do not recommend doing so in light of the inadequacies associated with using OLS/WLS-based Gini standard errors raised by Modarres and Gastwirth (2006), Ogwang (2004, 2006); Davidson (2009) and Langel and Tillé (2013), among others, in the context of the overall Gini. These inadequacies arise because OLS/WLS standard errors are based on ordered observations that are not independent even when the income series is independent identically distributed. Although ordering does not affect point estimates of the decompositions, it affects the OLS/WLS-based standard errors. To circumvent this problem, the use of resampling methods (e.g., the bootstrap or jackknife) in conjunction with the stochastic approach to obtain the standard errors of the decompositions is recommended, provided that these methods are applied correctly. For example, Davidson (2009) has noted, in the context of the overall Gini, that bootstrap standard errors can be reliable if applied correctly. More recently, Langel and Tillé (2013) demonstrated that jackknife standard errors of the Gini index can also be reliable if the randomness of the income ranks is properly taken into account by recalculating the ranks each time an observation is dropped in the computation of the jackknife standard error.

Shao and Tu (1995) provide a comprehensive overview of the bootstrap and jackknife techniques in general. Davidson (2009) describes the bootstrap and jackknife approaches to estimating the standard errors and confidence intervals of the overall Gini index. Since that jackknife confidence intervals of the subgroup decompositions of the Gini index are reported in one of the illustrative examples below, we provide a brief description of the jackknife approach to the computation of the standard errors of the subgroup decompositions. Let  $\hat{G}_W(n, k)$ ,  $\hat{G}_B(n, k)$  and  $\hat{G}_I(n, k)$  denote the estimates of the within-group, between-group and interaction components, respectively, of the Gini index based on the remaining  $(n - 1)$  observations after deleting the  $k$ th observation. Also, let  $\bar{G}_W(n) = n^{-1} \sum_{k=1}^n \hat{G}_W(n, k)$ ,  $\bar{G}_B(n) = n^{-1} \sum_{k=1}^n \hat{G}_B(n, k)$  and  $\bar{G}_I(n) = n^{-1} \sum_{k=1}^n \hat{G}_I(n, k)$  be the means of all the  $\hat{G}_W(n, k), k = 1, 2, \dots, n$ ,  $\hat{G}_B(n, k), k = 1, 2, \dots, n$ , and  $\hat{G}_I(n, k), k = 1, 2, \dots, n$ , respectively. The jackknife standard errors of the within-group, between-group and interaction components of the Gini index are given by

$$\begin{aligned} SE(\hat{G}_W) &= \sqrt{\left(\frac{n-1}{n}\right) \sum_{k=1}^n (\hat{G}_W(n, k) - \bar{G}_W(n))^2}, \\ SE(\hat{G}_B) &= \sqrt{\left(\frac{n-1}{n}\right) \sum_{k=1}^n (\hat{G}_B(n, k) - \bar{G}_B(n))^2} \quad \text{and} \\ SE(\hat{G}_I) &= \sqrt{\left(\frac{n-1}{n}\right) \sum_{k=1}^n (\hat{G}_I(n, k) - \bar{G}_I(n))^2}, \quad \text{respectively.} \end{aligned}$$

## 6. Illustrative Examples

### *Example 1: Artificial Data*

We first utilize a simple dataset comprising the incomes of  $n = 7$  individuals which are broken down into three subgroups with different mean incomes (i.e.,  $k = 3$ ) to illustrate



the steps involved in the empirical implementation of the proposed method. Silber (1989) and Sastry and Kelkar (1994) used the same data to demonstrate their Gini subgroup decomposition methods. The incomes constituting the first, second, and third subgroups are  $\{(1,3)$ ; mean income = 2 $\}$ ,  $\{(1,4,7)$ ; mean income = 4 $\}$ , and  $\{(6,10)$ ; mean income = 8 $\}$ , respectively. As mentioned above, the subgroups are arranged in ascending order of their mean incomes, in which case  $n_1 = n_3 = 2$  and  $n_2 = 3$ .

Table 1 shows the required ranks and their transformations, using the computational procedures and notations described in Sections 3 and 4. Non-zero entries for  $\hat{r}_{ij}^*$  in the last column of the table indicate the existence of a non-zero interaction component since there is an overlap in the income ranges for the first and second subgroups. Table 2 summarizes the empirical results. It is apparent from the table that, apart from rounding errors,  $\hat{\gamma} = \hat{\gamma}_w + \hat{\gamma}_B + 2\hat{\gamma}_I$  and  $\hat{G} = \hat{G}_W + \hat{G}_B + \hat{G}_I$ .

Table 3 compares our decomposition results with previous results based on the same dataset. The table indicates that our decomposition results are identical to Silber's (1989) results. However, Sastry and Kelkar's (1994) estimates of the between-group and interaction components differ from ours. The difference between Sastry and Kelkar's results and our results arises because Sastry and Kelkar assign the income-receiving units their original ranks in the computation of the between-group component.

According to our results, between-group inequality accounts for most of the observed inequality and the interaction term accounts for the smallest percentage of total inequality. Another notable aspect of Table 3 is the numerical equivalence between our subgroup decomposition and Dagum's (1997) decomposition. Specifically, our between-group and interaction components are numerically the same as Dagum's *net contribution between population subgroups* ( $G_{nb}$ ) and *the contribution of the income intensity of transvariation between subgroups* ( $G_I$ ), respectively. Hence, Dagum's approach to Gini subgroup decomposition only differs from the traditional approach in the interpretation of some components.

Example 2: Real Data

The methodology described above is also applied to the data on the total pre-tax post-transfer incomes, in Canadian dollars, of a random sample of 4,883 persons, derived from the Canadian Census 2006 Public Use Micro data Files.

Table 1. Computing the required ranks and their transformations\*

Group ( $i$ )	$y_{ij}$	$r_{ij}$	$r'_{ij}$	$\tilde{r}_{ij}$	$r_{ij}^*$	$r_{ij}^{*'} $	$\tilde{r}_{ij}^*$	$\hat{r}_{ij}^*$
1	1	1.5	1	1.5	−5	−1	−5	0.5
1	3	3	2	1.5	−2	1	−5	1
2	1	1.5	1	4	−5	−2	0	−1.5
2	4	4	2	4	0	0	0	0
2	7	6	3	4	4	2	0	1
3	6	5	1	6.5	2	−1	5	−1
3	10	7	2	6.5	6	1	5	0

\* $n_1 = n_3 = 2; n_2 = 3; n = 7; k = 3; r_{ij}^* = 2r_{ij} - n - 1; r_{ij}^{*'} = 2r'_{ij} - n_j - 1; \tilde{r}_{ij}^* = 2\tilde{r}_{ij} - n - 1; \hat{r}_{i1}^* = (r_{i1} - r'_{i1}), i = 1, n_1 = 2$  and  $\hat{r}_{ij}^* = (r_{ij} - r'_{ij} - \sum_{i=1}^{j-1} n_i), i = 1, 2, \dots, n_j; j = 2, k = 3$ .

Table 2. Computing the overall Gini index and its decompositions\*

Gini	Equation	$\hat{\gamma}$	$\hat{\gamma}_W$	$\hat{\gamma}_B$	$\hat{\gamma}_I$	$\hat{G}$	$\hat{G}_W$	$\hat{G}_B$	$\hat{G}_I$
Overall	12	2.625				0.375			
Within	13		0.563				0.080		
Between	14			1.875				0.268	
Interaction	15				0.094				0.027

\*  $n = 7$ ;  $\hat{G} = \hat{\gamma}/n$ ;  $\hat{G}_W = \hat{\gamma}_W/n$ ;  $\hat{G}_B = \hat{\gamma}_B/n$  and  $\hat{G}_I = 2\hat{\gamma}_I/n$ .

The fact that Canada has one of the highest per capita immigration rates in the world provides a strong case for studying the income differentials between immigrants and nonimmigrants in Canada. One of the common goals of such studies is to examine the factors that influence the intertemporal changes in income inequality between immigrants and nonimmigrants.

In this illustrative example, we consider a breakdown of the data into three nonoverlapping subgroups by immigration status (i.e., nonpermanent residents, nonimmigrants and immigrants). According to the 2006 Census definitions, nonimmigrants are people who are Canadian citizens by birth; immigrants are people who were, or had ever been, landed immigrants in Canada prior to the day of the 2006 Census; and nonpermanent residents are people from other countries who, at the time of the 2006 Census, held a work/study permit or claimed refugee status, as well as their family members living in Canada.

To provide some insights into the nature of the income distributions pertaining to the various subgroups, in Table 4, we report some relevant descriptive statistics pertaining to these subgroups and the full sample. The table indicates positive skewness of the distributions for all income ranges, as would be expected. Another notable feature of Table 4 is the significant overlap in the income ranges. Specifically, the income range [0,110000] for the nonpermanent resident subgroup, with the lowest mean income, lies entirely within the income range [0, 866340] for the immigrant subgroup, with the second lowest mean income, which in turn lies entirely within the income range [0,1285600] for the nonimmigrant subgroup, with the highest mean income. As mentioned above, the interaction component of the Gini index is non-zero if there are overlaps in the income ranges of some of the population subgroups, which is obviously the case in the present example. In fact, given the large overlaps in the income ranges under consideration, we would also expect the contribution of the interaction component to the overall Gini index to be large and statistically significant.

Table 3. Comparing decomposition methods

Method	Overall	Within group	Between group	Interaction
Silber (1989)	0.375	0.080	0.268	0.027
Sastry and Kelkar (1994)	0.375	0.080	0.205	0.089
Dagum (1997)	0.375	0.080	0.268*	0.027**
Present proposal	0.375	0.080	0.268	0.027

\*and \*\* denote estimates of Dagum’s (1997)  $G_{nb}$  and  $G_I$ , respectively.

Table 4. Income descriptive statistics by population subgroups (Canadian 2006 Census data)\*

Subgroup	Mean	Median	Min	Max	Skewness	Excess Kurtosis
Nonpermanent residents	26,487	13,000	0	110,000	1.1641	0.21446
Immigrants	33,338	21,000	0	866,340	7.4692	82.00
Nonimmigrants	34,410	25,000	0	1,285,600	8.6455	180.67
All groups	34,102	24,000	0	1,285,600	8.3156	147.43

\* The incomes are measured in Canadian dollars.

In order to gauge the extent of the overlap, we applied the computational procedures described in Sections 3 and 4 above to the 2006 Census data, the results of which are reported in Table 5. Several features of the table are particularly noteworthy. First, the estimate of the overall Gini index turns out to be 0.520 and is statistically significant at the conventional five percent level of significance, since the estimated 95 percent confidence interval does not include zero. Second, the estimates of the within-group, between-group and interaction components turn out to be 0.329, 0.007 and 0.184, respectively. However, only the within-group and interaction component estimates turn out to be statistically significant at five percent level of significance. The confidence interval for the between-group component turns out not to show statistical significance at five percent level of significance, since the associated 95 percent confidence interval includes zero.

It is also apparent from Table 5 that within-group inequality accounts for most of the observed inequality, followed by the interaction component, which is in turn followed by the between-group component. Owing to the huge overlap in the income ranges for the three subgroups, as mentioned above, it is not surprising that the contribution of the interaction component is larger than that of the between-group component. Clearly, these results are informative with respect to the nature of the observed income inequality and are consistent with the descriptive statistics reported in Table 4.

Another positive aspect of the approach adopted in this article is its ability to provide a more detailed breakdown of the subgroup decompositions to also include the contributions of the various subgroups to overall inequality as well as their standard errors. The issue of the contributions of the various subgroups to overall inequality has so far been largely ignored in the literature on Gini subgroup decomposition, which focuses more on the

Table 5. Gini subgroup decomposition results (Canadian 2006 Census data)

Component	Point estimate	95 percent confidence interval*
Within-group	0.3288	(0.3134 to 0.3442)
Between-group	0.0073	(− 0.0140 to 0.0285)
Interaction	0.1837	(0.1509 to 0.2165)
Overall	0.5198	(0.4998 to 0.5398)

\* The confidence intervals were constructed using the computed jackknife standard errors. To conserve space, the results of the intermediate computational procedures as described in Sections 3, 4 and 5 are not reported but are available from the author on request.

Table 6. A detailed breakdown of the Gini subgroup decompositions (Canadian 2006 Census data)

Immigration status (j)	Between-group and interaction pseudo-Ginis			Contributions of the various components to overall inequality				
	Within-group Gini $\hat{G}_{Wj}$	$\hat{G}_{Bj}$	$\hat{G}_{Ij}$	Pop. share $p_j$	Income share $s_j$	Within group $p_j s_j \hat{G}_{Wj}$	Between group $p_j s_j \hat{G}_{Bj}$	Interaction $2p_j s_j \hat{G}_{Ij}$
Non PR (j = 1)	0.6016	-124.2051	96.0433	0.0080	0.0062	0.0000	-0.0061	0.0095
Immigrant (j = 2)	0.5441	-3.3171	2.5051	0.2279	0.2228	0.0277	-0.1685	0.2545
Non-IM. (j = 3)	0.5112	0.3088	-0.0681	0.7641	0.7710	0.3011	0.1819	-0.0802
Total	-	-	-	1.0000	1.0000	0.3288**	0.0073**	0.1837**

\*The total contribution of subgroup j to overall inequality =  $p_j s_j \hat{G}_{Wj} + p_j s_j \hat{G}_{Bj} + 2p_j s_j \hat{G}_{Ij}$ , j = 1, 2, 3. For the nonpermanent resident group (j = 1), the total contribution is 0.0034; for the immigrant group (j = 2), the total contribution is 0.1136; and for the nonimmigrant group (j = 3) the total contribution is 0.4028.

\*\*The total contribution of the within-group component to overall inequality =  $\sum_{j=1}^3 p_j s_j \hat{G}_{Wj} = 0.3288$ ; the total contribution of the between-group component =  $\sum_{j=1}^3 p_j s_j \hat{G}_{Bj} = 0.0073$ ; and the total contribution of the interaction component =  $2 \sum_{j=1}^3 p_j s_j \hat{G}_{Ij} = 0.1837$ .

breakdown of the overall Gini index into the within-group, between-group and interaction components. Table 6 provides a two-way breakdown of the Gini subgroup decompositions. It is apparent from the entries in the table that the nonimmigrant subgroup makes the highest contribution to the observed inequality even though it has the lowest degree of within-group inequality, with an estimated Gini index of 0.511. Also, the nonpermanent resident subgroup makes the lowest contribution to overall inequality, yet it has the lowest degree of within-group inequality, with an estimated Gini index of 0.602. These results demonstrate one area of similarity between Gini subgroup decomposition, which also provides information on the contributions of the various population subgroups to overall inequality, and Gini income source decomposition, which provides information on the contributions of the various income sources to overall inequality.

The results presented in Table 6 confirm the possibility of negative between-group or interaction pseudo-Ginis as already alluded to above. For example, the table shows that both the immigrant and nonpermanent resident subgroups contribute negatively to between-group inequality, whereas the nonimmigrant subgroup is the sole negative contributor to the interaction component. In this regard, it is interesting to note that the nonpermanent resident and immigrant subgroups that contribute negatively to the between-group component are the ones whose income ranges fall entirely within the income range for the nonimmigrant subgroup, whose contribution to between-group inequality turns out to be positive. Furthermore, the nonimmigrant subgroup for which part of the income range does not overlap with the income ranges of the other two subgroups makes a negative contribution to the interaction component. Clearly, Table 6 illustrates the bidimensional nature of Gini subgroup decomposition, which could be fully exploited in future empirical studies. Specifically, it shows that Gini subgroup decomposition does not only entail a breakdown of overall inequality into within-group, between-group and interaction components but also a breakdown of the contributions of each subgroup to overall inequality, which is akin to the breakdown in income source decompositions.

## 7. Concluding Remarks

In this article we have extended previous literature in which the regression approach is used to construct the overall Gini index to provide a new way of viewing and decomposing the overall Gini index by population subgroups into within-group, between-group and interaction components. The methodology proposed entails specifying certain underlying “trick regression models” with known heteroscedastic structures related to income.

Although the stochastic approach proposed in this article provides accurate point estimates of the within-group, between-group and interaction components of the Gini decompositions, it would not be adequate to use the associated OLS/WLS standard errors owing to the problems mentioned in Section 5. In light of this difficulty, the use of jackknife or bootstrap standard errors of the subgroup decompositions is recommended.

Finally, we believe that this article demonstrates how the pseudo-Ginis that are computed in the intermediate stages of the subgroup decompositions of the Gini index can provide very useful insights into the contribution of each population subgroup to the within-group, between-group and interaction components of overall inequality and hence the total contribution of each subgroup to overall inequality.

## 8. References

- Bhattacharya, N. and Mahalanobis, B. (1967). Regional Disparities in Household Consumption in India. *American Statistical Association Journal*, 62, 143–161. DOI: <http://www.dx.doi.org/10.1080/01621459.1967.10482896>.
- Biewen, M. (2002). Bootstrap Inference for Inequality, Mobility and Poverty Measurement. *Journal of Econometrics*, 108, 317–342. DOI: [http://www.dx.doi.org/10.1016/S0304-4076\(01\)00138-5](http://www.dx.doi.org/10.1016/S0304-4076(01)00138-5).
- Cowell, F.A. (1989). Sampling Variance and Decomposable Inequality Measures. *Journal of Econometrics*, 42, 27–41. DOI: [http://www.dx.doi.org/10.1016/0304-4076\(89\)90073-0](http://www.dx.doi.org/10.1016/0304-4076(89)90073-0).
- Dagum, C. (1997). A New Approach to the Decomposition of the Gini Income Inequality Ratio. *Empirical Economics*, 22, 515–531. DOI: [http://www.dx.doi.org/10.1007/978-3-642-51073-1\\_4](http://www.dx.doi.org/10.1007/978-3-642-51073-1_4).
- Davidson, R. (2009). Reliable Inference for the Gini Index. *Journal of Econometrics*, 150, 30–40.
- Dixon, P.M., Weiner, J., Mitchel-Olds, T., and Woodley, R. (1987). Bootstrapping the Gini Coefficient of Inequality. *Ecology*, 68, 1548–1551. DOI: <http://www.dx.doi.org/10.2307/1939238>.
- Fei, J.C.H., Ranis, G., and Kuo, S.W.Y. (1978). Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 92, 17–53. DOI: <http://www.dx.doi.org/10.2307/1885997>.
- Giles, D. (2004). Calculating a Standard Error for the Gini Coefficient: Some Further Results. *Oxford Bulletin of Economics and Statistics*, 66, 425–433. DOI: <http://www.dx.doi.org/10.1111/j.1468-0084.2004.00086.x>.
- Gray, D., Mills, J.A., and Zandvakili, S. (2003). Statistical Inference of Inequality With Decompositions: the Canadian Experience. *Empirical Economics*, 28, 291–302. DOI: <http://www.dx.doi.org/10.1007/s001810200131>.
- Kanbur, R. (2006). The Policy Significance of Inequality Decompositions. *Journal of Economic Inequality*, 4, 367–374. DOI: <http://www.dx.doi.org/10.1007/s10888-005-9013-5>.
- Karoly, L.A. (1992). Changes in the Distribution of Individual Earnings in the United States: 1967–1986. *Review of Economics and Statistics*, 74, 107–115. Available at: <http://www.jstor.org/stable/2109548> (access September 1, 2011).
- Langel, M. and Tillé, Y. (2013). Variance Estimation of the Gini Index: Revisiting a Result Several Times Published. *Journal of the Royal Statistical Society Series A Statistics in Society*, 176, 521–540. DOI: <http://www.dx.doi.org/10.1111/j.1467-985X.2012.01048.x>.
- Maasoumi, E. (1994). Empirical Analysis of Inequality and Welfare. *Handbook of Applied Microeconomics*, P. Schmidt and H. Peasaran (eds). Oxford: Blackwell.
- Mills, J. and Zandvakili, S. (1997). Statistical Inference Via Bootstrapping for Measures of Economic Inequality. *Journal of Applied Econometrics*, 12, 133–150. DOI: [http://www.dx.doi.org/10.1002/\(SICI\)1099-1255\(199703\)12:2<133::AID-JAE433>3.0.CO;2-H](http://www.dx.doi.org/10.1002/(SICI)1099-1255(199703)12:2<133::AID-JAE433>3.0.CO;2-H).
- Modarres, R. and Gastwirth, J.L. (2006). A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality. *Oxford Bulletin of Economics and Statistics*, 68, 385–390. DOI: <http://www.dx.doi.org/10.1111/j.1468-0084.2006.00167.x>.

- Mussard, S. and Richard, P. (2012). Linking Yitzhaki's and Dagum's Gini Decompositions. *Applied Economics*, 44, 2997–3010. DOI: <http://www.dx.doi.org/10.1080/00036846.2011.568410>.
- Ogwang, T. (2000). A Convenient Method of Computing the Gini Index and its Standard Error. *Oxford Bulletin of Economics and Statistics*, 62, 123–129. DOI: <http://www.dx.doi.org/10.1111/1468-0084.00164>.
- Ogwang, T. (2004). Calculating a Standard Error for the Gini Coefficient: Some Further Results: Reply. *Oxford Bulletin of Economics and Statistics*, 66, 435–437. DOI: <http://www.dx.doi.org/10.1111/j.1468-0084.2004.00087.x>.
- Ogwang, T. (2006). A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality: Comment. *Oxford Bulletin of Economics and Statistics*, 68, 391–393. DOI: <http://www.dx.doi.org/10.1111/j.1468-0084.2006.00167.x>.
- Ogwang, T. (2007). Additional Properties of a Linear Pen's Parade for Individual Data Using the Stochastic Approach to the Gini Index. *Economics Letters*, 96, 369–374.
- Pyatt, G. (1976). On the Interpretation and Disaggregation of Gini Coefficients. *Economic Journal*, 86, 243–255. Available at: <http://www.jstor.org/stable/2230745>.
- Radaelli, P. (2010). On the Decomposition by Subgroups of the Gini Index and Zenga's Uniformity and Inequality Indexes. *International Statistical Review*, 78, 81–101. DOI: <http://www.dx.doi.org/10.1111/j.1751-5823.2010.00100.x>.
- Rao, V.M. (1969). Two Decompositions of Concentration Ratio. *Journal of the Royal Statistical Society Series A (General)*, 132, 418–425. Available at: <http://www.jstor.org/stable/2344120>.
- Sastry, D.V.S. and Kelkar, U.R. (1994). Note on the Decomposition of Gini Inequality. *Review of Economics and Statistics*, LXXVI, 584–586. Available at: <http://www.jstor.org/stable/2109984>.
- Shao, J. and Tu, D. (1995). *Jackknife and Bootstrap*. New York: Springer.
- Shorrocks, A.F. (1982). Inequality Decomposition by Factor Components. *Econometrica*, 50, 193–211. Available at: <http://www.jstor.org/stable/1912537>.
- Silber, J. (1989). Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality. *Review of Economics and Statistics*, 71, 107–115. Available at: <http://www.jstor.org/stable/1928057>.
- Yao, S. and Liu, J. (1996). Decomposition of the Gini Coefficient by Class: A New Approach. *Applied Economics Letters*, 3, 115–119.
- Yao, S. (1999). On the Decomposition of Gini Coefficients by Population Class and Income Source: A Spreadsheet Approach and Application. *Applied Economics*, 31, 1249–1264. DOI: <http://www.dx.doi.org/10.1080/000368499323463>.
- Yitzhaki, S. (1991). Calculating Jackknife Variance Estimators for Parameters of the Gini Method. *Journal of Business and Economic Statistics*, 9, 235–239. DOI: <http://www.dx.doi.org/10.1080/07350015.1991.10509849>.

Received October 2011

Revised May 2013

Accepted September 2013